

Bernardo Rolim Rangel

**A APLICAÇÃO DE *MACHINE LEARNING*
SUPERVISIONADA PARA A PREDIÇÃO DE CARBONO
ORGÂNICO NO ATLÂNTICO SUL A PARTIR DE DADOS
FISIOGRÁFICOS E TAXAS DE SEDIMENTAÇÃO**

**Trabalho Final de Curso
(Geologia)**

UFRJ
Rio de Janeiro
2021

Bernardo Rolim Rangel

**A APLICAÇÃO DE *MACHINE LEARNING* SUPERVISIONADA
PARA A PREDIÇÃO DE CARBONO ORGÂNICO NO ATLÂNTICO
SUL A PARTIR DE DADOS FISIAGRÁFICOS E TAXAS DE
SEDIMENTAÇÃO**

Trabalho Final de Curso de Graduação em Geologia do Instituto de Geociências, Universidade Federal do Rio de Janeiro – UFRJ, apresentado como requisito necessário para obtenção do grau de Geólogo.

Orientadores:

Patrick Francisco Führ Dal' Bó (UFRJ)
Andre Luiz Belem (UFF)

Co-orientador:

Fellippe Roberto Alves Bione (UFF)

Rio de Janeiro

Julho/2021

Bernardo Rolim Rangel

**A APLICAÇÃO DE *MACHINE LEARNING* SUPERVISIONADA
PARA A PREDIÇÃO DE CARBONO ORGÂNICO NO ATLÂNTICO
SUL A PARTIR DE DADOS FISIAGRÁFICOS E TAXAS DE
SEDIMENTAÇÃO**

Trabalho Final de Curso de Graduação em Geologia do Instituto de Geociências, Universidade Federal do Rio de Janeiro – UFRJ, apresentado como requisito necessário para obtenção do grau de Geólogo.

Orientadores:

Patrick Francisco Führ Dal' Bó (UFRJ)
Andre Luiz Belem (UFF)

Co-orientador:

Fellippe Roberto Alves Bione (UFF)

Aprovada em _____ de _____ de _____.
Por:

Orientador: Prof. Dr. Patrick Francisco Führ Dal' Bó (UFRJ)

Orientador: Prof. Dr. Andre Luiz Belem (UFF)

Co-orientador: MSc. Fellippe Roberto Alves Bione (UFF)

MSc. Jaqueline Torres de Souza (UFRJ)

MSc. Igor Viegas Alves Fernandes de Souza (CENPES/Petrobras)

Ao melhor pai do mundo, meu maior exemplo. Qualquer palavra é insuficiente para traduzir a minha gratidão, e as saudades são amenizadas pela certeza do nosso reencontro.

AGRADECIMENTOS

Os anos de geologia foram, sem dúvida, alguns dos melhores da minha vida. Sou grato à UFRJ e aos docentes de excelência que lecionam em nosso departamento que, mais do que professores, são guias que marcam nossas jornadas e enriquecem nossas experiências.

Agradeço imensamente aos meus orientadores: Andre Belem, que me encorajou a embarcar em no mundo até então desconhecido da programação, sem a qual este trabalho não seria possível; Patrick Dal' Bó, que além de excelente docente nas disciplinas de sedimentologia e estratigrafia, aceitou me orientar em um trabalho que estava, a princípio, sendo desenvolvido fora da UFRJ; e co-orientador Fellippe Bione pelas valiosas sugestões que enriqueceram este trabalho.

Agradeço às amizades que colecionei ao longo desses anos, que durarão por muito mais tempo do que o simples intervalo de uma graduação. Matheus Cabral, Pedro Duarte, Caíque Cabral, Gabriel de Oliveira, Rodrigo Freitas, Victor Menezes, João Pedro Bento e tantos(as) outros(as) que tornaram cada dia e cada campo mais felizes, muito obrigado!

Agradeço à minha namorada, Julia de Oliveira, por me apoiar durante essa trajetória, tendo muitas vezes que lidar com meu mau humor devido à labuta cansativa construída graças ao dueto trabalho/estudo.

Por fim, agradeço imensamente à parte mais importante da minha vida, a família! À minha mãe, Valéria; à minha segunda mãe e principal mentora para que eu me aventurasse na geologia, Ana Luiza; à minha irmã, Julia; aos meus avós, Lucia e Rogério que, por sinal, até hoje não compreendem a razão de eu ter feito duas graduações; e, especialmente, ao meu pai, Edison, que acompanhou e tornou essa jornada possível, mas que agora vê a conclusão desse ciclo pelos olhos do plano espiritual, orgulhoso e pleno da mesma forma caso estivesse entre nós. Eu amo todos vocês.

RESUMO

RANGEL, Bernardo Rolim. **A Aplicação de *Machine Learning* Supervisionada para a predição de carbono orgânico no Atlântico Sul a partir de dados fisiográficos e taxas de sedimentação**. 2021. 89 f. Trabalho Final de Curso (Geologia) – Departamento de Geologia, Instituto de Geociências, Universidade Federal do Rio de Janeiro, Rio de Janeiro.

Modelos de fácies orgânica são peças fundamentais no contexto da indústria do petróleo e gás, sobretudo na fase exploratória. Muitos dos alvos estratigráficos visados durante essa fase são do Cretáceo, período no qual o Oceano Atlântico teve origem e quando acumulações anômalas episódicas de carbono orgânico (CO) ocorreram nos sedimentos marinhos desta bacia oceânica recém-formada. Muitos fatores/forçantes são sabidamente relevantes para o acúmulo de carbono orgânico nos oceanos e fazem parte da modelagem de fácies orgânica. Entretanto, é desconhecido o papel da fisiografia e taxas de sedimentação na acumulação de carbono orgânico em bacias oceânicas. O objetivo deste trabalho foi avaliar a contribuição de fatores fisiográficos - latitude, paleobatimetria e distâncias para a linha de costa – e taxas de sedimentação na acumulação de carbono orgânico nas margens opostas do Atlântico Sul/Equatorial através do desenvolvimento e aplicação de modelo regressivo associado à *Machine Learning* supervisionada. Para tal, foram testados 19 modelos regressivos com base em dados em dados de 7 testemunhos obtidos em expedições do DSDP e ODP. Uma primeira análise dos padrões de distribuição do CO nos testemunhos estudados mostrou que a acumulação de carbono orgânico entre as duas margens do Atlântico Sul/Equatorial foi significativamente distinta, o que pode ser atribuído à atuação de processos específicos associados, sobretudo, à dinâmica de massas d'água, dinâmica atmosférica e evolução paleobatimétrica. O modelo preditivo com melhor performance ($R^2 = 0,69$; RMSE = 2,73) foi uma regressão de processo gaussiano, com função *kernel* racional quadrática. A aplicação desse modelo em *dataset* de teste resultou em menores coeficientes de determinação, com $R^2 = 0,37$. Apesar da boa performance do modelo regressivo escolhido e do seu potencial para predizer teores de carbono orgânico a partir de fatores fisiográficos e taxas de sedimentação, o mesmo demonstrou-se pouco flexível para adaptações diferentes do contexto do treinamento. A piora da performance da aplicação do modelo em *dataset* de teste sugere que os fatores fisiográficos influenciam indiretamente parâmetros significativos na modulação do acúmulo de CO, os quais são, também, afetados por outros fatores não fisiográficos. Compreende-se, dessa forma, o poder limitado do uso dos fatores fisiográficos como preditores únicos da acumulação de CO em sedimentos marinhos.

Palavras-chave: Cretáceo; *machine learning*; carbono orgânico

ABSTRACT

RANGEL, Bernardo Rolim. **The application of supervised Machine Learning for organic carbon prediction on South Atlantic based on physiographic factors and sedimentation rates**. 2021. 89 f. Trabalho Final de Curso (Geologia) – Departamento de Geologia, Instituto de Geociências, Universidade Federal do Rio de Janeiro, Rio de Janeiro.

Organic facies models are fundamental pieces in the context of the oil and gas industry, especially in the exploration phase. Many of the stratigraphic targets during this phase are from the Cretaceous, period in which the Atlantic Ocean originated and when episodic anomalous accumulations of organic carbon (OC) occurred in the marine sediments. Many factors/forcings are known to be relevant to the OC accumulation in the oceans and are part of organic facies modeling. However, the specific role of physiography factors and sedimentation rates in organic carbon accumulation in ocean basins is still poorly known. The objective of this work was to evaluate the contribution of physiographic factors - latitude, paleobathymetry and distances to the coastline - and sedimentation rates in the accumulation of organic carbon in both margins of the South/Equatorial Atlantic, through the development and application of a regressive model associated with supervised Machine Learning. For this purpose, 19 regressive models were tested based on data from 7 cores obtained in DSDP and ODP expeditions. A first analysis of the OC distribution patterns in the studied cores showed that this accumulation was significantly different between both margins, which can be attributed to specific processes associated with water masses and atmospheric dynamics, and paleobathymetric evolution. The best performing predictive model ($R^2 = 0.69$; RMSE = 2.73) was a Gaussian process regression with a quadratic rational kernel function. The application of this model in a test dataset resulted in lower coefficients of determination, with $R^2 = 0.37$. Despite the good performance of the chosen regression model and its potential to predict organic carbon contents from physiographic factors and sedimentation rates, it proved to be inflexible for different adaptations from the training circumstance. The poor performance of the model application in a test dataset suggests that physiographic factors indirectly influence significant parameters in the modulation of OC accumulation, which are also affected by other non-physiographic factors. Thus, the limited power of physiographic factors and sedimentation rates as unique predictors for OC accumulation in marine sediments is expected.

Key-words: Cretaceous; machine learning; organic carbon

LISTA DE FIGURAS

- Figura 1 - Diagramas esquemáticos do ciclo do carbono e mecanismos de deposição em ambiente marinho (a) o ciclo de longo prazo do carbono e (b) diagrama com os sistemas de feedback que modulam a deposição de CO em sedimentos marinhos (modificado de Berner, 2003). Setas normais ilustram respostas positivas e setas com círculos, respostas negativas..... 19
- Figura 2 - Distribuição temporal dos principais Eventos Anóxicos Oceânicos conhecidos, com destaque para o T-OAE, OAE 1a e OAE 2, em amarelo, considerados os eventos mais significativos em termos de abrangência espacial e magnitude, modificado de Jenkyns (2010). 22
- Figura 3 - Modelo conceitual dos processos principais envolvidos na ocorrência de OAEs (Robinson et al., 2017)..... 23
- Figura 4 - A compartimentação atual dos segmentos do Atlântico Sul a partir de lineamentos estruturais que se prolongam entre as margens conjugadas (Beniest *et al.*, 2017). 26
- Figura 5 - Esquematisação do processo de abertura do Atlântico Sul e fácies associadas a cada etapa da evolução tectônica (Wen et al., 2019)..... 28
- Figura 6 - Distribuição anual de ajustes convectivos na coluna d'água. Regiões com grande número de ajustes convectivos representam locais de formação de águas profundas. (a) 94 Ma, pólo norte; (b) 71 Ma, pólo norte; (c) 94 Ma, pólo sul e (d) 71 Ma, pólo sul (Donnadieu et al., 2016). 31
- Figura 7 - As vertentes de aprendizado em Machine Learning e as respectivas classes e tipos de modelos pertencentes a cada uma delas..... 33
- Figura 8 - Exemplos de métodos de validação de modelos. (a) Validação cruzada com *5-folds* e (b) Validação por *holdout* com 20% de dados separados para teste. 34
- Figura 9 - Localização geográfica dos 22 sites inicialmente selecionados nas bases de dados do portal IODP. Em vermelho, os 7 sites resultantes da aplicação dos critérios descritos no item 3.1. Pontos amarelos: testemunhos inicialmente triados; pontos vermelhos: testemunhos utilizados neste trabalho; estrela: testemunho para teste do modelo..... 37
- Figura 10 - Interface da ferramenta de busca da plataforma SEDIS. 38
- Figura 11 - Exemplo da interface da ferramenta cinemática dentro do software GPLates, utilizado para a obtenção de dados espaciais dos testemunhos trabalhados. 40
- Figura 12 - Exemplo da reconstrução paleogeográfica dos sites estudados (pontos vermelhos no mapa) a partir do uso do software GPLates para 3 idades distintas do Cretáceo Superior – Albiano (100 Ma), Turoniano (90 Ma) e Maastrichtiano (70 Ma) – plotados em paleomapas de Scotese (2016). 41
- Figura 13 – Exemplo de *struct* em Matlab com os dados armazenados para cada site estudado..... 41
- Figura 14 - Perfis geoquímicos de COT dos 7 testemunhos estudados, em ambas as margens do Atlântico Sul ao longo do Cretáceo Superior. As barras cinzas representam

o intervalo aproximado de ocorrência dos principais OAEs globais (Locklair et al., 2011; Jenkyns, 2010). 45

Figura 15 - Histogramas de probabilidade da distribuição de COT durante o Cretáceo Superior considerando (a) totalidade de dados e (b) a divisão por margens, com valores p do teste de normalidade de Shapiro-Wilk; (c) Boxplot da distribuição do percentil 25-75% com indicação das medianas (traço vermelho) e valor p do teste U..... 46

Figura 16 - (a) Mapa indicando sentidos de fluxo estimados para a DBW e APT e posições relativas dos testemunhos do segmento equatorial (P1 e P4), modificado de Berrecoso et al. (2010) com paleogeografia do Cenomaniano, modificado de Scotese (2008); (b) Dinâmica de circulação profunda das massas d'água DBW e APT na região do Platô de Demerara (sites ODP1258, 1260 e 1261) baseados em valores de ϵNd no intervalo Cenomaniano-Santoniano (b1), OAE2 (b2) e Campaniano (b3), modificado de Martin et al. (2012); (c) modelo conceitual para o processo de armadilha de nutrientes, modificado de Berrecoso et al. (2010). 49

Figura 17 - Mapa indicando a extensão de atuação das diferentes faixas de influência da Célula de Hadley e o posicionamento dinâmico da Zona de Convergência Intertropical (ITCZ) e a localização dos testemunhos P1, P4, P5 e P6 deste estudo (modificado de Wagner et al., 2013). Verde escuro: faixas tropicais; verde claro: faixas transicionais, onde a influência de ventos alísios e ressurgência são maiores; marrom: faixas subtropicais, onde a influência de monções e drenagem continental é mais pronunciada; C/T: Cenomaniano-Turoniano. 51

Figura 18 - Nova circulação oceânica integrando Oceano Austral, Atlântico Sul e Atlântico Norte a partir de meados do Campaniano e valores de ϵNd das respectivas bacias oceânicas (Robinson & Vance, 2012). 52

Figura 19 - Modelo conceitual de circulação oceânica entre Oceano Austral e segmento sul do Atlântico Sul a partir do Albiano, modificado de Dummann et al. (2020), indicando posição dos testemunhos P2, P3 e P7; GBG: passagem da Bacia da Geórgia. 53

Figura 20 - Erros residuais comparativos entre valores simulados e observados (a) Erro residual vs. Resposta observada; (b) boxplots dos erros residuais para os testemunhos P1-P7. 58

Figura 21 - Erro residual vs. Resposta observada com dados (a) normalizados entre 0-1 (range) e (b) normalizados por escore-Z..... 59

Figura 22 - Distribuição dos erros residuais ao longo das idades do Cretáceo Superior, com destaque para os intervalos temporais de ocorrências dos principais OAEs desse período. 61

Figura 23 - Distribuição das variáveis preditoras e variável de resposta para os testemunhos P1, P4, P5 e P6, com dados normalizados por escore-Z. Lat: latitude; PaleoBat: paleobatimetria; DistCost: distâncias para a linha de costa; Tsed: taxa de sedimentação; COT: carbono orgânico total. 62

Figura 24 - Distribuição das variáveis preditoras e variável de resposta normalizadas por escore-Z ao longo do Cretáceo Superior. As porcentagens representam a parcela de amostras dentro do conjunto amostral referentes a cada um dos testemunhos. Lat: latitude; PaleoBat: paleobatimetria; DistCost: distâncias para a linha de costa; Tsed: taxa de sedimentação; COT: carbono orgânico total. 64

Figura 25 - Distribuição dos valores de erro residual para 3 faixas de intervalo de COT: 0-5 wt%, 5-10 wt% e >10 wt%. A seta vermelha indica a tendência do modelo em superestimar valores mais baixos de COT e subestimar valores mais elevados, alcançando a simetria em valores intermediários. As porcentagens correspondem à proporção de cada faixa de COT dentro do espaço amostral. 65

Figura 26 - Resposta observada vs. resposta modelada, com coeficiente de determinação de 0,69, considerando o modelo regressivo Rational Quadratic GPR (validação de 10-folds) adotado neste trabalho..... 66

Figura 27 – Aplicação do modelo regressivo escolhido nos dados preditivos dos testemunhos de teste (T1-T4). (a) Resposta observada vs Resposta modelada; (b) Distribuição do erro residual vs Resposta observada..... 67

Figura 28 – Fluxograma com as etapas envolvidas no treinamento e aplicação do modelo regressivo associado à Machine Learning supervisionada..... 68

LISTA DE TABELAS

Tabela 1 - Classificação de potencial de geração de rochas imaturas com base no conteúdo de COT (modificado de Peters & Cassa, 1994).....	14
Tabela 2 - Nomenclatura e posição geográfica atual dos 7 <i>sites</i> selecionados no portal IODP para treinamento (P1-P7) e dos 4 testemunhos de teste da aplicação do modelo (T1-T4).....	35
Tabela 3 - Resultados de R ² , RMSE e gráficos de resposta observada vs resposta modelada para cada modelo de regressão treinado. As piores performances estão destacadas em vermelho e as melhores, em verde.	55

Sumário

AGRADECIMENTOS.....	v
RESUMO.....	vi
ABSTRACT.....	vii
LISTA DE FIGURAS.....	viii
LISTA DE TABELAS.....	xi
1. INTRODUÇÃO.....	14
2. BASE TEÓRICA.....	17
2.1. Ciclo do carbono: da atmosfera ao soterramento.....	17
2.1.1. A deposição de CO no Atlântico Sul durante os Eventos Anóxicos Oceânicos do Cretáceo.....	21
2.2. O processo de abertura do Atlântico Sul.....	25
2.2.1. Evolução tectônica.....	25
2.2.2. Aspectos da circulação oceânica no Cretáceo.....	29
2.2.3. Machine Learning.....	32
3. MATERIAIS & MÉTODOS.....	34
3.1. Escolha dos <i>sites</i> de estudo e banco de dados de COT.....	34
3.2. Modelos de idades.....	38
3.3. Taxas de sedimentação.....	39
3.4. Paleolatitudes e paleolongitudes.....	39
3.5. Paleobatimetria.....	42
3.6. Paleodistância para a linha de costa.....	42
3.7. Tratamento de dados e geoestatística.....	43
3.7.1. Estatísticas básicas.....	43
3.7.2. Aplicação de <i>Machine Learning</i> Supervisionada.....	43
4. RESULTADOS & DISCUSSÃO.....	43
4.1. Acumulação de Carbono Orgânico nas margens oeste e leste do Atlântico Sul durante o Cretáceo Superior.....	44

4.2. Aplicação de modelos regressivos em <i>Machine Learning</i> para avaliar o papel da fisiografia e taxas de sedimentação na acumulação de Carbono Orgânico no Atlântico Sul durante o Cretáceo Superior.....	54
5. CONCLUSÃO	68
6. REFERÊNCIAS BIBLIOGRÁFICAS.....	71
Anexo A – <i>Script</i> para Interpolação de idades	82
Anexo B – <i>Script</i> para criação do grid lat/lon	82
Anexo C – <i>Script</i> para obtenção das paleobatimetrias e distâncias para a linha de costa	83
Anexo D – <i>Script</i> para cálculo das taxas de sedimentação	85
Anexo E – <i>Script</i> para treinamento do modelo <i>Rational Quadratic GPR (10-folds)</i>	85
Anexo F – Linha de código para aplicação do modelo em <i>dataset</i> de teste	86

1. INTRODUÇÃO

O ciclo do carbono está intimamente relacionado a questões ecológicas, climáticas, geológicas e econômicas que permeiam a nossa vida (Battin *et al.*, 2009). Do ponto de vista ambiental, a produtividade primária promove a remoção do dióxido de carbono da atmosfera, sustenta cadeias tróficas e, em escalas geológicas de tempo, conduz à oxigenação atmosférica, possuindo importância significativa no controle do clima e da vida na Terra (Cartapanis *et al.*, 2016). Sob as perspectivas geológica e econômica, a deposição de matéria orgânica (MO) e, conseqüentemente, do carbono orgânico (CO), pode levar à formação de rochas geradoras de hidrocarbonetos, a depender da atuação conjunta de outros fatores como a qualidade da matéria orgânica depositada, as condições de preservação do ambiente deposicional, temperatura/tempo de maturação, dentre outros (Bergman *et al.*, 2020). Os teores de carbono orgânico total (COT) também exercem influência sobre o potencial de geração da rocha. Na sua vasta maioria, os teores de COT de sedimentos marinhos variam de 0,2 a 2 wt% (Schoepfer *et al.*, 2015), enquanto teores anômalos (> 3 wt%) são comumente – mas não exclusivamente - registrados em ambientes marinhos de circulação restrita (e.g. mares interiores) ou observados em intervalos específicos do tempo geológico, geralmente associados a litologias denominadas folhelhos negros, do inglês *black shales* (Dummann *et al.*, 2020). De forma mais detalhada, Peters & Cassa (1994) classificaram o potencial de geração de rochas imaturas com base no conteúdo de COT, conforme demonstrado na Tabela 1.

Tabela 1 - Classificação de potencial de geração de rochas imaturas com base no conteúdo de COT (modificado de Peters & Cassa, 1994).

Potencial de geração de óleo	COT (wt%)
Pobre	0-0,5
Intermediário	0,5-1
Bom	1-2
Muito bom	2-4
Excelente	>4

O potencial econômico e as implicações paleoambientais da existência de folhelhos negros ricos em carbono orgânico (>4 wt%, chegando até 30 wt%) são objeto de debate há muito tempo (Pettijohn, 1957). Até a década de 1960, acreditava-se que a gênese dessas rochas se dava exclusivamente em ambientes lacustres, marinhos de circulação restrita (e.g., Mar Negro) ou em bacias rasas e estratificadas associadas a fiordes ou estuários (e.g., Mar Báltico) (Arthur & Sageman, 1994). Embora estes ambientes sejam realmente propícios à sedimentação mais intensa e preservação de matéria orgânica, os modelos propostos não incluíam outras possibilidades, como a de ambientes marinhos francamente abertos. Isto se deve a dois fatores principais: (i) a limitação da existência de análogos modernos capazes de explicar a acumulação e preservação de matéria orgânica de maneira tão eficiente em ambiente marinho e; (ii) a inexistência, até então, de evidências de folhelhos negros associados a paleoambientes marinhos abertos.

Com o início das expedições dos programas internacionais de pesquisa *Deep Sea Drilling Project* (DSDP) na década de 1970 e *Ocean Drilling Program* (ODP), na década de 1980, foram descobertos intervalos de folhelhos negros marinhos depositados em platôs e regiões distais de bacias oceânicas diversas, sobretudo no Pacífico, associados a paleoambientes deposicionais que implicavam em configurações paleobatimétricas não condizentes com os modelos propostos até então. Os dados provenientes do DSDP também permitiram entender que esses horizontes ricos em matéria orgânica se distribuem global e sincronicamente, e foram depositados em curtos intervalos de tempo (<1 Myr). As constatações indicaram que esses intervalos não poderiam ser explicados apenas por condições de restrição de batimetria de bacia e, pela primeira vez, foi proposto o termo Evento Anóxico Oceânico (do inglês *Oceanic Anoxic Event* – OAE) (Schlanger & Jenkyns, 1976) de forma a definir as camadas de folhelhos negros marinhos ricos em carbono orgânico (CO).

A descoberta dos OAEs levantou hipóteses acerca dos mecanismos adicionais necessários para a ocorrência dos folhelhos negros, com dois pontos teóricos principais se destacando: (i) aumento da capacidade de preservação de matéria orgânica, causado pela estagnação de águas profundas e estratificação na coluna d'água marinha, com expansão das condições de anoxia (Demaison & Moore, 1980; Pedersen & Calvert, 1990) e (ii) aumento excessivo da produtividade primária marinha, com consequente expansão da zona de mínimo oxigênio (ZMO) e aumento do fluxo de carbono (Pedersen & Calvert, 1990). Apesar de ainda controversos, os modelos mais recentes propostos tendem a favorecer a segunda hipótese (Jenkyns, 2010; Mort *et al.*,

2008; Nederbragt *et al.*, 2004). Do ponto de vista geotectônico e paleoclimático, segundo (Kidder & Worsley, 2010), muitos dos picos de formação de rochas geradoras coincidiram com a ocorrência de grandes províncias ígneas (do inglês *Large Igneous Provinces* - LIP) e OAEs em meio a um cenário climático caracterizado por elevados teores de CO₂ atmosférico (>1000 ppmv) (Foster *et al.*, 2017), ausência de calotas polares e níveis eustáticos de até 250 m acima dos níveis atuais (Haq, 2014).

Rochas ricas em CO são encontradas em intervalos do Cretáceo médio/Superior de diversas bacias de margem passiva no Atlântico Sul e Equatorial, sendo consideradas geradoras marinhas importantes, como por exemplo na Bacia do Espírito Santo (folhelhos da Fm. Urucutuca; (França *et al.*, 2007) e na Bacia de Sergipe-Alagoas (folhelhos das Fm. Riachuelo e Cotinguiba; (ANP, 2017). Este é um dos motivos que despertam grande interesse nos meios científico e industrial sobre esta bacia oceânica (Beckmann *et al.*, 2008; Dummann *et al.*, 2021).

Com o apoio massivo da indústria de óleo e gás (O&G), modelos preditivos de fácies orgânica têm sido desenvolvidos desde o início da década de 1990 (Carpentier *et al.*, 1991; Passey *et al.*, 1990; Schwarzkopf, 1993) e aplicados em diversas bacias petrolíferas na tentativa de prever o conteúdo e a qualidade de CO depositado, bem como a distribuição temporal das principais rochas geradoras de hidrocarbonetos em ambientes marinhos, com foco principal na minimização dos riscos exploratórios. Dentre esses modelos, destacam-se o *OF-Mod (Organic Facies Model)* (Mann & Zweigel, 2008), e o *DORS (Dionisos Organic-Rich Sediments)* (Granjeon & Chauveau, 2014) desenvolvido pelo SINTEF-Noruega e IFPEN-França, respectivamente. Estas soluções enumeram diversos aspectos a serem considerados, tais como: (1) produtividade primária; (2) aporte de material orgânico terrígeno; (3) preservação diferencial do CO; (4) fluxo de CO; (5) eficiência de preservação do CO, dentre outros (Granjeon & Chauveau, 2014; Mann & Zweigel, 2008). Apesar dos fatores fisiográficos e as taxas de sedimentação afetarem direta e/ou indiretamente todos os aspectos considerados nos modelos de fácies orgânica, suas contribuições individuais para o acúmulo de CO nunca foram avaliadas. Dentre as limitações para esse tipo de avaliação está a disponibilidade de extensos bancos de dados (*Big Data*) que permitam o estabelecimento de uma maior robustez estatística. Essa limitação tem sido ultrapassada nos últimos anos em função das políticas de disponibilização de dados científicos (*Big Data Repositories*). Nesse novo contexto, ferramentas de Inteligência Artificial, dentre as quais incluem-se as técnicas de *Machine Learning*, têm sido cada vez mais empregadas em diversas áreas das geociências (Karpatne *et al.*, 2019a)

visando a identificação de padrões, feições, eventos e, até mesmo, o desenvolvimento de modelos preditivos (Alaudah *et al.*, 2019).

O objetivo deste trabalho consiste no treinamento de modelos regressivos associados à *Machine Learning* visando avaliar quantitativamente a influência de aspectos fisiográficos (nomeadamente latitude, paleobatimetria e paleodistâncias para a linha de costa) e taxas de sedimentação no acúmulo de CO no Atlântico Sul e Equatorial ao longo do Cretáceo Superior, o qual é marcado por importantes alterações fisiográficas decorrentes da abertura desta bacia oceânica (Donnadieu *et al.*, 2016). Assim sendo, por meio da análise de dados oriundos de 7 testemunhos recuperados durante expedições do DSPD e ODP, pretende-se trazer respostas às seguintes questões:

- A acumulação de CO no Atlântico Sul/Equatorial é homogênea durante o Cretáceo Superior ou existem diferenças entre as margens leste e oeste?
- Que tipo de modelo regressivo aplicado à *Machine Learning* é o mais adequado para prever os conteúdos de CO durante o Cretáceo Superior a partir de fatores fisiográficos e taxas de sedimentação?
- Com base no melhor modelo regressivo, qual a contribuição dos fatores fisiográficos e taxas de sedimentação para o acúmulo de CO no Atlântico Sul durante o Cretáceo Superior?

2. BASE TEÓRICA

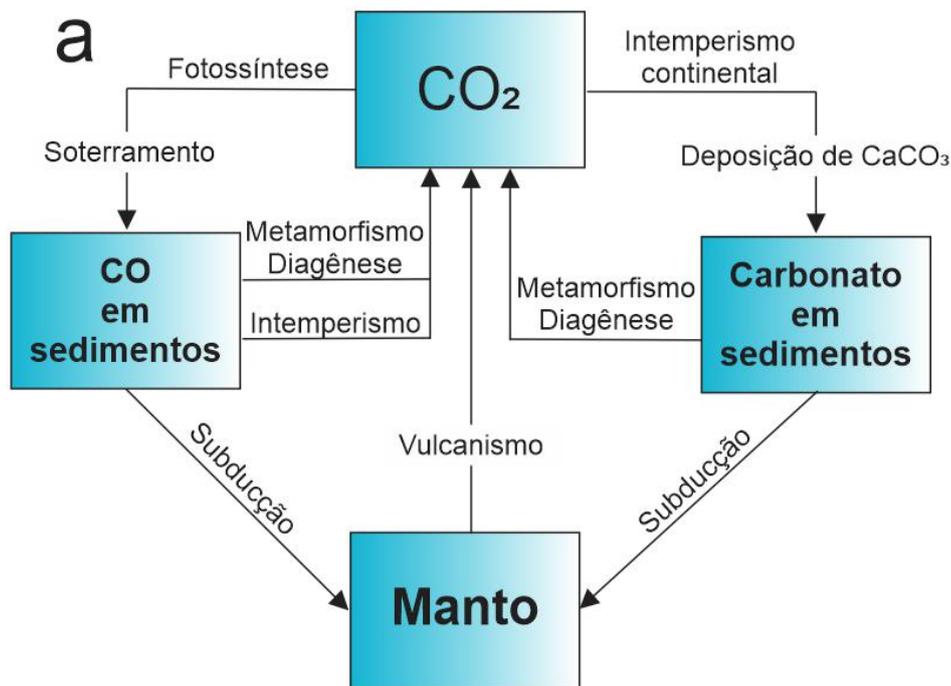
2.1. Ciclo do carbono: da atmosfera ao soterramento

O ciclo biogeoquímico do carbono, em escalas de tempo geológico, consiste em uma variedade de processos que podem ser simplificados estequiometricamente em duas equações distintas (Berner, 2003):

(1) $\text{CO}_2 + \text{CaSiO}_3 \leftrightarrow \text{CaCO}_3 + \text{SiO}_2$, que, da esquerda para a direita, representa a formação de carbonatos a partir do intemperismo de silicatos promovidos com a mobilização de CO_2 atmosférico e no sentido oposto, resulta na devolução de CO_2 à atmosfera a partir da degradação dos carbonatos soterrados na zona da diagênese;

(2) $\text{CO}_2 + \text{H}_2\text{O} \leftrightarrow \text{CH}_2\text{O} + \text{O}_2$, que é a reação da fotossíntese, fundamental para o entendimento da produtividade primária e a subsequente deposição de carbono orgânico em sedimentos. Esse material eventualmente transforma-se em querogênio, betume, óleo, gás e carvão, a depender do tipo de fácies orgânica em questão e da atuação dos processos de maturação térmica (Mendonça Filho *et al.*, 2012). No sentido inverso, a reação representa o intemperismo oxidativo da matéria orgânica exposta em ambientes subaéreos.

Conceitualmente, o ciclo de longo prazo do carbono pode ser sintetizado conforme o modelo abaixo (Fig. 1a):



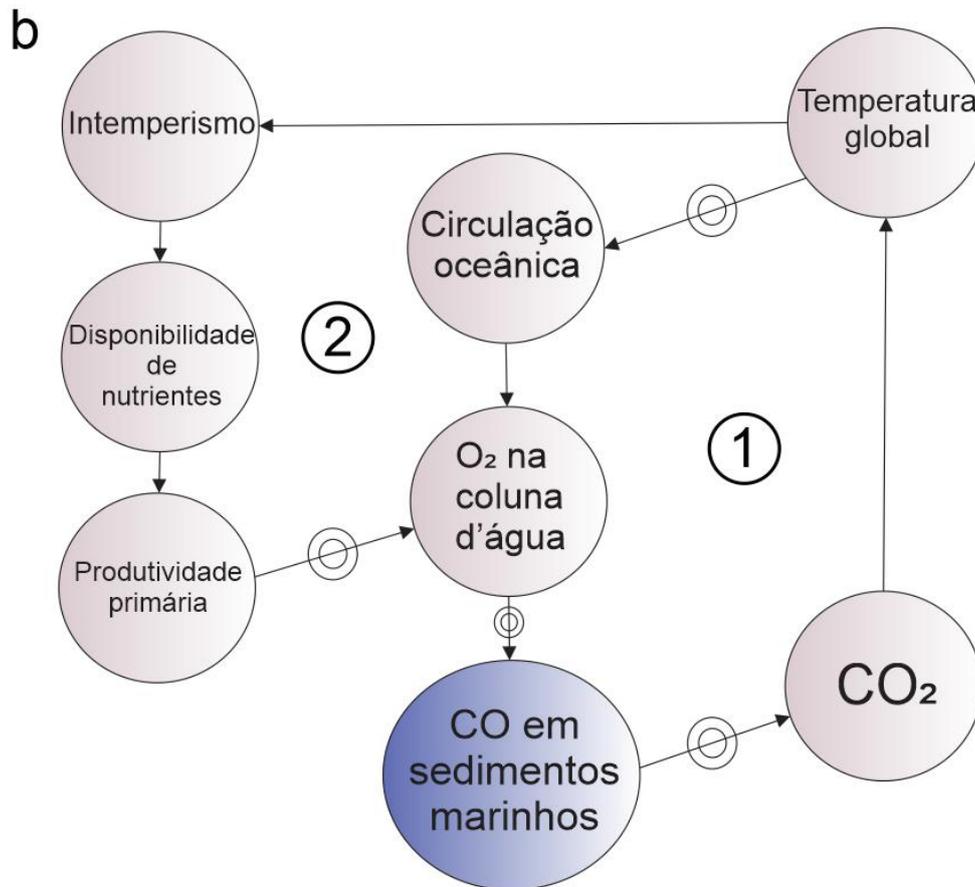


Figura 1 - Diagramas esquemáticos do ciclo do carbono e mecanismos de deposição em ambiente marinho (a) e o ciclo de longo prazo do carbono e (b) diagrama com os sistemas de feedback que modulam a deposição de CO em sedimentos marinhos (modificado de Berner, 2003). Setas normais ilustram respostas positivas e setas com círculos, respostas negativas.

Berner (2003) discute em seu estudo os complexos sistemas de *feedback* atuantes entre o soterramento de CO, ciclagem de nutrientes, clima e composição química atmosférica. A figura 1b, modificada deste trabalho, apresenta alguns dos diversos parâmetros que controlam a deposição de CO através de respostas positivas (setas normais), e negativas (setas com círculos). O tamanho das setas não está relacionado a nenhum tipo de avaliação quantitativa ou representação de maior/menor importância do parâmetro em questão. Analisando-se a figura, depreende-se, pelo ciclo 1, que o aumento do soterramento de CO nos oceanos promove a diminuição do CO₂ atmosférico e, conseqüentemente, a redução do efeito estufa global. O resfriamento do planeta aumenta o gradiente de temperatura entre altas e baixas latitudes, contribuindo para a intensificação da circulação/ventilação oceânica e limitando condições de anoxia fundamentais para o acúmulo de CO em sedimentos. De forma análoga, no ciclo 2, o resfriamento global contribui para a diminuição do intemperismo e, por conseqüência, do aporte de nutrientes aos oceanos. A queda da produtividade primária leva, em última

instância, à redução do acúmulo de CO em sedimentos marinhos. É importante ressaltar que, em ambos os *loops* descritos, o aumento inicial do soterramento de CO resulta, ao final de cada ciclo, em sua redução, evidenciando como este parâmetro se automodula e controla os níveis de dióxido de carbono atmosférico. Assim, pode-se admitir que o soterramento de CO em bacias oceânicas exerce controle sobre a química dos oceanos e da atmosfera (Garrels *et al.*, 1976).

Alguns dos parâmetros apresentados na figura 1b são, de forma indireta, influenciados ou relacionados a parâmetros fisiográficos e taxas de sedimentação que, em último grau, contribuem para o controle – ou para parte da explicação empírica - da deposição de CO da seguinte forma:

- (i) a latitude está intimamente associada a características climáticas que controlam índices de pluviosidade e, conseqüentemente, atuam sobre o intemperismo e o aporte de nutrientes/materiais terrígenos em bacias oceânicas através do escoamento superficial e descargas fluviais. Ademais, a latitude também modula a intensidade de insolação, produzindo efeitos sobre a produtividade primária (Mélières *et al.*, 1997);
- (ii) a paleobatimetria e sua evolução relaciona-se fortemente com a paleocirculação oceânica, uma vez que sua variação está ligada à abertura/fechamento de passagens oceânicas (ou *marine gateways* na literatura corrente) e elevação/subsidência do assoalho oceânico resultante de processos tectônicos (Pérez-Díaz & Eagles, 2017). Além disso, a paleobatimetria controla o tempo que o material orgânico levará para percorrer a coluna d'água até a sua deposição no assoalho oceânico, tendo efeito decisivo na preservação de CO (Emmel *et al.*, 2018);
- (iii) a distância em relação à costa está ligada empiricamente ao potencial de determinado ponto geográfico em receber aportes de material continental que contribuem para o inventário de carbono orgânico autóctone e alóctone. Os processos de contribuição continental e dispersão sobre o ambiente marinho estão vinculados sob um mesmo arcabouço físico. Em contrapartida, a entrada de material detrítico pode promover a diluição do carbono orgânico depositado, contribuindo para menores teores de carbono orgânico total (COT) nas rochas sedimentares; e

(iv) as taxas de sedimentação que, de maneira genérica, podem representar a propriedade de determinado ambiente deposicional em acumular material sedimentar. É interessante pontuar que, apesar de ser intuitivo pensar que a taxa de sedimentação varia de forma inversamente proporcional à distância para a linha de costa, mais fatores precisam ser levados em consideração. As especificidades oceanográficas de escala local e a topografia de fundo possuem forte influência na dispersão do material sedimentar e trazem complexidade para o entendimento dessa relação (Pérez-Díaz & Eagles, 2017). Além disso, as características químicas do material a ser depositado (lábil/refratário), associadas às condições redox da coluna d'água, por exemplo, também são fatores importantes a serem apontados e modulam significativamente os valores de taxas de sedimentação em bacias oceânicas (Bianchi *et al.*, 2017).

2.1.1. A deposição de CO no Atlântico Sul durante os Eventos Anóxicos Oceânicos do Cretáceo

Como já mencionado ao longo da introdução deste trabalho, o período Cretáceo reuniu características geológicas e paleoclimáticas específicas que culminaram em momentos de significativa deposição e preservação de carbono orgânico marinho durante intervalos curtos de tempo, que foram batizados de Eventos Anóxicos Oceânicos (Schlanger & Jenkyns, 1976). As ocorrências que levaram à definição de OAEs por esses autores resumiam-se aos intervalos marinhos do Aptiano-Albiano e Cenomaniano-Turoniano do Oceano Pacífico. Desde então, outros intervalos similares e não restritos ao Cretáceo foram descobertos e caracterizados como OAEs (Fig. 2; Jenkyns, 2010).

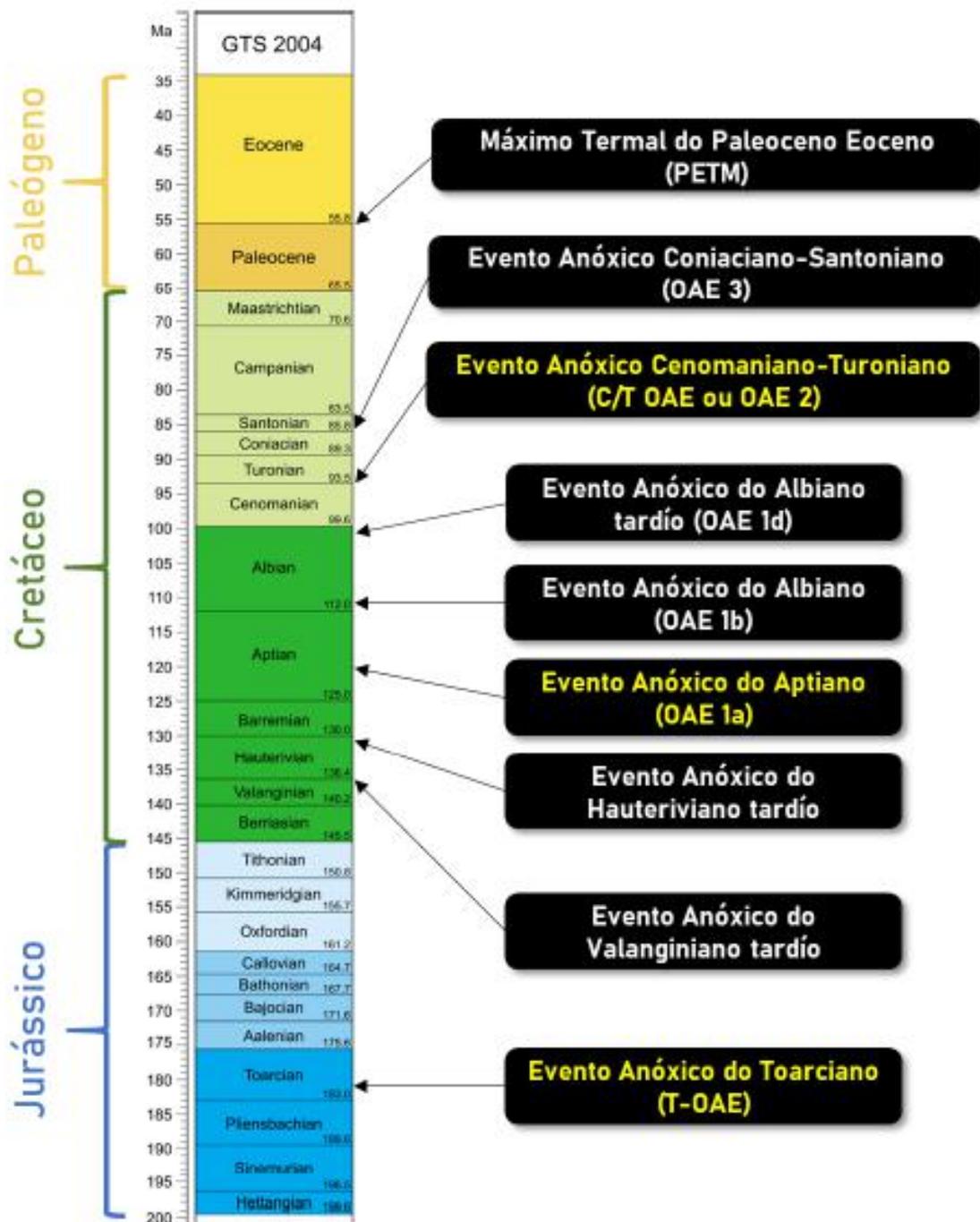


Figura 2 - Distribuição temporal dos principais Eventos Anóxicos Oceânicos conhecidos, com destaque para o T-OAE, OAE 1a e OAE 2, em amarelo, considerados os eventos mais significativos em termos de abrangência espacial e magnitude, modificado de Jenkyns (2010).

Devido à pluralidade de mecanismos e forçantes potencialmente envolvidos na ocorrência dos OAEs, há propostas de modelos conceituais generalistas para explicá-las (Fig. 3) (Robinson *et al.*, 2017). Porém, cada Evento pode ter maior ou menor participação de tais fatores principais, também sendo influenciados por condições paleogeográficas e de circulação específicas do intervalo temporal em questão

(Donnadieu *et al.*, 2016). Esses eventos representam situações bastante particulares no tempo geológico que culminaram em acumulações anômalas de carbono orgânico em horizontes de grande interesse econômico.

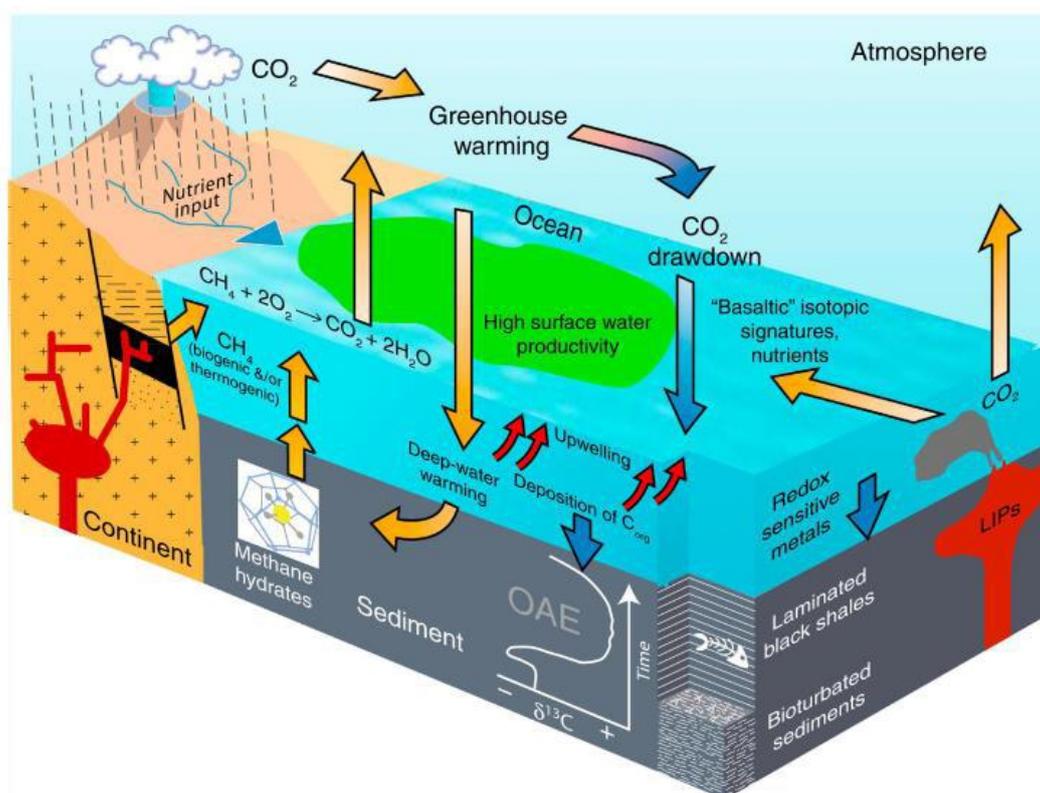


Figura 3 - Modelo conceitual dos processos principais envolvidos na ocorrência de OAEs (Robinson *et al.*, 2017).

Dentro do intervalo de tempo de interesse deste trabalho – o Cretáceo Superior (~100,5 – 66 Ma) – 3 OAEs globais são reconhecidos:

- (i) OAE 1d (Evento Breistroffer, Albiano-Cenomaniano, ~100 Ma), reconhecido inicialmente em um horizonte da Bacia Vocontiana, na França, onde há ocorrência de intercalações de folhelhos negros e aumento da população de foraminíferos e radiolários (Breistroffer, 1937; Erbacher & Thurow, 1997). Este evento possui outros equivalentes localizados no Atlântico e em calcários pelágicos do Complexo Franciscano, na Califórnia (Robinson *et al.*, 2008; Wilson & Norris, 2001). No contexto do Atlântico Sul, testemunhos provenientes de expedições do DSDP e ODP registram a ocorrência do OAE 1d de forma mais proeminente na região central (*sites* DSDP363 e 364) e equatorial (Site ODP962), e de forma menos óbvia em *sites* localizados na região mais ao sul (*sites* DSDP356, 327 e 511) (Wilson & Norris, 2001). Do

ponto de vista geoquímico, as observações do OAE 1d são mais bem caracterizadas no Atlântico Norte (Hofmann *et al.*, 2000; Wilson & Norris, 2001) mas indicam, através de dados de isótopos de oxigênio e carbono, que este foi um evento global, relacionado ao colapso térmico de massas d'água superficiais, proporcionando grandes acúmulos de carbono orgânico (COT >25 wt%) (Hofmann *et al.*, 2000). As extrapolações deste evento para as localidades do Atlântico Sul se dão, principalmente, por similaridades na variação litológica das colunas estratigráficas, partindo de padrões sutis de laminação em lamitos carbonáceos, para folhelhos negros bem desenvolvidos (Wilson & Norris, 2001);

- (ii) OAE 2 (Evento Bonarelli, Cenomaniano-Turoniano, ~93 Ma), um dos mais bem documentados OAEs, reconhecido já nas primeiras expedições do DSDP (Schlanger e Jenkyns, 1976) e correlacionado aos afloramentos do intervalo conhecido como *Livello Bonarelli* (do italiano, nível Bonarelli). Localizado nos Apeninos italianos, caracteriza-se por um intervalo de cerca de 1 metro de espessura de folhelho escuro com valores de COT chegando até a 30 wt% (Tsikos *et al.*, 2004). Algumas observações deste evento são identificadas no hemisfério Sul em *sites* do DSDP e nas bacias de Santos, Campos, Sergipe-Alagoas e Ceará (Arai, 1988; Forster *et al.*, 2008; Koutsoukos *et al.*, 1990; Mello *et al.*, 1989; Raven *et al.*, 2019; Valle *et al.*, 2019). Valle *et al.* (2019) sugerem que, na região da Bacia de Sergipe-Alagoas, além do OAE 2, outro evento de anoxia/disoxia também ocorreu no início do Cenomaniano (LCAE, do inglês *Lower Cenomanian Anoxic Event*). Nesta bacia, os episódios de anoxia são evidenciados por concentrações elevadas de pirita, diminuição de organismos bentônicos e excursões positivas de $\delta^{13}\text{C}_{\text{carbonato}}$, como tipicamente observado para este intervalo (Jenkyns, 2010). Em relação aos valores de COT, para o OAE 2 os valores estão próximos ao *background* (~1 wt%) do intervalo em fácies carbonáticas com picos de COT (~ 4 wt%) durante o LCAE (folhelhos e margas), evidenciando a importância do uso integrado de outros *proxies* para a identificação desses eventos a depender da variação litológica local; e
- (iii) OAE 3 (Coniaciano-Santoniano, ~86 Ma) que, diferentemente dos demais Eventos, possui ocorrência de maior duração (~5.5 Myr) (Locklair *et al.*, 2011) e excursões positivas de $\delta^{13}\text{C}_{\text{carbonato}}$ mais modestas do que a maioria

dos demais Eventos (~0.5 ‰). O OAE 3 ocorre em intervalos cronocorrelatos com aumento de teores de COT, bem distribuídos ao longo de todo o Atlântico Norte e Sul, tendo também já sido identificado em regiões do Oeste dos Estados Unidos, no Mar Interior Ocidental (Wagner *et al.*, 2004). Na costa africana, o OAE 3 é evidenciado em regiões *offshore* do Gabão (Katz *et al.*, 2000), Angola (Schoellkopf & Patterson, 2000) e Namíbia (Holtar & Forsberg, 2000); e em *sites* de expedições ODP (*site* 959) (Wagner *et al.*, 2004) e DSDP (*sites* 364, 530 e 361) (Stein *et al.*, 1989). Na margem brasileira, esses intervalos já foram observados no platô de São Paulo (DSDP356) (Stein *et al.*, 1989) e em intervalos das bacias de Campos, Sergipe-Alagoas e Ceará (Mello *et al.*, 1989). Observações também já foram feitas na região do Platô de Demerara (DSDP1261) (Wagreich, 2012). A ocorrência dos intervalos de folhelhos escuros relacionados a este Evento é explicada por episódios de anoxia de zona fótica (Wagner *et al.*, 2004). No entanto, comumente se intercalam intervalos de maior oxigenação da coluna d'água, evidenciados pela presença dos CORBs (do inglês *Cretaceous Oceanic Red Beds*) (Wagreich, 2012).

2.2. O processo de abertura do Atlântico Sul

2.2.1. Evolução tectônica

O processo de abertura do Atlântico Sul, que remonta à separação do supercontinente Gondwana, teve seu início no Cretáceo Inferior (~130 Ma) (Mohriak, 2003; Torsvik *et al.*, 2009) por meio da reativação de grandes sistemas de falhas transformantes promovida por plumas mantélicas ascendentes (Fromm *et al.*, 2015). No contexto desses sistemas, dois grandes lineamentos estruturais viabilizam a setorização do Atlântico Sul em 2 segmentos distintos e facilitam o entendimento da evolução do processo tectônico, a saber: a Zona de Fraturas de Florianópolis, que separa os segmentos sul e central do Atlântico Sul; e a Zona de Fraturas Romanche, que estabelece a divisão entre os segmentos central e norte ou equatorial. A porção inferior do segmento sul é, ainda, limitada pela Zona de Fraturas Agulhas-Falklands (Fig. 4) (Beniest *et al.*, 2017).

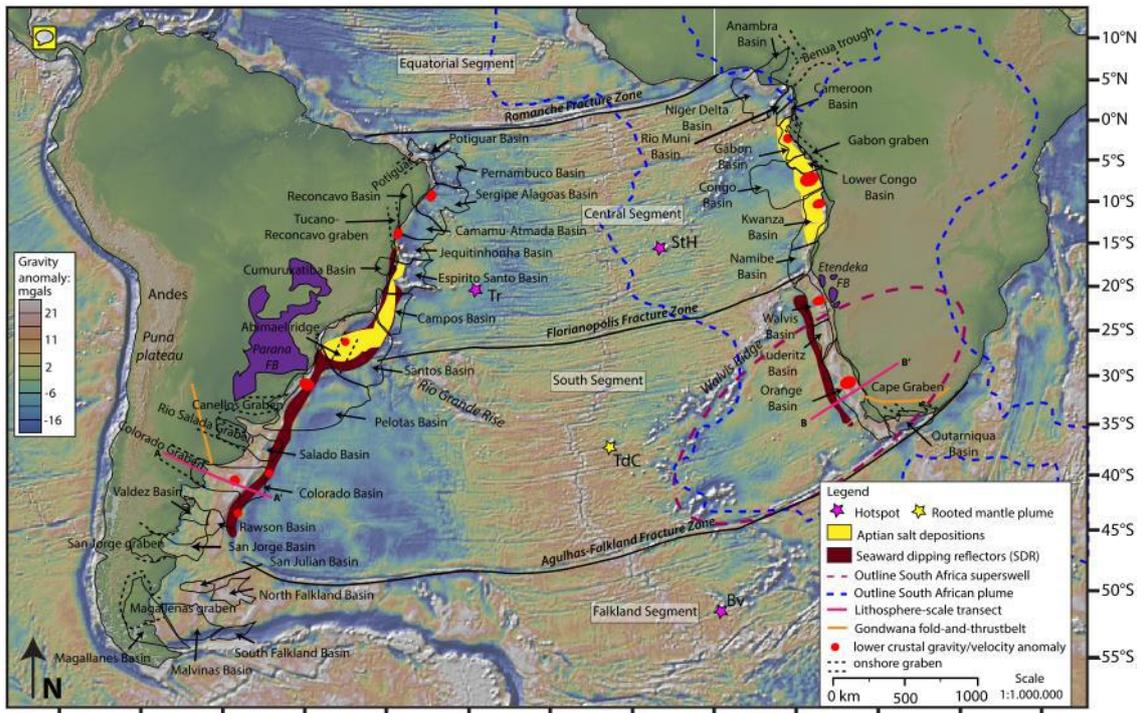


Figura 4 - A compartimentação atual dos segmentos do Atlântico Sul a partir de lineamentos estruturais que se prolongam entre as margens conjugadas (Beniest *et al.*, 2017).

De acordo com (Wen *et al.*, 2019), a evolução das margens conjugadas do Atlântico Sul pode ser dividida em 3 fases principais (Fig. 5):

- (i) Barremiano (129 – 125 Ma, Fig. 5a), com domínio de sistemas deposicionais continentais (fluviais e lacustres) em uma fase rifte com gradual transição para sistemas deposicionais marinhos no limite inferior do segmento sul, porção onde os esforços extensionais da separação de Gondwana se iniciaram e por onde houve conexão com o Oceano Índico. O rifte se desenvolveu em direção ao norte, com estabelecimento de sistemas de vales em rifte de orientação predominante N-S junto a uma série de falhamentos ortogonais, de orientação E-W, que criaram bacias continentais do tipo *pull-apart* sobretudo em latitudes próximas à fronteira entre os segmentos central e equatorial (e.g. bacias interiores do Nordeste). O desenvolvimento do sistema de falhas normais N-S relacionado à subsidência tectônica deu origem a lagos profundos (grábens) onde ocorreu a deposição de folhelhos associada a condições anóxicas que promoveram significativa preservação de material orgânico nos sedimentos (Karner *et al.*, 1997). Sob este contexto, e considerando as etapas subsequentes de maturação do material orgânico lacustre, as rochas sedimentares depositadas - dentro da extensão geográfica observada na

figura 5c – tornaram-se intervalos de rochas geradoras dos sistemas petrolíferos do pré-sal (Oliveira *et al.*, 2019).

(ii) Aptiano (~125-113 Ma, Figs. 5b, c), marcado por uma fase transicional, com incursão marinha significativa no segmento sul. O alto estrutural de direção aproximada E-W promovido pela Elevação do Rio Grande e Cadeia Walvis atuou como barreira para o avanço marinho, promovendo a ocorrência de ambiente lagunar restrito e sedimentação carbonática no segmento central do Atlântico Sul. No segmento norte, sob o controle dos cinturões transformantes E-W, o sistema de riftes era relativamente pequeno e estreito, ainda com o predomínio de fácies lacustres. Ao fim do Aptiano, evaporitos se formaram em larga escala no segmento central. Compostos predominantemente de halita e anidrita, essas rochas cobrem uma área de cerca de 1 milhão de km², ocorrendo desde a bacia de Santos até a bacia de Sergipe-Alagoas, e atingem espessuras de até 4.000 m (Kukla *et al.*, 2018). Dentro dos sistemas petrolíferos do pré-sal, essas extensas e espessas camadas de sal atuam como eficientes rochas selantes, limitando a migração do óleo gerado e permitindo que este se concentre nas rochas reservatório predominantemente carbonáticas (Mohriak *et al.*, 2012).

(iii) A partir do Albiano (<113 Ma, Fig. 5d), fase na qual o intervalo de interesse deste trabalho se inclui, formou-se no Atlântico Sul um centro de espalhamento de crosta oceânica *sensu strictu*, com separação efetiva das placas e começo da deriva continental (fase drifte). A separação dos dois continentes provocou aprofundamento progressivo da margem continental divergente e desenvolvimento de espessos depocentros (Mohriak, 2003). Do Albiano ao Maastrichtiano, um sistema transgressivo se desenvolveu com o nível eustático máximo de cerca de 250 m acima do nível atual ocorrendo na transição Cenomaniano-Turoniano (Haq, 2014), momento em que condições de anoxia contribuíram para a deposição e preservação de folhelhos negros sobretudo nos segmentos central e norte do Atlântico Sul (Wagner *et al.*, 2013). Desde o Paleoceno, a queda do nível eustático associada à existência de fontes suficientes de sedimentos em ambas as margens resultou em deltas progradantes e fluxos gravitacionais de águas profundas (turbiditos, majoritariamente), como se observa nos deltas do Níger, Foz do Amazonas e Pelotas (Wen *et al.*, 2019; Fig. 5e).

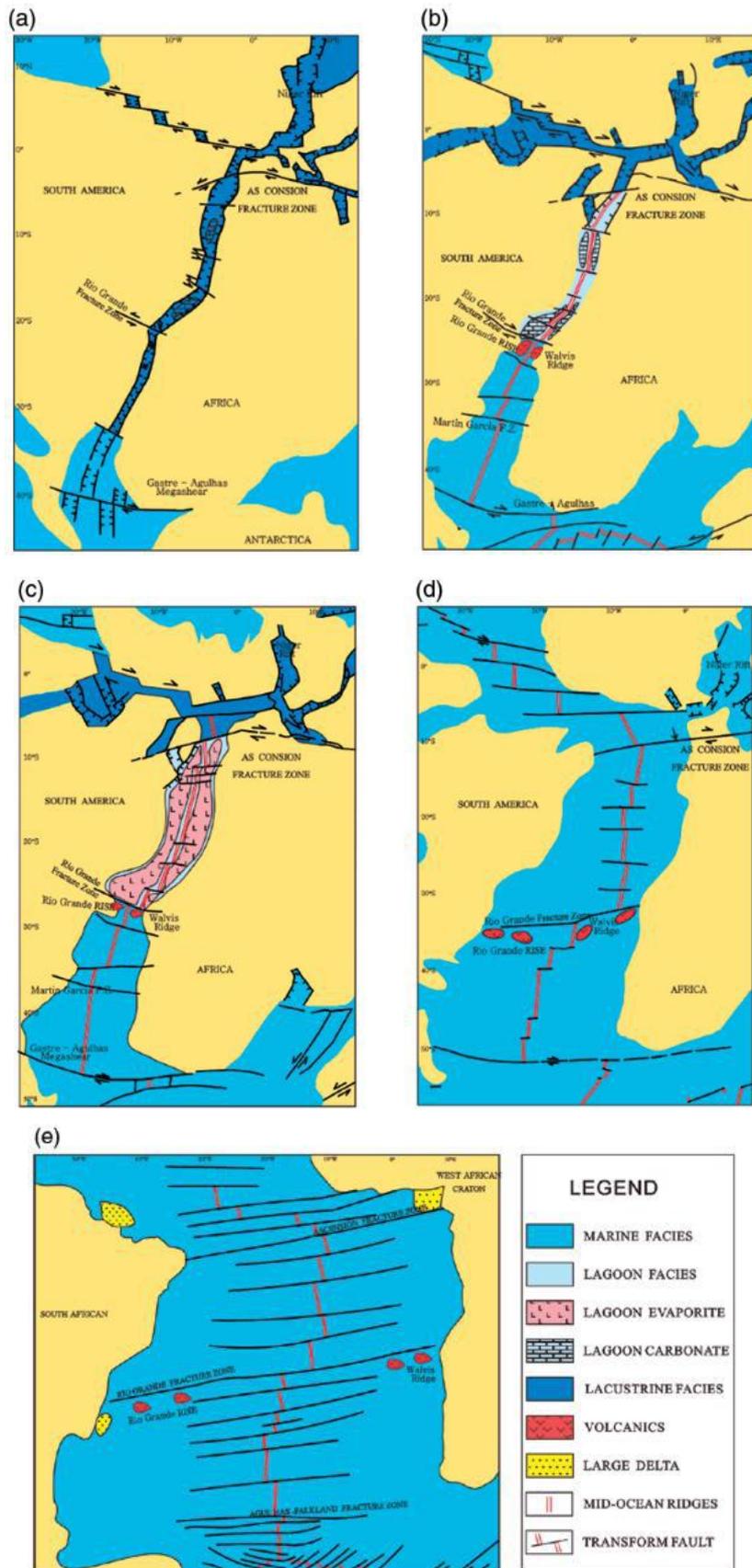


Figura 5 - Esquematização do processo de abertura do Atlântico Sul e fácies associadas a cada etapa da evolução tectônica (Wen *et al.*, 2019).

Atualmente, as margens conjugadas do Atlântico Sul possuem cerca de 25 bacias exploradas que, em conjunto, abrigam 111 campos de petróleo e gás com reservas recuperáveis de cerca de 256.000 MMboe (Wen *et al.*, 2019). Essa estimativa representa mais de 70% de todo óleo e gás equivalente descoberto em margens passivas do mundo todo (Thompson *et al.*, 2015).

2.2.2. Aspectos da circulação oceânica no Cretáceo

As características ambientais e paleogeográficas do planeta durante o Cretáceo contribuíram para que os padrões de circulação oceânica naquele tempo fossem bastante diversos dos observados atualmente. Segundo Hay (2008), a estabilidade dos sistemas modernos de ventos mantida pelas altas pressões polares era inexistente devido à ausência de calotas polares e do albedo promovido por elas, com consequência diretas na dinâmica de massas d'água oceânicas. Quanto à configuração dos continentes, a inexistência de um corredor oceânico bem desenvolvido no Atlântico Sul impedia a conexão profunda entre o Oceano Austral e o Atlântico Norte, atribuindo papel relevante à Célula de Revolvimento Meridional do Pacífico (ou *Pacific Meridional Overturning Circulation* – PMOC) como modo dominante da dinâmica oceânica daquele período (Ferreira *et al.*, 2018).

No Atlântico Sul, sugere-se que a anoxia oceânica que viabilizou a ocorrência de OAEs no Cretáceo só teria sido desenvolvida devido a uma restrição significativa da circulação oceânica, influenciada, sobretudo, pela configuração paleogeográfica daquele período (Poulsen *et al.*, 2001). Apesar da maior parte dos estudos apontarem para o importante papel do aumento das entradas de nutrientes e material terrígeno como promotor principal da anoxia e, conseqüentemente, gatilho para um OAE (Meyer & Kump, 2008; Monteiro *et al.*, 2012), outros estudos têm considerado o controle, mesmo que parcial, promovido pela circulação oceânica profunda no desencadeamento desses eventos (Donnadieu *et al.*, 2016; Dummann *et al.*, 2020; Robinson & Vance, 2012). Neste contexto, Hay (2009) mostrou que, a despeito do clima quente do Cretáceo favorecer o intemperismo e a erosão, o *runoff* fluvial de carbonato e fósforo era 30% menor comparado às taxas atuais, sugerindo ser improvável que uma bacia oceânica como o Atlântico Sul tenha se tornado anóxica sem que estivesse fortemente isolada dos demais oceanos.

Dummann *et al.* (2020), de fato, concluem que a paleogeografia é fator primordial para as mudanças de circulação oceânica observadas no intervalo Barremiano-Aptiano

(~124 – 110 Ma), no Cretáceo Inferior. Os autores demonstram, por meio de modelagem de correntes e valores de ϵNd – um *proxy* geoquímico utilizado na determinação da proveniência de massas d'água - que, no início do Aptiano (~118 Ma), a maior abertura do Platô das Falklands permitiu o começo da circulação de águas intermediárias provenientes do Atlântico Sul em direção ao Oceano Austral. A partir de 110 Ma, uma nova alteração paleogeográfica conduzindo a abertura de mais um *gateway* – a passagem da bacia da Geórgia (do inglês *Georgia Basin Gateway, GBG*) – permitiu o início da circulação de águas profundas que adentravam o Atlântico Sul a partir do Oceano Austral. Em ambos os momentos, a intensificação da circulação contribuiu para o aumento da oxigenação no segmento sul do Atlântico Sul, induzindo menor preservação de CO nesta porção. É importante ressaltar a existência de trabalhos que se baseiam em dados paleontológicos da margem continental brasileira para levantar a hipótese de que a sedimentação marinha aptiana tenha ocorrido vinda de norte para sul, sob domínio de águas oriundas do Mar de Tétis, se contrapondo à ideia comum de que a ingressão marinha tenha se iniciado a partir do sul (Arai, 2009, 2014).

Donnadieu *et al.* (2016), por sua vez, mostraram que, no Cretáceo Superior, a anoxia/disoxia do oceano profundo também esteve intimamente relacionada a mudanças fisiográficas ocorridas ao longo do processo de abertura do Atlântico Sul. Com a utilização de modelos de circulação oceânica considerando paleogeografias da transição Cenomaniano-Turoniano (~94 Ma) e Maastrichtiano (~71 Ma), esses autores sugeriram que a mudança da localidade de formação de águas profundas no Oceano Austral (Fig. 6) e a intensificação da circulação profunda/reversão do fluxo através da passagem do Atlântico Central e do Caribe, no Maastrichtiano, poderiam ter induzido mudanças importantes na dinâmica das correntes oceânicas. Como consequência, o estado geral de oxigenação do Atlântico Sul e Equatorial teria sido impactado, contribuindo para o encerramento de eventos anóxicos em grande escala. Os autores ressaltam, ainda, que apesar de serem determinantes na ocorrência de OAEs, a paleogeografia não é o único fator que ativaria o “gatilho” desses eventos.

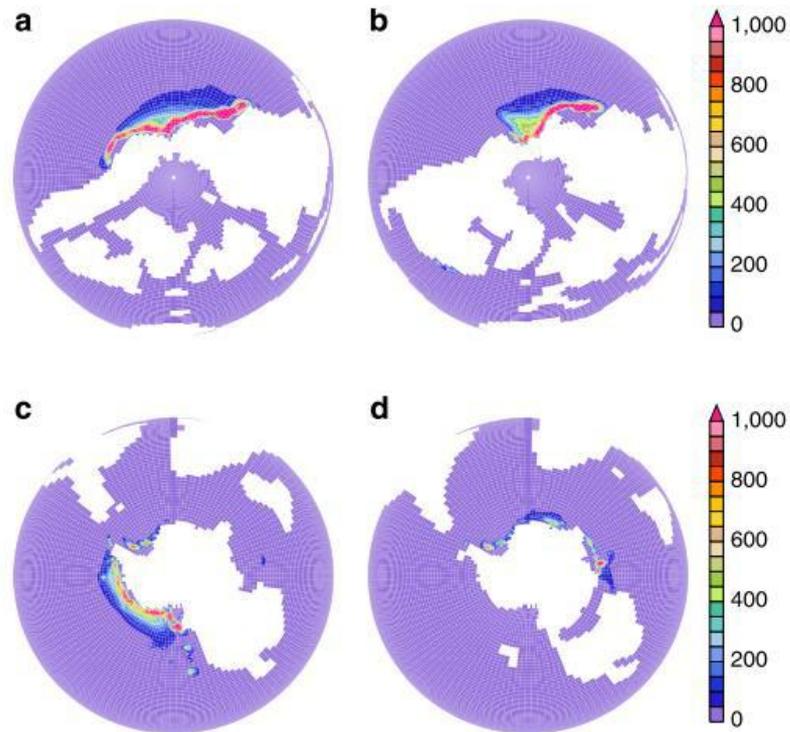


Figura 6 - Distribuição anual de ajustes convectivos na coluna d'água. Regiões com grande número de ajustes convectivos representam locais de formação de águas profundas. (a) 94 Ma, pólo norte; (b) 71 Ma, pólo norte; (c) 94 Ma, pólo sul e (d) 71 Ma, pólo sul (Donnadieu *et al.*, 2016).

Por meio da utilização do ϵNd para o Atlântico Norte e Sul, Robinson & Vance (2012) concluem que, no início/meados do Campaniano (~80 Ma), um novo padrão de circulação se iniciou nessas bacias oceânicas, integrando-as. De forma similar a Donnadieu *et al.* (2016), foi proposto que essa nova dinâmica oceânica, iniciada na segunda metade do Cretáceo Superior, teve forte contribuição de águas profundas formadas no Oceano Austral (e.g. Água do Componente Sul ou *Southern Component Water* – SCW) que fluíam em direção ao Atlântico Norte através da passagem do Atlântico Central já suficientemente desenvolvida para permitir a ocorrência da circulação de fundo. Essa configuração – águas profundas formadas nos setores Atlântico e Índico do Oceano Austral, e circulação fluindo em direção ao Atlântico Norte – caracterizaria a dinâmica oceânica da forma como conhecemos atualmente, ativando a Célula de Revolvimento Meridional do Atlântico (ou *Atlantic Meridional Overturning Circulation* – AMOC).

2.2.3. Machine Learning

Machine Learning ou aprendizagem de máquina consiste em um ramo de algoritmos computacionais vinculados ao conceito maior de inteligência artificial (Zhang & Ni, 2020). Desenvolvido para emular a inteligência humana com base no treinamento de modelos, esses algoritmos têm como objetivo fazer com que a máquina tome decisões e/ou faça previsões de uma variável qualitativa ou quantitativa - a variável de resposta - a partir de um conjunto de dados de entrada - as variáveis preditoras (Chang *et al.*, 2018). Técnicas relacionadas à *Machine Learning* estão em franca evolução e vêm sendo amplamente aplicadas em diversos campos do conhecimento como engenharia espacial, finanças, reconhecimento de padrões visuais, entretenimento, medicina, entre outros (El Naqa *et al.*, 2018).

Nas geociências, o avanço tecnológico das últimas décadas contribuiu para que essa disciplina ultrapassasse a fronteira entre o *Data-poor* e o *Big Data*, ou seja, os dados geocientíficos passaram a ser extensos demais para serem tratados por sistemas tradicionais (Karpatne *et al.*, 2019b). Neste sentido, técnicas de Machine Learning vêm sendo empregadas, sobretudo, no ramo da mineração (Rodríguez-Galiano *et al.*, 2015; Zuo, 2017), sismologia (Bergen *et al.*, 2019; Kong *et al.*, 2019), sensoriamento remoto (Lary *et al.*, 2016), geologia de engenharia (Goetz *et al.*, 2015; Marjanović *et al.*, 2011) e óleo e gás (Aliyuda *et al.*, 2020; Mohaghegh, 2020).

De forma geral, existem duas vertentes associadas à *Machine Learning* que diferem quanto à informação que é fornecida ao computador: a aprendizagem supervisionada e a aprendizagem não supervisionada. Na primeira modalidade, o usuário informa à máquina os valores ou classes da variável de resposta. No segundo caso, essa variável não é fornecida ao computador, o que faz com que a máquina agrupe os dados de entrada (clusterização) de acordo com padrões e/ou estruturas intrínsecas para prever resultados da variável de resposta (Ang *et al.*, 2016). A adoção por uma modalidade ou outra acarretará nas classes de modelo a serem utilizados na geração de um modelo preditivo (modelos de classificação, regressão ou clusterização, vide Fig. 7). Com base na escolha da classe de modelos desejada, existe, ainda, uma série de tipos de modelos que podem ser utilizados (Fig. 7).

Considerando a modalidade de aprendizagem supervisionada, os modelos de classificação são utilizados quando a variável de resposta é qualitativa. A classe de modelos regressivos, por sua vez, é utilizada quando a variável de resposta é quantitativa, ou seja, seus valores são numéricos.

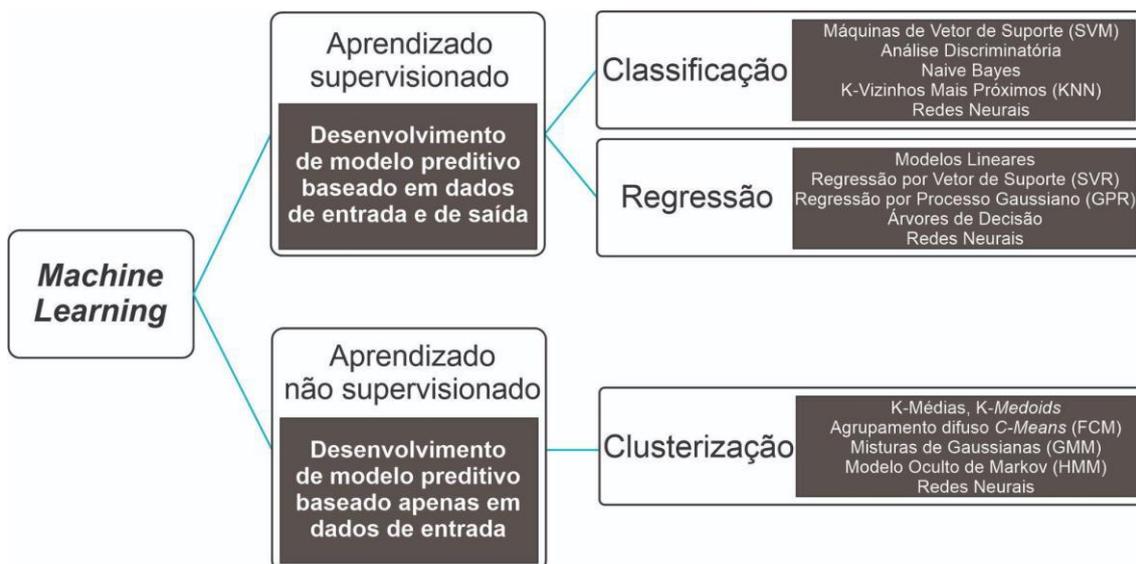


Figura 7 - As vertentes de aprendizado em Machine Learning e as respectivas classes e tipos de modelos pertencentes a cada uma delas¹.

Um passo mandatório para o treinamento de modelos é a seleção de uma técnica de validação, que consiste na forma como serão separadas as amostras que constituirão os grupos de treinamento e de teste. Dentre essas técnicas, destacam-se a validação cruzada e a validação por *holdout*. A validação cruzada consiste na definição de um valor k , que indica em quantas partes o conjunto amostral será subdividido. Em seguida, cada um desses subconjuntos será treinado, ao mesmo tempo em que são testados os $k-1$ conjuntos restantes, por k rodadas ou iterações (Fig. 8a). Os erros de cada rodada são computados e farão parte do cálculo do erro global do modelo. Este método contribui para que um modelo torne-se mais flexível quando aplicado em dados de teste, sendo mais generalista e reduzindo o efeito indesejado denominado *overfitting*. A validação por *holdout*, por sua vez, consiste em separar determinado percentual do conjunto amostral para testes do modelo, com o treinamento sendo realizado no percentual restante (Fig. 8b) (Yadav & Shukla, 2016).

¹ Figura modificada de <https://www.mathworks.com/help/stats/machine-learning-in-matlab.html>.

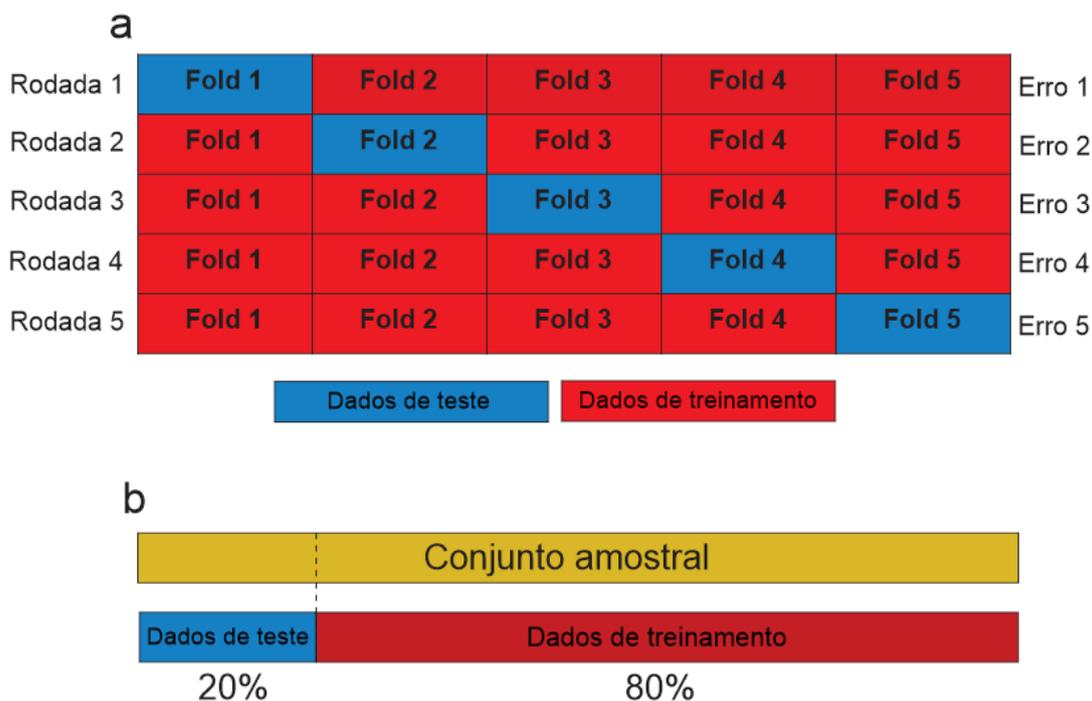


Figura 8 - Exemplos de métodos de validação de modelos. (a) Validação cruzada com 5-folds e (b) Validação por *holdout* com 20% de dados separados para teste.

3. MATERIAIS & MÉTODOS

As etapas a seguir detalham todos os procedimentos adotados para a obtenção, tratamento e análise dos dados deste trabalho, visando a interpretação de processos envolvidos no acúmulo de CO em ambas as margens do Atlântico Sul, bem como a geração de um modelo regressivo que elucide a contribuição dos fatores fisiográficos e taxas de sedimentação nesta acumulação.

3.1. Escolha dos *sites* de estudo e banco de dados de COT

Os dados utilizados no presente trabalho são resultados de análises de testemunhos obtidos durante expedições do DSDP. A seleção de *sites* obedeceu aos seguintes critérios iniciais:

- recuperação temporal que abrangesse totalmente (ou quase totalmente) o Cretáceo Superior (100,5 – 66 Ma), com o potencial de registrar diferentes etapas da fase drifte da abertura do Atlântico Sul que implicam em variações fisiográficas das bacias sedimentares geradas².

² As informações de cobertura temporal dos *sites* DSDP/ODP/IODP podem ser encontradas de forma compilada através do link http://iodp.tamu.edu/database/ages_lims.html;

- localização nas porções sul e equatorial do Oceano Atlântico, mais especificamente entre as latitudes 10° N e 52° S; e
- disponibilidade de dados de COT dentro do intervalo temporal de interesse, independente da resolução temporal destes dados;

Como resultado dos critérios de busca aplicados, 22 *sites* foram inicialmente encontrados no portal IODP (Fig. 9). Um único *site* (ES-1), utilizado como testemunho de teste, não é oriundo desse portal, tendo os dados de COT obtidos a partir da publicação de Santos *et al.* (submetido).

Dentre esses, selecionou-se aqueles com modelos de idade recentemente revistos e publicados (vide item 3.2), o que garante maior robustez na obtenção das idades absolutas a partir de profundidades, bem como no cálculo das taxas de sedimentação. Chegou-se, assim, em um total de 7 testemunhos (P1-P7, Tabela 2) para treinamento do modelo. Mesmo após essa etapa adicional de refinamento dos dados, garantiu-se que a cobertura espacial inicialmente proposta fosse mantida. Outros quatro testemunhos ainda foram selecionados para realização da aplicação do modelo de *Machine Learning* gerado, não sendo utilizado para o seu treinamento (T1-T4).

Tabela 2 - Nomenclatura e posição geográfica atual dos 7 *sites* selecionados no portal IODP para treinamento (P1-P7) e dos 4 testemunhos de teste da aplicação do modelo (T1-T4).

Número da expedição	Número do site	Código do site	Código adotado neste trabalho	Latitude	Longitude
207	1258	ODP1258	P1	9,4333	-54,7328
39	356	DSDP356	P2	-28,2870	-41,0880
71	511	DSDP511	P3	-51,0047	-46,9717
159	959	ODP959	P4	3,6276	-2,7358
40	364	DSDP364	P5	-11,5720	11,9717
75	530	DSDP530	P6	-19,1877	9,3858
40	361	DSDP361	P7	-35,0664	15,4485
207	1261	ODP1261	T1	9,0486	-28,3889
--	--	ES-1	T2	-20,4157	-39,4889
159	962	ODP962	T3	3,2514	-3,1816
40	363	DSDP363	T4	-19,6458	9,0467

Os *sites* selecionados encontram-se geograficamente dispostos de acordo com o mapa da Figura 9. Em um comparativo entre as margens do Atlântico Sul, 4 dos *sites* selecionados encontram-se na margem leste, enquanto os 3 restantes situam-se na borda oeste deste oceano. Para fins de simplificação de entendimento no momento da exposição de resultados e discussão e, ainda, considerando o grande número de *sites* estudados, adotou-se, neste trabalho, códigos em sequência (P1-P7, tabela 2) que remetem exatamente à ordenação da localização dos *sites* estudados, sendo P1-P3 localizados na margem oeste e P4-P7 situados na margem oposta (Fig. 9). Quanto aos testemunhos de teste, adota-se lógica similar na utilização dos códigos (T1-T4) e suas respectivas posições geográficas, com T1/T2 localizados na margem oeste e T3/T4 situados na margem leste.

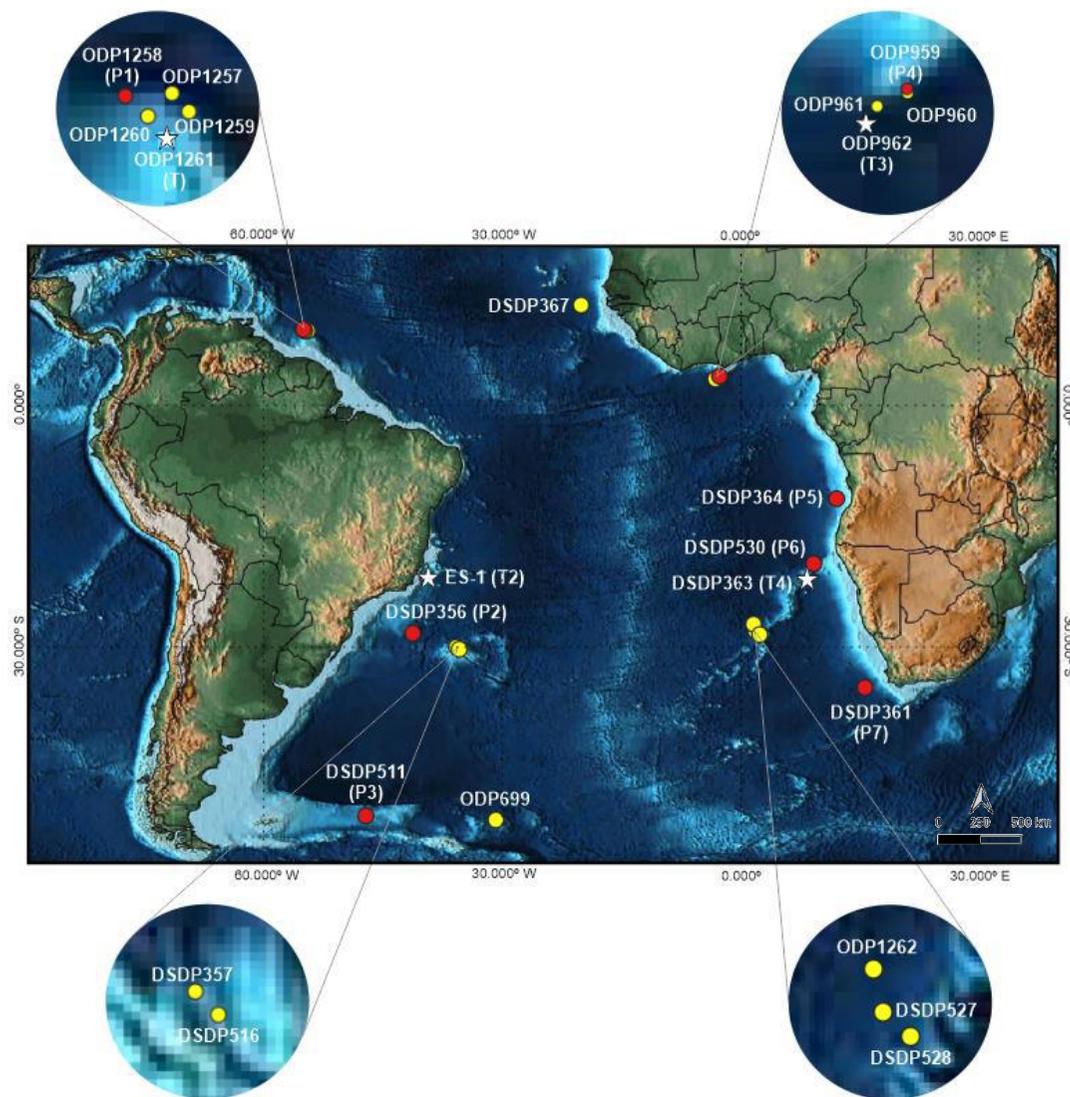


Figura 9 - Localização geográfica dos 22 *sites* inicialmente selecionados nas bases de dados do portal IODP. Em vermelho, os 7 *sites* resultantes da aplicação dos critérios descritos no item 3.1. Pontos amarelos: testemunhos inicialmente triados; pontos vermelhos: testemunhos utilizados neste trabalho; estrela: testemunhos para teste do modelo.

Os dados de COT foram obtidos a partir de buscas específicas para cada um dos *sites* (a partir da identificação de expedição/site) realizadas na plataforma SEDIS (Scientific Earth Drilling Information System), banco de dados que se insere dentro do portal IODP (*International Ocean Discovery Program*) e que concentra os dados compilados pós-moratória e de publicações resultantes de expedições DSDP/ODP/IODP. As ferramentas de busca do SEDIS podem ser acessadas através do endereço <http://sedis.iodp.org/search.php> (Fig. 10).

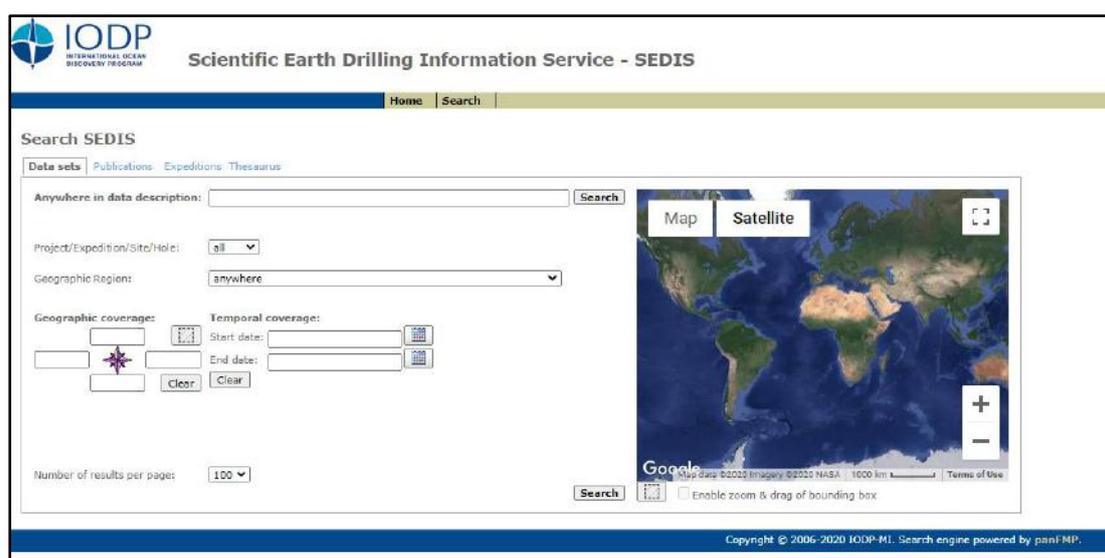


Figura 10 - Interface da ferramenta de busca da plataforma SEDIS.

3.2. Modelos de idades

Modelos de idades são essenciais para que as profundidades associadas ao parâmetro estudado (COT, nesta ocasião) sejam convertidas em idades absolutas, como via de regra em estudos paleoambientais. A existência de *data* ao longo dos testemunhos possibilita, ainda, a interpolação de idades entre o *datum* de idade mínima e o de idade máxima, permitindo que a todo esse trecho sejam atribuídas idades absolutas. Além disso, é a partir da relação profundidade-idade que as taxas de sedimentação são estimadas.

Os *data* que permitiram a construção de modelos de idades dos *sites* estudados foram obtidos, em sua totalidade, na publicação de Setoyama & Kanungo (2020) e baseiam-se em dados bioestratigráficos. Os detalhes dos modelos de idade apresentados por este estudo encontram-se no **Material Suplementar** deste trabalho.

Tendo como base os *data* supracitados, o passo seguinte consistiu na interpolação de idades nas profundidades associadas aos dados de COT. Esse procedimento foi realizado através do software Matlab (função '*griddedinterpolant*'), permitindo que os dados fossem interpolados linearmente. Dessa forma, todas as profundidades situadas entre dois *data* passam a receber valores de idade absoluta baseados na equação linear específica para tal intervalo, e assim sucessivamente para todo o intervalo compreendido entre o *datum* de idade mínima e o de idade máxima. Não foram atribuídas idades absolutas às profundidades que estivessem fora desse intervalo.

Para o caso específico do testemunho de teste ES-1 (T2), o modelo de idades é oriundo da cicloestratigrafia e obtido a partir da publicação de Santos *et al.*, (submetido).

3.3. Taxas de sedimentação

As taxas de sedimentação consistem em uma medida da deposição sedimentar em determinada localidade, normalizada por unidade de tempo. Neste estudo, as taxas de sedimentação foram calculadas com base nos dados de idade e profundidade, após aplicação dos modelos de idade para obtenção das idades absolutas associados a cada dado de COT. Um exemplo de cálculo pode ser descrito abaixo:

Profundidade (m)	Idade (Ma)	
P_1	I_1	\longrightarrow
P_2	I_2	
...	...	
P_n	I_n	

$$\begin{aligned}
 t_{sed_1} &= \frac{P_2 - P_1}{I_2 - I_1} \\
 &\cdot \\
 &\cdot \\
 &\cdot \\
 t_{sed_n} &= \frac{P_n - P_{n-1}}{I_n - I_{n-1}}
 \end{aligned}$$

As taxas de sedimentação obtidas neste trabalho são expressas em $m. My^{-1}$.

3.4. Paleolatitudo e paleolongitudo

Todos os dados extraídos dos bancos de dados são oriundos de testemunhos e, dessa forma, possuem uma única dimensão (1-D). A análise fisiográfica demanda, necessariamente, que os dados sejam espacializados (2-D) ao longo do tempo geológico, de forma a permitir que parâmetros paleogeográficos façam parte do estudo. A representação 3D (latitude, longitude, tempo) foi construída por meio da utilização de

ferramentas de reconstruções cinemáticas de placas tectônicas disponíveis no *software* GPlates v2.2.0 (Müller *et al.*, 2018), gerando para cada *site* estudado as respectivas paleoposições (paleolatitudes e paleolongitudes) contidas dentro do intervalo temporal de interesse (Fig. 11).

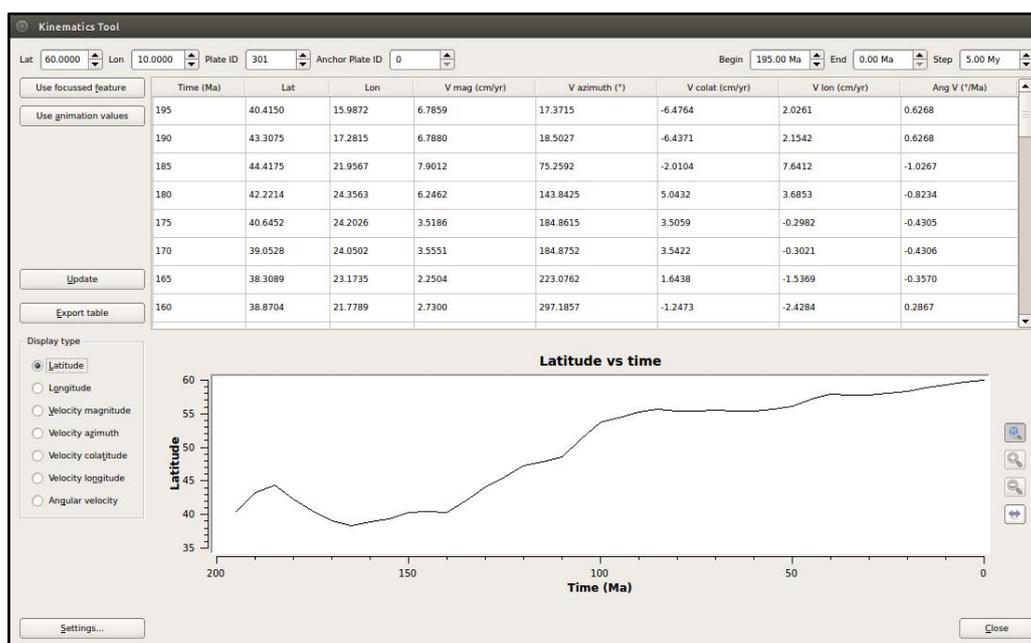


Figura 11 - Exemplo da interface da ferramenta cinemática dentro do software GPlates, utilizado para a obtenção de dados espaciais dos testemunhos trabalhados.

A obtenção de paleoposições só é possível com a utilização de um modelo de rotação, arquivo que reconstrói as movimentações relativas entre placas tectônicas ao longo do tempo geológico. O modelo de rotação utilizado no GPlates para este trabalho é parte do pacote de dados do projeto PALEOMAP PaleoAtlas (Scotese, 2016) que traz, além deste modelo em resolução temporal de 1 Ma, mapas paleogeográficos com resolução de 5 Ma (gerados a partir de paleotopografias e paleobatimetrias descritas no item 3.4). Além disso, foi necessário importar no GPlates a localização atual de cada *site* em formato *shapefile* (.shp) e identificá-los de acordo com o seu *Plate ID*, ou seja, informar ao *software* a qual placa tectônica cada ponto estudado pertence e atrelá-los a um ponto de referência (neste caso específico, a cadeia mesoceânica do Atlântico) para que todas as regressões cinemáticas pudessem ser computadas (Fig. 12). Os mapas apresentados foram produzidos por meio do software *QGIS 3.12.3*.

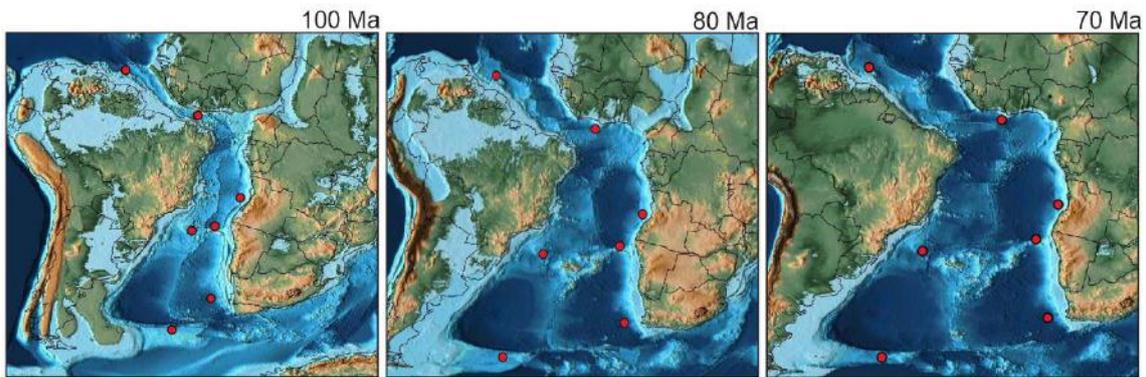


Figura 12 - Exemplo da reconstrução paleogeográfica dos *sites* estudados (pontos vermelhos no mapa) a partir do uso do software GPlates para 3 idades distintas do Cretáceo Superior – Albiano (100 Ma), Turoniano (90 Ma) e Maastrichtiano (70 Ma) – plotados em paleomapas de Scotese (2016).

Todos os dados extraídos de bancos de dados (ex.: COT) bem como os dados gerados/calculados (ex.: paleoposições, idades interpoladas, dados cinemáticos) até aqui foram agrupados dentro de uma estrutura de dados que permitisse a sua manipulação de forma mais rápida e prática. O formato adotado foi a *struct*, que possibilita que dados sejam armazenados em campos específicos na forma de vetores, matrizes, caracteres etc. A criação (arquivo com extensão *.mat*) e utilização dessa *struct* foi feita no software Matlab (Fig. 13) e todas as etapas posteriores do trabalho utilizaram os dados nela contidos.

Variables - myIDP8test2

myIDP8test2

1x11 struct with 23 fields

...	core	depthtoc	dephageMa	mindepth	maxdepth	matrixlabel	matrixdata	AgeMaDep	AgeDepthToc	plate_time	plate_lat	plate_lon
1	'207/1258'	494x2 double	11515x2 double	0	485	1x28 cell	486x28 double	18x2 double	494x3 double	491x1 double	491x1 double	491x1 double
2	'39/356'	65x2 double	494x2 double	0	738	1x28 cell	739x28 double	16x2 double	65x3 double	56x1 double	56x1 double	56x1 double
3	'71/511'	455x2 double	105x2 double	0	630	1x28 cell	631x28 double	19x2 double	455x3 double	272x1 double	272x1 double	272x1 double
4	'159/959'	2055x2 double	29698x2 double	0	1159	1x28 cell	1160x28 double	26x2 double	2055x3 double	2048x1 double	2048x1 dou...	2048x1 double
5	'40/364'	127x2 double	21x2 double	0	1083	1x28 cell	1084x28 double	21x2 double	127x3 double	110x1 double	110x1 double	110x1 double
6	'75/530'	544x2 double	24x2 double	0	1120	1x28 cell	1121x28 double	24x2 double	544x3 double	491x1 double	491x1 double	491x1 double
7	'40/361'	142x2 double	12x2 double	0	1305	1x28 cell	1306x28 double	12x2 double	142x3 double	85x1 double	85x1 double	85x1 double
8	'1-ESS-86A'	[]	[]	[]	[]	[]	[]	140070x2 dou...	485x3 double	485x1 double	485x1 double	485x1 double
9	'207/1261'	[]	[]	[]	[]	[]	[]	[]	337x3 double	337x1 double	337x1 double	338x1 double
10	'40/363'	103x2 double	28x2 double	0	707	1x28 cell	708x28 double	28x2 double	103x3 double	99x1 double	99x1 double	99x1 double
11	'159_962'	160x2 double	13x2 double	0	394	1x28 cell	395x28 double	13x2 double	160x3 double	115x1 double	115x1 double	115x1 double

Figura 13 – Exemplo de *struct* em Matlab com os dados armazenados para cada site estudado.

3.5. Paleobatimetria

Um modelo digital de elevação (MDE) é uma representação matricial (em *grid*, onde para cada lat/lon existe um valor de batimetria associado) da paleotopografia e da paleobatimetria que foram reconstruídas ao longo do tempo geológico. Os paleoMDE utilizados neste trabalho foram gerados com base em dados geológicos (litofácies, história tectônica) e geofísicos, cujos detalhes estão descritos em Scotese & Wright (2018). Originalmente, possuem resolução espacial de 1° x 1°, resolução temporal de 5 Ma e foram produzidos, a exemplo dos mapas paleogeográficos já citados, como parte do projeto PALEOMAP³, e estão disponíveis aos usuários na forma de arquivos *netcdf* de extensão *.nc* (formato também reconhecido e manipulável dentro do Matlab).

Neste trabalho, os arquivos de paleoMDEs originais foram reinterpolados dentro do Matlab para que as resoluções temporal e espacial fossem reajustadas para 0,5 Ma e 0,25° x 0,25°, respectivamente. A importância dessa etapa está na redução do erro quando se utiliza a função de interpolação (*interp3*, com método cúbico de interpolação) para obter as paleobatimetrias dos *sites* estudados a partir de suas respectivas latitudes e longitudes.

3.6. Paleodistância para a linha de costa

As paleodistâncias para a linha de costa foram calculadas a partir dos grids paleobatimétricos, onde estabeleceu-se que a linha de costa seria representada pelo intervalo de batimetria -20m/20m. Por meio da função *dsearchn* existente no Matlab, calculou-se as menores distâncias euclidianas ($d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$) entre os pontos de interesse e a linha de costa estabelecida. A esses valores, multiplicou-se o raio da Terra (6371 Km) para que as distâncias finais fossem obtidas.

³ Os dados do projeto PALEOMAP podem ser baixados em <https://www.earthbyte.org/paleodem-resource-scotese-and-wright-2018/>.

3.7. Tratamento de dados e geoestatística

3.7.1. Estatísticas básicas

O tratamento inicial dos dados por meio de estatística descritiva consistiu no estabelecimento dos parâmetros básicos (espaço amostral, médias, desvios padrões e medianas); construção de *boxplots*; histogramas de probabilidade; teste de normalidade de Shapiro-Wilk e no teste não-paramétrico de Mann Whitney (*teste U*, utilizados para distribuições não normais) para a comparação estatística dos dados de COT entre as margens do Atlântico Sul/Equatorial.

3.7.2. Aplicação de *Machine Learning* Supervisionada

A geração do modelo preditivo de COT para o Atlântico Sul durante o Cretáceo Superior foi obtida através do uso da ferramenta *Regression Learner*, pertencente à suíte de recursos do Matlab (*toolbox Deep Learning & Machine Learning*). Com esta ferramenta, treinou-se todos os 19 modelos disponíveis e o critério de escolha do melhor modelo baseou-se na menor raiz do erro quadrático médio ou RMSE (do inglês *Root Mean Squared Error*) resultante. Tendo em vista a heterogeneidade da variância dos valores da COT (vide **Resultados & Discussão**, Fig. 14) os modelos foram treinados em validação cruzada *k-fold*, com 5, 10 e 15 *folds*. Os dados de entrada não sofreram normalização e não houve redução dimensional por aplicação prévia de PCA ou qualquer outro método de clusterização. As variáveis de entrada (preditoras) utilizadas foram latitude, paleobatimetria, distâncias para a linha de costa e taxas de sedimentação; a variável de resposta utilizada foi o COT.

4. RESULTADOS & DISCUSSÃO

Os resultados utilizados para o desenvolvimento deste trabalho estão disponibilizados integralmente no **Material Suplementar**. A seguir, serão apresentados os resultados e discutidas as questões fundamentais que representam os objetivos centrais deste estudo.

4.1. Acumulação de Carbono Orgânico nas margens oeste e leste do Atlântico Sul durante o Cretáceo Superior

Ao longo do Cretáceo Superior, entre 100,5 e 66 Ma, elevados teores de COT (>4 a 29 wt%) foram registrados em ambas as margens do Atlântico Sul (Fig. 14). Na margem oeste, os maiores teores de CO foram observados na borda equatorial (P1), enquanto na margem leste ocorreram na Bacia de Angola (P5 e P6) e na borda equatorial (P4). De forma pontual e com menor resolução temporal, o testemunho P2, localizado no Platô de São Paulo, também registrou picos pontuais com COT atingindo teores de até cerca de 5 wt% (Fig. 14). Interessante notar que a ocorrência dos maiores teores de COT nas bordas equatoriais de ambas as margens se deu de forma assíncrona, tendo ocorrido inicialmente (Cenomaniano-Turoniano) na margem oeste e, posteriormente (Coniaciano-Santoniano), na margem leste. Apesar desse padrão de assincronia ter sido observado para as bordas equatoriais, alguns dos altos teores de COT nos testemunhos P5 e P6, Bacia de Angola, ocorreram de forma síncrona à margem oeste, cobrindo o intervalo Cenomaniano-Santoniano. Após o fim do Turoniano não houve mais deposição de altos teores de COT (>4 wt%) na margem oeste, enquanto na margem leste (P4, P5 e P6), a deposição permaneceu em patamares elevados até meados do Santoniano e apresentou níveis intermediários (entre 1 e 3 wt%) até o fim do Maastrichtiano (P4). Por fim, os testemunhos P3 (Platô das Falklands, margem oeste) e P7 (Bacia do Cabo, margem leste) não registraram, em nenhum momento do Cretáceo Superior, acúmulos de COT superiores a 2 wt%, com valores predominantes menores do que 1 wt%. De uma forma geral, os altos valores de COT observados em ambas as margens do Atlântico Sul são coincidentes com a ocorrência dos principais OAEs dentro do período Cretáceo Superior - OAE 1d, OAE 2 e OAE 3 (Fig. 14). No entanto, teores elevados de COT (> 4 wt%) também ocorreram fora das janelas temporais de OAEs. Assim, tendo em vista que os intervalos destacados representam eventos de anoxia oceânica de caráter global (Jenkyns, 2010), a ocorrência de fatores característicos locais ou regionais não podem ser descartados para explicar altas acumulações de COT. Por esse ponto de vista, pode-se sugerir que a gama de processos atuantes para viabilizar a ocorrência de um OAE pode, também, estar contribuindo para o acúmulo anômalo de COT nos testemunhos estudados, embora não seja a pretensão deste estudo associar tais picos anômalos a determinado OAE. Caso esta fosse a intenção, seria necessário lançar mão de ferramentas geoquímicas adicionais ao COT como, por exemplo, isótopos de carbono ($\delta^{13}\text{C}$) visando uma melhor caracterização desses eventos (Boudinot *et al.*, 2020).

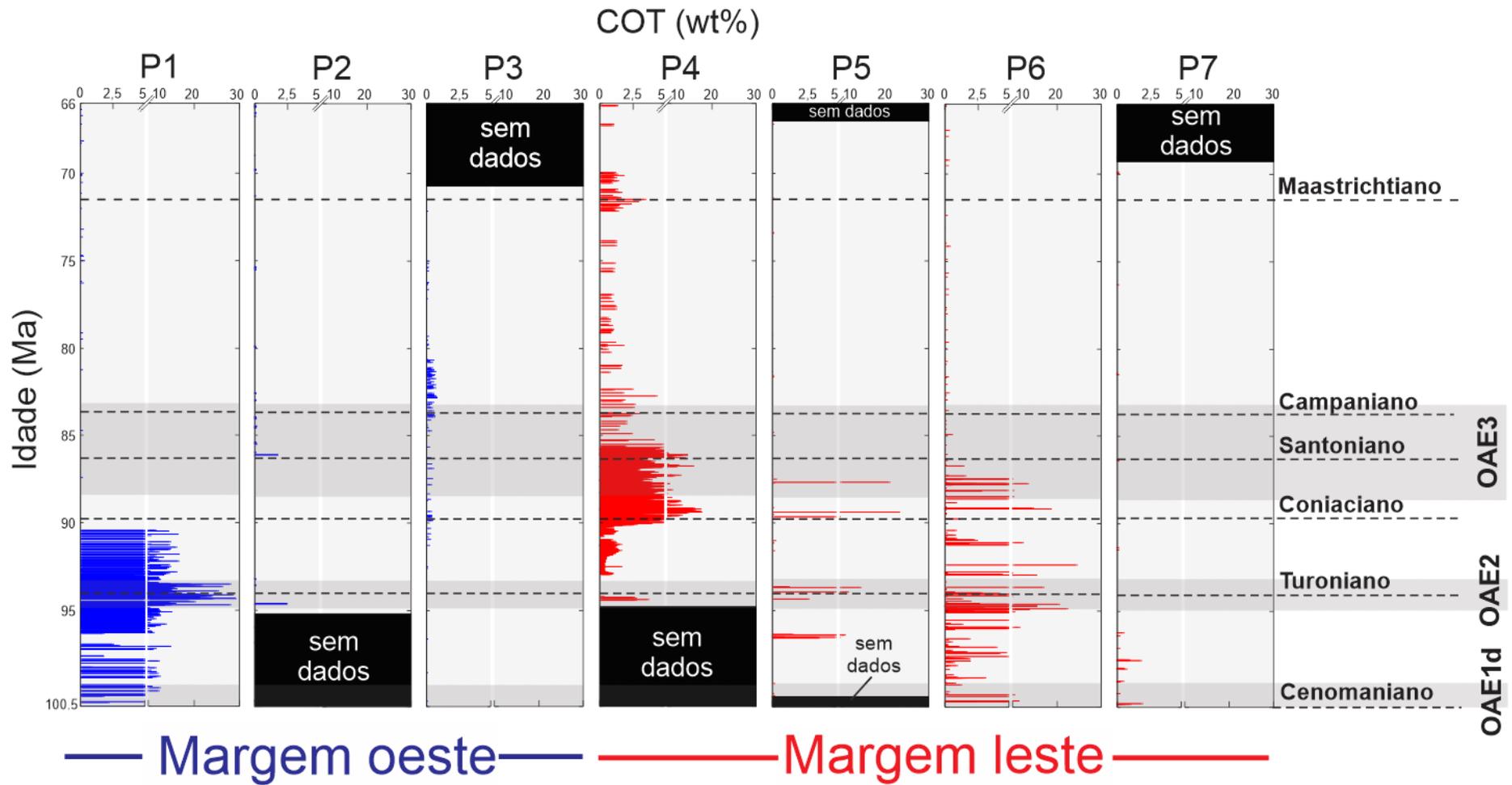


Figura 14 - Perfis geoquímicos de COT dos 7 testemunhos estudados, em ambas as margens do Atlântico Sul ao longo do Cretáceo Superior. As barras cinzas representam o intervalo aproximado de ocorrência dos principais OAEs globais (Locklair *et al.*, 2011; Jenkyns, 2010).

A distribuição de COT em ambas as margens do Atlântico Sul durante o Cretáceo Superior não é gaussiana, ou seja, não há simetria de distribuição em torno da média. A não normalidade da distribuição amostral pode ser percebida visualmente e também comprovada pelos valores p obtidos no teste de Shapiro-Wilk, aplicado no conjunto de dados de ambas as margens (Fig. 15a, b). Em decorrência disso, a comparação estatística do acúmulo de COT entre as margens se deu por meio da utilização de um teste não-paramétrico (teste U ou teste de Mann-Whitney). Assim sendo, a análise estatística de comparação entre as medianas da acumulação de COT entre o Cenomaniano e Maastrichtiano (Cretáceo Superior) nas margens leste e oeste do Atlântico Sul confirmou que, com um nível de significância de 5% ($\alpha=0,05$), a deposição de CO entre as margens foi significativamente desigual (Fig. 15c).

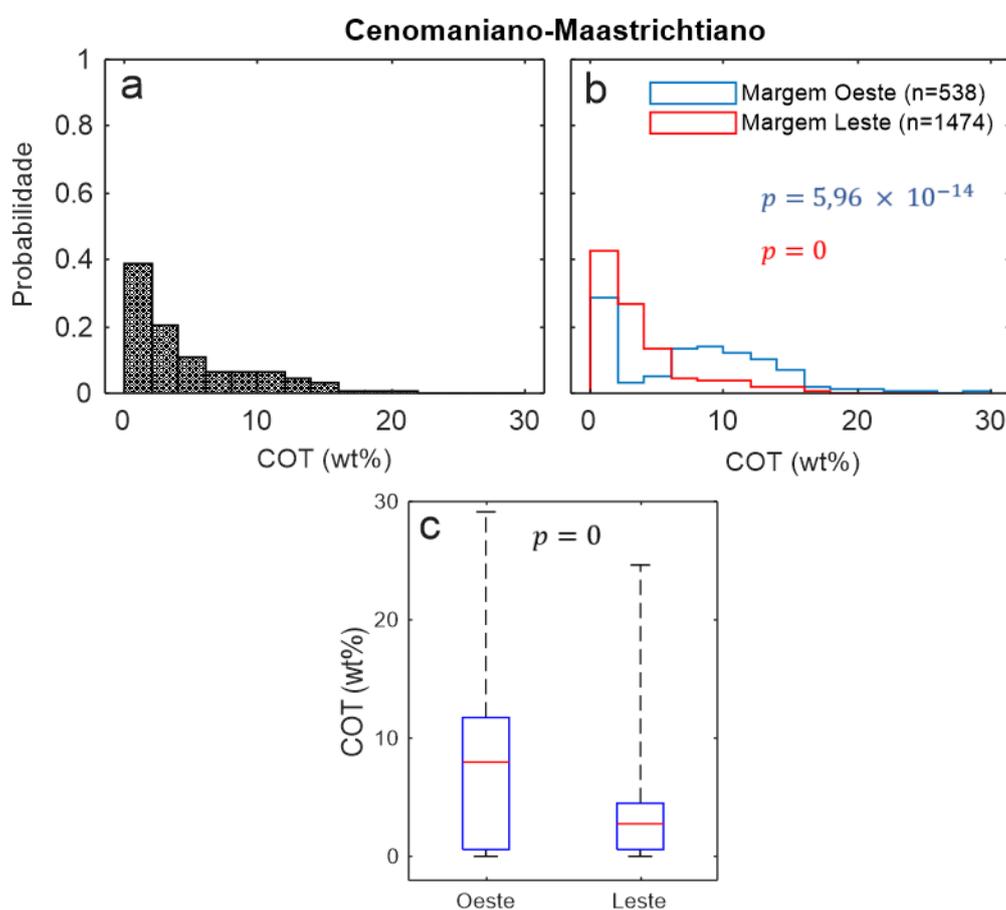


Figura 15 - Histogramas de probabilidade da distribuição de COT durante o Cretáceo Superior considerando (a) totalidade de dados e (b) a divisão por margens, com valores p do teste de normalidade de Shapiro-Wilk; (c) Boxplot da distribuição do percentil 25-75% com indicação das medianas (traço vermelho) e valor p do teste U .

É possível observar que, em termos de espaço amostral, a margem leste é melhor representada, possuindo mais do que o dobro do número de amostras para todo o intervalo do Cretáceo Superior. O intervalo Cenomaniano-Santoniano é o mais amostrado em ambas as margens ($n=439$ na margem oeste e $n=1326$ na margem leste), o que era de se esperar pelo fato desta faixa temporal possuir horizontes com médias de teores de COT mais elevadas ($9,32\pm 5,51$ wt% para a margem oeste e $3,88\pm 3,91$ wt% para a margem leste) e abrigar as janelas de ocorrência dos OAEs indicados. Considerando todo o Cretáceo Superior, nota-se, nas duas margens, o grande desvio em relação aos valores médios de COT ($7,66\pm 6,08$ wt% para a margem oeste e $3,57\pm 3,82$ wt% para a margem leste), reforçando a ideia de que picos de COT são episódicos e valores reduzidos de *background* (0-1 wt%) são bastante frequentes (28% das ocorrências na margem oeste e 31% na margem leste), conforme pode ser observado no histograma da Figura 15b. Tendo em vista esses resultados, foi possível afirmar que existem diferenças significativas na deposição de CO entre as margens oeste e leste do Atlântico Sul durante o Cretáceo Superior. No entanto, cabe ainda discutir os potenciais processos envolvidos nas diferenças registradas.

Dentre as diferenças mais marcantes está a assincronia observada na ocorrência de altos teores de COT no segmento equatorial de ambas as margens (P1 e P4) pelo menos ao longo do Turoniano, considerando que não há dados de P4 para grande parte do Cenomaniano (Fig. 14). Nesse contexto, diversos estudos de paleocirculação oceânica, especificamente ligados à dinâmica da formação e circulação de águas profundas no segmento equatorial descrevem processos que, potencialmente, explicam as diferenças observadas nessa região através da formação de uma armadilha de nutrientes durante o intervalo Cenomaniano-Santoniano (Berrocoso *et al.*, 2010; Martin *et al.*, 2012; Trabucho Alexandre *et al.*, 2010).

Com base na utilização do ϵNd , marcador geoquímico de proveniência de massas d'água, Martin *et al.* (2012) propuseram um padrão de circulação equatorial que envolve a presença de duas massas d'água profundas, a saber: a Água de Fundo de Demerara (do inglês *Demerara Bottom Water, DBW*), e águas profundas formadas no Tétis/Atlântico Norte (aqui designada como Água Profunda do Tétis, APT). Durante o intervalo Cenomaniano-Santoniano (excluindo o OAE2, Fig. 16b1), os autores sugeriram que a DBW e a APT interagem ortogonalmente (Fig. 16a), dando origem a uma "armadilha de nutrientes dinâmica" em que os nutrientes de origem continental - carregados pela DBW - misturavam-se com a APT e eram empurrados para sudoeste, sendo levados para regiões proximais à costa norte da América do Sul. A ocorrência de

ressurgência promovida pela ação de ventos permitiria que esses nutrientes aflorassem na superfície estimulando a produção primária e contribuindo para os altos teores de COT nos sedimentos marinhos daquela região (Berrecoso *et al.*, 2010). A alta produtividade induzida pela ressurgência teve, potencialmente, como consequência o estabelecimento de uma zona de anoxia capaz de favorecer a preservação do CO depositado (Trabucho Alexandre *et al.*, 2010). Durante o OAE2 e as condições extremas de efeito estufa deste intervalo, os valores de ϵNd sugerem maior influência da APT devido à intensificação de sua produção no Tétis e/ou redução de densidade e convecção da DBW, promovendo intensificação da ressurgência e consequente aumento na deposição/preservação de CO na região do testemunho P1 (Martin *et al.*, 2012; Fig. 16b). Após o OAE2, a DBW retomou, paulatinamente, a dominância na região, até se consolidar como única água de fundo a partir do Campaniano até meados/final do Maastrichtiano, interrompendo o processo de armadilha de nutrientes, uma vez que a DBW flui paralelamente à linha de costa e em batimetrias elevadas. É importante ressaltar que não há deposição significativa de CO durante o Coniaciano-Santoniano no segmento equatorial da margem oeste (P1, Fig. 14). Apesar do processo de armadilha de nutrientes ainda estar ativo no referido intervalo temporal, o testemunho P1 (*site* ODP1258) é o mais distal dos testemunhos ODP retirados no Platô de Demerara, o que pode, potencialmente, explicar a interrupção antecipada (ao fim do Turoniano) da ocorrência de altos teores de COT neste ponto, quando comparados a *sites* proximais da mesma região (e.g. ODP1260 e 1261, como descrito em Martin *et al.*, 2012). Conforme o mapa da Fig. 16a, o testemunho P4 não sofre influência destes processos, o que explicaria, a princípio, a ausência de ocorrência de altos teores de COT no segmento equatorial da margem leste do Atlântico Sul durante o Turoniano.

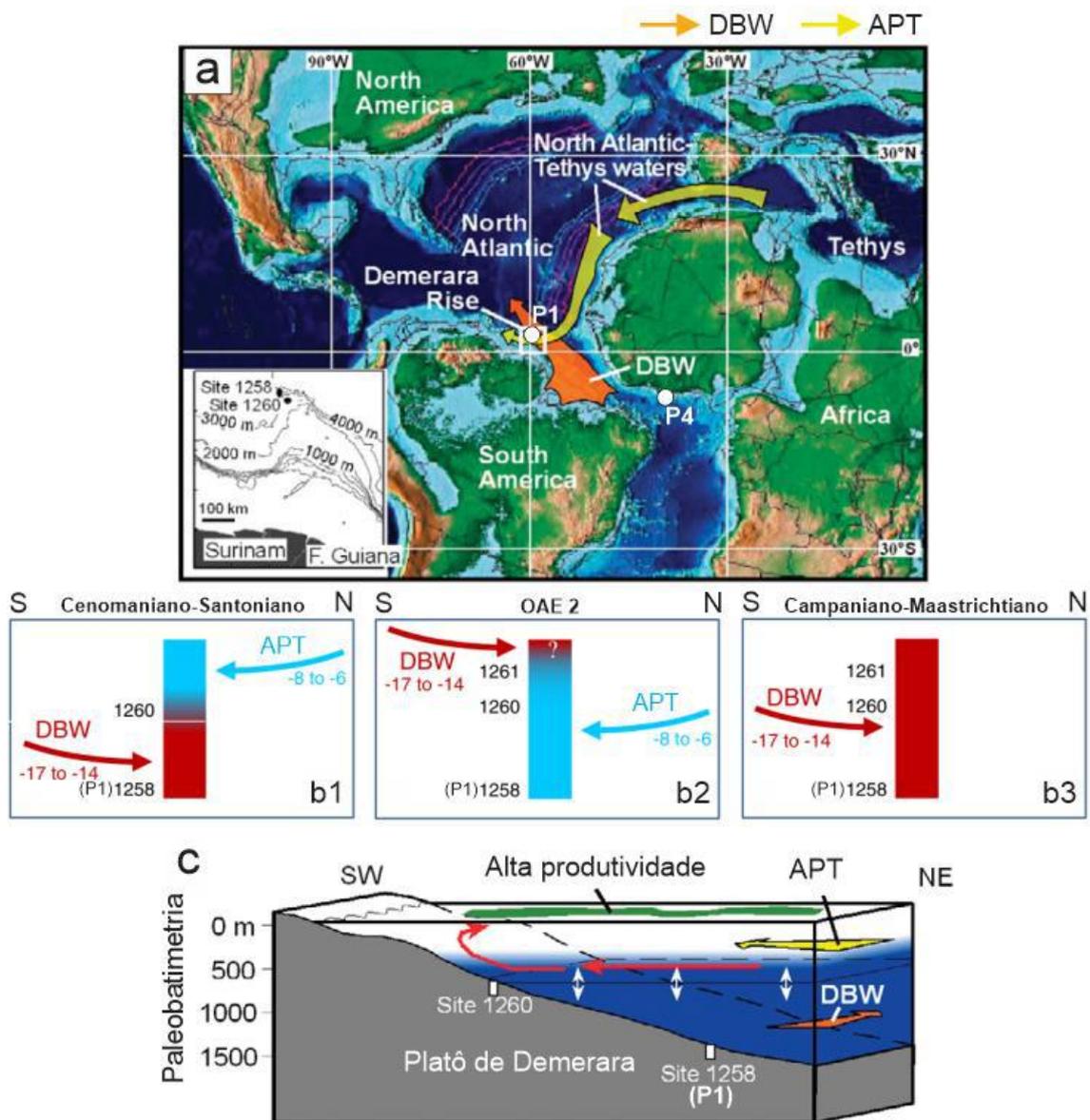


Figura 16 - (a) Mapa indicando sentidos de fluxo estimados para a DBW e APT e posições relativas dos testemunhos do segmento equatorial (P1 e P4), modificado de Berrecoso *et al.* (2010) com paleogeografia do Cenomaniano, modificado de Scotese (2008); (b) Dinâmica de circulação profunda das massas d'água DBW e APT na região do Platô de Demerara (*sites* ODP1258, 1260 e 1261) baseados em valores de δNd no intervalo Cenomaniano-Santoniano (b1), OAE2 (b2) e Campaniano (b3), modificado de Martin *et al.* (2012); (c) modelo conceitual para o processo de armadilha de nutrientes, modificado de Berrecoso *et al.* (2010).

Wagner *et al.* (2013), estudando diversos registros DSDP/ODP distribuídos no Atlântico Norte e Sul (dentre os quais os testemunhos P1, P4, P5 e P6 deste estudo), discutem que, a expansão da Célula de Hadley, decorrente do aumento das condições de efeito estufa durante o Cenomaniano-Turoniano (C/T na Fig. 17), teria produzido um alargamento da faixa subtropical, gerando condições favoráveis para o fortalecimento dos ventos alísios e, conseqüentemente, corroborando a intensificação dos processos

de ressurgência nesta região. Essas condições paleoclimáticas, potencialmente, explicariam os altos teores de COT observados nos testemunhos P1, P5 e P6 (Fig. 14). Nesse mesmo sentido, Kruijs & Barron (1990) confirmaram, por meio de estudo de modelagem, que a região da Bacia de Angola (P5 e P6) estaria, durante o Cenomaniano-Turoniano, submetida à ação de ressurgência sazonal durante o verão, o que teria se potencializado, ainda mais, durante o OAE2, favorecendo a intensificação das zonas de mínimo oxigênio (ZMO) e, conseqüentemente, a acumulação de CO sedimentar (Pancost *et al.*, 2004). O desenvolvimento de uma “mega-ZMO” que cobriu a margem leste do Atlântico Sul durante o intervalo Cenomaniano-Turoniano tem sido sugerido para explicar a presença dos *black shales* na região dos testemunhos P5 e P6. A grande variabilidade observada nos registros de COT, sobretudo no testemunho P6, tem sido explicada como resultante de processos de retrabalhamento de sedimentos ricos em CO oriundos da plataforma continental ou do flanco superior da Cadeia Walvis e seu subsequente transporte e redeposição em região submetida a condições menos anóxicas do fundo da Bacia de Angola decorrentes de correntes de turbidez (Forster *et al.*, 2008).

O testemunho P4 apresenta posicionamento favorável para assumir os efeitos da ação do fortalecimento dos ventos alísios e conseqüente intensificação dos processos de ressurgência e ZMO. De acordo com o modelo de idades aqui estabelecido, entretanto, não há dados de COT disponíveis para grande parte do Cenomaniano nesse testemunho. Ao fim do Cenomaniano, quando a série de dados de COT passa a existir para o P4, há registro de teores moderados de cerca de 2,5 wt% no Turoniano e teores elevados (>5 wt%) ao longo do Coniaciano e início do Santoniano, cuja ocorrência pode estar associada às condições paleoambientais estabelecidas pelo OAE3.

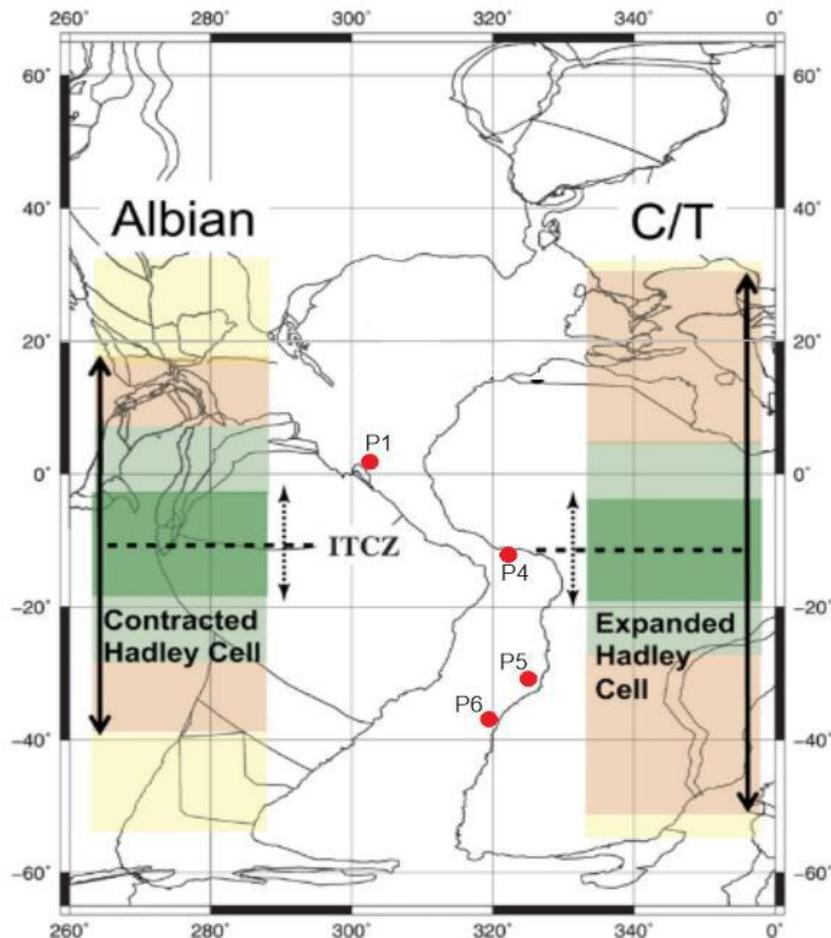


Figura 17 - Mapa indicando a extensão de atuação das diferentes faixas de influência da Célula de Hadley e o posicionamento dinâmico da Zona de Convergência Intertropical (ITCZ) e a localização dos testemunhos P1, P4, P5 e P6 deste estudo (modificado de Wagner *et al.*, 2013). Verde escuro: faixas tropicais; verde claro: faixas transicionais, onde a influência de ventos alísios e ressurgência são maiores; marrom: faixas subtropicais, onde a influência de monções e drenagem continental é mais pronunciada; C/T: Cenomaniano-Turoniano.

Nesse contexto, Flögel *et al.* (2008) indicaram que o segmento equatorial das margens oeste e leste do Atlântico Sul durante o Cretáceo Superior responderam de forma diferente às forçantes climáticas orbitais⁴. Uma análise espectral das frequências da acumulação de CO na região do Platô de Demerara (P1) durante o Coniaciano-Santoniano identificou as bandas de excentricidade (~400 e 101 ka) como dominantes, enquanto na margem leste (P4) as bandas orbitais de obliquidade e precessão seriam as mais marcantes, o que difere, consideravelmente, das condições atuais que condicionam padrões similares entre ambas as margens do segmento equatorial do

⁴ Forçantes orbitais ou astronômicas do clima consistem em mudanças periódicas do eixo de rotação da Terra ao longo da trajetória ao redor do Sol. Essas variações dependem da interação gravitacional entre o Sol, a Terra, a lua e outros planetas do nosso sistema solar (de Boer & Smith, 2009).

Atlântico Sul. Dessa forma, a ausência das montanhas andinas e do sistema de drenagem do Rio Amazonas durante o Cretáceo Superior, seriam fatores de estresse atmosférico/orográfico que alteraram de forma significativa o paleoclima da América do Sul nesse intervalo e, provavelmente, contribuíram para as assincronias observadas nas deposições de CO entre os testemunhos P1 e P4.

Em contraponto à ideia de uma circulação oceânica profunda vigorosa no segmento equatorial durante o Cretáceo Superior, acima descrita e defendida por Berrecoso *et al.* (2010), Trabuco Alexandre *et al.* (2010) e Martin *et al.* (2012), outros autores suportam ideia contrária, ou seja, de que o segmento equatorial do Atlântico Sul possuía uma circulação oceânica restrita, promovendo a anoxia oceânica e propiciando maior preservação da MO depositada (Murphy & Thomas, 2013; Robinson *et al.*, 2010; Robinson & Vance, 2012). De acordo com esses pesquisadores, somente a partir do Campaniano, com a abertura da passagem do Atlântico Central, águas profundas formadas no Oceano Austral (*e.g.*, SCW) fluíram em direção ao Atlântico Norte, ventilando essas bacias oceânicas e interrompendo as condições de anoxia/euxínia (Fig. 18). Essa hipótese explicaria os baixos conteúdos de COT observados em todos os testemunhos estudados (Fig. 14, P1-P7) a partir do início do Campaniano.

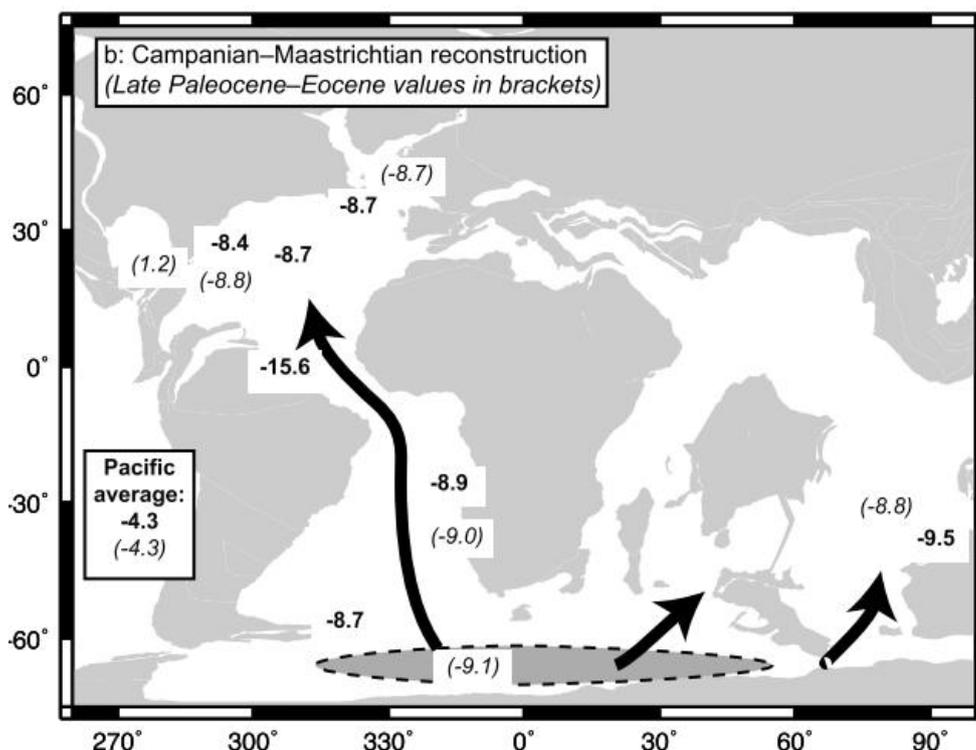


Figura 18 - Nova circulação oceânica integrando Oceano Austral, Atlântico Sul e Atlântico Norte a partir de meados do Campaniano e valores de ϵNd das respectivas bacias oceânicas (Robinson & Vance, 2012).

Pérez-Díaz & Eagles (2017) corroboram essa ideia, constatando que a passagem do Atlântico Central tenha atingido uma paleobatimetria suficiente para viabilizar, no início do Campaniano, o intercâmbio de águas profundas (~ 3500 m) entre os segmentos central e equatorial do Atlântico Sul, promovendo maior oxigenação dessas porções oceânicas. Segundo esses autores, entretanto, a subsidência da cadeia Rio Grande-Walvis, até então, não havia sido suficiente para permitir a passagem de águas profundas oriundas do sul. Isto pode ter se dado no Campaniano (Robinson & Vance, 2012), no Maastrichtiano (Donnadieu *et al.*, 2016) ou até mesmo no Paleoceno (Voigt *et al.*, 2013). É, de fato, ponto de debate o momento exato em que a evolução paleobatimétrica do Atlântico Sul gerou a abertura da passagem do Atlântico Central e permitiu o fluxo da SCW para norte, alterando de forma marcante o padrão de circulação termohalina global.

Condições análogas de baixa acumulação de COT (<2 wt%) perduraram durante todo o Cretáceo Superior em ambas as margens do Atlântico Sul, tendo sido observadas nos testemunhos P2 (margem oeste), P3 e P7 (margem leste), posicionados nos segmentos sul e central. De fato, Dummann *et al.* (2020) mostraram que, desde o final do Albiano (~110 Ma), a configuração paleogeográfica já possibilitava a circulação profunda entre o Oceano Austral e o Atlântico Sul, limitado pela Cadeia Rio Grande-Walvis (Fig. 19), promovendo a ventilação desta bacia oceânica e desfavorecendo a preservação de matéria orgânica.

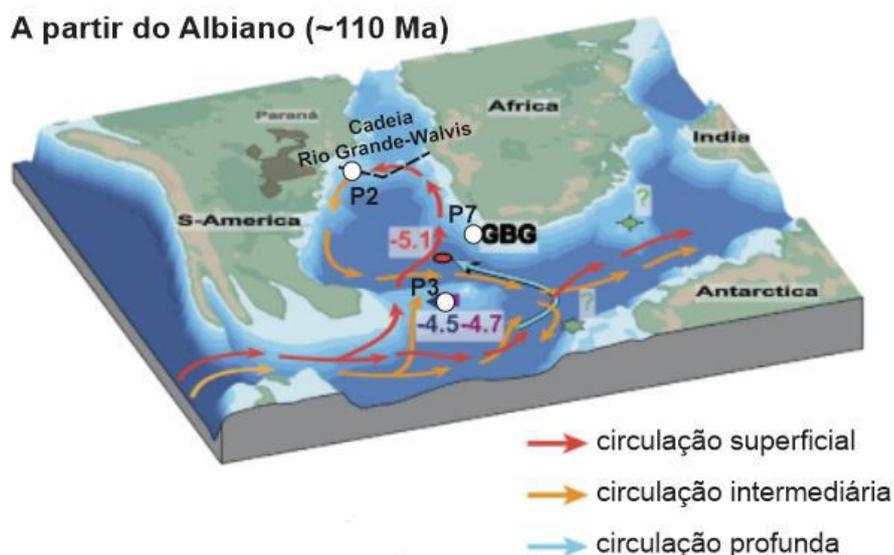


Figura 19 - Modelo conceitual de circulação oceânica entre Oceano Austral e segmento sul do Atlântico Sul a partir do Albiano, modificado de Dummann *et al.* (2020), indicando posição dos testemunhos P2, P3 e P7; GBG: passagem da Bacia da Geórgia.

Com base na discussão acima apresentada, é possível perceber que fatores fisiográficos, especialmente relacionados à evolução tectônica que condicionou a abertura do Oceano Atlântico, tiveram um papel fundamental na acumulação de CO no fundo marinho durante o Cretáceo Superior. No entanto, uma avaliação mais precisa da magnitude da contribuição desses fatores ainda não foi devidamente realizada. Dessa forma, esse estudo apresentará a seguir uma discussão sobre a aplicação de modelos regressivos em *Machine Learning* visando mensurar a parcela de explicação atribuída aos fatores fisiográficos e taxas de sedimentação no acúmulo de CO em ambientes marinhos do Atlântico Sul ao longo do Cretáceo Superior.

4.2. Aplicação de modelos regressivos em *Machine Learning* para avaliar o papel da fisiografia e taxas de sedimentação na acumulação de Carbono Orgânico no Atlântico Sul durante o Cretáceo Superior

No presente estudo, a aplicação da técnica de *Machine Learning* tem o objetivo de mensurar, em termos percentuais, a contribuição dos fatores fisiográficos - batimetria, distâncias para a linha de costa, latitude - e taxas de sedimentação na explicação da variância de CO acumulado em sedimentos marinhos. Como explicado na seção **Materiais & Métodos**, adotou-se aqui um algoritmo de regressão, uma vez que a variável de resposta (COT, também chamada de variável dependente) utilizada é quantitativa. Dentro dos algoritmos de regressão, optou-se, ainda, por testar os modelos que pertencem à vertente da *Machine Learning* supervisionada, devido ao fato da variável de resposta estar sendo informada ao computador para ser predita a partir das variáveis independentes (ou variáveis preditoras). No caso hipotético onde a resposta desejada não é informada à máquina, deve-se utilizar modelos pertencentes à vertente não-supervisionada de *Machine Learning* (Osisanwo *et al.*, 2017).

Nesse sentido, 19 modelos regressivos foram treinados com validação cruzada *K-fold* (5, 10 e 15-*folds*). Tendo em vista a distribuição irregular dos valores de COT (Fig. 14), entendeu-se que a validação cruzada por *holdout* não seria adequada, uma vez que este método requer uma menor variância nos dados de resposta. No total, produziu-se 114 resultados de performance (R^2 e RMSE), apresentados na Tabela 3.

Tabela 3 - Resultados de R², RMSE e gráficos de resposta observada vs resposta modelada para cada modelo de regressão treinado. As piores performances estão destacadas em vermelho e as melhores, em verde.

Tipo	Modelo	Validação Cruzada <i>K-fold</i>					
		<i>5-folds</i>		<i>10-folds</i>		<i>15-folds</i>	
		R ²	RMSE	R ²	RMSE	R ²	RMSE
Linear	Linear	0,27	4,17	0,27	4,17	0,27	4,16
	Interactions Linear	0,35	3,93	0,35	3,94	0,36	3,93
	Robust Linear	0,23	4,30	0,23	4,30	0,23	4,29
	Stepwise Linear	0,35	3,94	0,35	3,94	0,36	3,93
Tree	Fine Tree	0,58	3,19	0,59	3,15	0,59	3,15
	Medium Tree	0,61	3,05	0,63	2,99	0,63	2,97
	Coarse Tree	0,55	3,28	0,57	3,21	0,58	3,15
Support Vector Machine (SVM)	Linear SVM	0,23	4,30	0,21	4,36	0,22	4,31
	Quadratic SVM	-57,49	37,43	-220,93	73,00	-125,55	55,02
	Cubic SVM	-7,69x10 ⁶	13577	-1,39x10 ⁷	18317	-1,14x10 ⁸	52193
	Fine Gaussian SVM	0,48	3,53	0,48	3,54	0,48	3,53

Tipo	Modelo	Validação Cruzada <i>K-fold</i>					
		<i>5-folds</i>		<i>10-folds</i>		<i>15-folds</i>	
		R ²	RMSE	R ²	RMSE	R ²	RMSE
Support Vector Machine (SVM)	Medium Gaussian SVM	0,39	3,81	0,39	3,82	0,39	3,81
	Coarse Gaussian SVM	0,29	4,14	0,29	4,14	0,29	4,12
Ensemble	Boosted Trees	0,60	3,10	0,60	3,09	0,60	3,09
	Bagged Trees	0,66	2,85	0,67	2,82	0,67	2,83
Support Vector Machine (SVM)	Squared Exponential GPR-SE	0,57	3,22	0,59	3,15	0,58	3,16
	Matern 5/2 GPR-M 5/2	0,59	3,12	0,61	3,07	0,60	3,09
	Exponential GPR-E	0,65	2,90	0,66	2,84	0,66	2,86
	Rational Quadratic GPR-RQ	0,67	2,80	0,69	2,73	0,69	2,73

A análise de performance dos treinamentos realizados mostrou que os modelos gaussianos *Rational Quadratic* (GPR-RQ)⁵, validados por 10 e 15-*folds*, apresentaram resultados semelhantes com os maiores coeficientes de determinação (R²=0,69) e menores valores da raiz do erro quadrático médio (RMSE=2,73). No entanto, considerando que a validação por 10-*folds* realiza menos iterações e, portanto, gera um menor custo operacional para reprodutibilidade (tempo de máquina), decidiu-se adotar o modelo gaussiano GPR-RQ validado por 10-*folds* como o modelo selecionado neste

⁵ A função de base dos modelos de processo gaussiano testados foi constante; o kernel isotrópico também foi adotado.

estudo. Dessa forma, a seguir serão apresentados parâmetros de performance deste modelo frente aos dados de distribuição de COT dos testemunhos P1-P7 ao longo do Cretáceo Superior, visando discutir os pontos fortes e as limitações da aplicação deste modelo nos dados em questão.

A análise do erro residual, que consiste na diferença entre resposta observada e resposta modelada para cada um dos testemunhos estudados (P1-P7), mostrou que, em termos gerais, grande parte da dispersão (mais precisamente 61% dos dados), encontra-se próximo do resíduo zero, variando entre +/- 0,5 COT wt%, com uma expressiva maioria ocupando o intervalo interquartil, o qual também se encontra próximo ao eixo (Fig. 20). Esse fato é refletido pelo bom coeficiente de determinação encontrado ($R^2=0,69$) que atesta o bom desempenho deste modelo. Por outro lado, a distribuição do resíduo confrontado com a resposta observada (dados medidos de COT) mostra uma clara tendência (ou viés) de superestimar os baixos valores e subestimar os altos valores de COT (Fig. 20a, seta vermelha). Além disso, a variância dos erros observada entre os testemunhos é bastante heterogênea (Fig. 20b), o que pode estar relacionado à tendência anteriormente enfatizada. Esse padrão heterogêneo das variâncias indica uma condição conhecida como heterocedasticidade, a qual pode apontar para uma potencial diminuição na eficiência do modelo (Bollerslev, 1986) apesar de não invalidar a aplicação dos preditores estudados (Maia, 2019). Dentre as possíveis causas para a heterocedasticidade e o viés observados estão:

- a. ausência de normalização dos dados de entrada;
- b. falhas na especificação do modelo, a qual seria condicionada pela ausência de outras potenciais variáveis preditoras que não foram considerados no treinamento;
- c. presença de valores extremos da variável de resposta (COT); e
- d. natureza das variáveis preditoras.

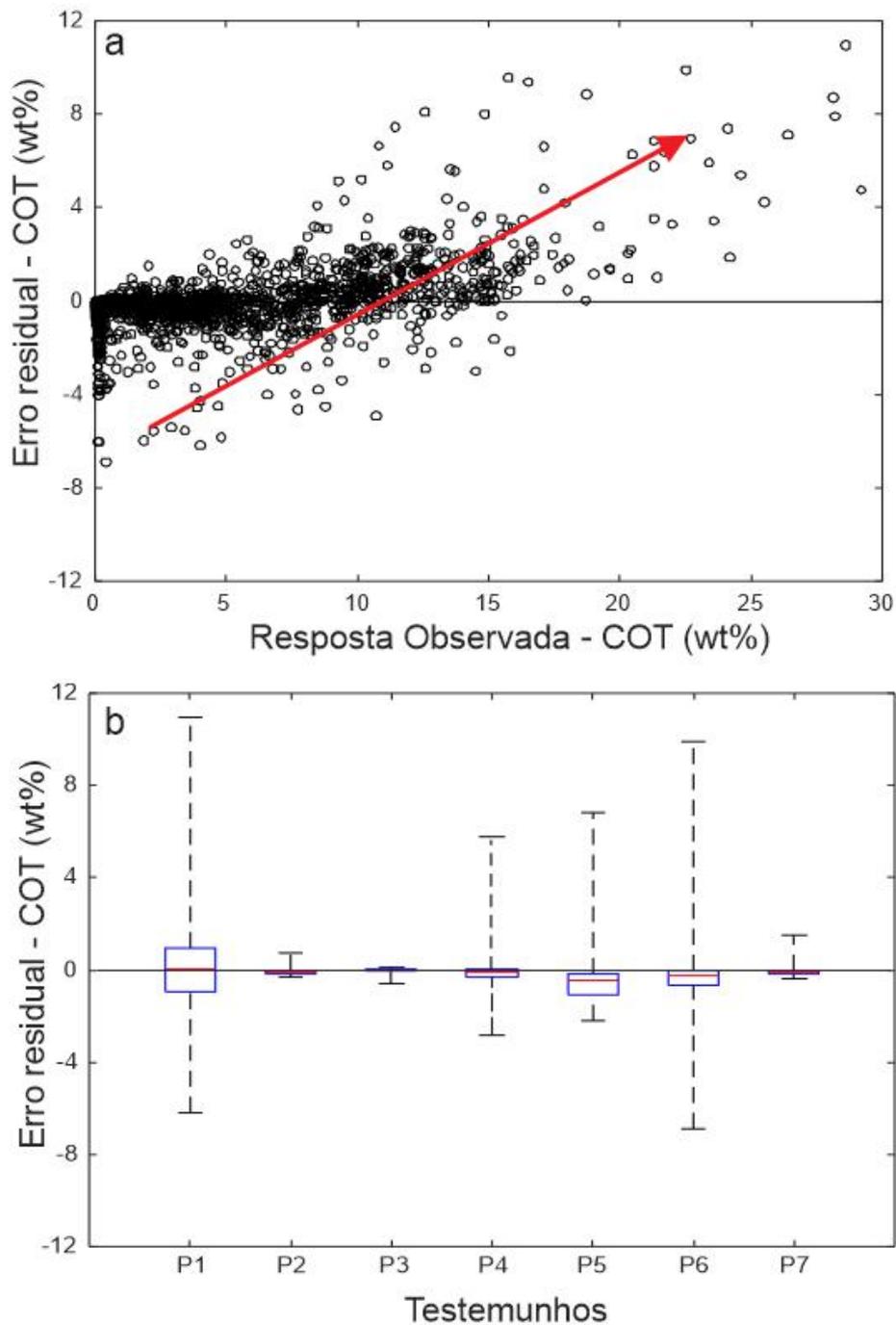


Figura 20 - Erros residuais comparativos entre valores simulados e observados (a) Erro residual vs. Resposta observada; (b) boxplots dos erros residuais para os testemunhos P1-P7.

Nesse sentido, se faz necessária uma avaliação mais aprofundada sobre o papel das possíveis causas (itens *a-d*, acima listados) da heterocedasticidade na distribuição de erros residuais do modelo:

a. Normalização dos dados - A fim de testar a influência da falta de normalização dos dados, foram testados 2 tipos de normalização, a saber: *escore-Z* e *range 0-1*. A primeira consiste em subtrair de cada amostra a média da distribuição total de determinada variável e dividir essa diferença pelo desvio padrão. A segunda realoca todos os valores entre 0 e 1. Os resultados obtidos com o treinamento do modelo com dados normalizados mostraram a ineficácia da normalização para a extinção das tendências e fim da heterocedasticidade (Fig. 21a, b).

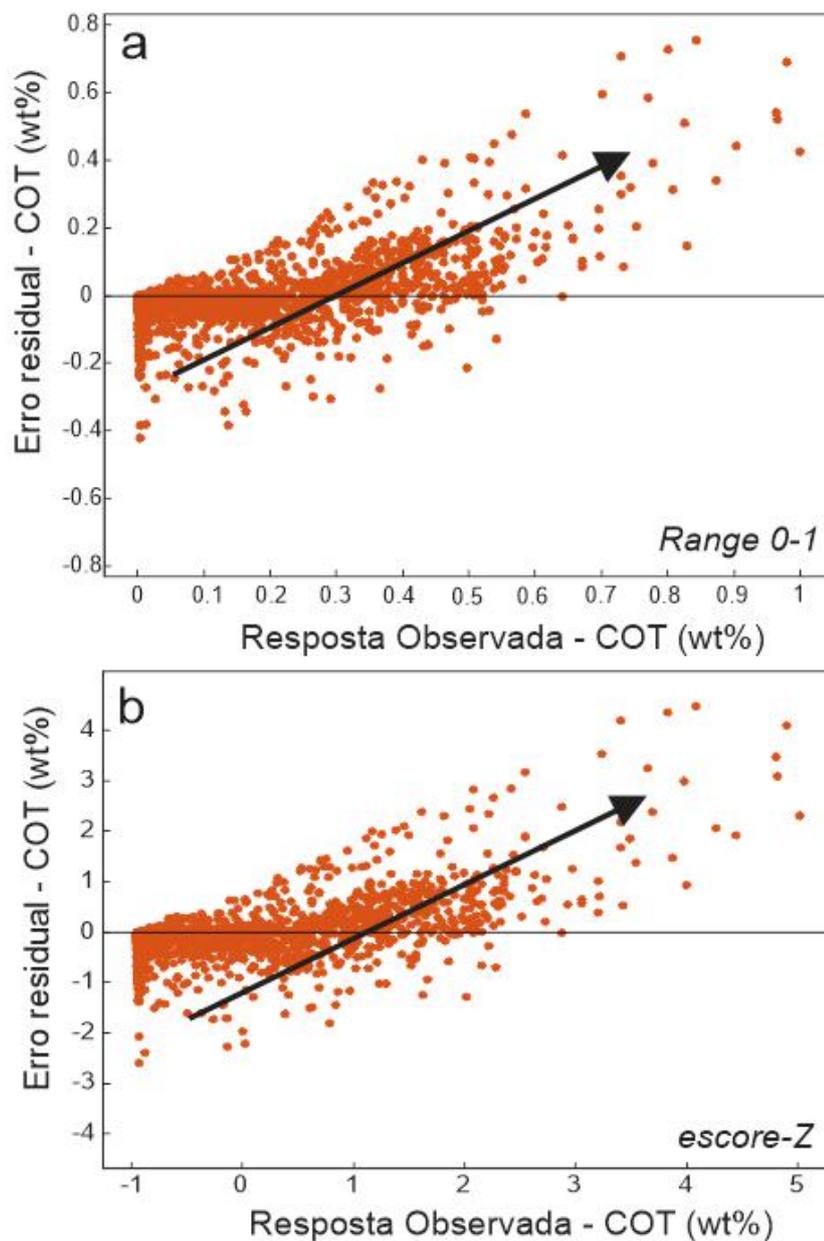


Figura 21 - Erro residual vs. Resposta observada com dados (a) normalizados entre 0-1 (range) e (b) normalizados por *escore-Z*.

Efetuuou-se, também, a normalização logarítmica dos dados. Entretanto, a logaritmização dos dados de latitude e paleobatimetria resultou em números complexos, inviabilizando a leitura desses valores pela ferramenta de treinamento em modelos regressivos.

b. Falhas de especificação do modelo - Essa falha diz respeito ao fato de as variáveis preditivas (i.e., latitude, paleobatimetria, paleodistância da costa e taxa de sedimentação) não serem capazes por responder integralmente por toda a variabilidade da variável de resposta, no caso o COT. Além disso, a tendência clara marcada pela seta na Figura 21a sugere, de fato, que o *viés de especificação* esteja ocorrendo, ou seja, o modelo pode estar não considerando variáveis ou termos de interação importantes para a predição. Apesar dessa condição ser, potencialmente, responsável por adicionar um fator de erro no modelo produzido, o objetivo deste estudo é mesmo conhecer a parcela de contribuição dos fatores fisiográficos e taxas de sedimentação para a acumulação de CO, sem jamais assumir que essa acumulação é integralmente explicada por tais preditores. Dentro da proposta do presente estudo, portanto, essa condição não pode ser remediada.

c. Valores extremos da variável resposta (COT) - A distribuição dos valores de COT ao longo dos testemunhos estudados (P1-P7, Fig. 14) é bastante heterogênea variando entre um mínimo de 0,01 e um máximo de 29,2% COT wt%, com uma distribuição de valores que pode ser observada no histograma de frequência de valores ilustrado na figura 15. Vale notar, no entanto, que os maiores valores de COT (> 4 wt%) foram registrados apenas nos P1, P4, P5 e P6, os quais também demonstram as maiores distribuições dos erros residuais (Fig. 20b). Essa condição confirma que, potencialmente, a alta variabilidade observada nos dados de resposta (COT), com a ocorrência de valores extremos, pode, de fato, estar contribuindo para a heterocedasticidade e o viés observados. No entanto, cabe lembrar que as acumulações anormalmente altas de CO no fundo oceânico durante intervalos temporais restritos, em especial durante os eventos anóxicos (OAEs), refletem condições atípicas, porém naturais dos sistemas marinhos. Assim, uma análise da distribuição temporal dos erros residuais demonstra uma grande variabilidade com maiores amplitudes entre o

Cenomaniano-Santoniano, o qual coincide com a ocorrência de eventos anóxicos oceânicos (Fig. 22). Esse contexto de distribuição temporal irregular dos erros decorrentes de alterações abruptas do sistema, indica a inabilidade dos modelos regressivos em representar sistemas naturais de alta complexidade e variabilidade.

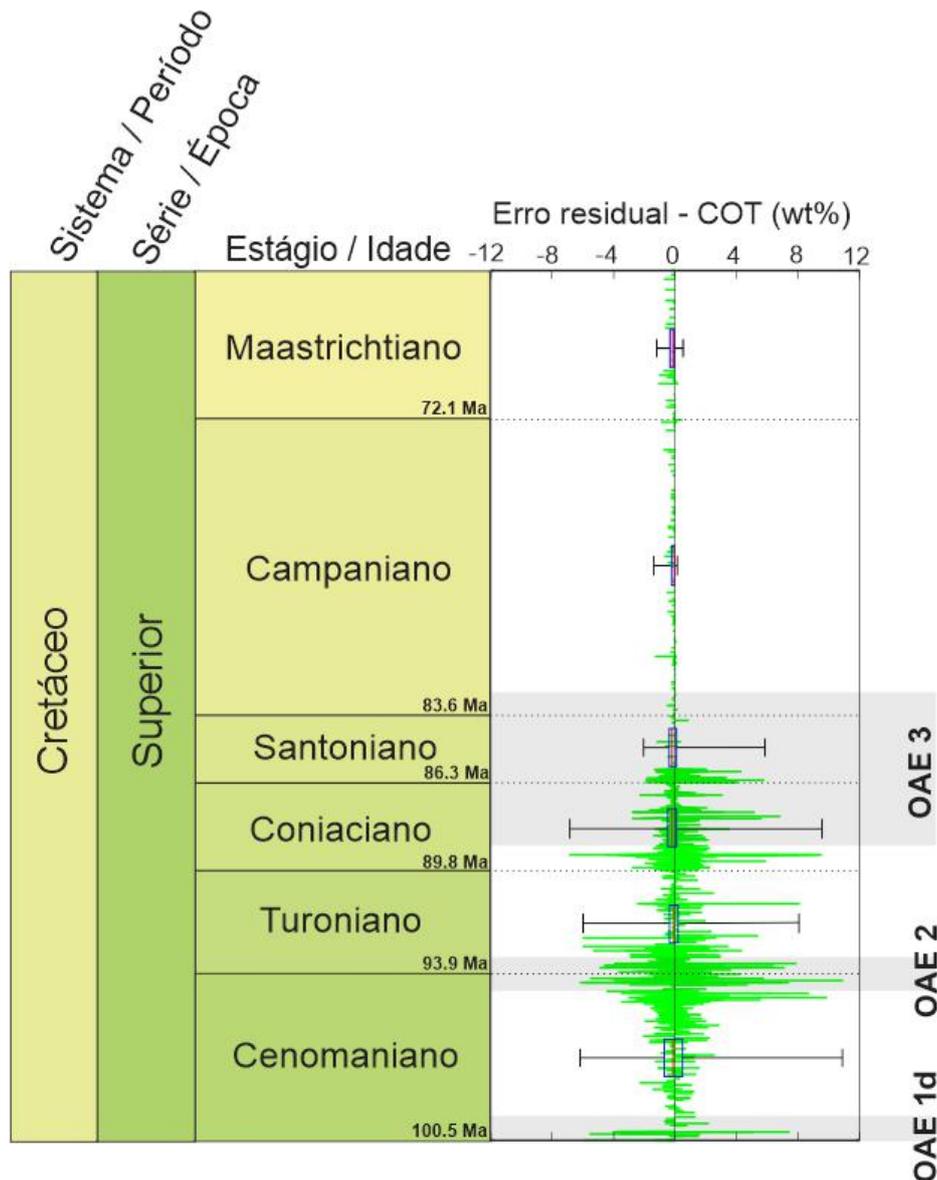


Figura 22 - Distribuição dos erros residuais ao longo das idades do Cretáceo Superior, com destaque para os intervalos temporais de ocorrências dos principais OAEs desse período.

d. Natureza das variáveis predictoras - A comparação entre a magnitude das variâncias das variáveis predictoras e da variável resposta demonstra com clareza uma fonte para a heterocedasticidade observada no modelo (Fig. 23). Como ilustrado na figura 20b, os testemunhos P1, P4, P5 e P6 - que juntos representam

cerca de 92% do espaço amostral - apresentam grande variância no erro residual, além de concentrarem os valores extremos de COT. Assim, uma vez que a variância do COT é expressivamente maior do que a variância das preditoras, o modelo encontra dificuldades para simulá-lo, em especial nos extremos do intervalo da variância, causando a superestimação dos baixos valores e subestimação dos altos valores de COT (Fig. 23).

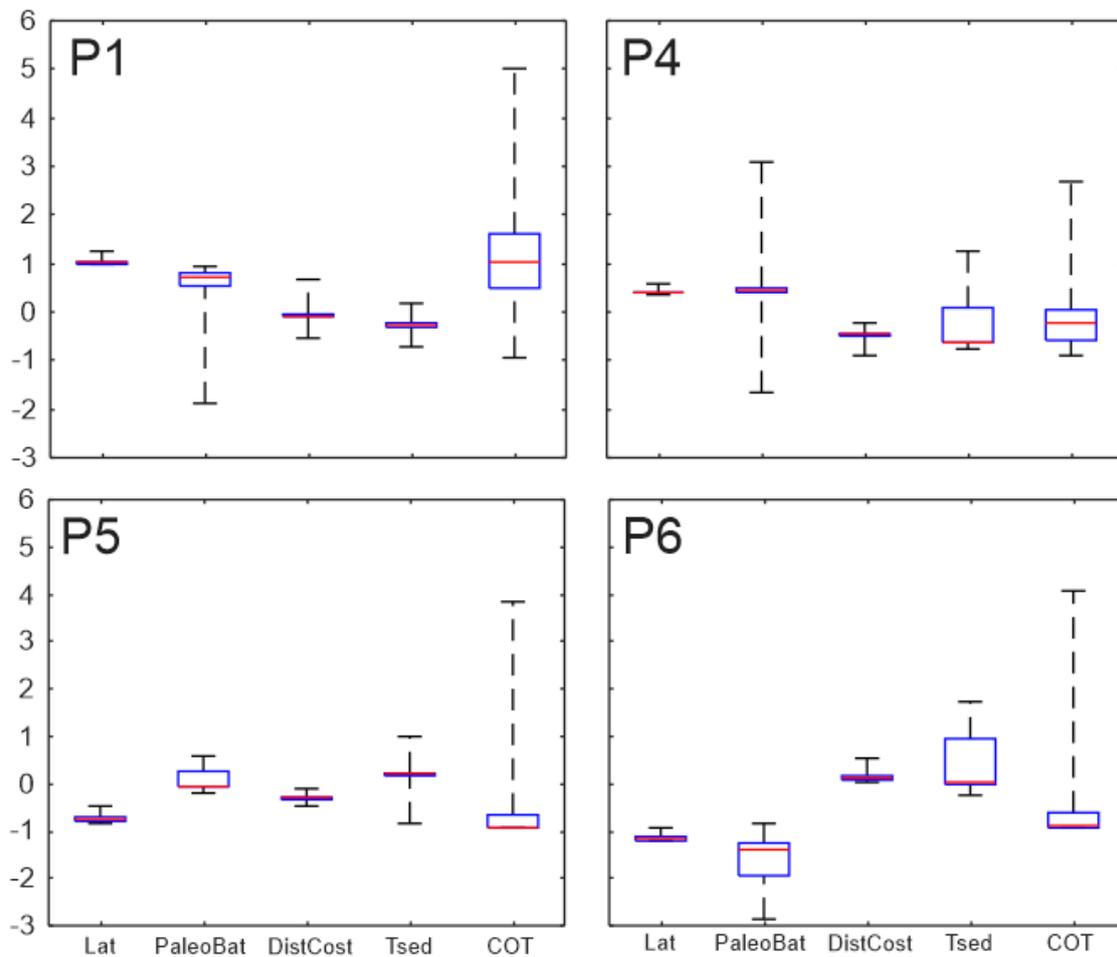


Figura 23 - Distribuição das variáveis preditoras e variável de resposta para os testemunhos P1, P4, P5 e P6, com dados normalizados por escore-Z. Lat: latitude; PaleoBat: paleobatimetria; DistCost: distâncias para a linha de costa; Tsed: taxa de sedimentação; COT: carbono orgânico total.

De forma análoga, a relação entre as variáveis preditoras e a variável de resposta pode ser mais detalhadamente analisada ao se observar suas distribuições normalizadas (escore-Z) ao longo do tempo (Fig. 24), o que representa a forma que o modelo considera dos dados durante o treinamento.

A despeito do Cretáceo Superior representar um intervalo de tempo em que houve a consolidação do Atlântico Sul como bacia oceânica francamente aberta, observa-se que, em linhas gerais, os fatores fisiográficos apresentam variações discretas. Uma vez que as alterações fisiográficas decorrem de processos tectônicos (Pérez-Díaz & Eagles, 2017) e variações eustáticas/euribáticas (Haq *et al.*, 2014) que agem paulatinamente no decorrer do tempo geológico, essas variações sutis eram esperadas. Por outro lado, considerando que a deposição de COT não responde unicamente às forçantes fisiográficas, o registro de valores anômalos ocorre a despeito de qualquer variação igualmente brusca da fisiografia (variáveis preditoras). Assim, durante os OAEs, por exemplo, ocorrem bruscas alterações ambientais que condicionam a acumulação de altos teores de CO₂, entretanto, sem acarretar mudanças igualmente significativas nas variáveis preditoras. Fora das janelas temporais em que valores anômalos de COT ocorrem, a variação sutil das variáveis preditoras (fisiografia) têm como resposta valores de COT próximos a zero (Fig. 24).

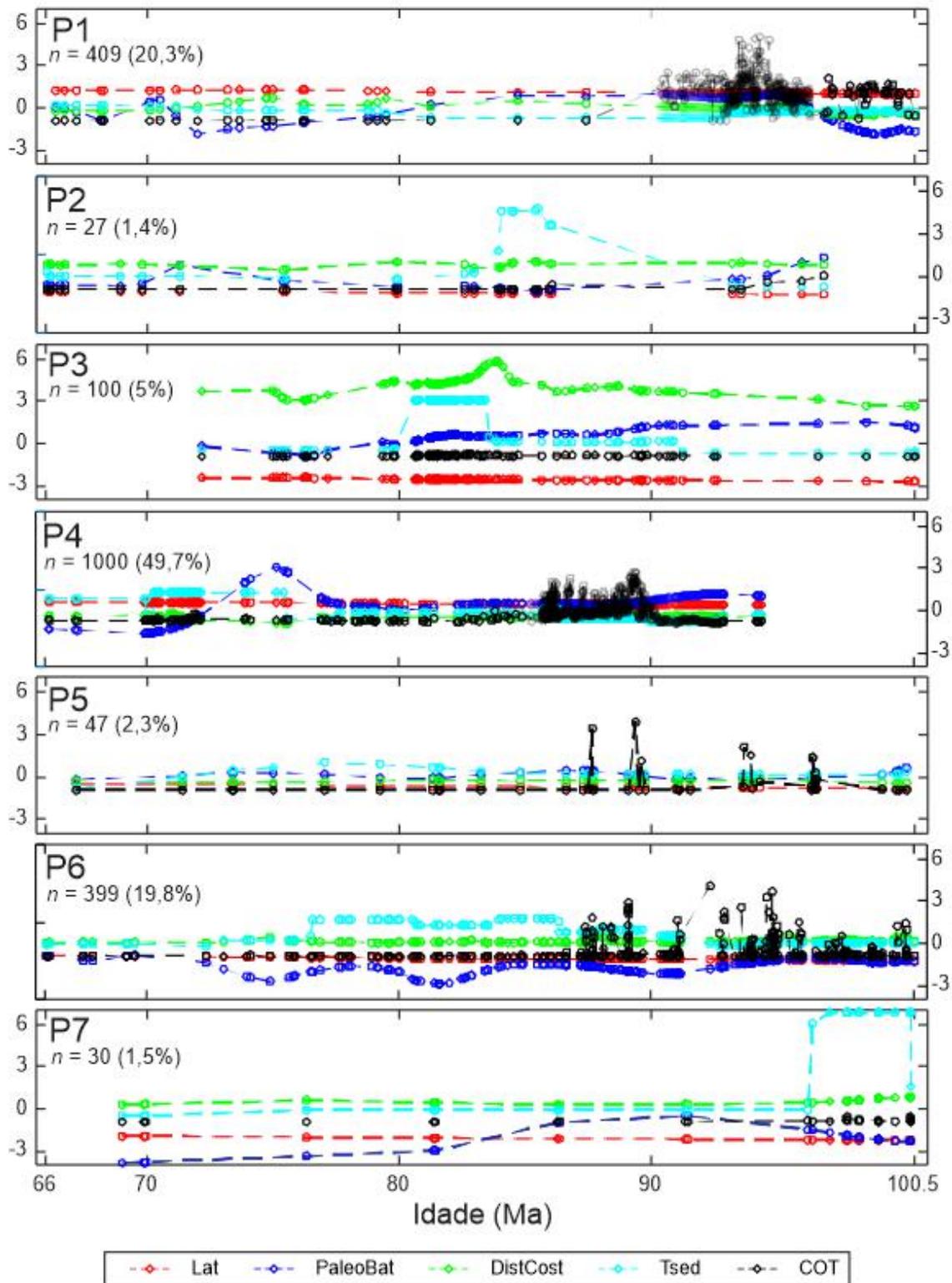


Figura 24 - Distribuição das variáveis preditoras e variável de resposta normalizadas por escore-Z ao longo do Cretáceo Superior. As porcentagens representam a parcela de amostras dentro do conjunto amostral referentes a cada um dos testemunhos. Lat: latitude; PaleoBat: paleobatimetria; DistCost: distâncias para a linha de costa; Tsed: taxa de sedimentação; COT: carbono orgânico total.

Levando-se em conta a natureza desproporcional das variâncias das preditoras (fisiografia) em relação à variável de resposta, ou seja, COT (Fig. 23), bem como o comportamento paradoxal entre essas variáveis em momentos específicos do tempo geológico (Fig. 24), fica evidente que outras forçantes não consideradas neste trabalho atuam na modulação do acúmulo de CO no assoalho marinho do Atlântico Sul ao longo do Cretáceo Superior (Behrooz *et al.*, 2018; Brumsack, 2006; McAnena *et al.*, 2013). Além disso, essas observações esclarecem, em parte, o desequilíbrio da performance do modelo, o qual resulta em um padrão de superestimação dos baixos valores de COT (0-5 wt%) e subestimação de valores elevados (>10 wt%), com melhores previsões (simetria do erro em torno da reta $E_{\text{residual}} = 0$) em valores intermediários (5-10 wt%), como se observa na figura 25.

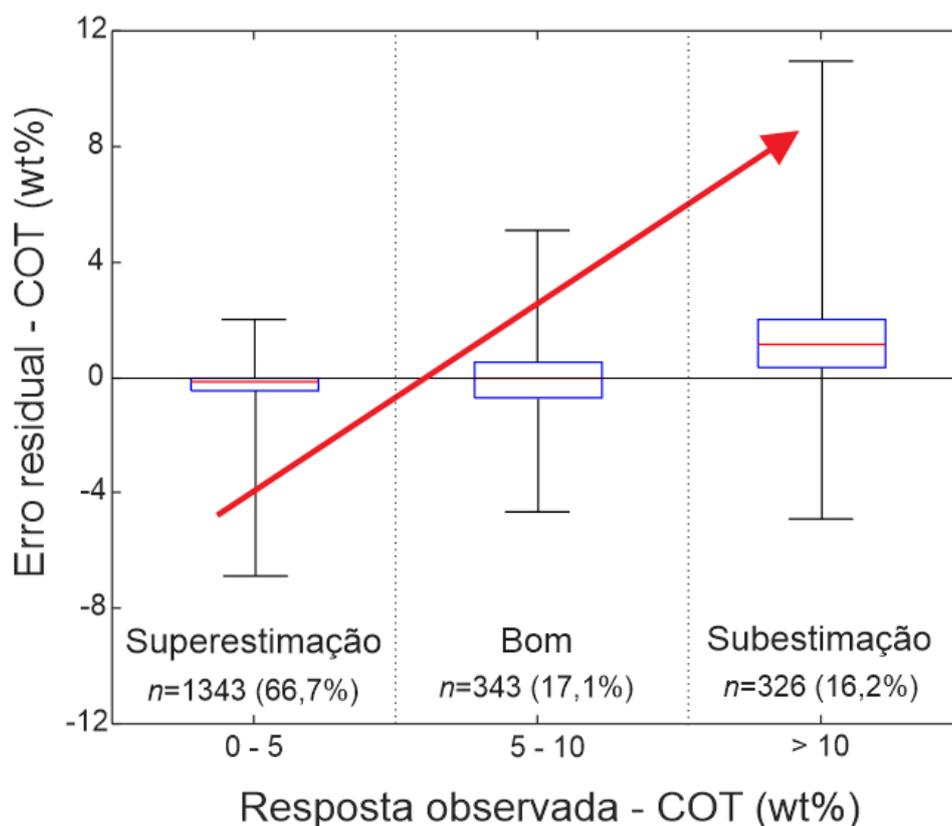


Figura 25 - Distribuição dos valores de erro residual para 3 faixas de intervalo de COT: 0-5 wt%, 5-10 wt% e >10 wt%. A seta vermelha indica a tendência do modelo em superestimar valores mais baixos de COT e subestimar valores mais elevados, alcançando a simetria em valores intermediários. As porcentagens correspondem à proporção de cada faixa de COT dentro do espaço amostral.

Apesar do n significativamente maior dos dados de COT entre 0 e 5 wt%, testou-se extrair 335 dados (média entre os n das faixas de COT 5-10 e >10 wt%) randomicamente do conjunto de dados dessa faixa de COT para se avaliar o efeito do

número de dados na tendência de superestimação da resposta modelada. Ainda assim, o resultado apresentou um padrão de superestimação da resposta modelada, descartando a hipótese de que o espaço amostral original desta faixa de COT estivesse influenciando tal tendência.

A despeito das limitações acima discutidas sobre a utilização de modelos regressivos associados à *Machine Learning* para avaliar o papel das variáveis fisiográficas e taxas de sedimentação na acumulação de CO durante o Cretáceo Superior no Atlântico Sul, foi possível observar que a fisiografia e as taxas de sedimentação explicam, com base no treinamento do modelo, 69% da variância dos dados observados de COT (Fig. 27).

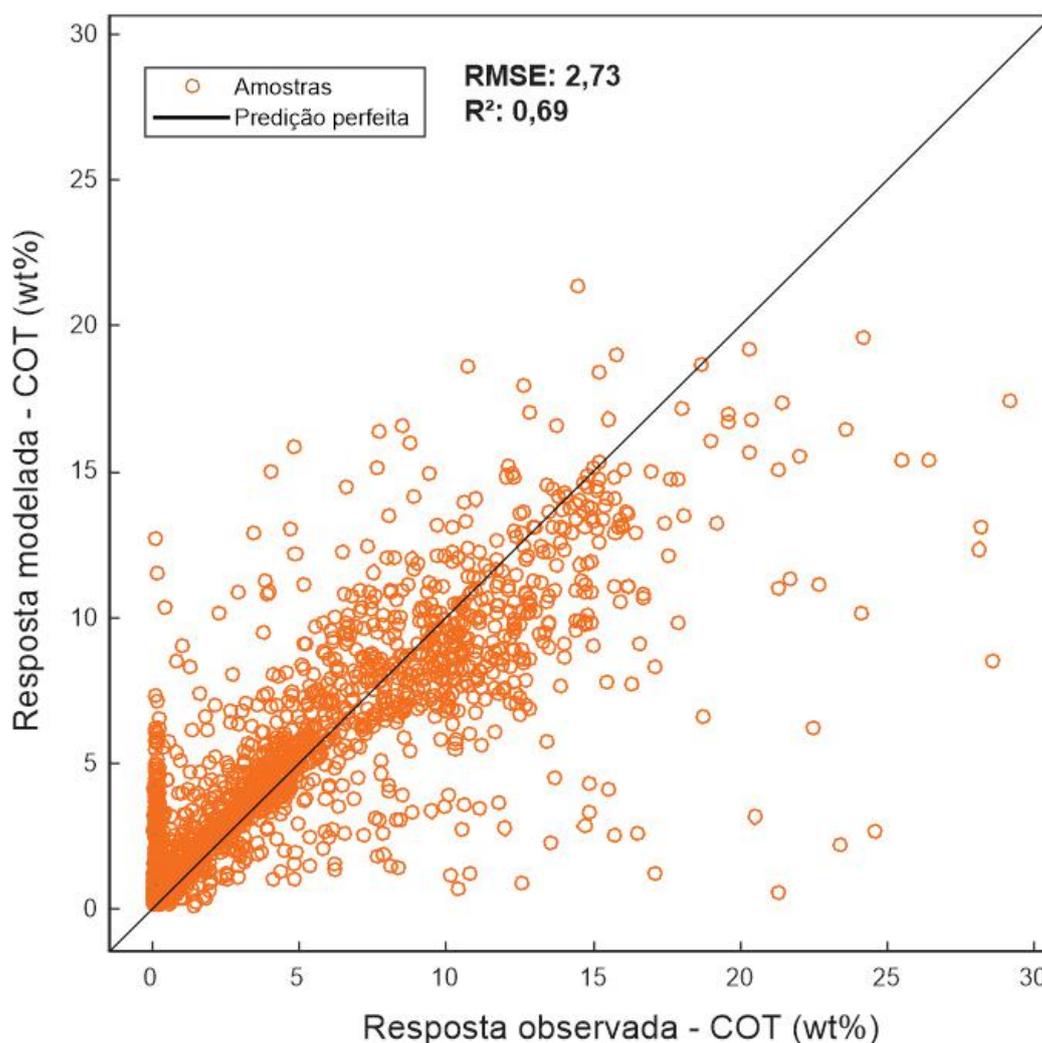


Figura 26 - Resposta observada vs. resposta modelada, com coeficiente de determinação de 0,69, considerando o modelo regressivo Rational Quadratic GPR (validação de 10-folds) adotado neste trabalho.

No entanto, para o real entendimento do poder preditivo do COT a partir dos fatores fisiográficos e taxas de sedimentação, se faz necessária a aplicação do modelo produzido em um conjunto de dados preditores independente do *dataset* de treinamento. Para tanto, utilizou-se os testemunhos de teste (T1-T4), cuja distribuição encontra-se descrita na Fig. 9 (**Material & Métodos**). Como resultado deste teste, obteve-se um coeficiente de determinação (R^2) de 0,37, indicando piora em relação aos dados obtidos na validação do treinamento do modelo. Além disso, a distribuição dos resíduos acompanha a tendência já observada no treinamento, ou seja, superestimação dos valores mais baixos da distribuição (<3 COT wt%) e subestimação dos valores mais elevados (>6 COT wt%) (Fig. 27). A diminuição da variância do COT explicada pelos preditores utilizados era esperada, uma vez que trata-se de um *dataset* independente do treinamento. O modelo regressivo associado à *Machine Learning* supervisionada demonstra-se, assim, pouco flexível para aplicação em dados novos (*overfitting*).

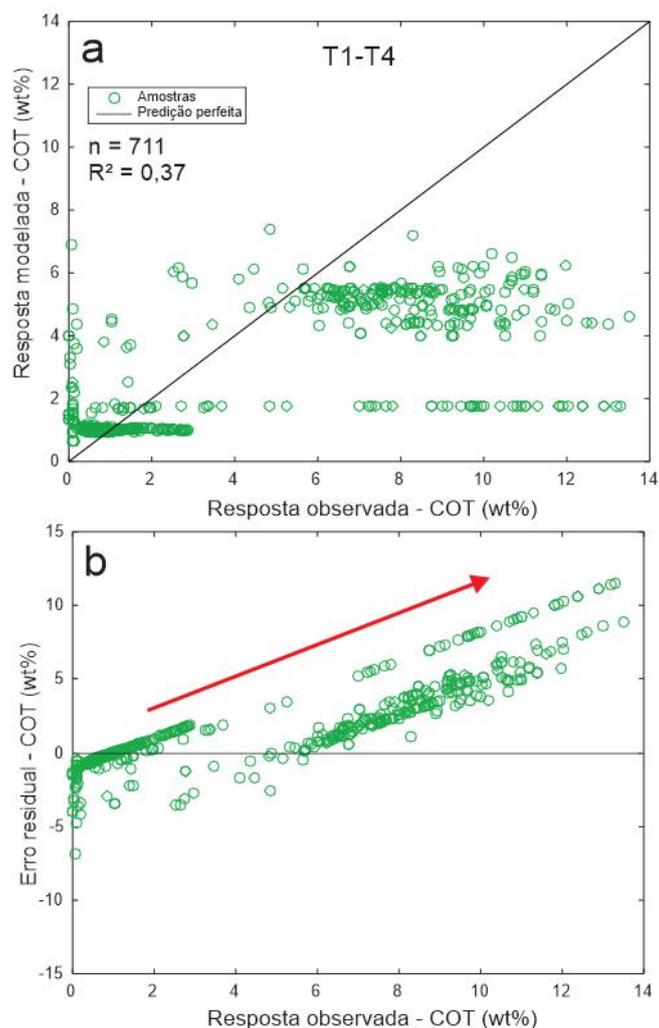


Figura 27 – Aplicação do modelo regressivo escolhido nos dados preditivos dos testemunhos de teste (T1-T4). (a) Resposta observada vs Resposta modelada; (b) Distribuição do erro residual vs Resposta observada.

Além dos potenciais motivos já citados anteriormente, podem ser considerados também para essa constatação uma ou mais das seguintes razões:

- A piora da performance na aplicação do modelo em *dataset* de teste sugere que os fatores fisiográficos influenciam indiretamente parâmetros significativos na modulação do acúmulo de CO, como o aporte de nutrientes, produtividade primária, condições redox na coluna d'água, etc, os quais são, também, afetados por outros fatores não fisiográficos, como o clima, por exemplo. Compreende-se, dessa forma, o poder limitado do uso dos fatores fisiográficos e taxas de sedimentação como preditores únicos da acumulação de CO em sedimentos marinhos do Atlântico Sul ao longo do Cretáceo Superior;
- A despeito das taxas de sedimentação serem diretamente associadas ao acúmulo de CO e, por isso, amplamente utilizadas em modelos preditivos, a acuidade das estimativas desses valores está relacionada ao modelo de idades utilizado, de forma que poucos *data* interpolados linearmente (como é o caso deste estudo) podem causar um *bias* em sua variância, afetando a sua correlação com o COT.

5. CONCLUSÃO

Este estudo utilizou modelos regressivos associados à *Machine Learning* para mensurar o papel de variáveis fisiográficas (latitude, paleobatimetria, paleodistâncias da linha de costa) e taxas de sedimentação na acumulação de CO ao longo do Cretáceo Superior em 7 testemunhos coletados no Atlântico Sul e Equatorial, seguindo a seguinte ordem lógica de trabalho (Fig. 28):

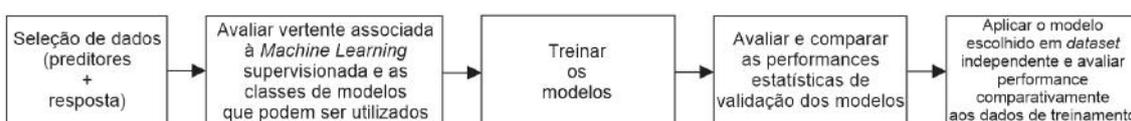


Figura 28 – Fluxograma com as etapas envolvidas no treinamento e aplicação do modelo regressivo associado à *Machine Learning* supervisionada.

Para tanto, um conjunto de 10.060 dados (2.012 dados para cada variável) foi selecionado em diferentes bases de dados, e confrontados entre si. A comparação entre os dados de COT mostrou diferenças estatisticamente significativas no acúmulo de carbono orgânico entre as margens leste e oeste do Atlântico Sul, o que pode ser atribuído à atuação de processos específicos associados, sobretudo, à dinâmica de massas d'água (e.g. processo de armadilha de nutrientes na região do Alto de Demerara), dinâmica atmosférica (e.g. expansão temporal das células de Hadley) e evolução paleobatimétrica (e.g. circulação de fundo no setor Sul do Atlântico Sul a partir do final do Cretáceo Inferior).

A partir do entendimento dos processos que conduziram a acumulação de CO em ambas as margens do Atlântico Sul, treinou-se 19 modelos regressivos associados à *Machine Learning* supervisionada para prever o conteúdo de COT (variável de resposta) em função das variáveis preditoras (latitude, paleobatimetria, paleodistâncias para a linha de costa e taxas de sedimentação). O modelo regressivo *Rational Quadratic GPR* teve a melhor performance, resultando em uma raiz do erro quadrático média de 2,73 e uma explicação de 69% da variância do COT. No entanto, um aprofundamento nas limitações desse modelo mostrou uma heterogeneidade e um viés dos erros residuais, a qual pode ser explicada, em princípio, pela natureza das variáveis preditoras face à variável de resposta, onde se observa maior variância desta se comparada àquelas. Essa constatação, aliada ao coeficiente de determinação obtido sugere que, como esperado, outros fatores/forçantes não considerados neste estudo, atuam como controladores da deposição do carbono orgânico no assoalho oceânico, incluindo possivelmente condições paleoambientais atípicas relacionadas ao estabelecimento de OAEs. A aplicação do modelo escolhido em conjunto de dados de teste, independentes do treinamento, reforça essa necessidade, uma vez que a variância do COT explicada pelos preditores resultou em 37%. Esse resultado sugere que o modelo é pouco flexível para aplicação em dados novos. Entende-se, também, o poder limitado do uso dos fatores fisiográficos e taxas de sedimentação como preditores únicos da acumulação de CO em sedimentos marinhos do Atlântico Sul ao longo do Cretáceo Superior. Além disso, artefatos associados às interpolações lineares das taxas de sedimentação potencialmente mascaram a sua real variância frente ao COT, impedindo que o seu verdadeiro papel na acumulação de CO seja avaliado.

A adoção de outros fatores fisiográficos, sedimentológicos, geoquímicos (e.g. $\delta^{13}\text{C}$) ou parâmetros de *logging* de poço (e.g. densidade, *gamma ray* espectral, resistividade, potencial espontâneo) como variáveis preditoras, bem como a exploração

de novas ferramentas supervisionadas e não supervisionadas de *Machine Learning* (e.g. Análise de Componentes Principais – *PCA*, misturas gaussianas) podem refinar o entendimento do acúmulo de COT em bacias oceânicas, bem como fornecer subsídios fundamentais para a melhoria de modelos de fácies orgânica e, conseqüentemente, contribuir para a minimização de riscos exploratórios e o aprofundamento do conhecimento científico nesse tema.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- Alaudah, Y., Michałowicz, P., Alfarraj, M., & AlRegib, G. (2019). A machine-learning benchmark for facies classification. *Interpretation*, 7(3), SE175–SE187. <https://doi.org/10.1190/int-2018-0249.1>
- Aliyuda, K., Howell, J., & Humphrey, E. (2020). Impact of geological variables in controlling oil-reservoir performance: An insight from a machine-learning technique. *SPE Reservoir Evaluation and Engineering*, 23(4), 1314–1327. <https://doi.org/10.2118/201196-PA>
- Ang, J. C., Mirzal, A., Haron, H., & Hamed, H. N. A. (2016). Supervised, unsupervised, and semi-supervised feature selection: A review on gene selection. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 13(5), 971–989. <https://doi.org/10.1109/TCBB.2015.2478454>
- ANP. (2017). *Bacia de Sergipe-Alagoas-Terra: Sumário Geológico e Setores em Oferta, 14a Rodada de Licitações*. http://rodadas.anp.gov.br/arquivos/Round14/Mapas/sumarios/Sumario_Geologico_R14_SEAL_Terra.pdf
- Arai, M. (1988). Geochemical reconnaissance of the Mid-Cretaceous anoxic event in the Santos basin, Brazil. *Revista Brasileira de Geociências*, 18(3), 273–282.
- Arai, M. (2009). Paleogeografia do Atlântico Sul no Aptiano: um novo modelo a partir de dados micropaleontológicos recentes. *Boletim de Geociências Da Petrobras*, 17(2), 331–351.
- Arai, M. (2014). Aptian/albian (early cretaceous) paleogeography of the south atlantic: A paleontological perspective. *Brazilian Journal of Geology*, 44(2), 339–350. <https://doi.org/10.5327/Z2317-4889201400020012>
- Arthur, M. A., & Sageman, B. B. (1994). Marine Black Shales: Depositional Mechanisms and Environments of Ancient Deposits. *Annual Review of Earth and Planetary Sciences*, 22(1), 499–551. <https://doi.org/10.1146/annurev.ea.22.050194.002435>
- Battin, T. J., Luysaert, S., Kaplan, L. A., Aufdenkampe, A. K., Richter, A., & Tranvik, L. J. (2009). The boundless carbon cycle. *Nature Geoscience*, 2(9), 598–600. <https://doi.org/10.1038/ngeo618>
- Beckmann, B., Hofmann, P., März, C., Schouten, S., Sinninghe Damsté, J. S., & Wagner, T. (2008). Coniacian-Santonian deep ocean anoxia/euxinia inferred from molecular and inorganic markers: Results from the Demerara Rise (ODP Leg 207). *Organic Geochemistry*, 39(8), 1092–1096. <https://doi.org/10.1016/j.orggeochem.2008.03.019>

Behrooz, L., Naafs, B. D. A., Dickson, A. J., Love, G. D., Batenburg, S. J., & Pancost, R. D. (2018). Astronomically Driven Variations in Depositional Environments in the South Atlantic During the Early Cretaceous. *Paleoceanography and Paleoclimatology*, 33(8), 894–912. <https://doi.org/10.1029/2018PA003338>

Beniest, A., Koptev, A., & Burov, E. (2017). Numerical models for continental break-up: Implications for the South Atlantic. *Earth and Planetary Science Letters*, 461, 176–189. <https://doi.org/10.1016/j.epsl.2016.12.034>

Bergen, K. J., Chen, T., & Li, Z. (2019). Preface to the focus section on machine learning in seismology. In *Seismological Research Letters* (Vol. 90, Issue 2 A, pp. 477–480). Seismological Society of America. <https://doi.org/10.1785/0220190018>

Bergman, S. C., Eldrett, J. S., & Minisini, D. (2020). *Phanerozoic Large Igneous Province, Petroleum System, and Source Rock Links*. <https://doi.org/10.1002/9781119507444.ch9>

Berner, R. a. (2003). Fuels and Atmospheric Composition. *Nature*, 426(November), 323–326. <http://www.ncbi.nlm.nih.gov/pubmed/14628061>

Berrocso, Á. J., MacLeod, K. G., Martin, E. E., Bourbon, E., Londoño, C. I., & Basak, C. (2010). Nutrient trap for Late Cretaceous organic-rich black shales in the tropical North Atlantic. *Geology*, 38(12), 1111–1114. <https://doi.org/10.1130/G31195.1>

Bianchi, T. S., Cui, X., Blair, N. E., Burdige, D. J., & Eglinton, T. I. (2017). Centers of Organic Carbon Burial and Oxidation at the Land-Ocean Interface Organic Geochemistry Centers of organic carbon burial and oxidation at the land-ocean interface. *Organic Geochemistry*, 115(October), 138–155. <https://doi.org/10.1016/j.orggeochem.2017.09.008>

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307–327. [https://doi.org/10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1)

Boudinot, F. G., Dildar, N., Leckie, R. M., Parker, A., Jones, M. M., Sageman, B. B., Bralower, T. J., & Sepúlveda, J. (2020). Neritic ecosystem response to Oceanic Anoxic Event 2 in the Cretaceous Western Interior Seaway, USA. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 546. <https://doi.org/10.1016/j.palaeo.2020.109673>

Breistroffer, M. (1937). Sur la stratigraphie du Crétacé moyen en Chartreuse. *Comptes Rendus de l'Academie Des Sciences*, 202, 1691–1693.

Brumsack, H. J. (2006). The trace metal content of recent organic carbon-rich sediments: Implications for Cretaceous black shale formation. In *Palaeogeography, Palaeoclimatology, Palaeoecology* (Vol. 232, Issues 2–4, pp. 344–361). <https://doi.org/10.1016/j.palaeo.2005.05.011>

- Carpentier, B., Huc, A., & Bessereau, G. (1991). Wirelogging and source rocks- Estimation of organic carbon content by the CARBOLOG Method. Fossil fuels View project Caprock View project. *The Log Analyst*, 32(3), 279–297. <https://www.researchgate.net/publication/269873585>
- Cartapanis, O., Bianchi, D., Jaccard, S. L., & Galbraith, E. D. (2016). Global pulses of organic carbon burial in deep-sea sediments during glacial maxima. *Nature Communications*, 7. <https://doi.org/10.1038/ncomms10796>
- Chang, S., Cohen, T., & Ostdiek, B. (2018). What is the machine learning? *Physical Review D*, 97(5). <https://doi.org/10.1103/PhysRevD.97.056009>
- de Boer, P. L., & Smith, D. G. (2009). Orbital Forcing and Cyclic Sequences. In *Orbital Forcing and Cyclic Sequences* (Vol. 19, pp. 1–14). Blackwell Publishing Ltd. <https://doi.org/10.1002/9781444304039.ch1>
- Demaison, G. J., & Moore, G. T. (1980). Anoxic Environments and Oil Source Bed Genesis. *Organic Geochemistry*, 2(1), 9–31. [https://doi.org/10.1016/0146-6380\(80\)90017-0](https://doi.org/10.1016/0146-6380(80)90017-0)
- Donnadieu, Y., Pucéat, E., Moiroud, M., Guillocheau, F., & Deconinck, J. F. (2016). A better-ventilated ocean triggered by Late Cretaceous changes in continental configuration. *Nature Communications*, 7. <https://doi.org/10.1038/ncomms10316>
- Dummann, W., Steinig, S., Hofmann, P., Flögel, S., Osborne, A. H., Frank, M., Herrle, J. O., Bretschneider, L., Sheward, R. M., & Wagner, T. (2020). The impact of Early Cretaceous gateway evolution on ocean circulation and organic carbon burial in the emerging South Atlantic and Southern Ocean basins. *Earth and Planetary Science Letters*, 530(May), 115890. <https://doi.org/10.1016/j.epsl.2019.115890>
- Dummann, W., Steinig, S., Hofmann, P., Lenz, M., Kusch, S., Flögel, S., Herrle, J. O., Hallmann, C., Rethemeyer, J., Kasper, H. U., & Wagner, T. (2021). Driving mechanisms of organic carbon burial in the Early Cretaceous South Atlantic Cape Basin (DSDP Site 361). *Climate of the Past*, 17(1), 469–490. <https://doi.org/10.5194/cp-17-469-2021>
- el Naqa, I., Ruan, D., Valdes, G., Dekker, A., McNutt, T., Ge, Y., Wu, Q. J., Oh, J. H., Thor, M., Smith, W., Rao, A., Fuller, C., Xiao, Y., Manion, F., Schipper, M., Mayo, C., Moran, J. M., & ten Haken, R. (2018). Machine learning and modeling: Data, validation, communication challenges. *Medical Physics*, 45(10), 834–840. <https://doi.org/10.1002/mp.12811>
- Emmel, B., Baskoro, A., de Jager, G., Grøver, A., & Roli, O. A. (2018). The influence of paleo-bathymetry on total organic carbon distribution tested in the Cretaceous Hammerfest Basin, Barents Sea. *Marine and Petroleum Geology*, 98, 491–506. <https://doi.org/10.1016/j.marpetgeo.2018.09.003>
- Erbacher, J., & Thurow, J. (1997). Influence of oceanic anoxic events on the evolution of mid-Cretaceous radiolaria in the North Atlantic and western Tethys. *Marine Micropaleontology*, 30(1–3), 139–158. [https://doi.org/10.1016/S0377-8398\(96\)00023-0](https://doi.org/10.1016/S0377-8398(96)00023-0)

- Ferreira, D., Cessi, P., Coxall, H. K., de Boer, A., Dijkstra, H. A., Drijfhout, S. S., Eldevik, T., Harnik, N., McManus, J. F., Marshall, D. P., Nilsson, J., Roquet, F., Schneider, T., & Wills, R. C. (2018). Atlantic-Pacific Asymmetry in Deep Water Formation. *Annual Review of Earth and Planetary Sciences*, *46*, 327–352. <https://doi.org/10.1146/annurev-earth-082517-010045>
- Flögel, S., Beckmann, B., Hofmann, P., Bornemann, A., Westerhold, T., Norris, R. D., Dullo, C., & Wagner, T. (2008). Evolution of tropical watersheds and continental hydrology during the Late Cretaceous greenhouse; impact on marine carbon burial and possible implications for the future. *Earth and Planetary Science Letters*, *274*(1–2), 1–13. <https://doi.org/10.1016/j.epsl.2008.06.011>
- Forster, A., Kuypers, M. M. M., Turgeon, S. C., Brumsack, H. J., Petrizzo, M. R., & Sinninghe Damsté, J. S. (2008). The Cenomanian/Turonian oceanic anoxic event in the South Atlantic: New insights from a geochemical study of DSDP Site 530A. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *267*(3–4), 256–283. <https://doi.org/10.1016/j.palaeo.2008.07.006>
- Foster, G. L., Royer, D. L., & Lunt, D. J. (2017). Future climate forcing potentially without precedent in the last 420 million years. *Nature Communications*, *8*. <https://doi.org/10.1038/ncomms14845>
- França, R. L., del Rey, A. C., Tagliari, C. V., Brandão, J. R., & Fontanelli, P. de R. (2007). Bacia do Espírito Santo. *B. Geoci. Petrobras*, *15*(2), 501–509. https://www.researchgate.net/publication/279624221_Espirito_santo_basin_Bacia_do_Espirito_Santo
- Fromm, T., Planert, L., Jokat, W., Ryberg, T., Behrmann, J. H., Weber, M. H., & Haberland, C. (2015). South Atlantic opening: A plume-induced breakup? *Geology*, *43*(10), 931–935. <https://doi.org/10.1130/G36936.1>
- Garrels, R. M., Lerman, A., & Mackenzie, F. T. (1976). Controls of atmospheric O₂ and CO₂: Past present and future Carbon dioxide-carbonic acid system in seawater View project. *American Scientist*, *63*(3), 306–315. <https://www.researchgate.net/publication/230890608>
- Goetz, J. N., Brenning, A., Petschko, H., & Leopold, P. (2015). Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Computers and Geosciences*, *81*, 1–11. <https://doi.org/10.1016/j.cageo.2015.04.007>
- Granjeon, D., & Chauveau, B. (2014). *Sedimentary basin development method using stratigraphic simulation coupled with an organic matter production and degradation model* (Patent No. US 2014/0163883 A1). Granjeon D. & Chauveau, B. <https://patentimages.storage.googleapis.com/05/d5/4b/b12cddd6a33a34/US20140163883A1.pdf>
- Haq, B. U. (2014). Cretaceous eustasy revisited. *Global and Planetary Change*, *113*(January 2013), 44–58. <https://doi.org/10.1016/j.gloplacha.2013.12.007>

- Hay, W. W. (2008). Evolving Ideas About the Cretaceous Climate and Ocean Circulation. *Cretaceous Research*, 29(5–6), 725–753. <https://doi.org/10.1016/j.cretres.2008.05.025>
- Hay, W. W. (2009). Cretaceous Oceans and Ocean Modeling. In *Cretaceous Ocean Red Beds: Stratigraphy, Composition, Origins, and Paleooceanographic and Paleoclimatic Significance* (Issue February, pp. 243–271). <https://doi.org/10.2110/sepmsp.091.233>
- Hofmann, P., Ricken, W., Schwark, L., & Leythaeuser, D. (2000). Carbon-sulfur-iron relationships and $\delta^{13}\text{C}$ of organic matter for late Albian sedimentary rocks from the North Atlantic Ocean: Paleooceanographic implications. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 163(3–4), 97–113. [https://doi.org/10.1016/S0031-0182\(00\)00147-4](https://doi.org/10.1016/S0031-0182(00)00147-4)
- Holtar, E., & Forsberg, A. W. (2000). Postrift development of the Walvis Basin, Namibia: Results from the exploration campaign in Quadrant 1911. In M. R. Mello & B. J. Katz (Eds.), *Petroleum Systems of South Atlantic Margins* (pp. 429–446). Am. Assoc. of Petrol. Geol.
- Jenkyns, H. C. (2010). Geochemistry of oceanic anoxic events. *Geochemistry, Geophysics, Geosystems*, 11(3), 1–30. <https://doi.org/10.1029/2009GC002788>
- Karner, G. D., Driscoll, N. W., McGinnis, J. P., Brumbaugh, W. D., & Cameron, N. R. (1997). Tectonic significance of syn-rift sediment packages across the Gabon - Cabinda continental margin. *Marine and Petroleum Geology*, 14(7–8), 973–1000. [https://doi.org/10.1016/s0264-8172\(97\)00040-8](https://doi.org/10.1016/s0264-8172(97)00040-8)
- Karpatne, A., Ebert-Uphoff, I., Ravela, S., Babaie, H. A., & Kumar, V. (2019a). Machine Learning for the Geosciences: Challenges and Opportunities. *IEEE Transactions on Knowledge and Data Engineering*, 31(8), 1544–1554. <https://doi.org/10.1109/TKDE.2018.2861006>
- Karpatne, A., Ebert-Uphoff, I., Ravela, S., Babaie, H. A., & Kumar, V. (2019b). Machine Learning for the Geosciences: Challenges and Opportunities. *IEEE Transactions on Knowledge and Data Engineering*, 31(8), 1544–1554. <https://doi.org/10.1109/TKDE.2018.2861006>
- Katz, B. J., Dawson, W. C., Liro, L. M., Robison, V. D., & Stonebraker, J. D. (2000). Petroleum systems of the Ogooué Delta, offshore Gabon. In M. R. Mello & B. J. Katz (Eds.), *Petroleum Systems of South Atlantic Margins* (pp. 247–256). Am. Assoc. of Petrol. Geol.
- Kidder, D. L., & Worsley, T. R. (2010). Phanerozoic Large Igneous Provinces (LIPs), HEATT (Haline Euxinic Acidic Thermal Transgression) episodes, and mass extinctions. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 295(1–2), 162–191. <https://doi.org/10.1016/j.palaeo.2010.05.036>

- Kong, Q., Trugman, D. T., Ross, Z. E., Bianco, M. J., Meade, B. J., & Gerstoft, P. (2019). Machine learning in seismology: Turning data into insights. *Seismological Research Letters*, 90(1), 3–14. <https://doi.org/10.1785/0220180259>
- Koutsoukos, E. A. M., Leary, P. N., & Hart, M. B. (1990). Latest Cenomanian-earliest Turonian low-oxygen tolerant benthonic foraminifera: a case-study from the Sergipe basin (N.E. Brazil) and the western Anglo-Paris basin (southern England). *Palaeogeography, Palaeoclimatology, Palaeoecology*, 77(2), 145–177. [https://doi.org/10.1016/0031-0182\(90\)90130-Y](https://doi.org/10.1016/0031-0182(90)90130-Y)
- Kruijs, E., & Barron, E. (1990). Climate Model Prediction of Paleoproductivity and Potential Source-Rock Distribution. In A. Y. Huc (Ed.), *Deposition of Organic Facies*. American Association of Petroleum Geologists. <https://doi.org/10.1306/St30517C13>
- Kukla, P. A., Strozyk, F., & Mohriak, W. U. (2018). South Atlantic salt basins – Witnesses of complex passive margin evolution. *Gondwana Research*, 53(April), 41–57. <https://doi.org/10.1016/j.gr.2017.03.012>
- Lary, D. J., Alavi, A. H., Gandomi, A. H., & Walker, A. L. (2016). Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, 7(1), 3–10. <https://doi.org/10.1016/j.gsf.2015.07.003>
- Locklair, R., Sageman, B., & Lerman, A. (2011). Marine carbon burial flux and the carbon isotope record of Late Cretaceous (Coniacian-Santonian) Oceanic Anoxic Event III. *Sedimentary Geology*, 235(1–2), 38–49. <https://doi.org/10.1016/j.sedgeo.2010.06.026>
- Maia, A. G. (2019). *Econometria - Conceitos e Aplicações* (1st ed.). Saint Paul Editora.
- Mann, U., & Zweigel, J. (2008). Modelling Source-Rock Distribution and Quality Variations: The Organic Facies Modelling Approach. In *Analogue and Numerical Modelling of Sedimentary Systems: From Understanding to Prediction*. <https://doi.org/10.1002/9781444303131.ch11>
- Marjanović, M., Kovačević, M., Bajat, B., & Voženílek, V. (2011). Landslide susceptibility assessment using SVM machine learning algorithm. *Engineering Geology*, 123(3), 225–234. <https://doi.org/10.1016/j.enggeo.2011.09.006>
- Martin, E. E., MacLeod, K. G., Jiménez Berrocso, A., & Bourbon, E. (2012). Water mass circulation on Demerara Rise during the Late Cretaceous based on Nd isotopes. *Earth and Planetary Science Letters*, 327–328, 111–120. <https://doi.org/10.1016/j.epsl.2012.01.037>
- McAnena, A., Flögel, S., Hofmann, P., Herrle, J. O., Griesand, A., Pross, J., Talbot, H. M., Rethemeyer, J., Wallmann, K., & Wagner, T. (2013). Atlantic cooling associated with a marine biotic crisis during the mid-Cretaceous period. *Nature Geoscience*, 6(7), 558–561. <https://doi.org/10.1038/ngeo1850>

Mélières, M. A., Rossignol-Strick, M., & Malaizé, B. (1997). Relation between low latitude insolation and $\delta^{18}\text{O}$ change of atmospheric oxygen for the last 200 kyrs, as revealed by Mediterranean sapropels. *Geophysical Research Letters*, 24(10), 1235–1238. <https://doi.org/10.1029/97GL01025>

Mello, M. R., Koutsoukos, E. A. M., Hart, M. B., Brassell, S. C., & Maxwell, J. R. (1989). Late cretaceous anoxic events in the Brazilian continental margin. *Organic Geochemistry*, 14(5), 529–542. [https://doi.org/10.1016/0146-6380\(89\)90033-8](https://doi.org/10.1016/0146-6380(89)90033-8)

Mendonça Filho, J. G., Rêgo, T., Oliveira Mendona, J. de, de Oliveira, A. D., da Silva, T. F., Franco, N., & da Silva, F. S. (2012). Organic Facies: Palynofacies and Organic Geochemistry Approaches. In *Geochemistry - Earth's System Processes* (Issue May 2016). InTech. <https://doi.org/10.5772/47126>

Meyer, K. M., & Kump, L. R. (2008). Oceanic euxinia in Earth history: Causes and consequences. *Annual Review of Earth and Planetary Sciences*, 36, 251–288. <https://doi.org/10.1146/annurev.earth.36.031207.124256>

Mohaghegh, S. D. (2020, April 1). *Subsurface analytics: Contribution of artificial intelligence and machine learning to reservoir engineering, reservoir modeling, and reservoir management*. Petroleum Exploration and Development; Elsevier B.V. [https://doi.org/10.1016/S1876-3804\(20\)60041-6](https://doi.org/10.1016/S1876-3804(20)60041-6)

Mohriak, W. U. (2003). Bacias Sedimentares da Margem Continental Brasileira. In L. A. Bizzi, C. Schobbenhaus, R. M. Vidotti, & J. H. Gonçalves (Eds.), *Bacias Sedimentares da Margem Continental Brasileira* (pp. 87–94). CPRM. http://www.cprm.gov.br/publique/media/recursos_minerais/livro_geo_tec_rm/capIII-a.pdf

Mohriak, W. U., Szatmari, P., & Anjos, S. (2012). Salt: Geology and tectonics of selected Brazilian basins in their global context. *Geological Society Special Publication*, 363(1), 131–158. <https://doi.org/10.1144/SP363.7>

Monteiro, F. M., Pancost, R. D., Ridgwell, A., & Donnadieu, Y. (2012). Nutrients as the dominant control on the spread of anoxia and euxinia across the Cenomanian-Turonian oceanic anoxic event (OAE2): Model-data comparison. *Paleoceanography*, 27(4). <https://doi.org/10.1029/2012PA002351>

Mort, H. P., Adatte, T., Keller, G., Bartels, D., Föllmi, K. B., Steinmann, P., Berner, Z., & Chellai, E. H. (2008). Organic carbon deposition and phosphorus accumulation during Oceanic Anoxic Event 2 in Tarfaya, Morocco. *Cretaceous Research*, 29(5–6), 1008–1023. <https://doi.org/10.1016/j.cretres.2008.05.026>

Murphy, D. P., & Thomas, D. J. (2013). The evolution of Late Cretaceous deep-ocean circulation in the Atlantic basins: Neodymium isotope evidence from South Atlantic drill sites for tectonic controls. *Geochemistry, Geophysics, Geosystems*, 14(12), 5323–5340. <https://doi.org/10.1002/2013GC004889>

Nederbragt, A. J., Thurow, J., Vonhof, H., & Brumsack, H.-J. (2004). Modelling oceanic carbon and phosphorus fluxes: implications for the cause of the late Cenomanian Oceanic Anoxic Event (OAE2). *Journal of the Geological Society*, 161(4), 721–728. <https://doi.org/10.1144/0016-764903-075>

Oliveira, V. C. B. de, Silva, C. M. de A., Borghi, L. F., & Carvalho, I. de S. (2019). Lacustrine coquinas and hybrid deposits from rift phase: Pre-Salt, lower Cretaceous, Campos Basin, Brazil. *Journal of South American Earth Sciences*, 95(July), 102254. <https://doi.org/10.1016/j.jsames.2019.102254>

Osisanwo, F. Y., Akinsola, J. E. T., Awodele, O., Olakanmi, J. O., & Akinjobi, J. (2017). Supervised Machine Learning Algorithms: Classification and Comparison. *International Journal of Computer Trends and Technology*, 48(3), 128–138. <https://doi.org/10.14445/22312803/IJCTT-V48P126>

Pancost, R. D., Crawford, N., Magness, S., Turner, A., Jenkyns, H. C., & Maxwell, J. R. (2004). Further evidence for the development of photic-zone euxinic conditions during Mesozoic oceanic anoxic events. *Journal of the Geological Society*, 161(3), 353–364. <https://doi.org/10.1144/0016764903-059>

Passey, Q. R., Moretti, F. J., Kulla, J. B., Creaney, S., & Stroud, J. D. (1990). A practical model for organic richness from porosity and resistivity logs. *AAPG Bulletin*, 74(12), 1777–1794. <https://www.osti.gov/biblio/5682561>

Pedersen, T. F., & Calvert, S. E. (1990). Anoxia vs. Productivity: What Controls the Formation of Organic-Carbon-Rich Sediments and Sedimentary Rocks? *AAPG Bulletin*, 74. <https://doi.org/10.1306/0C9B232B-1710-11D7-8645000102C1865D>

Pérez-Díaz, L., & Eagles, G. (2017). South Atlantic paleobathymetry since early Cretaceous. *Scientific Reports*, 7(1), 11819. <https://doi.org/10.1038/s41598-017-11959-7>

Peters, K. E., & Cassa, M. R. (1994). Applied Source Rock Geochemistry. In L. B. Magoon & W. G. Dow (Eds.), *The Petroleum System. From Source to Trap* (pp. 93–120).

Pettijohn, F. J. (1957). Sedimentary Rocks. In *Geological Magazine* (Issue 6). Edward Arnold. <https://doi.org/10.1017/S0016756800070254>

Poulsen, C. J., Barron, E. J., Arthur, M. A., & Peterson, W. H. (2001). Response of the Mid-Cretaceous global oceanic circulation to tectonic and CO₂ forcings. *Paleoceanography*, 16(6), 576–592. <https://doi.org/10.1029/2000PA000579>

Raven, M. R., Fike, D. A., Bradley, A. S., Gomes, M. L., Owens, J. D., & Webb, S. A. (2019). Paired organic matter and pyrite $\delta^{34}\text{S}$ records reveal mechanisms of carbon, sulfur, and iron cycle disruption during Ocean Anoxic Event 2. *Earth and Planetary Science Letters*, 512, 27–38. <https://doi.org/10.1016/j.epsl.2019.01.048>

Robinson, S. A., Clarke, L. J., Nederbragt, A., & Wood, I. G. (2008). Mid-Cretaceous oceanic anoxic events in the Pacific Ocean revealed by carbon-isotope stratigraphy of the Calera Limestone, California, USA. *Bulletin of the Geological Society of America*, *120*(11–12), 1416–1426. <https://doi.org/10.1130/B26350.1>

Robinson, S. A., Heimhofer, U., Hesselbo, S. P., & Petrizzo, M. R. (2017). Mesozoic climates and oceans – a tribute to Hugh Jenkyns and Helmut Weissert. *Sedimentology*, *64*(1), 1–15. <https://doi.org/10.1111/sed.12349>

Robinson, S. A., Murphy, D. P., Vance, D., & Thomas, D. J. (2010). Formation of “Southern Component Water” in the Late Cretaceous: Evidence from Nd-isotopes. *Geology*, *38*(10), 871–874. <https://doi.org/10.1130/G31165.1>

Robinson, S. A., & Vance, D. (2012). Widespread and synchronous change in deep-ocean circulation in the North and South Atlantic during the Late Cretaceous. *Paleoceanography*, *27*(1), 1–8. <https://doi.org/10.1029/2011PA002240>

Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geology Reviews*, *71*, 804–818. <https://doi.org/10.1016/j.oregeorev.2015.01.001>

Santos, T. P., Bione, F. R., Venancio, I. M., Bernardes, M. C., Belem, A. L., Lisboa, L. P., Franco, D. R., Diaz, R. A., Moreira, M., Souza, I. V., & Albuquerque, A. L. S. (n.d.). Late Cretaceous astrochronology, organic carbon evolution and paleoclimate inferences for the subtropical western South Atlantic, Espirito Santo Basin. *Cretaceous Research*.

Schlanger, S. O., & Jenkyns, H. C. (1976). Cretaceous Oceanic Anoxic Events: Causes and Consequences. *Geologie in Mijnbouw*, *55*(3–4), 179–184. <https://www.researchgate.net/publication/27710997>

Schoellkopf, N. B., & Patterson, B. A. (2000). Petroleum systems of offshore Cabinda. In M. R. Mello & B. J. Katz (Eds.), *Petroleum Systems of South Atlantic Margins* (pp. 362–376). Am. Assoc. of Petrol. Geol.

Schoepfer, S. D., Shen, J., Wei, H., Tyson, R. v., Ingall, E., & Algeo, T. J. (2015). Total organic carbon, organic phosphorus, and biogenic barium fluxes as proxies for paleomarine productivity. In *Earth-Science Reviews* (Vol. 149, pp. 23–52). Elsevier. <https://doi.org/10.1016/j.earscirev.2014.08.017>

Schwarzkopf, T. A. (1993). Model for prediction of organic carbon content in possible source rocks. *Marine and Petroleum Geology*, *10*(5), 478–492. [https://doi.org/10.1016/0264-8172\(93\)90049-X](https://doi.org/10.1016/0264-8172(93)90049-X)

Scotese, C. (2016). *Paleomap Paleoatlas for Gplates and the Paleodataplotter Program*. <https://doi.org/10.1130/abs/2016nc-275387>

Scotese, C. R., & Wright, N. M. (2018). *PALEOMAP Paleodigital Elevation Models (PaleoDEMS) for the Phanerozoic* (pp. 1–26). Research School of Earth Sciences, Australian National University. <https://www.earthbyte.org/paleodem-resource-scotese-and-wright-2018/>

Setoyama, E., & Kanungo, S. (2020). Mesozoic biochronostratigraphy and paleoenvironment of the South Atlantic: A revised framework based on 20 DSDP and ODP deep-water sites. *Journal of South American Earth Sciences*, 99(February), 102511. <https://doi.org/10.1016/j.jsames.2020.102511>

Stein, R., Rullkötter, J., & Welte, D. H. (1989). Changes in paleoenvironments in the Atlantic Ocean during Cretaceous times: results from black shales studies. *Geologische Rundschau*, 78(3), 883–901. <https://doi.org/10.1007/BF01829328>

Thompson, D. L., Stilwell, J. D., & Hall, M. (2015). Lacustrine carbonate reservoirs from Early Cretaceous rift lakes of Western Gondwana: Pre-Salt coquinas of Brazil and West Africa. *Gondwana Research*, 28(1), 26–51. <https://doi.org/10.1016/j.gr.2014.12.005>

Torsvik, T. H., Rousse, S., Labails, C., & Smethurst, M. A. (2009). A new scheme for the opening of the South Atlantic Ocean and the dissection of an Aptian salt basin. *Geophys. J. Int.*, 177, 1315–1333. <https://doi.org/10.1111/j.1365-246X.2009.04137.x>

Trabucho Alexandre, J., Tuenter, E., Henstra, G. A., van der Zwan, K. J., van de Wal, R. S. W., Dijkstra, H. A., & de Boer, P. L. (2010). The mid-Cretaceous North Atlantic nutrient trap: Black shales and OAEs. *Paleoceanography*, 25(4), 1–14. <https://doi.org/10.1029/2010PA001925>

Tsikos, H., Jenkyns, H. C., Walsworth-Bell, B., Petrizzo, M. R., Forster, A., Kolonic, S., Erba, E., Premoli Silva, I., Baas, M., Wagner, T., & Sinninghe Damsté, J. S. (2004). Carbon-isotope stratigraphy recorded by the Cenomanian-Turonian Oceanic Anoxic Event: Correlation and implications based on three key localities. *Journal of the Geological Society*, 161(4), 711–719. <https://doi.org/10.1144/0016-764903-077>

Valle, B., Dal' Bó, P. F., Mendes, M., Favoreto, J., Rigueti, A. L., Borghi, L., de Oliveira Mendonça, J., & Silva, R. (2019). The expression of the Oceanic Anoxic Event 2 (OAE2) in the northeast of Brazil (Sergipe-Alagoas Basin). *Palaeogeography, Palaeoclimatology, Palaeoecology*, 529(May), 12–23. <https://doi.org/10.1016/j.palaeo.2019.05.029>

Wagner, T., Hofmann, P., & Flögel, S. (2013). Marine black shale deposition and Hadley Cell dynamics: A conceptual framework for the Cretaceous Atlantic Ocean. *Marine and Petroleum Geology*, 43, 222–238. <https://doi.org/10.1016/j.marpetgeo.2013.02.005>

Wagner, T., Sinninghe Damsté, J. S., Hofmann, P., & Beckmann, B. (2004). Euxinia and primary production in Late Cretaceous eastern equatorial Atlantic surface waters fostered orbitally driven formation of marine black shales. In *Paleoceanography* (Vol. 19, Issue 3). <https://doi.org/10.1029/2003PA000898>

Wagreich, M. (2012). "OAE 3" - Regional Atlantic organic carbon burial during the Coniacian-Santonian. *Climate of the Past*, 8(5), 1447–1455. <https://doi.org/10.5194/cp-8-1447-2012>

Wen, Z., Jiang, S., Song, C., Wang, Z., & He, Z. (2019). Basin evolution, configuration styles, and hydrocarbon accumulation of the South Atlantic conjugate margins. *Energy Exploration and Exploitation*, 37(3), 992–1008. <https://doi.org/10.1177/0144598719840751>

Wilson, P. A., & Norris, R. D. (2001). Warm tropical ocean surface and global anoxia during the mid-Cretaceous period. *Nature*, 412(6845), 425–429. <https://doi.org/10.1038/35086553>

Yadav, S., & Shukla, S. (2016). Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification. *Proceedings - 6th International Advanced Computing Conference, IACC 2016*, 78–83. <https://doi.org/10.1109/IACC.2016.25>

Zhang, Y., & Ni, Q. (2020). Recent advances in quantum machine learning. *Quantum Engineering*, 2(1). <https://doi.org/10.1002/que2.34>

Zuo, R. (2017). Machine Learning of Mineralization-Related Geochemical Anomalies: A Review of Potential Methods. In *Natural Resources Research* (Vol. 26, Issue 4, pp. 457–464). Springer New York LLC. <https://doi.org/10.1007/s11053-017-9345-4>

Anexo A – Script para Interpolação de idades

```
load ('IODPnew2')

idx = find(strcmp('inserir nome do testemunho',{myIODP8test2.core}));

if isempty(idx)
    fprintf(1, '[ERROR] InterpIODP: Core não encontrado\n')
    return
end
[mindepth, imin] = min(myIODP8test2(idx).AgeMaDep(:,1));
[maxdepth, imax] = max(myIODP8test2(idx).AgeMaDep(:,1));
minage = myIODP8test2(idx).AgeMaDep(imin,2);
maxage = myIODP8test2(idx).AgeMaDep(imax,2);

[~, ia, ~] = unique(myIODP8test2(idx).AgeMaDep(:,1), 'stable');

xdata = myIODP8test2(idx).AgeMaDep(ia,:);
xdata = sortrows(xdata);

Fage = griddedInterpolant(xdata(:,1), xdata(:,2), 'linear', 'none');

if (~isempty(find(myIODP8test2(8).AgeDepthToc(:,1) < mindepth)))
    fprintf(1, '[WARNING] InterpIODP: Profundidades fora do modelo de idade atual\n')
    fprintf(1, 'Depth Min: %6.1f (age min %6.2f)\n', ...
        mindepth, minage)
end
if (~isempty(find(myIODP8test2(8).AgeDepthToc(:,1) > maxdepth)))
    fprintf(1, '[WARNING] InterpIODP: Profundidades fora do modelo de idade atual\n')
    fprintf(1, 'Depth Max: %6.1f (age max %6.2f)\n', ...
        maxdepth, maxage)
end
age = Fage(myIODP8test2(8).AgeDepthToc(:,1));

end
```

Anexo B – Script para criação do grid lat/lon

```
bounds = [-70 25 -60 15];

filesScottese = dir('./Scotese_Wright_2018_Maps_1-88_1degX1deg_PaleoDEMS_nc_v2/*.nc');

AgeScottese = 0:5:120;

filenc = strcat(filesScottese(1).folder, '/', filesScottese(1).name);

latScottese = ncread(filenc, 'lat');
lonScottese = ncread(filenc, 'lon');

[~, lonmin] = min(abs(lonScottese - bounds(1)));
[~, lonmax] = min(abs(lonScottese - bounds(2)));
```

```

[~,latmin] = min(abs(latScottese-bounds(3)));
[~,latmax] = min(abs(latScottese-bounds(4)));

[lonS,latS] = meshgrid(lonScottese,latScottese);
[glon,glat] =
meshgrid(lonScottese(lonmin:lonmax),latScottese(latmin:latmax));

[X,Y] = meshgrid(glon(1,:),glat(:,1));

for i=1:length(AgeScottese)

    PaleoBAT2(i).glat = [];
    PaleoBAT2(i).glon = [];
    PaleoBAT2(i).AgeMa = AgeScottese(i);
    PaleoBAT2(i).xwd = [];

    filenc =
strcat(filesScottese(i).folder,'/',filesScottese(i).name);
    xz = ncread(filenc,'z');
    xz = xz';

    newbat = interp2(lonS,latS,xz,glon,glat);

    PaleoBAT2(i).xwd = newbat;

end

PaleoBAT2(1).glat = glat;
PaleoBAT2(1).glon = glon;

```

Anexo C – Script para obtenção das paleobatimetrias e distâncias para a linha de costa

```

load('PaleoBAT.mat')

mylat = PaleoBAT2(1).glat;
mylon = PaleoBAT2(1).glon;
myage = [PaleoBAT2.AgeMa];

fprintf(1,'Criando o meshgrid para interpolação ...\n')
[X,Y,Z] = meshgrid(mylon(1,:),mylat(:,1),myage);
mybat =
reshape([PaleoBAT2.xwd],length(mylat(:,1)),length(mylon(1,:)),[]); %

res = 0.25;
tres = 0.5;

a = minmax(mylon(1,:));
b = minmax(mylat(:,1));
c = minmax(myage);
[Xq,Yq,Zq] = meshgrid(a(1):res:a(2),b(1):res:b(2),c(1):tres:c(2)); %

fprintf(1,'Interpolando o novo grid...\n')

```

```

mybat_new = interp3(X,Y,Z,mybat,Xq,Yq,Zq, 'cubic');

fprintf(1, 'Preparando o grid para cálculo das distâncias...\n')
a = minmax(mylon(1,:));
b = minmax(mylat(:,1)');
[Xd,Yd] = meshgrid(a(1):res:a(2),b(1):res:b(2));

radius = 6371;

mydist = nan(size(mybat_new));

for i=1:size(mybat_new,3)
    fprintf(1, 'calculando para time index %d [de %d]...\n',i,size(mybat_new,3))
    nowbat = mybat_new(:,:,i);
    [ii] = find(nowbat>=-20&nowbat<=20);
    Pbat = [Xd(ii),Yd(ii)];
    Qbat = [Xd(:),Yd(:)];
    [k,dist]=dsearchn(Pbat,Qbat);
    mydist(:,:,i) = reshape(dist,size(Xd,1),[])*radius;
end

load ('IODPnew2.mat')

comp = 1:length(myIODP8test2)

for i=1:length(comp)
    fprintf(1, 'IODP core %d [de %d] : %s calculando...\n',i,length(myIODP5),myIODP5(comp(i)).core)
    xage = myIODP8test2(comp(i)).AgeDepthToc(:,1);
    idx = find(~isnan(xage));

    myIODP5age = xage(idx);

    myIODP5lat =
    interp1([myIODP8test2(i).plate_time],[myIODP8test2(i).plate_lat],myIODP5age);
    myIODP5lon =
    interp1([myIODP8test2(i).plate_time],[myIODP8test2(i).plate_lon],myIODP5age);

    myPaleobat =
    interp3(Xq,Yq,Zq,mybat_new,myIODP5lon,myIODP5lat,myIODP5age);
    myPaleodist =
    interp3(Xq,Yq,Zq,mydist,myIODP5lon,myIODP5lat,myIODP5age);

    myIODP8test2(i).plate_time = myIODP5age;
    myIODP8test2(i).plate_lat = myIODP5lat;
    myIODP8test2(i).paleobat = myPaleobat;
    myIODP8test2(i).distcost = myPaleodist;
end

```

Anexo D – Script para cálculo das taxas de sedimentação

```
load IODPnew2.mat

srates = [];
age = [];
time=[];

for i=11

    for j=1:1:(length(myIODP8test2(i).plate_time(:,1))-1)

        x = (myIODP8test2(i).AgeDepthToc(j+1,2) -
myIODP8test2(i).AgeDepthToc(j,2))./(myIODP8test2(i).AgeDepthToc(j+1,1)
-myIODP8test2(i).AgeDepthToc(j,1));
        id = myIODP8test2(i).AgeDepthToc(j+1,1);

        srates = cat(1,srates,x);
        age = cat(1,age,id);

    end

    srates = [age srates];

    myIODP8test2(i).tsed = srates(:,2);

    h = find(isnan(myIODP8test2(i).plate_lat));
    myIODP8test2(i).tsed(h,:)= NaN;

srates = [];
age = [];
time =[];

end
```

Anexo E – Script para treinamento do modelo *Rational Quadratic GPR* (10-folds)

```
inputTable = array2table(trainingData, 'VariableNames', {'column_1',
'column_2', 'column_3', 'column_4', 'column_5', 'column_6'});

predictorNames = {'column_1', 'column_2', 'column_3', 'column_4'};
predictors = inputTable(:, predictorNames);
response = inputTable.column_5;
isCategoricalPredictor = [false, false, false, false];

regressionGP = fitrgp(...
    predictors, ...
    response, ...
    'BasisFunction', 'constant', ...
    'KernelFunction', 'rationalquadratic', ...
    'Standardize', true);
```

```

predictorExtractionFcn = @(x) array2table(x, 'VariableNames',
predictorNames);
gpPredictFcn = @(x) predict(regressionGP, x);
trainedModel.predictFcn = @(x)
gpPredictFcn(predictorExtractionFcn(x));

inputTable = array2table(trainingData, 'VariableNames', {'column_1',
'column_2', 'column_3', 'column_4', 'column_5', 'column_6'});

predictorNames = {'column_1', 'column_2', 'column_3', 'column_4'};
predictors = inputTable(:, predictorNames);
response = inputTable.column_5;
isCategoricalPredictor = [false, false, false, false];

partitionedModel = crossval(trainedModel.RegressionGP, 'KFold', 10);

validationPredictions = kfoldPredict(partitionedModel);

validationRMSE = sqrt(kfoldLoss(partitionedModel, 'LossFun', 'mse'));

```

Anexo F – Linha de código para aplicação do modelo em *dataset* de teste

```
Yfit_teste = trainedModel.predictFcn = (data_teste(:,1:4));
```

Os scripts utilizados para geração das figuras não estão descritos, mas podem ser disponibilizados mediante contato com o autor através do e-mail **berglz@gmail.com**. A *struct* com todos os dados utilizados na realização deste estudo, bem como o arquivo contendo o modelo podem ser disponibilizados da mesma forma.