

Isabelle Sousa de Moura

Combinação de Modelos de Previsão Climática

Natal - RN

1 de Junho de 2018

Isabelle Sousa de Moura

Combinação de Modelos de Previsão Climática

Monografia de Graduação apresentada ao Departamento de Estatística do Centro de Ciências Exatas e da Terra da Universidade Federal do Rio Grande do Norte como requisito parcial para a obtenção do grau de Bacharel em Estatística.

Universidade Federal do Rio Grande do Norte

Centro de Ciências Exatas e da Terra

Departamento de Estatística

Orientador: Prof. Dr. Marcus Alexandre Nunes

Natal - RN

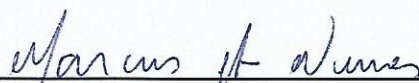
1 de Junho de 2018

Isabelle Sousa de Moura

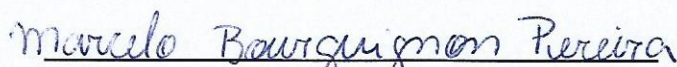
Combinação de Modelos de Previsão Climática

Monografia de Graduação apresentada ao Departamento de Estatística do Centro de Ciências Exatas e da Terra da Universidade Federal do Rio Grande do Norte como requisito parcial para a obtenção do grau de Bacharel em Estatística.

Aprovada em 14 de junho de 2018 .



Prof. Dr. Marcus Alexandre Nunes
Orientador



Prof. Dr. Marcelo Bourguignon
Pereira
Examinador



Prof. Dr. André Luís Santos de Pinho
Examinador

Natal - RN

1 de Junho de 2018

Dedico a Elisabeth, Elisângela e Veneranda.

Agradecimentos

Agradeço às mulheres da minha vida, minha mãe Elisabeth Vilma, minha tia Elisângela Veneranda e minha avó Maria Veneranda, que sempre foram minha maior fonte de inspiração, força e dedicação. Sem vocês eu não teria chegado onde cheguei.

Sou grata ao meu tio Kleber Xavier, meu padrasto Alexander Medeiros, meu irmão Gabriel Sousa, meu pai Edson Lima e minha avó Maria Piedade, por acreditarem e me apoiarem em tudo.

Aos amigos que fiz ao longo do curso e me ajudaram tanto nesses 4 anos Rayane, Mariana, José Edson, Igor, Ariane, Ruanderson, Antony, Rodrigo, Taynná, Adryan e Lucas. À minha amiga Samara, muito obrigado pelo carinho. Agradeço todos os meus professores Damião, André, Luz, Mariana, Paulo Roberto, Iloneide, Marcelo, Pledson, Dione e Carla, principalmente ao professor Marcus Alexandre Nunes, que fez toda a diferença no meu curso, não me deixando desistir do meu sonho e por todos os ensinamentos e conselhos.

Agradeço a Naurinete Jesus da Costa Barreto, por ter disponibilizado os dados analisados neste trabalho.

Também agradeço a David Mendes, pelas discussões científicas.

“Todas as vitórias ocultam uma abdicação.”

Simone de Beauvoir

Resumo

A precipitação é uma das variáveis mais importantes para descrever o clima futuro, pois descreve qualquer tipo de fenômeno relacionado à queda de água do céu. Existem diversos modelos de simulação climática previsão da precipitação hoje. Os modelos adotados neste trabalho serão o BCC-CSM1.1, CCSM4, CESM1, CPC-NOAA, NorESM1-ME e MRI-CGCM3 em três regiões distintas Amazônia, Bacia do Rio da Prata e o Nordeste Brasileiro. São considerados três tipos de cenários diferentes: RCP 4.5, RCP 6.0 e RCP 8.5. Cenários possuem diferentes características em relação às variáveis de emissão de gases, concentração de gases de efeito estufa, e informações de tipo de cobertura terrestre. Estudos anteriores mostram que a combinação entre modelos torna a previsão mais precisa. Devido a isso, foram adotados dois métodos de combinação de modelos, chamados *Random Forest* e *SVM* (*Support Vector Machine* ~ Máquina de Vetores Suporte). A raiz do erro quadrático médio foi utilizada para a comparação entre os dois métodos, após essa comparação foi observado que os dois métodos conseguem fazer uma previsão aproximada dos valores reais, sendo o *Random Forest* tendo o menor erro quadrático médio em todos os diferentes tipos de cenários entre as regiões escolhidas.

Palavras-chave: Aprendizado de Máquina. Floresta Aleatória. Máquina de vetores de suporte.

Abstract

Precipitation is one of the most important variables to describe the future climate, since it describes any type of phenomenon related to the fall of water from the sky. There are several models of climate simulation to forecast precipitation today. The models adopted in this work will be BCC-CSM1.1, CCSM4, CESM1, CPC-NOAA, NorESM1-ME and MRI-CGCM3 in three distinct regions Amazonia, Rio de la Plata Basin and Northeast Brazil. Three different types of scenarios are considered: RCP 4.5, RCP 6.0, and RCP 8.5. These scenarios have different characteristics in relation to the variables of emission of gases, concentration of gases of greenhouse effect, and information of type of terrestrial coverage. Previous studies show that the combination of models makes forecasting more accurate. Due to this, two methods of combining models, Random Forest llamods and Support Vector Machine (SVM) were adopted. Following the mean square error was used for the comparison between the two methods, after this comparison it was observed that the two methods can make an approximate prediction of the real values, with Random Forest having the lowest mean square error in all the different types of scenarios and considered regions.

Keywords: Machine Learning. Random Forest. Support Vector Machine.

Lista de ilustrações

Figura 2.1 – Famílias de modelos.	18
Figura 3.1 – Cénarios de acordo com Relatório Especial de Cenários de Emissões (SRES).	22
Figura 3.2 – Cénarios visando caminhos de concentração representativos (RCPs).	22
Figura 3.3 – Previsão Baseada em Florestas Aleatórias.	25
Figura 3.4 – Conjunto de treinamento binário e três diferentes hipóteses.	26
Figura 3.5 – Validação Cruzada divididos em k subgrupos.	27
Figura 4.1 – Correlação entre os Modelos na Região da Amazônia.	30
Figura 4.2 – Correlação entre os Modelos na Região da Bacia do Rio da Prata.	31
Figura 4.3 – Correlação entre os Modelos na Região do Nordeste Brasileiro.	32
Figura 4.4 – Combinação de modelos para a região da Amazônia utilizando Random Forest.	33
Figura 4.5 – Combinação de modelos para a região da Amazônia utilizando Support Vector Machine.	33
Figura 4.6 – Combinação de modelos para a região da Bacia do Rio da Prata utilizando Random Forest.	34
Figura 4.7 – Combinação de modelos para a região da Bacia do Rio da Prata utilizando Support Vector Machine.	34
Figura 4.8 – Combinação de modelos para a região do Nordeste Brasileiro utilizando Random Forest.	35
Figura 4.9 – Combinação de modelos para a região do Nordeste Brasileiro utilizando Support Vector Machine.	35

Lista de tabelas

Tabela 2.1 – Modelos do CMIP5 utilizados neste estudo com a descrição das resoluções horizontais aproximadas	19
Tabela 2.2 – Resumo das principais características dos RCPs utilizados pelo CMIP5(RCP 2.6, RCP 4.5, RCP 6.0, RCP 8.5).	20
Tabela 4.1 – RMSE das regiões do Brasil de acordo com os modelos de previsão e variando os cenários.	36

Sumário

1	INTRODUÇÃO	11
2	REVISÃO BIBLIOGRÁFICA	13
3	MODELAGEM	21
3.1	Random Forest	24
3.2	Suport Vector Machine	25
3.3	Validação Cruzada	26
3.4	Caret	27
4	RESULTADOS	29
4.1	Estatística Descritiva	29
4.2	Combinação de Modelos	30
5	CONSIDERAÇÕES FINAIS	37
A	ANEXO	38
A.1	Código para Análise dos Dados	38
	REFERÊNCIAS	53

1 Introdução

De acordo com o Intergovernmental Panel on Climate Change (IPCC) existem 4 tipos de cenários possíveis para a previsão do clima, A1, A2, B1 e B2. Os cenários A1 e A2 são mais pessimistas. Eles descrevem um mundo futuro com crescimento econômico muito rápido com a utilização cada vez maior de combustíveis fósseis, agravando por consequência o clima. Os cenários B1 e B2 por sua vez acreditam que a sustentabilidade e as soluções ambientais serão aplicadas, sendo assim uma forma mais otimista de prever o clima futuro.

A precipitação é uma das variáveis mais importantes para descrever o clima futuro sendo uma medida contínua não uniforme com todos os valores expressos em milímetros (mm) para um determinado intervalo de tempo. Essa medição é feita por amostragem nas estações meteorológicas, radares ou modelagem. No nosso caso vamos trabalhar com os dados da precipitação observada e com a modelagem de precipitação. Os modelos de precipitação utilizados neste trabalho serão o BCC-CSM1.1, CCSM4, CESM1, CPC-NOAA, NorESM1-ME e MRI-CGCM3. Os dados foram disponibilizados pelo site Program for Climate Model Diagnosis & intercomparison.(PCMDI- <https://pcmdi.llnl.gov/>).

Existem vários modelos de previsão computacional. Uma maneira de aprimorar essa previsão é combinando estes modelos. A combinação de modelos de previsão climática é útil quando não se tem certeza qual método é mais preciso, quando se quer evitar grandes erros e também quando não se tem certeza sobre a situação, segundo ARMSTRONG, J. S. (2001). Os métodos que irão ser utilizados para combinar modelos de previsão climática nesse trabalho serão o *Random Forest* e *SVM (Support Vector Machine - Máquina de Vetores Suporte)*. Estes modelos se ajustam aos dados, reconhecendo padrões, criando assim uma combinação capaz de realizar melhores previsões. Por fim, tendo os resultados a respeito da melhor combinação de modelos, podemos ter uma previsão climática mais precisa

O presente trabalho procurou identificar qual método de previsão melhor consegue prever a precipitação nas regiões da Amazônia, Bacia do Rio da Prata e Nordeste Brasileiro utilizando a combinação de modelos em três tipos de cenários diferentes: RCP 4.5, RCP 6.0 e RCP 8.5. Utilizando o pacote *caret* para conseguir prever a precipitação e dois métodos específicos dentro do pacote *caret* que foram o *Random Forest* e *Support Vector Machine*.

Este trabalho está dividido da seguinte maneira: no capítulo 2 apresentamos a forma para aprimorar modelos usando a combinação de modelos, métodos para calcular o intervalo de incerteza e uma medida de confiabilidade das mudanças climáticas e para classificar e para medir se os modelos de previsão estavam conseguindo prever, além

de explicar que todos os modelos possuem o mesmo descendente. No capítulo 3, apresentamos os métodos que foram utilizados para fazer a previsão dos modelos e explicar melhor sobre o pacote utilizado e explicar como são classificados os cenários utilizados. No capítulo 4 mostramos os resultados das simulações para cada método de previsão nas regiões selecionadas para cada cenário e a correlação entre modelos. Por fim, no capítulo 5, são feitas as análises finais para definir qual método conseguiu melhor prever os dados.

2 Revisão Bibliográfica

Uma maneira de aprimorar modelos de previsão é a combinação entre modelos independentes. Analisar diferentes dados e utilizar diferentes métodos podem gerar esses modelos independente para assim combinados. A combinação de previsões climáticas é útil quando não se tem certeza sobre qual método é mais preciso, quando se quer evitar grandes erros e também quando não se tem certeza sobre a situação de qual modelo é mais preciso. O levantamento realizado revelou três procedimentos a respeito da combinação de modelos de previsão:

- Usar diferentes métodos, pois as previsões são melhoradas se feitas de forma independente.
- Quando viável ao pesquisador é aconselhável utilizar pelo menos 5 previsões, porém com o aumento na quantidade de métodos o aperfeiçoamento da combinação será reduzido.
- Manuseando pesos diferentes para cada previsão, porém deve-se evitar pesos de julgamento nos casos em que os pesquisadores que fazem a ponderação não têm informações sobre a precisão relativa de fontes alternativas de previsões nesta situação uma solução são esquemas de ponderação mecânica que podem ajudar a proteger contra vieses. Se não houver evidências de que um determinado modelo é melhor para as previsões do que o outro, deve se adotar ponderações iguais para todos os modelos.

Conclui-se que a combinação de modelos de previsões são mais precisas do que todas as previsões estudadas, superando até o melhor modelo, ou seja, a combinação nunca será menos precisa do que a previsão independente Armstrong (2001).

O *Reliability Ensemble Averaging* (REA) é um método para calcular o intervalo de incerteza e uma medida de confiabilidade das mudanças climáticas. O critério de confiabilidade é o desempenho do modelo em reproduzir o clima atual e as mudanças simuladas através do modelo. Esse método leva em consideração dois critérios de confiabilidade:

- O desempenho do modelo na reprodução do clima atual.
- Convergência das mudanças simuladas entre os modelos .

Esse método é aplicado para medir a média de temperatura e precipitação de mudanças sazonais para o século XXI, ou seja mudanças que ocorrem sempre em um

determinado intervalo de tempo no século XXI. Levando em consideração a diferença dos cenários A1 e A2 e a temperatura média do clima prevista para o clima atual, separadas nos meses de verão (dezembro, janeiro e fevereiro) e nos meses de inverno (junho, julho e agosto). Nesse estudo, os autores Giorgi F. e Mearns (2002) sugeriram que as informações coletivas são mais confiáveis do que qualquer modelo individual, reforçando o que foi comprovado por Armstrong (2001).

Foram aplicados ao método REA 22 regiões de escala subcontinental que foram simuladas por 9 modelos de previsão (*AOGCMs-Atmosphere-Ocean Global Circulation Models*) em dois cenários diferentes (A1 e A2). O objetivo dessa aplicação era minimizar o peso das simulações que apresentam um desempenho inferior na representação do clima atual. Como os dados não seguem uma distribuição probabilística preferiu se empregar um método quantitativo, para não fazer suposições sobre a distribuição de fatores. Os critérios aplicados foram para cada região separadamente, pois os modelos podem se ajustar a algumas regiões e a outras não. Nessa análise não obtiveram modelos com um desempenho inferior ou superior aos demais modelos, tanto em modelos para temperatura tanto quanto em modelos para precipitação. Porém o intervalo de incerteza usando o REA foi menor que o intervalo usando o desvio médio quadrado, devido minimização dos *outliers*, que eram insatisfatórios para o desempenho do modelo.

Caracterizar e quantificar a incerteza nas projeções de mudanças climáticas é de importância fundamental para abordagens estratégicas de adaptação e mitigação. A incerteza na mudança climática futura deriva de três fontes principais:

- Forçar a incerteza, ou seja, fatores externos que não se tem total domínio sobre eles que influenciam no clima, como emissões futuras de gases de efeito estufa (GEE), concentrações de ozônio estratosférico etc.
- Resposta do modelo, pois diferentes modelos podem produzir respostas diferentes para as mesmas forças externas.
- Variabilidade Interna é a variabilidade natural do sistema climático que ocorre na ausência da força externa e inclui processos intrínsecos à atmosfera, ao oceano e ao sistema acoplado oceano-atmosfera.

A variabilidade atmosférica interna também pode ser denominada como “ruído climático” surge a partir de processos dinâmicos não-lineares intrínsecos à atmosfera. Embora a atmosfera contenha pouca memória além de algumas semanas, ela exhibe uma variabilidade de escala de longo tempo característica de um processo estocástico aleatório.

O CMIP3 oferece uma estimativa para a incerteza na mudança climática, que foi avaliada para um número de variáveis climáticas, incluindo a temperatura do ar, a precipitação e a circulação atmosférica em larga escala. Essas incertezas, bem como aquelas

baseadas em integrações de controle de modelo longo, também foram usadas para estimar a contribuição da força externa às mudanças climáticas observadas ao longo do século XX. Foram analisados pelo Deser C. (2012) um novo conjunto de 40 membros para o período 2000-2060 realizado com o Modelo de Sistema Comunitário de Clima Versão 3 (CCSM3) pertencente ao CMIP3 no cenário “A1B” com três parâmetros básicos, temperatura do ar da superfície, precipitação e pressão do nível do mar, para uma visão ampla da resposta climática. Destacando as diferenças entre o inverno e o verão. No artigo foi utilizado dois métodos para calcular a resposta climática, os dois métodos produzem resultados virtualmente similares.

- diferenças de época entre os últimos 10 anos (2051–2060) e os primeiros 10 anos (2005–2014).
- tendências lineares dos mínimos quadrados ajustadas ao período 2005–2060.

O estudo de Deser C. (2012) concluiu que o agrupamento CCSM3 de 40 membros é caracterizada por uma precipitação diminuindo dentro dos trópicos e subtropicais, mais aquecimento sobre a terra do que sobre o oceano, e mais forte aquecimento no Ártico e continentes adjacentes de alta latitude no inverno. um padrão geral de diminuição da pressão do nível do mar em altas latitudes e aumento da pressão do nível do mar em latitudes médias, exceto para o verão.

Para fazer combinações de projeções de clima futuro para diferentes modelos climáticos, Chandler (2015) diz que é necessário uma análise da estrutura estatística. Tal estrutura leva em consideração que os modelos disponíveis são imperfeitos e não abrangem todas as possíveis decisões na modelagem climática. Reconhecendo seus pontos fortes e fracos. As informações de modelos individuais são automaticamente ponderadas, juntamente com as observações históricas e do conhecimento prévio. Os pesos para um simulador dependem da sua variabilidade interna, do seu consenso esperado com outros modelos, da variabilidade interna do clima real e da propensão de os simuladores coletivamente se desviarem da realidade.

Projeções de clima futuro são geralmente derivadas de modelos climático. Apesar dos modelos apresentarem uma melhor compreensão disponível hoje do sistema climático, elas variam entre elas na medida em que a escolha do modelo é às vezes a fonte dominante de incerteza em tais projeções. O usuário que for aplicar as projeções de mudanças climáticas precisa considerar informações de um conjunto de modelos múltiplos. Porém a interpretação de um conjunto de modelos múltiplos não é simples pois diferentes esquemas de ponderação, cada um baseado em critérios plausíveis, podem produzir resultados diferentes e sem a justificação de qual ponderação deve se escolher, o que deveria ajudar acaba atrapalhando pois agora se tornaram uma fonte adicional de incerteza Chandler (2015).

Para quantificar a incerteza probabilisticamente, é necessário usar técnicas estatísticas modernas para derivar funções de densidade de probabilidade para quantidades de interesse. Essa abordagem requer a especificação de um modelo estatístico para representar a estrutura do conjunto e os dados gerados. Usa-se uma estrutura Bayesiana e, em alguns casos, o modelo estatístico implica em um conjunto de pesos a serem anexados aos simuladores individuais. Outros pesquisadores já tentaram encontrar o modelo estatístico associado aos pesos da REA. No entanto, dificuldades permanecem com a interpretação das probabilidades mesmo dessas abordagens sofisticadas Chandler (2015).

Cada modelo climático criado até hoje foi evoluído e melhorado, porém não são independentes uns dos outros e independentes da geração anterior, como uma árvore genealógica. Uma das razões para os modelos serem parecidos é porque todos compartilham o código comum, como mostrado na Figura 2.1. Outra explicação para a semelhança de modelos sucessivos em uma instituição pode ser que diferentes centros se preocupem com diferentes aspectos do clima e usem diferentes conjuntos de dados e métricas para avaliar a qualidade do modelo durante o desenvolvimento. A confiança nas projeções de modelo não se dá devido a grande quantidade de dados e sim pelos os aspectos da mudança climática que entendemos e que podem ser ligados a processos físicos conhecidos e a modelos e conceitos mais simples. A nova geração de modelos climáticos globais no CMIP5 apoia a ideia de um processo evolutivo bastante gradual, através do qual os modelos melhoram com o tempo. O CMIP5 parece ser um melhor CMIP3 em vez de um conjunto radicalmente novo, também em sua resposta às mudanças climáticas. Apesar da complexidade de muitos dos modelos do CMIP5, é talvez uma indicação de que essa estratégia de desenvolvimento de usar melhorias no poder computacional para adicionar complexidade é bem-sucedida na melhor representação do atual sistema climático Knutti (2013).

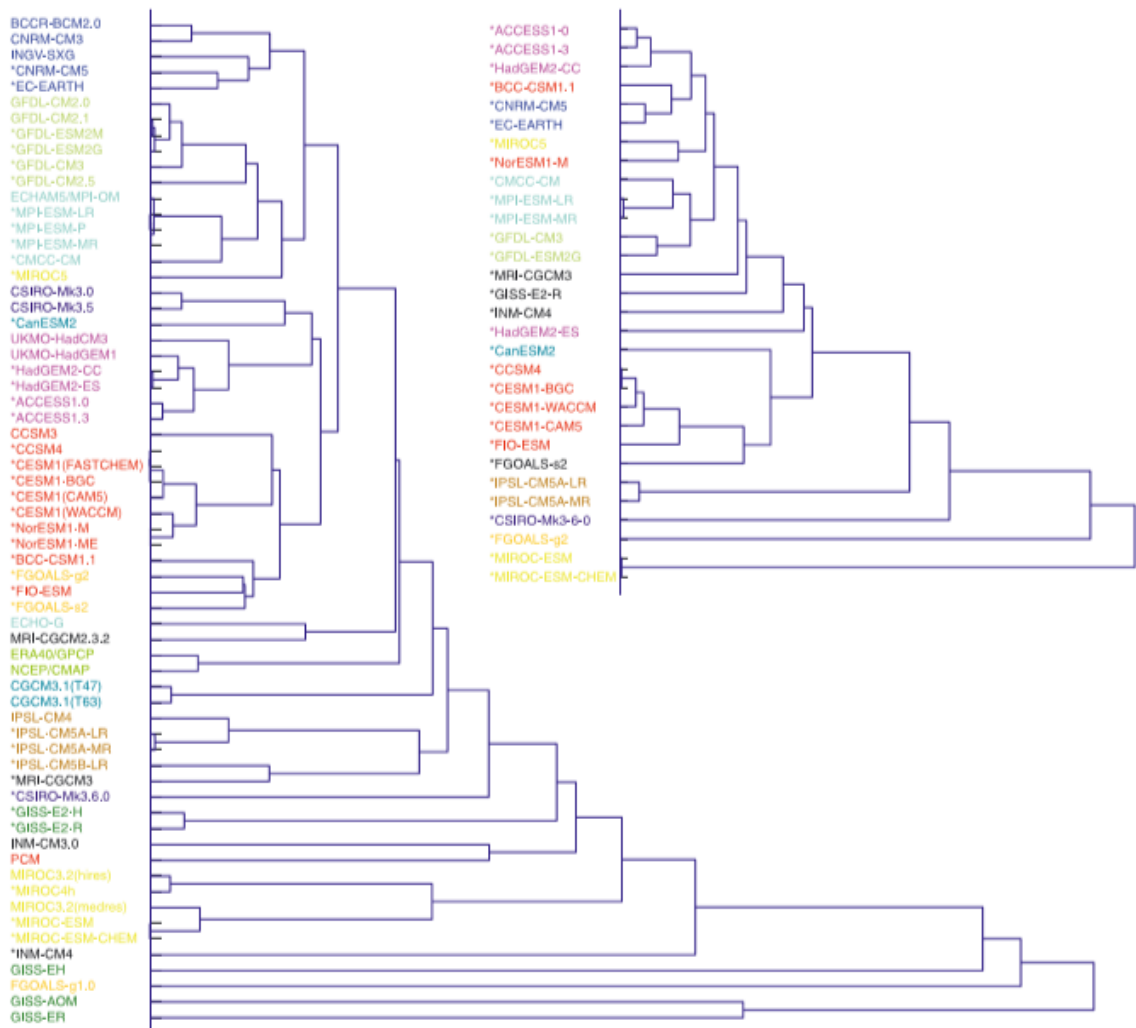
A modelagem computacional é um instrumento utilizado para a melhor compreensão sobre o clima global e por assim conseguindo prever as suas consequências. É necessário conseguir também prever as incertezas do modelo em calcular projeções.

Santos (2015) utilizou técnicas de *dowscaling* para diminuir as incertezas com base nos modelos CMIP5 utilizando regressão linear múltipla por componentes principais e Redes Neurais como aprimoramento dessas incertezas visto que ele por si só já tira as incertezas do modelo. As características dos cenários utilizados estão na Tabela 2.2 pertencentes ao *Special Report on Emission Scenarios*, no trabalho as projeções de cenários utilizadas foram o RCP 2.6 e 8.5 e as simulações referentes ao cenário histórico avaliadas nas sub-regiões Amazônia, Nordeste do Brasil e da Bacia do Prata.

Nas análises por Santos (2015) foram analisadas pelo REA, que aborda a mudança esperada em um variável baseando em uma combinação de projeções de vários modelos. Com os critérios de confiabilidade do REA medir o desempenho do modelo climático em ser preciso na previsão do clima e estimar a probabilidade das mudanças climáticas futuras

projetas pelo conjunto de modelos climáticos globais. Outra técnica utilizada para medir se os modelos de previsão estavam conseguindo prever foram as redes neurais artificiais (RNA) sendo um sistema baseado na operação de uma rede neural biológica que tem a habilidade para armazenar conhecimento e torná-lo disponível para o uso. As redes neurais artificiais fundamentam-se nos estudos sobre a estrutura do cérebro humano para tentar emular sua forma inteligente de processar informações, ou seja, uma rede neural é capaz de produzir saídas adequadas para entradas que não estavam presentes durante o treinamento. Algumas características dos RNAs são importantes como a não-linearidade, mapeamento de entrada-saída, adaptabilidade, resposta a evidências e informação contextual. Após comparar as duas técnicas, Santos (2015) conclui que foi possível ver que houve uma queda na média das projeções de mudança de precipitação e temperatura ao comparar as simulações pelo método REA com a média aritmética simples dos modelos de circulação geral, ou seja, As RNAs podem ser usadas para ajudar a compreensão das mudanças climáticas, tornando as projeções mais confiáveis, notou-se também que para os cenários RCP 2.6 e 8.5 existiram um padrão bastante similar, variando entre 1 a 5 mm de precipitação.

Figura 2.1 – Famílias de modelos.



Fonte: Knutti R. (2013).

Tabela 2.1 – Modelos do CMIP5 utilizados neste estudo com a descrição das resoluções horizontais aproximadas

Modelo	Instituição	Resolução	Referência Principal
ACCESS1.0	Commonwealth scientific and industrial Research Organization/Bureau of Meteorology	1.3° x 1.9°	Bi et al(2003)
BCC-CSM1-1	Beijing Climate Center	2.8° x 2.8°	Xiao-Ge, Tong-Wen e Jie (2013)
CCSM4	National Center for Atmospheric Research	1.25° x 0.94°	Graig Vetenstein e Jacob (2012)
CRNM-CM5	Centre National de Recherches Meteorologiques / Centre Europeen de Recherche et Formation Avancees en Calcul Scientifique (CNRM/CERFACS)	1.4° x 1.4°	Voltaire et al. (2013)
CSIRO-Mk3.6.0	Commonwealth, Scientific and Industrial Research Organisation in collaboration with the Queensland Climate Change Centre of Excellence	1.875° x 1.875°	Collier et al. (2011)
HadGEM2-ES	Met Office Hadley Centre	1.875° x 1.25°	Collins et al. (2011)
INMCM4	Institute for Numerical Mathematics	1.5° x 2.0°	Volodin, Dianskii e Gusev(2010)
MPI-ESM-LR	Max Planck Institute for Meteorology	1.8° x 1.8°	Giorgetta et al.(2013)
MRI-CGCM3	meteorological Research Institute	1.1° x 1.1°	Yukimoto et al. (2012)
NorESM1	Norwegian Climate Centre (NCC)	1.9° x 2.5°	Bentsen et al. (2013)

Fonte: Santos T. S. (2015).

Tabela 2.2 – Resumo das principais características dos RCPs utilizados pelo CMIP5(RCP 2.6, RCP 4.5, RCP 6.0, RCP 8.5).

Variáveis	RCP 8.5	RCP 6.0	RCP 4.5	RCP 2.6
Forçante Radiativa ($E = \text{wm}^{-2}$)	>8.5 em 2100 e continua subindo	6 em 2100, com uma estabilização após 200	4.5 em 2100, com uma estabilização após 2100	Um pico de 3 antes de 2100, e após o pico declina
Concentração de CO_2 equivalente (ppm)	>1370 em 2100	850 com uma estabilização após 2100	650 estabilizando após 2100	Pico em 490, antes de 2100 e declina
Poluição do ar	Média - alta	Média	Média	Média - Baixa
Emissão de gases do efeito estufa	Alta	Média à alta (Mitigação)	Média à baixa (Mitigação)	Muito baixa
Crescimento Populacional	Alto	Médio	Baixo	Média - Pasto e cultivo
Área Agrícola	Média - pasto e cultivo	Média - cultivo e muito baixa para pasto(total muito baixo)	Baixa - pasto e cultivo	

Fonte: Jones e Carvalho (2013), Vuuren et al. (2011).

3 Modelagem

O Painel Intergovernamental de Mudanças Climáticas (IPCC) elaborou um Relatório Especial de Cenários de Emissões (SRES), para melhor conseguir fazer uma previsão climática, aplicando os impulsos humanos, mudanças naturais e crescimento demográfico ao longo do tempo nos cenários projetados. Cenários são expectativas alternativas de como o futuro poderá se tornar, existindo uma possibilidade de que qualquer cenário possa ocorrer.

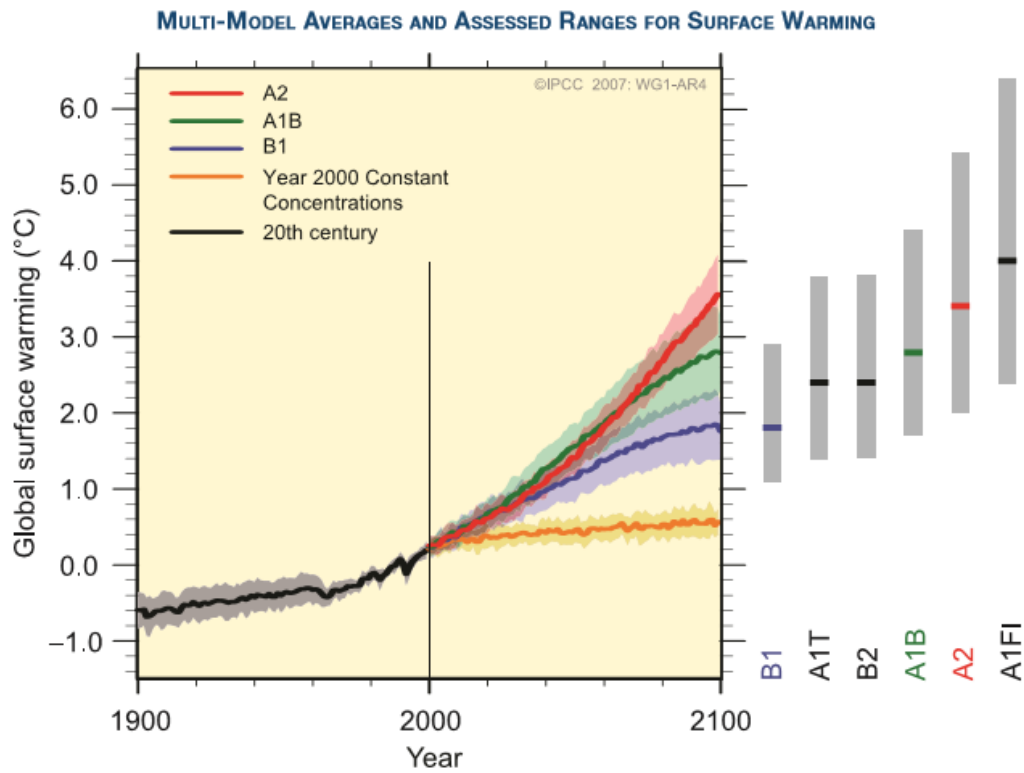
De acordo com ALLEY, R. B. et al.,(2007), pode-se descrever a história da família de cenário A1 com um crescimento econômico acelerado, onde a população global consegue atingir o topo e declina logo em seguida. Existem subgrupos no cenário A1. Os subgrupos são o A1FI, AIT e A1B8, que são diferenciados apenas pela afetação de cada intensidade da tecnologia no clima. A família A2 tem no seu conceito um mundo heterogêneo, onde a autosuficiência nacional e a preservação de identidade são os fundamentos essenciais. Tem o mesmo conceito da família A1, porém a família A1 é um cenário global, a família A2 tem um cenário mais regional.

A história do cenário B1 relata o início parecido com o A1, porém com rápidas mudanças na estrutura econômica em direção a sustentabilidade econômica, social e ambiental. No cenário B2 descreve um mundo onde a ênfase está em soluções locais. Na Figura 3.1 pode se ver mais claramente o processo de cada cenário em forma de gráfico. Em 2014 um novo conceito foi utilizado para substituir o SRES. Os caminhos de concentração representativos (RCPs) possui o mesmo fundamento dos cenários com a diferença de que suas trajetórias são realizadas pela concentração de gases do efeito estufa, diferente do seu antecessor que realizava apenas a trajetórias dada a emissão de gases, como mostrado na Figura 3.2. No novo relatório são calculados com base na capacidade de dissipar calor em cada um dos cenários.

Os novos cenários possuem uma escala de projeção variando de 2.6 a 8.5. A figura 3.2 descreve o cenário destas projeções. A projeção RCP2.6 assumiria o papel no cenário B1, utilizando de projeções mais otimistas, atingindo um nível de radiação elevado até 2050 e reduzindo antes de 2100. Na projeção RCP4.5 o cenário de estabilização é previsto assim como o RCP2.6 antes de 2100, tendo uma diferença pequena de tempo. O cenário RCP6.0 força a radiação após 2100 sem estratégias claras para reduzir o efeito estufa. O RCP8.5 é o cenário mais pessimista, se caracterizando pelo aumento da emissão de gases de efeito estufa e em uma maior concentração.

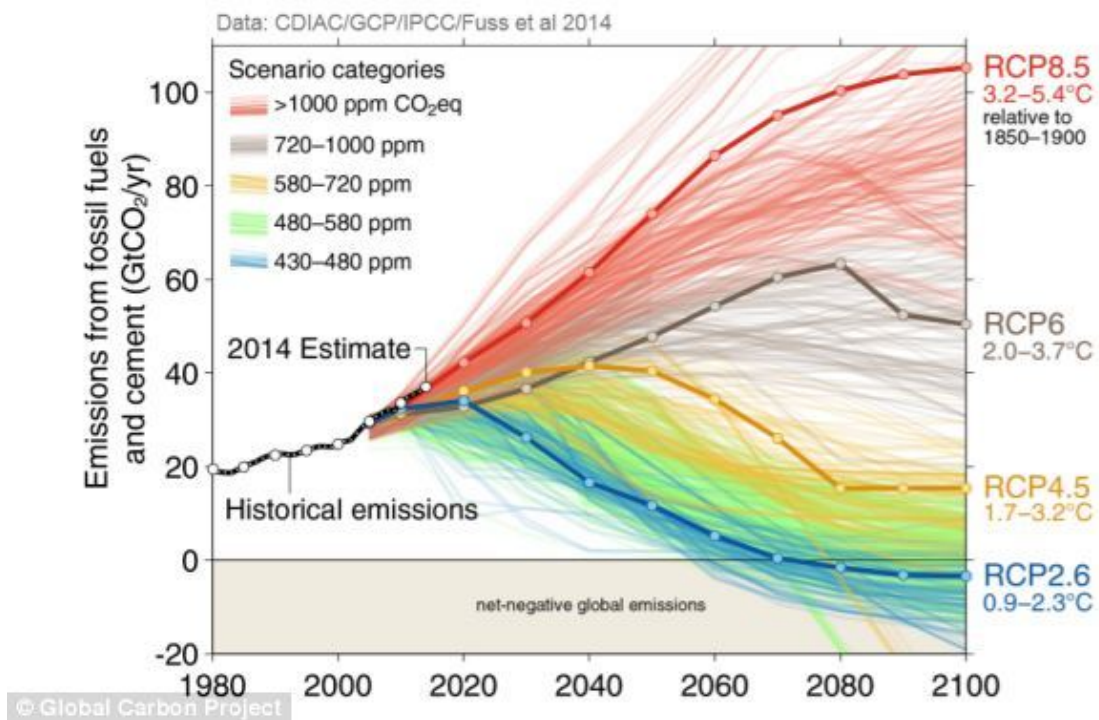
Porém os cenários não influenciaram na hora de coletar os dados. Os dados foram coletados no *site program for climate Model Diagnosis intercomparison* (<https://esgf->

Figura 3.1 – Cénários de acordo com Relatório Especial de Cénários de Emissões (SRES).



Fonte: Disponível em : <https://www.ipcc.ch/publications_and_data.html>

Figura 3.2 – Cénários visando caminhos de concentração representativos (RCPs).



Fonte: Fuss et al. (2014)

node.llnl.gov/projects/cmip5/). Neste site pode-se escolher o modelo e baixar os dados simulados referentes a ele. Pode-se ainda escolher qual tipo de variável se deseja trabalhar, precipitação ou temperatura, em nosso caso escolhemos a precipitação. (Santos, T.S. , Mendes, D e Torres R. R. (2015)) Obtivemos algumas dificuldades na hora que baixamos esses dados. Um dos problemas foi a diferença de resolução espacial de cada modelo. Para obter a combinação dos modelos, era necessário colocar em uma mesma resolução. Outro problema encontrado foi que os dados vieram com observações de todo o mundo. Para o nosso estudo só foi necessário utilizar as sub-regiões Amazônia, o Nordeste do Brasileiro e a Bacia do Prata. Os modelos selecionados foram sugeridos por especialistas da área de ciências climáticas. Estes foram os selecionados:

- BCC-CSM1.1: Significa *Beijing Climate Center Climate System Model*, um modelo desenvolvido no Centro Climático de Pequim . O BCC-CSM1.1 faz parte dos modelos do CMIP5. É um modelo climático global que inclui uma vegetação diversificada e ciclo de carbono global. Os dados do modelo simulados foram gerados trocando a atmosfera e o oceano uma vez por dia simulado e a troca de carbono atmosférico com a biosfera terrestre é calculada em cada etapa de tempo do modelo (20 min) (XIN Y.-J. et al., 2012).
- CCSM4: O *Community Climate System Model* ou Modelo de Sistema Climático Comunitário foi recentemente modifica para melhor da sua versão anterior (CCSM3). No novo modelo foi melhorado os desenvolvimentos de todos os componentes do CCSM. Usa-se no modelo a atmosfera padrão, do modelo anterior para o atual o que não mudaram foram os vieses da temperatura da superfície do mar nas principais regiões de ressurgência. Com tais mudanças a variabilidade da El Ninó se torna muito mais realista, embora a amplitude seja muito grande em comparação com as observações (GENT et al., 2011).
- CPC-NOAA: O *Climate Prediction Center at the National Centers for Environmental Prediction*, NOAA, é um conjunto de dados que foi modelado nos anos de 1963 a 1993. Os dados são análises das variações diurnas, diárias, mensais-sazonais e interanuais na precipitação Higgins (1996). Os dados podem ser baixados diretamente do site (<http://www.cpc.ncep.noaa.gov/products/JAWFMonitoring/Brazil/index.shtml>).
- NorESM1-M: O *Norwegian Earth System Model*, ou o Sistema de Modelo Norueguês. foi aplicada com resolução espacial média para fornecer resultados para o CMIP5. a diferença desse modelo é uma resolução horizontal de aproximadamente 2 para a atmosfera e componentes terrestres e 1 para os componentes oceânicos e de gelo, tendo uma sensibilidade climática de equilíbrio menor entre os modelos que contribuem para o CMIP5 (IVERSEN et al., 2012).

- MRI-CGCM3: Foi desenvolvido no Instituto de Pesquisa Meteorológica (MRI). Este modelo é uma atualização geral do antigo modelo climático MRI-CGCM2 da MRI. É composta de modelos de atmosfera-terra, aerossol e gelo oceânico. O modelo reproduz o clima médio geral, incluindo a variação sazonal em vários aspectos da atmosfera e dos oceanos. A variabilidade no clima simulado também é avaliada e é considerada realista, incluindo o El Niño e a Oscilação Sul e as oscilações do Ártico e Antártico. O mesmo exibe um comportamento estável sem desvios climáticos, pelo menos no balanço de radiação, na temperatura próxima à superfície e nos principais índices de circulação oceânica (YUKIMOTO et al., 2012).

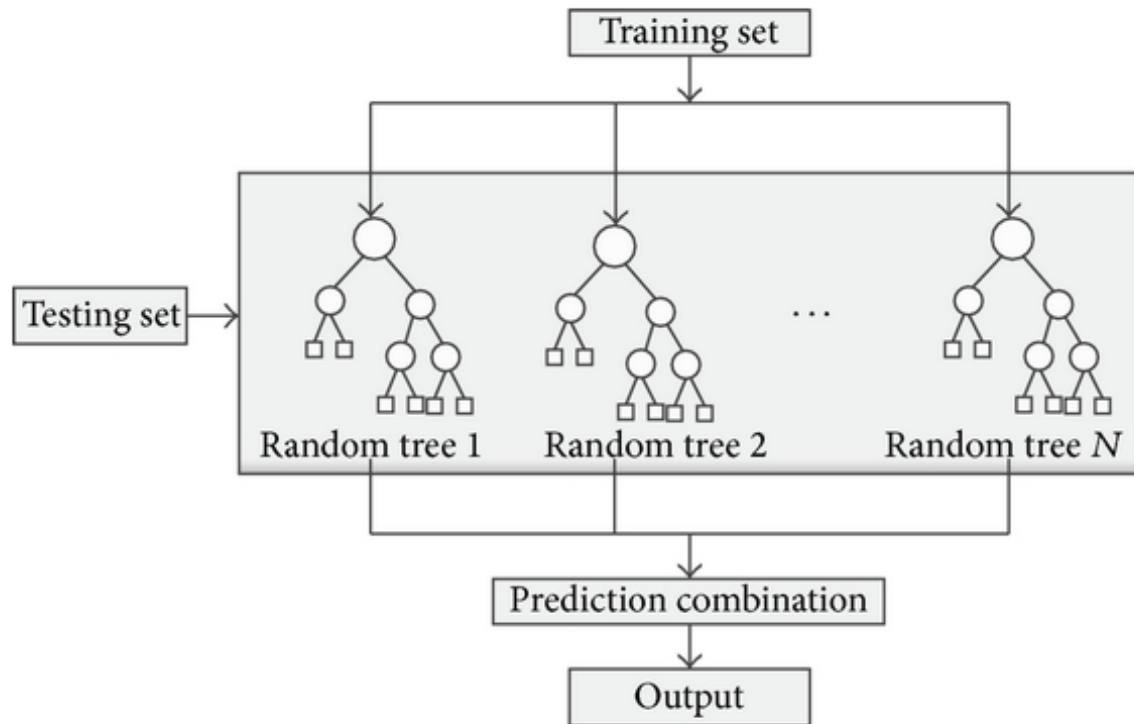
Com os dados preparados, os métodos utilizados para fazer a combinação entre os 6 modelos foram o Random Forest e SVM . O Random Forest são uma maneira de ordenar várias árvores de decisão descorrelacionadas, treinadas em várias partes do mesmo conjunto de treinamento, para no final reduzir a correlação entre as variáveis. Este método também nos permite encontrar a importância de cada variável nos modelos acima. O SVM é um classificador, que reconhece padrões, classifica-os e realiza a análise de regressão. Os métodos foram rodados dentro do software R, que possui um pacote chamado ‘caret’. Este pacote possui a função `traincontrol`, responsável pelo treinamento do modelo, utilizamos o método de validação cruzada para escolher o melhor modelo para os dados. Após realizarmos o treinamento, verificamos através da função `test`, se ele não estava superajustado.

3.1 Random Forest

O Random Forest é um algoritmo de árvores de classificação $h(x, \Theta_k)$, $k=1, \dots, \Theta_k$ é um vetor aleatório gerado, independente do vetor aleatório passado $, \Theta_1, \dots, \Theta_{k-1}$, mas com a mesma distribuição para cada k e x é o valor de entrada (BREIMAN, 2001).

Cada árvore de decisão usa um vetor aleatório que é gerado a partir de alguma distribuição de probabilidade fixa. Um vetor aleatório pode ser incorporado no processo de desenvolvimento da árvore de diversas maneiras. O primeiro passo é criar vetores aleatórios, ou seja, selecionar aleatoriamente W características de entrada para dividir em cada nó da árvore de decisão. Os dados de entrada para o desenvolvimento da técnica são selecionados de forma aleatória utilizando a técnica de *bagging* para gerar amostras de *bootstrap*. A força e a correlação de florestas aleatórias podem depender do tamanho de W . Se W for pequeno o suficiente, as árvores serão menos correlacionadas. E quanto maior o número de características melhor será o classificador. Após usar esses vetores randômicos para construir as árvores de decisão o próximo passo é combinar todos. Por esse motivo as florestas aleatórias são excelentes preditores (TAN et al., 2006).

Figura 3.3 – Previsão Baseada em Florestas Aleatórias.



Fonte: Zhang Fan Min (2014)

De forma prática as árvores de decisão geram regras de classificação do tipo SE e ENTÃO, ou seja, em cada nó, a variável que mais se discrimina entre as respostas e cada nó bifurca a árvore em dois ramos.

Há uma hierarquia: primeiro se utiliza a variável mais discriminante; em seguida, em cada ramo, escolhe-se, para cada novo nó, outra variável mais discriminante e assim é a construção das árvores.

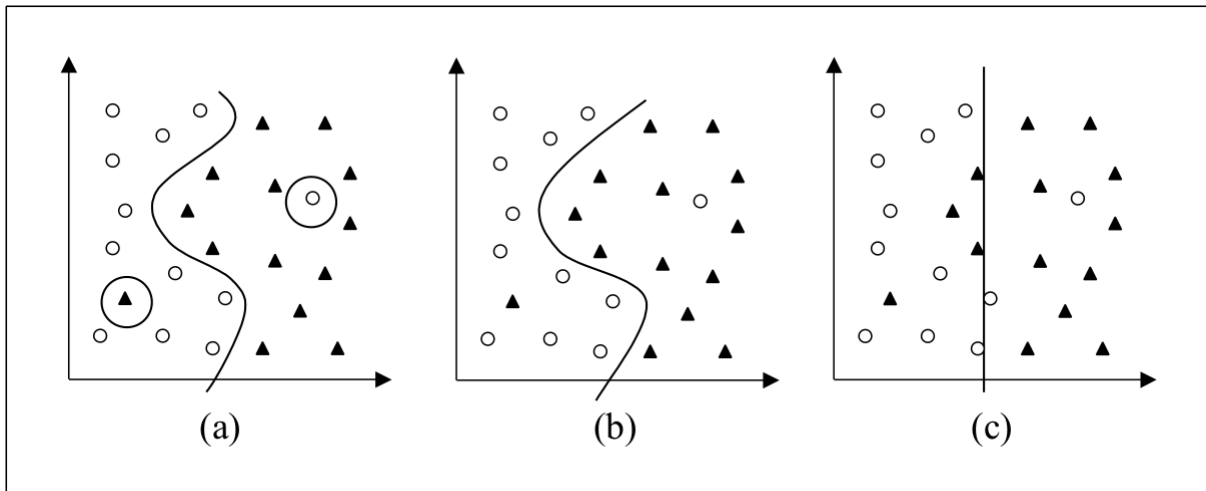
3.2 Suport Vector Machine

Igual o Random Forest, o SVM é supervisionado, que utilizam dados de treinamento para melhor ajustar os dados e assim aplicar a todo o conjunto de observações. Um requisito importante para as técnicas de machine learning é que elas sejam capazes de lidar com os ruídos e minimizar a influência de *outliers* no processo de indução.

Seja f um classificador e F o conjunto de todos os classificadores que um determinado algoritmo de machine learning pode gerar. O algoritmo, durante o processo de aprendizado, utiliza um conjunto de treinamento T , composto de n pares (x_i, y_i) , para gerar um classificador particular $\hat{f} \in F$.

Na Figura 3.4 vemos alguns classificadores para o conjunto de treinamento. Pode-se

Figura 3.4 – Conjunto de treinamento binário e três diferentes hipóteses.



Fonte: Lorena (2007)

observar que no gráfico A e no gráfico B se obteve uma melhor separação dos dados, no gráfico A existe um super ajustamento dos dados, onde ele separa cada observação, esse super ajuste da para observar pelas observações selecionadas em círculo e pela curva do ajustamento esta delimitando de forma mais acentuada. Porém o gráfico B é o melhor classificador, pois na inserção de novos dados ele melhor que ajusta por ter uma variabilidade maior.

O SVM e faz essa classificação através de hiperplanos de margem máxima. Os hiperplanos são os limites de decisão, base no conjunto de treinamento, com um conjunto L tal que

$$L = X : f(x) = W'X - b = 0.$$

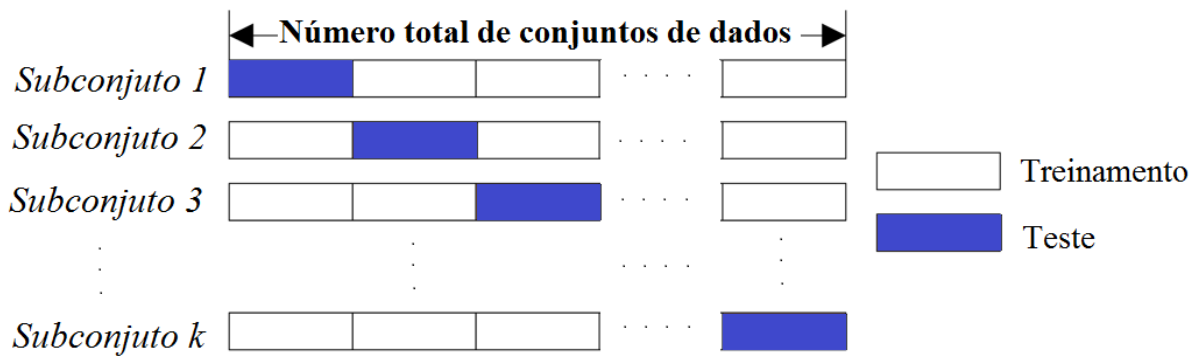
Para quaisquer dois pontos X_1 e X_2 em L , em que W deve ser ortogonal à superfície L .

$$W'(X_1 - X_2) = 0.$$

3.3 Validação Cruzada

A validação cruzada também é um processo de aprendizado de máquina supervisionado. Na Validação Cruzada usa a abordagem de que cada registro dos dados é utilizado o mesmo número de vezes para treinamento e uma vez para teste.

O algoritmo funciona da seguinte maneira divide o número de dados em k subconjuntos de tamanho iguais. Após a separação escolhe-se uma partição para treinamento e os outros para teste. Após essa escolha, troca-se os papéis dos subconjuntos de modo que a partição que era de treinamento se torne o de teste. Este procedimento é repetido k vezes de modo que cada partição seja usada como teste exatamente uma vez, como demonstrado

Figura 3.5 – Validação Cruzada divididos em k subgrupos.

Fonte: Autorial Própria

na Figura 3.5. O erro total é encontrado pela média dos resultados obtidos em cada etapas até k . Criando assim uma estimativa da qualidade do modelo de conhecimento gerado e permitindo análises estatísticas.

3.4 Caret

Caret é uma abreviação para *classification and regression training*, classificação e treinamento de regressão. O pacote em si foi criado para eliminar diferenças sintáticas entre muitas das funções para construir e prever modelos para desenvolver um conjunto de abordagens quase automatizadas e razoáveis para otimizar os valores dos parâmetros de ajuste para muitos desses modelos e criar um pacote que possa ser facilmente estendido para sistemas de processamento paralelo (KUHN, 2008).

O pacote em si tem varias funções vantajosas tanto para o processo de análise dos dados, pré-processamento, treinamento, seleção e validação dos modelos, além de criar uma área compartilhada para a utilização de machine learning, podendo assim trabalhar com mais de um tipo de machine learning.

A função `traincontrol` gera parâmetros que controlam ainda mais como os modelos são criados, com valores possíveis. Podendo escolher os principais métodos entre:

- “boot” que refere ao método bootstrap;
- “cv” ou mais cross validation que foi explicado na Seção 3.3 mais detalhadamente;
- “LOOCV” significa Leave-one-subject-out cross validation;
- “LGOCV” é o leave-group-out cross validation;
- “none” esse método é melhor aproveitado quando só se encaixa um modelo para todo o conjunto de treinamento;

- “oob” é utilizado principalmente para floresta aleatória, análise discriminante flexível ensacada ou modelos de florestas de árvores condicionais;

A função *train* pode ser usada para selecionar valores dos parâmetros de ajuste do modelo e estimar o desempenho do modelo usando a reamostragem podendo ser bootstrapping, *k*-fold cross-validation, leave-one-out cross-validation e leave-group-out cross-validation. Neste trabalho, os métodos utilizados para o treinamento do modelo Random Forest e “svmRadial” Support Vector Machine.

4 Resultados

Neste capítulo vamos demonstrar a aplicação dos métodos que foram descritos nos Capítulos 3.1 e 3.2 e também avaliar descritivamente os modelos selecionados para a pesquisa.

4.1 Estatística Descritiva

A correlação é uma medida que mede a dependência entre variáveis. Essa classificação tem as intensidades variando de fraca, moderada ou forte. O que determina esse nível é o valor da correlação, ρ , e que varia no intervalo $-1 \leq \rho \leq 1$. Quanto mais próxima de 1, ou -1, mais forte é essa correlação. Quanto mais próxima de zero o seu nível, menor a correlação. Ou seja, quanto mais próximo o valor da correlação estiver das extremidades (-1 ou 1), mais correlacionadas estão as variáveis estudadas.

Usamos a correlação para avaliar a dependência entre os modelos escolhidos. Pode-se observar da Figura 4.1 até a Figura 4.3 que existe uma correlação forte entre os modelos. Os modelos possuem correlações fortes, principalmente nas regiões da Bacia do Rio da Prata e do Nordeste Brasileiro, onde todas as correlações foram acima de 0,6. Um dos motivos que possa explicar esse fenômeno é que todos os modelos possuem um ancestral em comum, como foi mostrado na Figura 2.1.

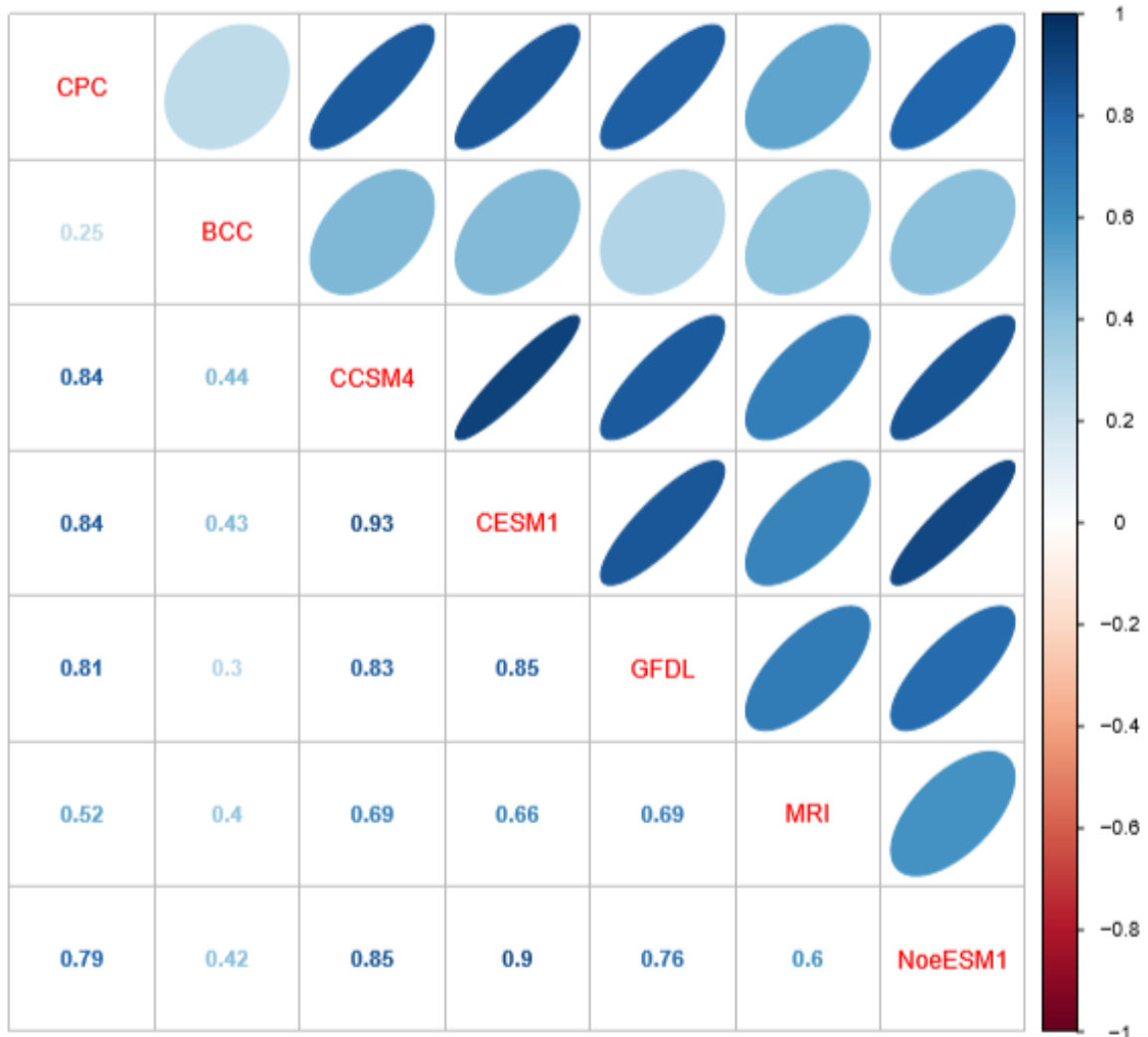
Na Figura 4.1, podemos concluir que o modelo BCC foi o modelo que menos teve uma correlação forte, com todas as correlações com os outros modelos abaixo de 0,5 o segundo modelo menos correlacionado foi o modelo MRI. Os modelos na região da Amazônia obtiveram um resultado inferior comparado com os das outras regiões.

Diferentemente da Figura 4.1, a Figura 4.2, que retrata a região da Bacia do Rio da Prata teve todos os modelos com correlação forte, todos acima de 0,8. Ou seja, todos os modelos nessa região são dependentes um do outro.

A região do Nordeste Brasileiro tem uma semelhança com a região da Bacia do Rio da Prata em relação aos modelos serem correlacionados, como se pode observar na Figura 4.3. O modelo que obteve uma correlação maior com os outros foi o BCC, sendo o oposto do que aconteceu na correlação da região da Amazônia.

Os modelos que são mais correlacionados entre si nas regiões escolhidas são os modelos CCSM4 e o modelo CESM1. Em todas as regiões a sua correlação foi acima de 0,9. As correlações dos modelos na região da Amazônia variaram de correlações moderadas a correlações fortes.

Figura 4.1 – Correlação entre os Modelos na Região da Amazônia.



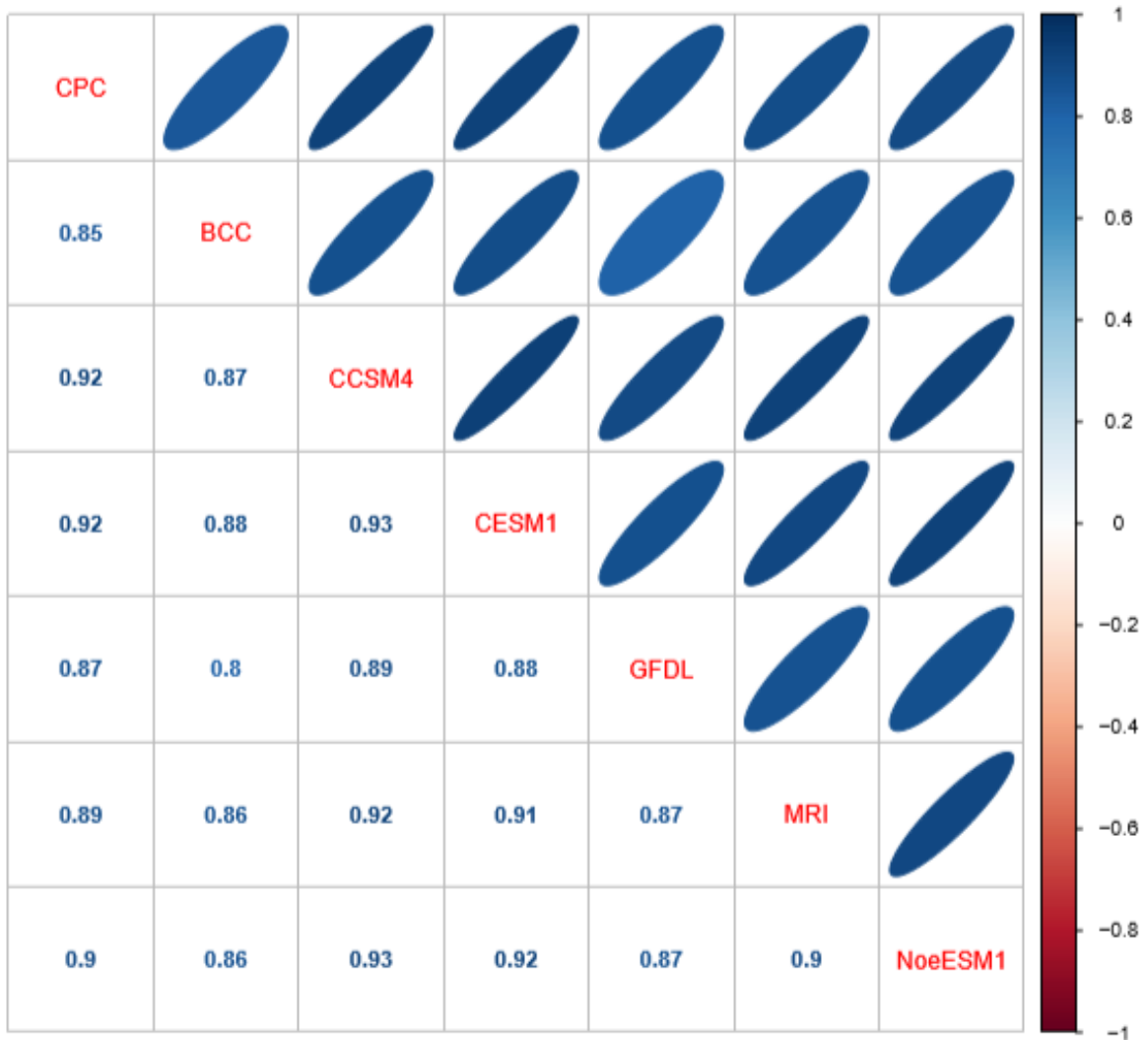
Fonte: Autoria Própria

4.2 Combinação de Modelos

A Raíz do Erro Quadrático Médio (REQM) é uma medida comum usada para medir a qualidade do ajuste do modelo que foi utilizado. Geralmente é utilizado para verificar resultados experimentais. O cálculo é feito se fazendo a diferença entre os dados que foram gerados pelos modelos de previsão e os dados que foram observados/reais, de acordo com a fórmula $EQM = E(\hat{\theta} - \theta)^2$. O ideal para o desvio médio quadrático é ser muito próximo de zero, para o modelo de previsão seja isento de erros e consequentemente mais preciso. Neste trabalho, o *RMSE* foi calculado da seguinte maneira:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2}$$

Figura 4.2 – Correlação entre os Modelos na Região da Bacia do Rio da Prata.



Fonte: Autoria Própria

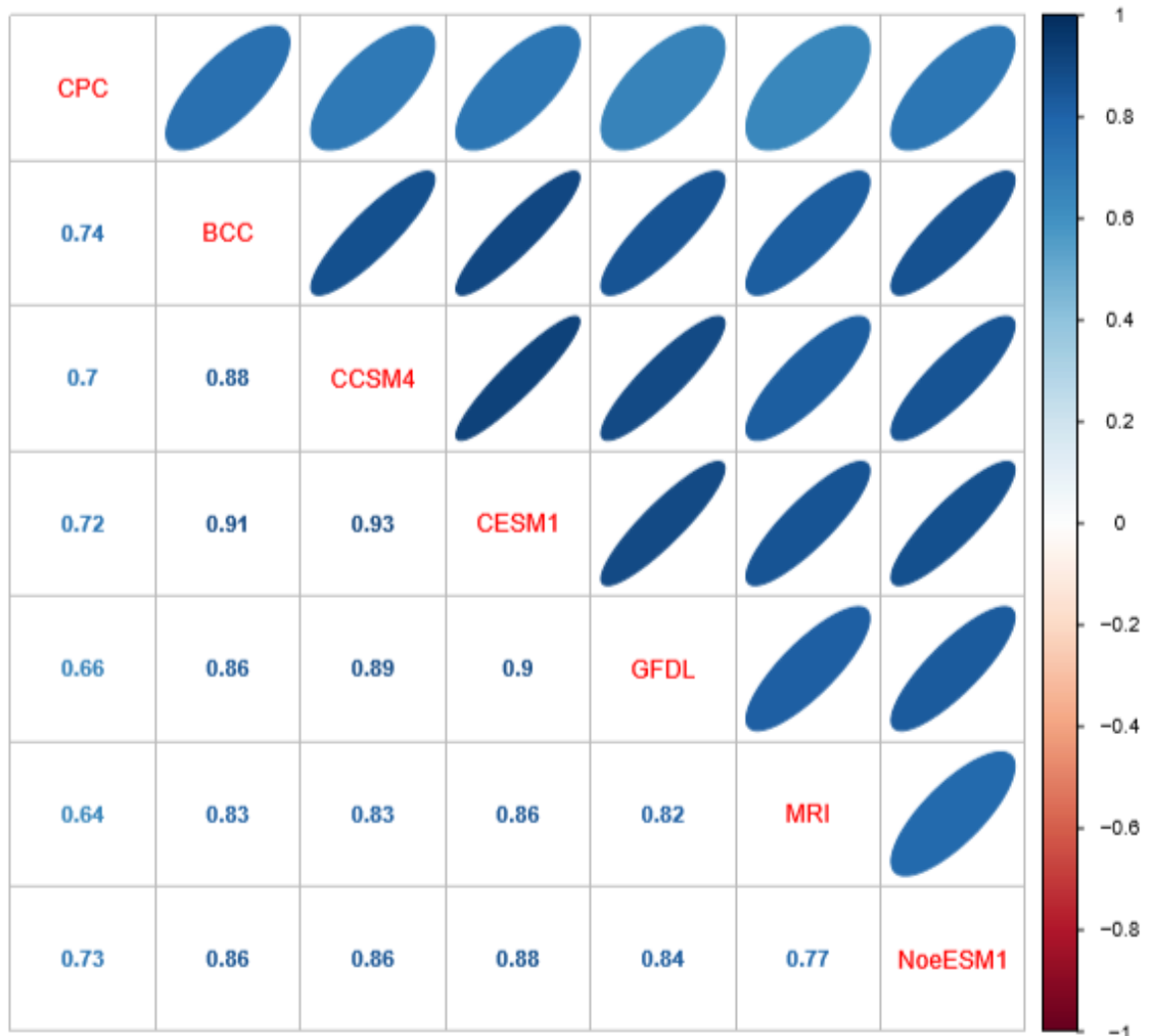
em que

- \widehat{Y}_i = Valor previsto para a observação i .
- Y_i = Valor observado para a observação i .
- n = Tamanho amostral.

De acordo com a Tabela 4.1, podemos observar que os resultados obtidos são baixos, mas o Random Forest é um modelo que obtém melhor desempenho do que o SVM, conseguindo um menor erro na Região da Bacia do Rio da Prata, conclusão que também pode ser tiradas pelos gráficos da Figura 4.5 até a Figura 4.8.

Em uma análise prévia, observamos que o Nordeste Brasileiro tem uma precipitação menor do que a Amazônia e a Bacia do Rio da Prata. Nos gráficos, o modelo MRIRCP,

Figura 4.3 – Correlação entre os Modelos na Região do Nordeste Brasileiro.



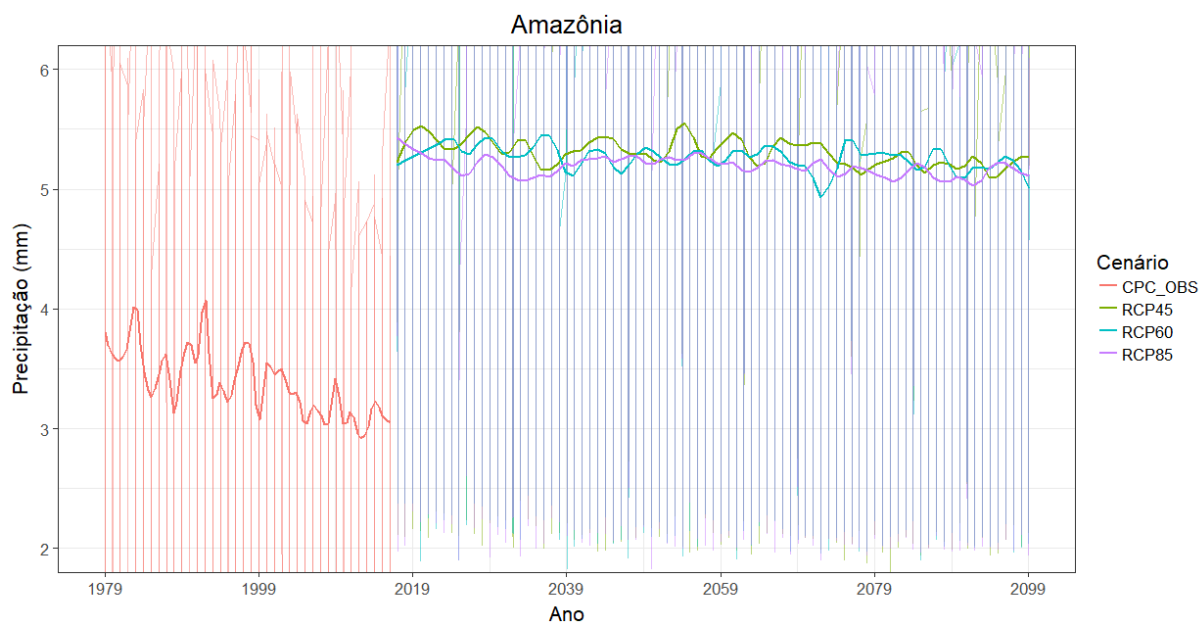
Fonte: Autoria Própria

em quaisquer dos cenários, é o que mais se distancia do observado, o modelo mais próximo ao observado foi o modelo NorESM1-ME. A Bacia do Rio da Prata foi a região em que os modelos tiveram um comportamento mais homogêneo entre eles, apresentando uma menor variabilidade. Na região da Amazônia observamos que a precipitação foi a maior entre as três regiões e apresentou maior variabilidade entre os modelos, sendo o modelo MRIRCP que também se apresenta mais distante dos dados observados.

No gráfico das Figuras 4.4 e 4.5 pode-se observar que nos modelos da região da Amazônia não obtiveram um desempenho esperado em nenhum dos cenários. Isso pode ser explicado por anomalias da convergência do ar em baixos níveis atmosféricos associadas ao aquecimento do continente.

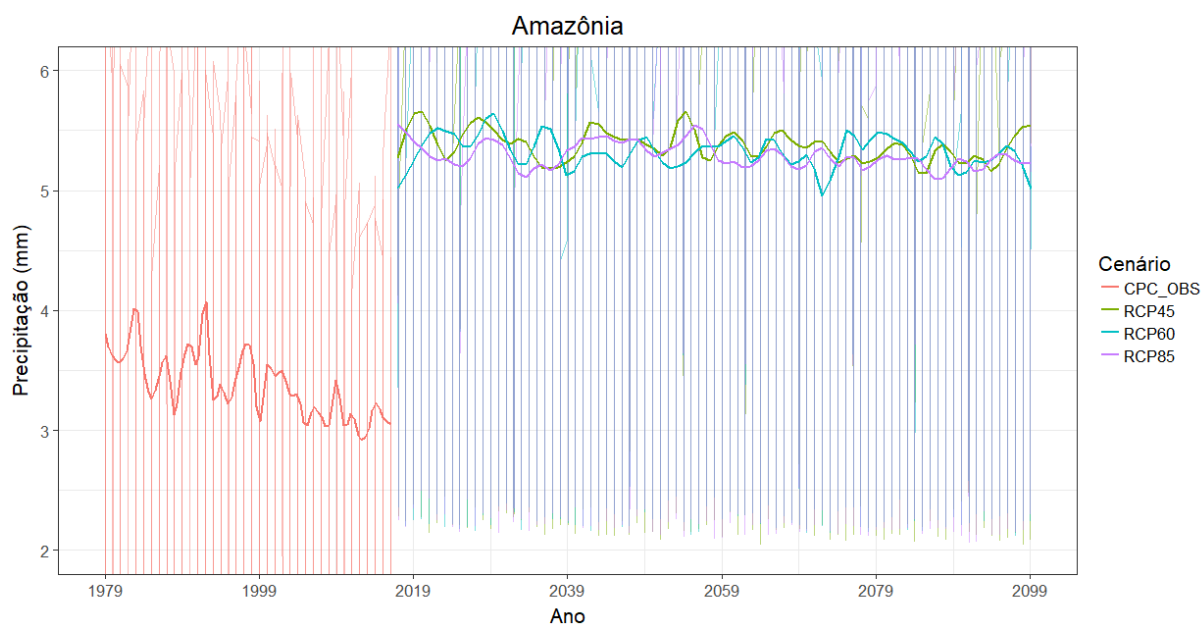
A previsão da precipitação da região na Bacia do Rio da Prata foi a que melhor conseguiu se aproximar dos dados observados, tendo mais distinção entre cenários no

Figura 4.4 – Combinação de modelos para a região da Amazônia utilizando Random Forest.



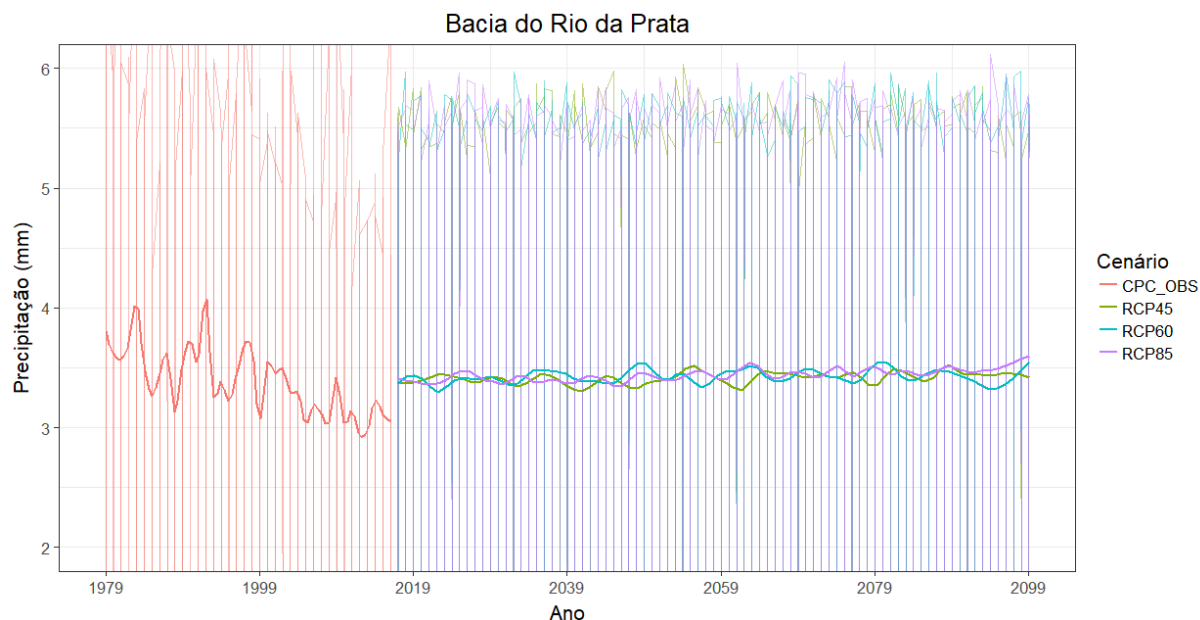
Fonte: Autoria Própria

Figura 4.5 – Combinação de modelos para a região da Amazônia utilizando Support Vector Machine.



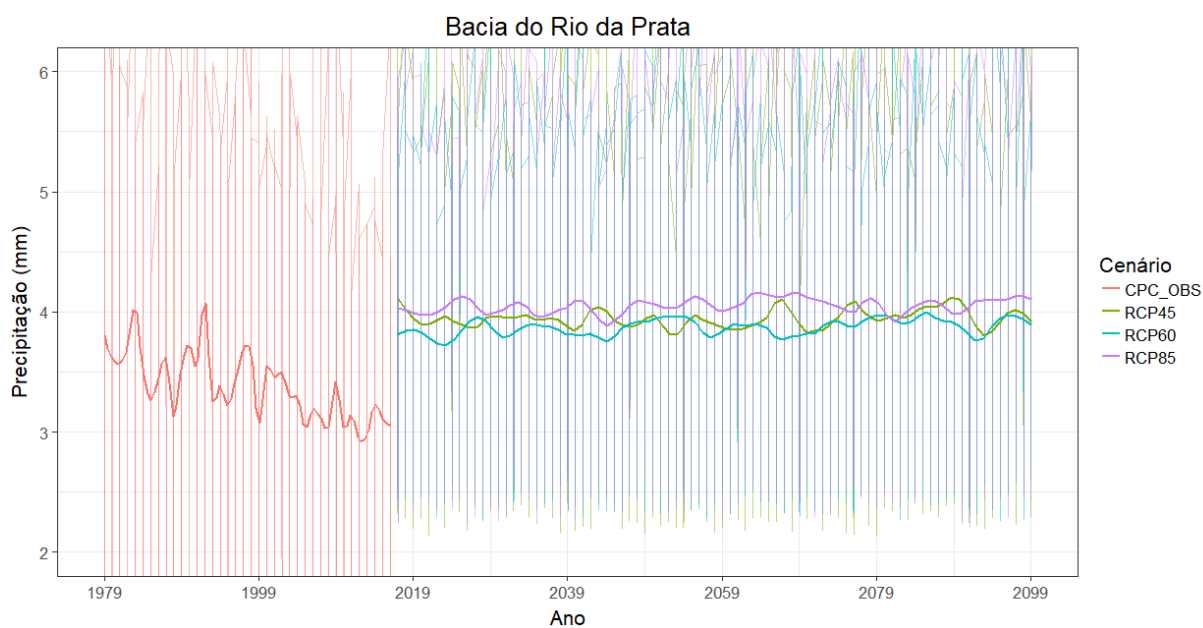
Fonte: Autoria Própria

Figura 4.6 – Combinação de modelos para a região da Bacia do Rio da Prata utilizando Random Forest.



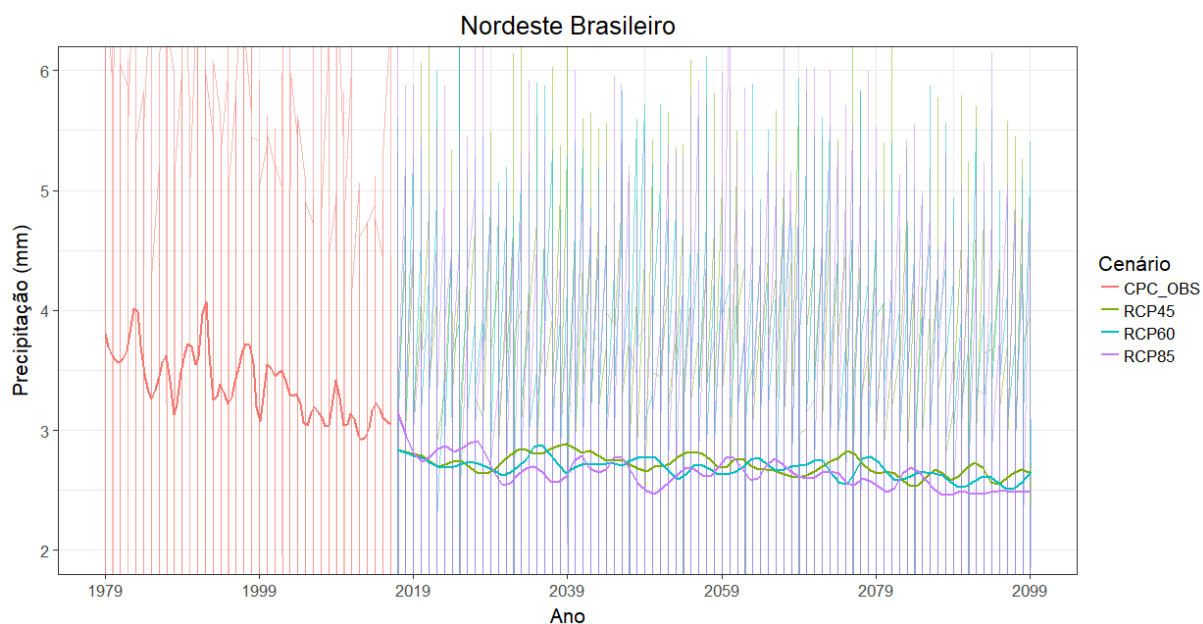
Fonte: Autoria Própria

Figura 4.7 – Combinação de modelos para a região da Bacia do Rio da Prata utilizando Support Vector Machine.



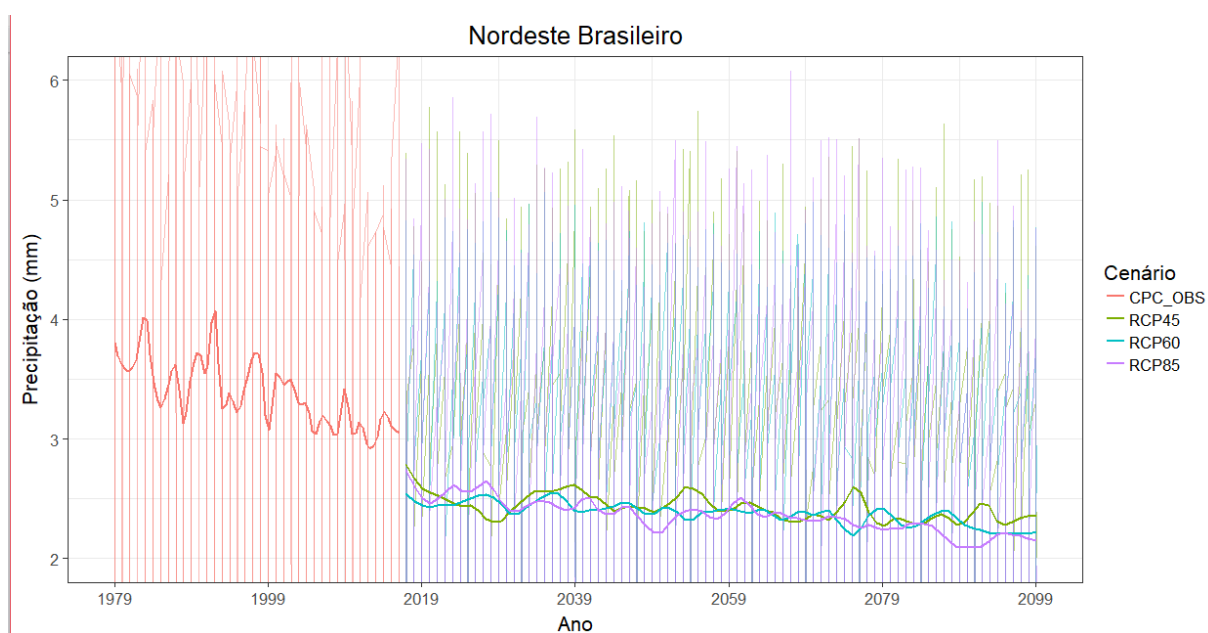
Fonte: Autoria Própria

Figura 4.8 – Combinação de modelos para a região do Nordeste Brasileiro utilizando Random Forest.



Fonte: Autoria Própria

Figura 4.9 – Combinação de modelos para a região do Nordeste Brasileiro utilizando Support Vector Machine.



Fonte: Autoria Própria

Tabela 4.1 – RMSE das regiões do Brasil de acordo com os modelos de previsão e variando os cenários.

	RCP 4.5		RCP 6.0		RCP8.5	
	Random Forest	SVM	Random Forest	SVM	Random Forest	SVM
Amazônia	1,26	1,31	1,28	1,30	1,31	1,30
Bacia do Rio da Prata	0,63	0,64	0,63	0,62	0,66	0,65
Nordeste Brasileiro	1,43	1,50	1,47	1,51	1,47	1,52

Fonte: Autoria Própria.

método do *SVM* (Figura 4.7). No método do *Random Forest* obtivemos uma distinção moderada, nota-se que a previsão entre os cenários foi mais homogênea, como visto na Figura (4.7). A região do Nordeste Brasileiro teve uma previsão aproximada, parecida com a previsão da região da Bacia do Rio da Prata. Observamos esse fenômeno nas Figuras 4.8 e 4.9.

5 Considerações Finais

O presente trabalho procurou identificar qual método de previsão melhor consegue prever a precipitação nas regiões da Amazônia, Bacia do Rio da Prata e Nordeste Brasileiro utilizando a combinação de modelos para melhor ajustar os dados e obter uma boa previsão. Utilizamos o pacote *caret* para prever a precipitação e dois métodos específicos dentro do pacote *caret* que foram o *Random Forest* e *Support Vector Machine*.

Dentre os modelos disponibilizados no IPCC, foram escolhidos 5 modelos do CMIP5, sendo eles o BCC-CSM1.1, CCSM4, CPC-NOAA, NorESM1-M e MRI-CGCM3. Após a escolha dos modelos, colocamos os modelos na mesma resolução, ou seja, foram simulados com um espaçamento maior entre as estações meteorológicas que coletam a precipitação. Após as simulações estarem no mesmo grid de igual espaçamento, de modo que todos os modelos pudessem ser comparados e combinados.

Na análise descritiva, percebemos que os modelos possuem uma correlação alta entre si. Assim, todos possuem dependência com o outro, o que pode ser explicado com facilidade pois todos os modelos derivam da mesma família de modelos.

Analisando o erro quadrático médio podemos observar que tanto para o Random Forest quanto para o SVM os erros foram muito pequenos, sugerindo que os modelos conseguiram fazer uma previsão dos anos de 2017 até 2099 da precipitação nas regiões da Amazônia, Bacia do Rio da Prata e Nordeste Brasileiro bem próximas da precipitação observada de 1978 até 2016. A região com menor erro de previsão foi a da Bacia do Rio da Prata, seguida por Nordeste Brasileiro pela Amazônia.

No fim da combinação entre os modelos, o Random Forest e SVM conseguiram fazer previsões próximas aos dados observados. Nos gráficos pode-se ver que a região da Bacia do Rio da Prata teve um desempenho maior, com um baixo erro quadrático médio, na sua previsão observa que os cenários apesar de se passarem em situações diferentes, obtiveram um resultado semelhante. Já a região da Amazônia obteve o pior desempenho entre as 3 regiões, dando uma previsão da precipitação da combinação dos modelos em cenários selecionados mais discrepante do que as demais. Os cenários obtiveram resultados semelhantes, apesar das variações de emissões de combustíveis fósseis emitidos na Atmosfera. De acordo com os gráficos, podemos chegar a conclusão que não haverá mudança brusca na precipitação em relação aos cenários, porém a precipitação será afetada de diferentes formas em cada região.

A Anexo

A.1 Código para Análise dos Dados

Abaixo está o código utilizado para a análise dos dados da Amazônia. De modo análogo, o código abaixo também foi rodado para os dados da Bacia do Rio da Prata e para o Nordeste Brasileiro.

```

library(randomForest)
library(ggplot2)
library(caret)
library (WriteXLS)
library (e1071)
library(mice)
library(ModelMetrics)
library(corrplot)
library(MASS)
library(dplyr)
library(kernlab)
library(lubridate)

### AMZ

##### Lendo os modelos para previsão historico #####

CPChistorico<-read.table(file="CPC_historico.txt",header=T,sep=";")
BCChistorico<-read.table(file="C:/Users/Isabelle/Desktop/marcus-ipcc-data/
modelos novos/BCC_historico.txt",header=T,sep=";")
CCSM4historico<-read.table(file="C:/Users/Isabelle/Desktop/marcus-ipcc-data/
modelos novos/CCSM4_historico.txt",header=T,sep=";")
CESM1historico<-read.table(file="C:/Users/Isabelle/Desktop/marcus-ipcc-data/
modelos novos/CESM1_historico.txt",header=T,sep=";")
GFDLhistorico<-read.table(file="C:/Users/Isabelle/Desktop/marcus-ipcc-data/
modelos novos/GFDL_historico.txt",header=T,sep=";")
MRIhistorico<-read.table(file="C:/Users/Isabelle/Desktop/marcus-ipcc-data/
modelos novos/MRI_historico.txt",header=T,sep=";")
NorESM1historico<-read.table(file="C:/Users/Isabelle/Desktop/marcus-ipcc-data/

```

```
modelos novos/NorESM1_historico.txt",header=T,sep=";")
```

```
CPChistoricoAMZ<-(CPChistorico%>%  
  select(Year,Month,AMZ)%>%  
  filter( Year > "1978" & Year < "2006")  
)
```

```
BCChistoricoAMZ<-(BCChistorico%>%  
  select(Year,Month,AMZ)%>%  
  filter( Year > "1978" & Year <"2006")  
)
```

```
CCSM4historicoAMZ<-(CCSM4historico%>%  
  select(Year,Month,AMZ)%>%  
  filter( Year > "1978" & Year <"2006")  
)
```

```
CESM1historicoAMZ<-(CESM1historico%>%  
  select(Year,Month,AMZ)%>%  
  filter( Year > "1978" & Year <"2006")  
)
```

```
MRChistoricoAMZ<-(MRChistorico%>%  
  select(Year,Month,AMZ)%>%  
  filter( Year > "1978" & Year <"2006")  
)
```

```
GFDLhistoricoAMZ<-(GFDLhistorico%>%  
  select(Year,Month,AMZ)%>%  
  filter( Year > "1978" & Year <"2006")  
)
```

```
NorESM1historicoAMZ<-(NorESM1historico%>%  
  select(Year,Month,AMZ)%>%  
  filter( Year > "1978" & Year <"2006")  
)
```

```
Data<- paste(NorESM1historicoAMZ$Year, NorESM1historicoAMZ$Month, "01", sep="-")
Data <- ymd(Data)
is.Date(Data)
```

```
AMZhistorico<-data.frame(CPC =CPChistoricoAMZ$AMZ,BCC = BCChistoricoAMZ$AMZ,
  CCSM4=CCSM4historicoAMZ$AMZ, CESM1=CESM1historicoAMZ$AMZ,
  GFDL=GFDLhistoricoAMZ$AMZ,MRI=MRHistoricoAMZ$AMZ,
  NoeESM1=NorESM1historicoAMZ$AMZ, Data=Data)
```

```
##### RCP 45 #####
```

```
BCC45T<-read.table(file="C:/Users/Isabelle/Desktop/marcus-ipcc-data/
modelos novos/MODELOS 09.11.17/RCP45/modelos_ BCC_RCP45.csv",header=T,sep=";")
CCSM445T<-read.table(file="C:/Users/Isabelle/Desktop/marcus-ipcc-data/
modelos novos/MODELOS 09.11.17/RCP45/modelos_CCSM4_RCP45.csv",header=T,sep=";")
CESM145T<-read.table(file="C:/Users/Isabelle/Desktop/marcus-ipcc-data/
modelos novos/MODELOS 09.11.17/RCP45/modelos_CESM1_RCP45.csv",header=T,sep=";")
GFDL45T<-read.table(file="C:/Users/Isabelle/Desktop/marcus-ipcc-data/
modelos novos/MODELOS 09.11.17/RCP45/modelos_GFDL_RCP45.csv",header=T,sep=";")
MRI45T<-read.table(file="C:/Users/Isabelle/Desktop/marcus-ipcc-data/
modelos novos/MODELOS 09.11.17/RCP45/modelos_MRI_RCP45.csv",header=T,sep=";")
NorESM145T<-read.table(file="C:/Users/Isabelle/Desktop/marcus-ipcc-data/
modelos novos/MODELOS 09.11.17/RCP45/modelos_NorESM1_RCP45.csv",header=T,sep=";")
```

```
BCCAMZ45<- (BCC45T%>%
  select(Year,AMZ)%>%
  filter( Year > "2016" & Year <"2100")
)
```

```
CCSM4AMZ45<- (CCSM445T%>%
  select(Year,AMZ)%>%
```

```
    filter( Year > "2016" & Year <"2100")
)

CESM1AMZ45<-(CESM145T%>%
  select(Year,AMZ)%>%
  filter( Year > "2016" & Year <"2100")
)

MRIAMZ45<-(MRI45T%>%
  select(Year,AMZ)%>%
  filter( Year > "2016" & Year <"2100")
)

GFDLAMZ45<-(GFDL45T%>%
  select(Year,AMZ)%>%
  filter( Year > "2016" & Year <"2100")
)

NorESM1AMZ45<-(NorESM145T%>%
  select(Year,AMZ)%>%
  filter( Year > "2016" & Year <"2100")
)

RCP45AMZ<-data.frame(BCC = BCCAMZ45$AMZ, CCSM4=CCSM4AMZ45$AMZ,
  CESM1=CESM1AMZ45$AMZ, GFDL=GFDLAMZ45$AMZ, MRI=MRIAMZ45$AMZ,
  NoeESM1=NorESM1AMZ45$AMZ)

###Adicionando as Ultimas linhas

CPC1<-(CPChistorico%>%
  select(Year,Month,AMZ)%>%
  filter( Year > "2005" & Year <"2017")
)

BCC451<-(BCC45T%>%
  select(Year,Month,AMZ)%>%
  filter( Year > "2005" & Year <"2017")
```

```
)

CCSM4451<- (CCSM445T%>%
  select(Year,Month,AMZ)%>%
  filter( Year > "2005" & Year <"2017")
)

CESM1451<- (CESM145T%>%
  select(Year,Month,AMZ)%>%
  filter( Year > "2005" & Year <"2017")
)

MRI451<- (MRI45T%>%
  select(Year,Month,AMZ)%>%
  filter( Year > "2005" & Year <"2017")
)

GFDL451<- (GFDL45T%>%
  select(Year,Month,AMZ)%>%
  filter( Year > "2005" & Year <"2017")
)

NorESM1451<- (NorESM145T%>%
  select(Year,Month,AMZ)%>%
  filter( Year > "2005" & Year <"2017")
)

Data<- paste(NorESM1451$Year,NorESM1451$Month, "01", sep="-")
Data <- ymd(Data)
is.Date(Data)

ULTIMASLINHAAMZ45<-data.frame(CPC = CPC1$AMZ,BCC = BCC451$AMZ,
CCSM4=CCSM4451$AMZ,CESM1=CESM1451$AMZ,
GFDL=GFDL451$AMZ,MRI=MRI451$AMZ,NoeESM1=NorESM1451$AMZ,Data=Data)

AMZhistoricocompleto45 <- rbind(AMZhistorico, ULTIMASLINHAAMZ45)

##### RCP 60 #####
```

```
BCC60T<-read.table(file="C:/Users/Isabelle/Desktop/marcus-ipcc-data/
modelos novos/MODELOS 09.11.17/RCP60/modelos_BCC_RCP60.csv",header=T,sep=";")
CCSM460T<-read.table(file="C:/Users/Isabelle/Desktop/marcus-ipcc-data/
modelos novos/MODELOS 09.11.17/RCP60/modelos_CCSM4_RCP60.csv",header=T,sep=";")
CESM160T<-read.table(file="C:/Users/Isabelle/Desktop/marcus-ipcc-data/
modelos novos/MODELOS 09.11.17/RCP60/modelos_CESM1_RCP60.csv",header=T,sep=";")
GFDL60T<-read.table(file="C:/Users/Isabelle/Desktop/marcus-ipcc-data/
modelos novos/MODELOS 09.11.17/RCP60/modelos_GFDL_RCP60.csv",header=T,sep=";")
MRI60T<-read.table(file="C:/Users/Isabelle/Desktop/marcus-ipcc-data/
modelos novos/MODELOS 09.11.17/RCP60/modelos_MRI_RCP60.csv",header=T,sep=";")
NorESM160T<-read.table(file="C:/Users/Isabelle/Desktop/marcus-ipcc-data/
modelos novos/MODELOS 09.11.17/RCP60/modelos_NorESM1_RCP60.csv",header=T,sep=";")
```

```
BCCAMZ60<-(BCC60T%>%
  select(Year,AMZ)%>%
  filter( Year > "2016" & Year <"2100")
)
```

```
CCSM4AMZ60<-(CCSM460T%>%
  select(Year,AMZ)%>%
  filter( Year > "2016" & Year <"2100")
)
```

```
CESM1AMZ60<-(CESM160T%>%
  select(Year,AMZ)%>%
  filter( Year > "2016" & Year <"2100")
)
```

```
MRIAMZ60<-(MRI60T%>%
  select(Year,AMZ)%>%
  filter( Year > "2016" & Year <"2100")
)
```

```
GFDLAMZ60<-(GFDL60T%>%
  select(Year,AMZ)%>%
  filter( Year > "2016" & Year <"2100")
)
```

```
NorESM1AMZ60<-(NorESM160T%>%  
  select(Year,AMZ)%>%  
  filter( Year > "2016" & Year <"2100")  
)
```

```
RCP60AMZ<-data.frame(BCC = BCCAMZ60$AMZ, CCSM4=CCSM4AMZ60$AMZ,  
CESM1=CESM1AMZ60$AMZ, GFDL=GFDLAMZ60$AMZ,  
MRI=MRIAMZ60$AMZ,NoeESM1=NorESM1AMZ60$AMZ )
```

###Adicionando as Ultimas linhas

```
CPC1<-(CPChistorico%>%  
  select(Year,Month,AMZ)%>%  
  filter( Year > "2005" & Year <"2017")  
)
```

```
BCC601<-(BCC60T%>%  
  select(Year,Month,AMZ)%>%  
  filter( Year > "2005" & Year <"2017")  
)
```

```
CCSM4601<-(CCSM460T%>%  
  select(Year,Month,AMZ)%>%  
  filter( Year > "2005" & Year <"2017")  
)
```

```
CESM1601<-(CESM160T%>%  
  select(Year,Month,AMZ)%>%  
  filter( Year > "2005" & Year <"2017")  
)
```

```
MRI601<-(MRI60T%>%  
  select(Year,Month,AMZ)%>%  
  filter( Year > "2005" & Year <"2017")  
)
```

```
GFDL601<-(GFDL60T%>%
  select(Year,Month,AMZ)%>%
  filter( Year > "2005" & Year <"2017")
)

NorESM1601<-(NorESM160T%>%
  select(Year,Month,AMZ)%>%
  filter( Year > "2005" & Year <"2017")
)

Data<- paste(NorESM1601$Year,NorESM1601$Month, "01", sep="-")
Data <- ymd(Data)
is.Date(Data)

ULTIMASLINHAAMZ60<-data.frame(CPC = CPC1$AMZ,BCC = BCC601$AMZ,
CCSM4=CCSM4601$AMZ, CESM1=CESM1601$AMZ,
GFDL=GFDL601$AMZ,MRI=MRI601$AMZ,NoeESM1=NorESM1601$AMZ,Data=Data)

AMZhistoricocompleto60 <- rbind(AMZhistorico, ULTIMASLINHAAMZ60)

##### RCP 85 #####

BCC85T<-read.table(file="C:/Users/Isabelle/Desktop/marcus-ipcc-data/
modelos novos/MODELOS 09.11.17/RCP85/modelos_BCC_RCP85.csv",header=T,sep=";")
CCSM485T<-read.table(file="C:/Users/Isabelle/Desktop/marcus-ipcc-data/
modelos novos/MODELOS 09.11.17/RCP85/modelos_CCSM4_RCP85.csv",header=T,sep=";")
CESM185T<-read.table(file="C:/Users/Isabelle/Desktop/marcus-ipcc-data/
modelos novos/MODELOS 09.11.17/RCP85/modelos_CESM1_RCP85.csv",header=T,sep=";")
GFDL85T<-read.table(file="C:/Users/Isabelle/Desktop/marcus-ipcc-data/
modelos novos/MODELOS 09.11.17/RCP85/modelos_GFDL_RCP85.csv",header=T,sep=";")
MRI85T<-read.table(file="C:/Users/Isabelle/Desktop/marcus-ipcc-data/
modelos novos/MODELOS 09.11.17/RCP85/modelos_MRI_RCP85.csv",header=T,sep=";")
NorESM185T<-read.table(file="C:/Users/Isabelle/Desktop/marcus-ipcc-data/
modelos novos/MODELOS 09.11.17/RCP85/modelos_NorESM1_RCP85.csv",header=T,sep=";")

BCCAMZ85<-(BCC85T%>%
  select(Year,AMZ)%>%
```

```
    filter( Year > "2016" & Year <"2100")
  )

CCSM4AMZ85<-(CCSM485T%>%
  select(Year,AMZ)%>%
  filter( Year > "2016" & Year <"2100")
)

CESM1AMZ85<-(CESM185T%>%
  select(Year,AMZ)%>%
  filter( Year > "2016" & Year <"2100")
)

MRIAMZ85<-(MRI85T%>%
  select(Year,AMZ)%>%
  filter( Year > "2016" & Year <"2100")
)

GFDLAMZ85<-(GFDL85T%>%
  select(Year,AMZ)%>%
  filter( Year > "2016" & Year <"2100")
)

NorESM1AMZ85<-(NorESM185T%>%
  select(Year,AMZ)%>%
  filter( Year > "2016" & Year <"2100")
)

RCP85AMZ<-data.frame(BCC = BCCAMZ85$AMZ, CCSM4=CCSM4AMZ85$AMZ,
  CESM1=CESM1AMZ85$AMZ, GFDL=GFDLAMZ85$AMZ,
  MRI=MRIAMZ85$AMZ,NoeESM1=NorESM1AMZ85$AMZ )

###Adicionando as Ultimas linhas

CPC1<-(CPChistorico%>%
  select(Year,Month,AMZ)%>%
  filter( Year > "2005" & Year <"2017")
```

```
)
```

```
BCC851<- (BCC85T%>%  
  select(Year,Month,AMZ)%>%  
  filter( Year > "2005" & Year <"2017")  
)
```

```
CCSM4851<- (CCSM485T%>%  
  select(Year,Month,AMZ)%>%  
  filter( Year > "2005" & Year <"2017")  
)
```

```
CESM1851<- (CESM185T%>%  
  select(Year,Month,AMZ)%>%  
  filter( Year > "2005" & Year <"2017")  
)
```

```
MRI851<- (MRI85T%>%  
  select(Year,Month,AMZ)%>%  
  filter( Year > "2005" & Year <"2017")  
)
```

```
GFDL851<- (GFDL85T%>%  
  select(Year,Month,AMZ)%>%  
  filter( Year > "2005" & Year <"2017")  
)
```

```
NorESM1851<- (NorESM185T%>%  
  select(Year,Month,AMZ)%>%  
  filter( Year > "2005" & Year <"2017")  
)
```

```
Data<- paste(NorESM1851$Year,NorESM1851$Month, "01", sep="-")  
Data <- ymd(Data)  
is.Date(Data)
```

```
ULTIMASLINHAAMZ85<-data.frame(CPC = CPC1$AMZ,BCC = BCC851$AMZ,  
CCSM4=CCSM4851$AMZ, CESM1=CESM1851$AMZ,  
GFDL=GFDL851$AMZ,MRI=MRI851$AMZ,NoeESM1=NorESM1851$AMZ,Data=Data)
```

```
AMZhistoricocompleto85 <- rbind(AMZhistorico, ULTIMASLINHAAMZ85)
```

```
##### Trabalhando com os dados
```

```
#Variando o grid
```

```
rfGrid<- expand.grid(mtry = c(1,2, 3, 4, 5 ,6))
```

```
svmRadialGrid <- expand.grid(C = 2^(-10:3), sigma = 2^(-10:0))
```

```
#traincontrol do svm e random forest
```

```
trcontrolrfbaciarf <- trainControl( method = "cv",
```

```
number = 5,repeats = 10,p=0.80,savePred = TRUE)
```

```
trcontrolbaciasvm<- trainControl( method = "cv",
```

```
number = 5,repeats = 10,p=0.80,savePred = TRUE)
```

```
#Train do svm e random forest
```

```
#RF
```

```
trainrfbaciarf45<- train(CPC ~ BCC+NoeESM1+MRI+CCSM4+CESM1+GFDL,
```

```
data=AMZhistoricocompleto45,method = "rf",
```

```
trControl = trcontrolrfbaciarf,verbose=F,tuneGrid=rfGrid)
```

```
trainrfbaciarf60<- train(CPC ~ BCC+NoeESM1+MRI+CCSM4+CESM1+GFDL,
```

```
data=AMZhistoricocompleto60,method = "rf",
```

```
trControl = trcontrolrfbaciarf,verbose=F,tuneGrid=rfGrid
```

```
)
```

```
trainrfbaciarf85<- train(CPC ~ BCC+NoeESM1+MRI+CCSM4+CESM1+GFDL,
```

```
data=AMZhistoricocompleto85,method = "rf",
```

```
trControl = trcontrolrfbaciarf,verbose=F,tuneGrid=rfGrid)
```

```
#SVM
```

```
trainbaciasvm45<- train(CPC ~ BCC+NoeESM1+MRI+CCSM4+CESM1+GFDL,
```

```
data=AMZhistoricocompleto45,method = "svmRadial",
```

```
trControl = trcontrolbaciasvm,verbose=F,tuneGrid=svmRadialGrid)
```

```
trainbaciasvm60<- train(CPC ~ BCC+NoeESM1+MRI+CCSM4+CESM1+GFDL,
data=AMZhistoricocompleto60,method = "svmRadial",
trControl = trcontrolbaciasvm,verbose=F,tuneGrid=svmRadialGrid)
trainbaciasvm85<- train(CPC ~ BCC+NoeESM1+MRI+CCSM4+CESM1+GFDL,
data=AMZhistoricocompleto85,method = "svmRadial",
trControl = trcontrolbaciasvm,verbose=F,tuneGrid=svmRadialGrid)
```

```
#Predizendo valores para os cenários RCP45, RCP60 e RCP85
```

```
#RF
```

```
predicaoAMZRCP45rf<- predict(trainrfbaciarf45,RCP45AMZ)
predicaoAMZRCP60rf<- predict(trainrfbaciarf60,RCP60AMZ)
predicaoAMZRCP85rf<- predict(trainrfbaciarf85,RCP85AMZ)
```

```
write.csv(predicaoAMZRCP45rf,"predicaoAMZRCP45rf.csv")
write.csv(predicaoAMZRCP60rf,"predicaoAMZRCP60rf.csv")
write.csv(predicaoAMZRCP85rf,"predicaoAMZRCP85rf.csv")
write.csv(CPChistoricoAMZ,"CPChistoricoAMZ.csv")
```

```
head(predicaoAMZRCP45rf)
head(predicaoAMZRCP60rf)
head(predicaoAMZRCP85rf)
```

```
#SVM
```

```
predicaoAMZRCP45svm<- predict(trainbaciasvm45,RCP45AMZ)
predicaoAMZRCP60svm<- predict(trainbaciasvm60,RCP60AMZ)
predicaoAMZRCP85svm<- predict(trainbaciasvm85,RCP85AMZ)
```

```
write.csv(predicaoAMZRCP45svm,"predicaoAMZRCP45svm.csv")
write.csv(predicaoAMZRCP60svm,"predicaoAMZRCP60svm.csv")
write.csv(predicaoAMZRCP85svm,"predicaoAMZRCP85svm.csv")
```

```
head(predicaoAMZRCP45svm)
head(predicaoAMZRCP60svm)
head(predicaoAMZRCP85svm)
```

```
##### GRÁFICO
```

```
#RF
```

```
plotarrfAMZ<-read.table(file="C:/Users/Isabelle/Desktop/marcus-ipcc-data/
modelos novos/MODELOS 09.11.17/plotarrfAMZ.csv",header=T,sep=";")
str(plotarrfAMZ)
```

```
plotarrfAMZ$Datas = as.Date(plotarrfAMZ$Datas, format = "%m/%d/%Y")
```

```
ggplot(plotarrfAMZ, aes(x=Datas, y=precipitacao, colour=tipo)) +
  geom_line(aes(colour=tipo), alpha=0.5) +
  coord_cartesian(ylim = c(2, 6)) +
  theme_bw(base_size = 16, base_family = "")+
  scale_x_date(breaks=seq(from=dmy("01/01/1979"), to=dmy("01/01/2099"),
by="20 year"), date_labels = "%Y") +
  geom_smooth(method="loess", se=FALSE, span=0.1) +
  labs(x="Ano", y="Precipitação (mm)", colour="Cenário", title="Amazônia") +
  theme(plot.title = element_text(hjust = 0.5))
```

```
#SVM
```

```
plotarsvmAMZ<-read.table(file="C:/Users/Isabelle/Desktop/marcus-ipcc-data/
modelos novos/MODELOS 09.11.17/plotarsvmAMZ.csv",header=T,sep=";")
str(plotarsvmAMZ)
```

```
plotarsvmAMZ$Datas = as.Date(plotarsvmAMZ$Datas, format = "%m/%d/%Y")
```

```
ggplot(plotarsvmAMZ, aes(x=Datas, y=precipitacao, colour=tipo)) +
  geom_line(aes(colour=tipo), alpha=0.5) +
  coord_cartesian(ylim = c(2, 6)) +
  scale_x_date(breaks=seq(from=dmy("01/01/1979"), to=dmy("01/01/2099"), by="5 year")
  geom_smooth(method="loess", se=FALSE, span=0.1) +
  labs(x="Ano", y="precipitação (mm)", colour="Cenário", title="Amazônia") +
  theme(plot.title = element_text(hjust = 0.5))
```

```
###CALCULO DO RMSE
```

```
#RMSE- erro médio quadratico
```

```
(RMSE45LPB<-rmse(CPChistoricoLPB$LPB,predicaoLPBRCP45rf))
```

```
(RMSE60LPB<-rmse(CPChistoricoLPB$LPB,predicaoLPBRCP60rf))
```

```
(RMSE85LPB<-rmse(CPChistoricoLPB$LPB,predicaoLPBRCP85rf))
```

```
(RMSE45AMZ<-rmse(CPChistoricoAMZ$AMZ,predicaoAMZRCP45rf))
```

```
(RMSE60AMZ<-rmse(CPChistoricoAMZ$AMZ,predicaoAMZRCP60rf))
```

```
(RMSE85AMZ<-rmse(CPChistoricoAMZ$AMZ,predicaoAMZRCP85rf))
```

```
(RMSE45NEB<-rmse(CPChistoricoNEB$NEB,predicaoNEBRCP45rf))
```

```
(RMSE60NEB<-rmse(CPChistoricoNEB$NEB,predicaoNEBRCP60rf))
```

```
(RMSE85NEB<-rmse(CPChistoricoNEB$NEB,predicaoNEBRCP85rf))
```

```
#SVM
```

```
(RMSE45LPBSVM<-rmse(CPChistoricoLPB$LPB,predicaoLPBRCP45svm))
```

```
(RMSE60LPBSVM<-rmse(CPChistoricoLPB$LPB,predicaoLPBRCP60svm))
```

```
(RMSE85LPBSVM<-rmse(CPChistoricoLPB$LPB,predicaoLPBRCP85svm))
```

```
(RMSE45AMZSVM<-rmse(CPChistoricoAMZ$AMZ,predicaoAMZRCP45svm))
```

```
(RMSE60AMZSVM<-rmse(CPChistoricoAMZ$AMZ,predicaoAMZRCP60svm))
```

```
(RMSE85AMZSVM<-rmse(CPChistoricoAMZ$AMZ,predicaoAMZRCP85svm))
```

```
(RMSE45NEBSVM<-rmse(CPChistoricoNEB$NEB,predicaoNEBRCP45svm))
```

```
(RMSE60NEBSVM<-rmse(CPChistoricoNEB$NEB,predicaoNEBRCP60svm))
```

```
(RMSE85NEBSVM<-rmse(CPChistoricoNEB$NEB,predicaoNEBRCP85svm))
```

```
AMZCORR<-read.table(file="C:/Users/Isabelle/Desktop/marcus-ipcc-data/modelos novos/
```

```
NEBCORR<-read.table(file="C:/Users/Isabelle/Desktop/marcus-ipcc-data/modelos novos/
```

```
LPBCORR<-read.table(file="C:/Users/Isabelle/Desktop/marcus-ipcc-data/modelos novos/
```

```
AMZCORR2 <- cor(AMZCORR[2:8])
```

```
(corrplot.mixed(AMZCORR2, lower = "number", upper = "ellipse",  
t1.col = "black",number.digits=4 ))
```

Referências

- ARMSTRONG, J. S. Combining forecasts. *Kluwer Academic Publishers*, p. 417–440, 2001.
- BREIMAN, L. Random forest. *Kluwer Academic Publishers.*, v. 45, 2001.
- CHANDLER, R. E. Exploiting strength, discounting weakness: combining information from multiple climate simulators. *Royal Society Publishing*, 2015.
- DESER C., P. A. e. B. V. Uncertainty in climate change projections: the role of internal variability. *Climate Dynamics*, 2012.
- FUSS, S. et al. Betting on negative emissions. *Nature Climate Change* 4, 2014.
- GENT, M. M. P. R. et al. The Community Climate System Model Version 4. *Journal of climate*, v. 24, 2011.
- GIORGI F. E MEARNS, L. O. Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the "Reliability Ensemble Averaging"(REA) method. *Journal of Climate*, 2002.
- HIGGINS, W. R. *A gridded hourly precipitation data base for the United States (1963-1993)*. [S.l.]: :Camp Springs, MD : U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, National Weather Service, 1996.
- IVERSEN, T. et al. The Norwegian Earth System Model, NorESM1-M – Part 2: Climate response and scenario projections. *Geoscientific Model Development Discussions*, v. 5, 2012.
- KNUTTI R., M. D. e. G. A. Climate model genealogy : Generation CMIP5 and how we got there. *Geophysical Research Letters*, 2013.
- KUHN, M. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, v. 28, 2008.
- LORENA, A. C. P. L. F. d. C. A. C. Uma Introdução às Support Vector Machine. 2007.
- SANTOS T. S., M. D. e. T. R. R. Artificial neural networks and multiple linear regression model using principal components to estimate rainfall over South America. 2015.
- TAN, P.-N. et al. Introduction to DATAMINING. *Person Education*, 2006.
- XIN Y.-J., Y.-J. C. F. W. et al. Asymmetry of surface climate change in RCP2.6 projected by multiple CMIP5 models. *Adv. Atmos.Sci.*, 2012.
- YUKIMOTO, S. et al. A New Global Climate Model of the Meteorological Research Institute: MRI-CGCM3 —Model Description and Basic Performance—. *Journal of the Meteorological Society of Japan*, v. 90A, 2012.
- ZHANG FAN MIN, d. X. H. H.-R. Aggregated Recommendation through Random Forests. *The Scientific World Journal*, 2014.