A técnica de KDE, matematicamente equivalente à Krigagem de Regressão (*Regression* *Kriging*), caracteriza-se quando comparada a outros algoritmos geoestatísticos primariamente pela utilização de uma função de regressão linear aplicada a variável modelada, e o uso dos erros da regressão para estimativa dos pesos da krigagem (Hengl et al., 2007).

Ao mesmo tempo, simulações gaussianas permitem melhor desempenho na estimativa de propriedades na porção não-estruturada do variogramas (Hansen et al., 2006), evitando feições suavizadas características de algoritmos baseados primariamente em krigagem (Schlumberger, 2015).

O estudo caracteriza-se por uma etapa inicial de processamento de dados de sísmica de reflexão para o realce e identificação automática de feições estruturais no reservatório. Servindo como entrada em uma etapa seguinte de modelamento estrutural onde dados de horizontes são deformados por estas estruturas. Com um modelo estrutural definido, segue uma etapa de população do modelo com valores oriundos de medidas simples e algoritmos de geoestatística. Obtendo a distribuição espacial das propriedades de poço modeladas.

Para a etapa de classificação de reservatório e cálculos de volume, são utilizadas técnicas computacionais de aprendizado de máquina não supervisionado como Mapas Auto Organizáveis (Self Organizing Maps, SOM), K-Means e Modelos de Mistura Gaussiana (Gaussian Mixture Models, GMM) para a separação de áreas do reservatório de acordo com as propriedades modeladas anteriormente.

Técnicas de aprendizado de máquina não supervisionado mostram-se como uma alternativa eficiente para a classificação de reservatório e tem sido frequentemente utilizadas para lidar com problemas onde não existe classificação prévia (Taner et al, 2001). Kuznetsova et al. (2016) chega a apresentar resultados comparáveis à métodos de aprendizado de máquina supervisionado como *random* *forest* e *support* *vector* *machine* utilizando métodos não supervisionados para classificação de fácies em logs de poço.

Os métodos de aprendizado de máquina não supervisionados também são estabelecidos como efetivos para estimativas de fácies, em especial eletrofácies, derivadas de dados petrofísicos (Al-Mudhafar et al., 2015; Aragão & Sava, 2019).

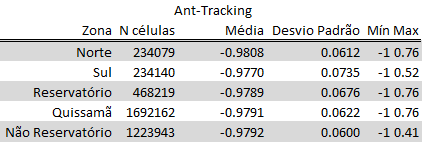
Kuroda et al. (2012) realiza um estudo aplicando Mapas Auto-Organizáveis e K-Means no espaço obtido do mapa para classificação de fácies em logs de poço na Bacia de Campos. O fluxo de trabalho desenvolvido neste estudo baseia-se no estudo de Kuroda et al. (2012), aplicando o método desenvolvido aos modelos tridimensionais obtidos por algoritmos geoestatísticos em vez de logs de poço de maneira direta.

O algoritmo de Mistura Gaussiana, ao apresentar saídas probabilísticas para cada classe, permite quantificar as probabilidades de cada ponto pertencer a cada classe (Viroli & McLachlan, 2019). Como o cálculo de volumes, objetivo final deste estudo, passa pela estimativa da razão Net-to-Gross, também medida entre 0 e 1 (Egbele et al., 2015), modelos de Mistura Gaussiana são utilizados para estimar as razões nos cálculos de volume.

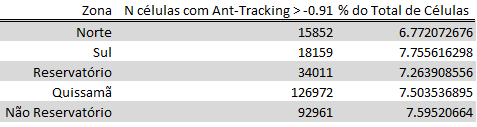
A base de dados deste estudo inclui horizontes no domínio da profundidade para o topo e base do reservatório além de outras unidades geológicas estratigraficamente próximas, volume 3D de sísmica de reflexão no domínio da profundidade recortado para a área do reservatório e logs de poço com diferentes disponibilidades de logs para 47 poços diferentes interceptando ou próximos ao reservatório, dados todos fornecidos pela Agência Nacional de Petróleo – ANP.

o maior número de células com elevados valores de Ant-Tracking na Zona Sul (Figura 21; Figura 22) bem como a maior porcentagem de suas células se apresentarem com valores elevados de Ant-Tracking (Tabela 6; Tabela 7) fornece certa evidência para a divisão observada por Tomaso et al. (2013).

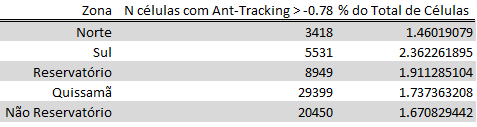
**Tabela 5 - Distribuição dos valores de Ant-Tracking nas diferentes Zonas do volume sísmico. Embora a Zona Sul mostre um valor média de Ant-Tracking superior tanto à Zona Norte quanto a região não reservatório do Quissamã, a diferença observada na média é menor do que o desvio padrão destas medidas.**

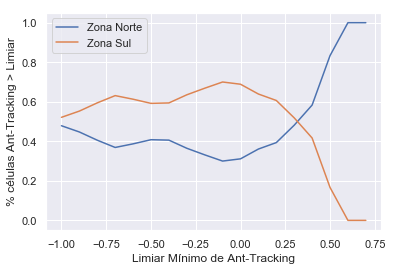


**Tabela 6 – Número de células por zona com valores de Ant-Tracking superiores a -0.91. O valor de -0.91 corresponde ao valor médio de Ant-Tracking na região sobre o contato óleo-água dentro da Formação Quissamã adicionado ao desvio padrão do valor de Ant-Tracking nesta mesma região.**

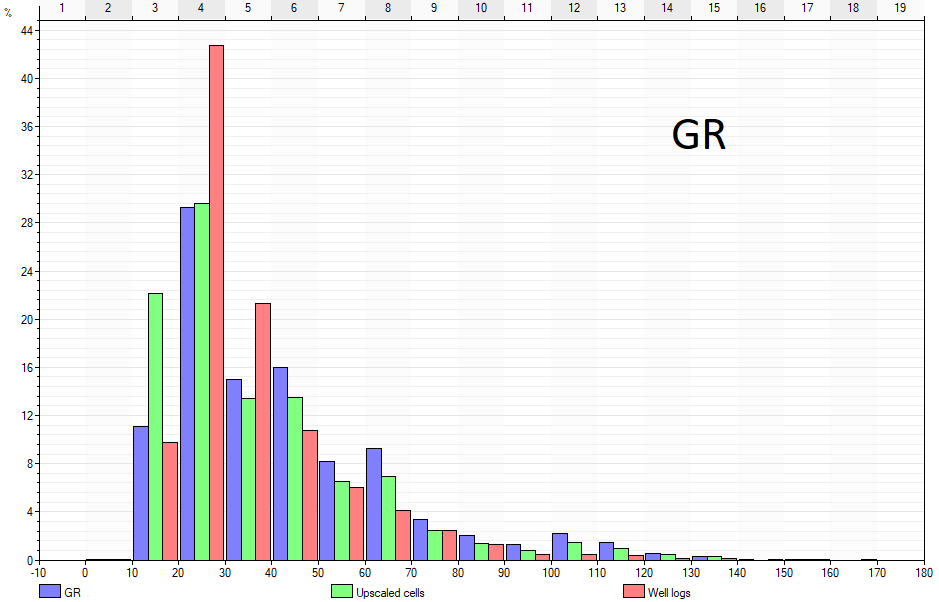


**Tabela 7 - Número de células por zona com valores de Ant-Tracking superiores a -0.78. O valor de -0.78 corresponde ao valor médio de Ant-Tracking na região sobre o contato óleo-água dentro da Formação Quissamã adicionado à três vezes o desvio padrão do valor de Ant-Tracking nesta mesma região.**

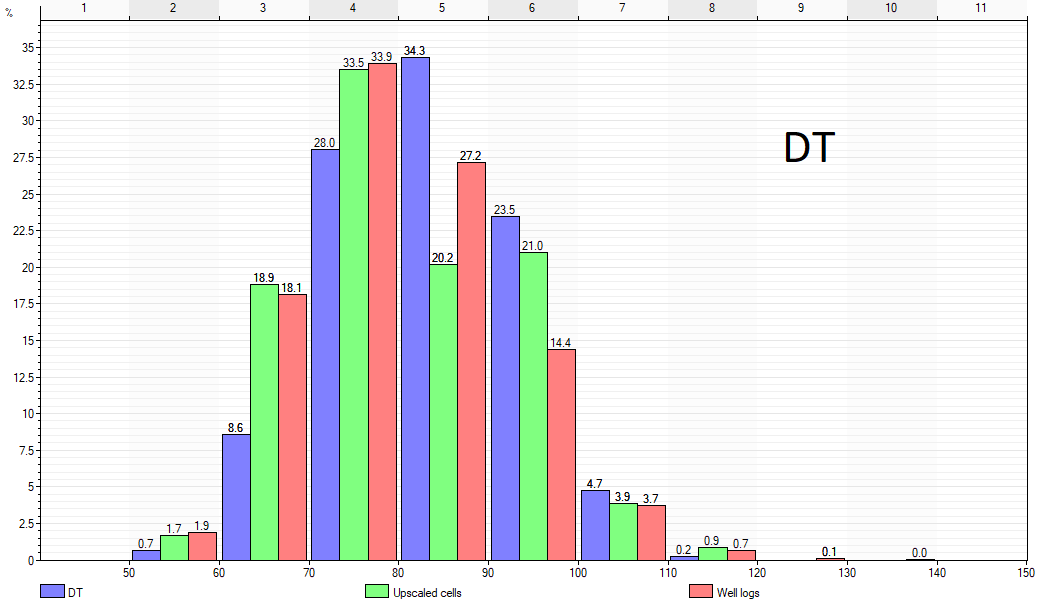




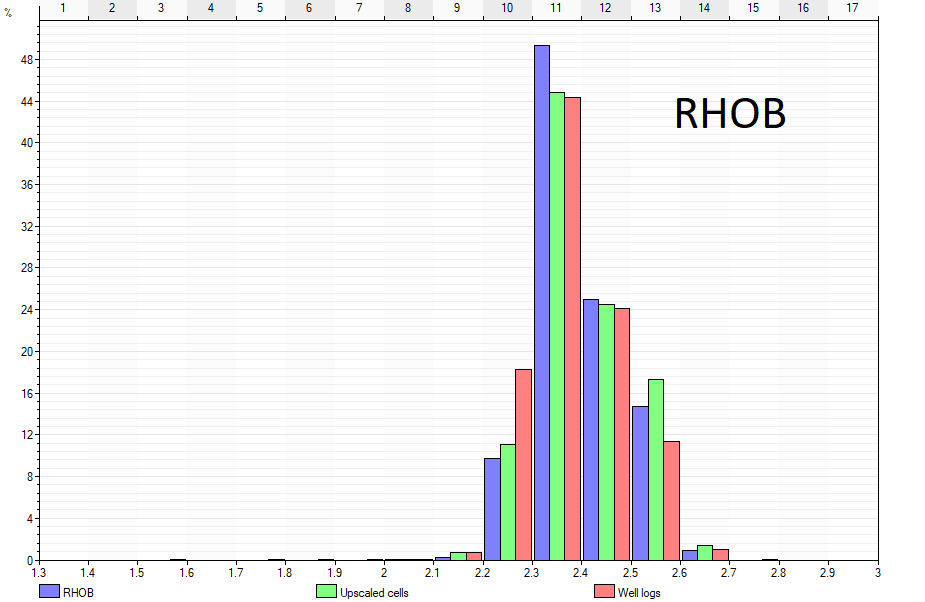
**Figura 22 - Porcentagem de células localizadas na Zona Norte ou Sul para cada valor mínimo de Ant-Tracking (Limiar). Nota-se que para Ant-Tracking de -1,0 os valores se distribuem igualmente entre as duas zonas, refletindo a separação das duas zonas pela Latitude mediana do reservatório. Para valores entre -0.8 e 0, valores que se correlacionam com elevadas intensidades de fraturamento, 60% a 75% dos valores encontram-se na Zona Sul. Para os poucos (n = 151) valores positivos de Ant-Tracking, estes concentram-se fortemente na Zona Norte. Assim indicando que as descontinuidades ocorrem de maneira mais disseminada pela Zona Sul enquanto concentram-se em uma área de alta descontinuidade na Zona Norte.**



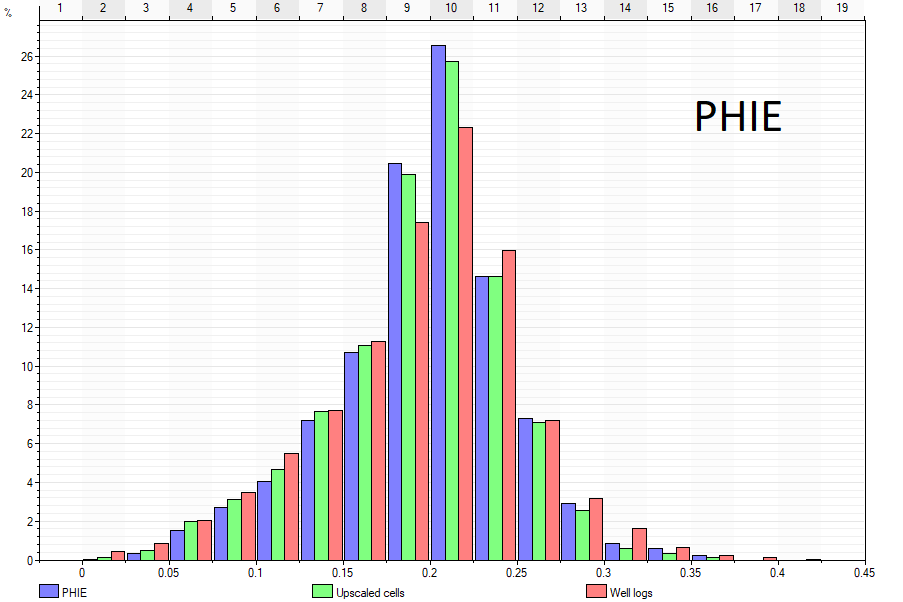
**Figura 23 - Distribuição de dados de**



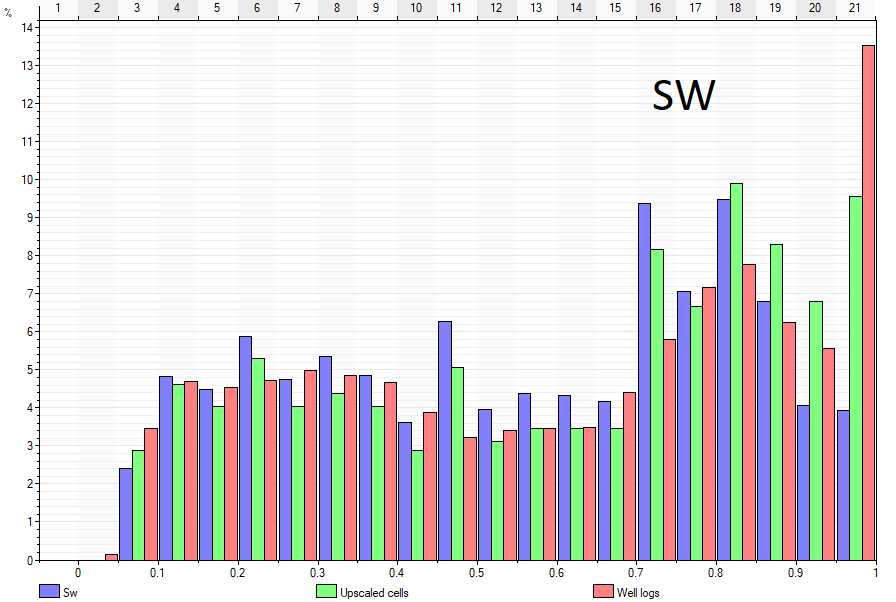
**Figura 24 - Distribuição de dados de**



**Figura 25 – Distribuição de dados de**



**Figura 26 – Distribuição de dados de**



**Figura 27 – Distribuição de dados de**

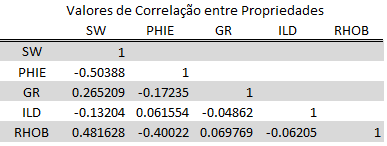
O modelo de saturação de água (Figura 28) tem medidas distribuídas de forma similar ao longo dos valores modelados, exceto pelo pico para valores de aproximadamente 60% observados na distribuição dos dados (Figura 23). O Modelo de saturação de água também apresenta uma transição suave de valores devido à utilização do algoritmo de KDE no modelamento. Também se observa valores de saturação de água menores na zona sul (média de 53%) quando comparados aos valores na Zona Norte (média de 62%). Os altos valores para a porção não reservatório do Quissamã (média de 73%) são esperados e fornecem evidência de uma correta determinação do contato óleo-água. Os valores mais altos observados mais próximos do contato óleo-água e do topo da Formação Quissamã também mostram a delimitação das fronteiras do reservatório.

Os modelos de porosidade efetiva, raio gamma, resistividade e densidade ocorrem com valores variando de forma consideravelmente mais abrupta (Figura 29; Figura 30; Figura 31; Figura 32). Este efeito se dá devido ao GRFS honrar variações locais em médias e variâncias dos dados, ao contrário das medidas estáticas do modelo de krigagem.

Nota-se no modelo de porosidade que não há grande variação nos valores médios de porosidade entre as diferentes zonas, com valores de aproximadamente 20% de porosidade para toda a área da Formação Quissamã. O modelo de porosidade mostra uma distribuição de valores mais altos (25% a 30%) em camadas horizontais em meio a valores de baixa porosidade (10% a 15%). Embora não seja possível inferir sobre a distribuição litológica das fácies da Formação Quissamã com base nestes dados, a ocorrência de lentes mais grossas em meio a material menor poroso explicaria a distribuição observada.

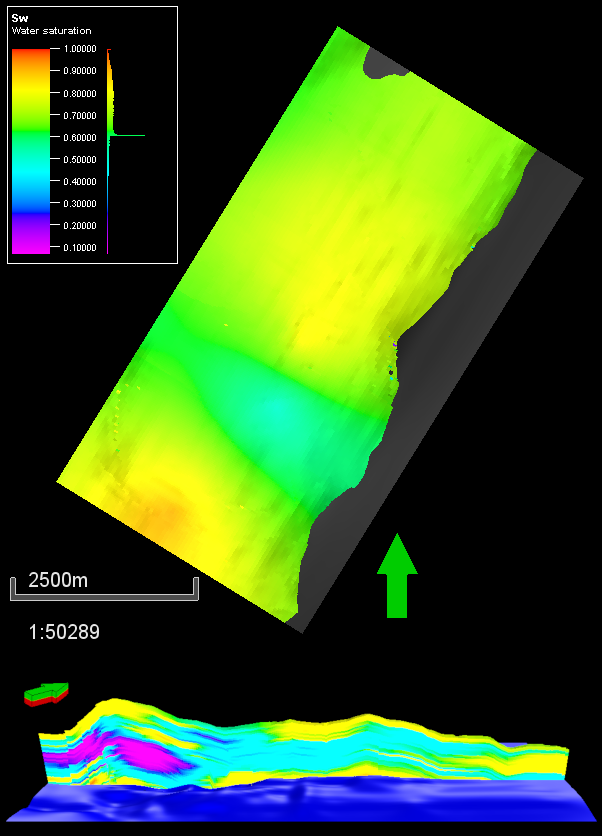
Nota-se também que nos locais de ocorrência de baixos valores de saturação de água também tendem a ser observados altos valores de porosidade. Tal comportamento é esperado visto que a porosidade é uma variável na equação de Archie utilizada para estimar as saturações, no entanto, a resistividade também é uma variável da equação de Archie, e enquanto é observada correlação de apenas -13% entre saturação e resistividade, é observada uma correlação de -50% entre saturação e porosidade (Tabela 7). Indicando que regiões mais porosas do reservatório não só possuem maior espaço poroso para acumulação de fluidos, mas que estes fluidos são compostos em menor parte por água.

**Tabela 7 – Valores de correlação entre as propriedades modeladas. Tabela tem valores repetidos omitidos.**

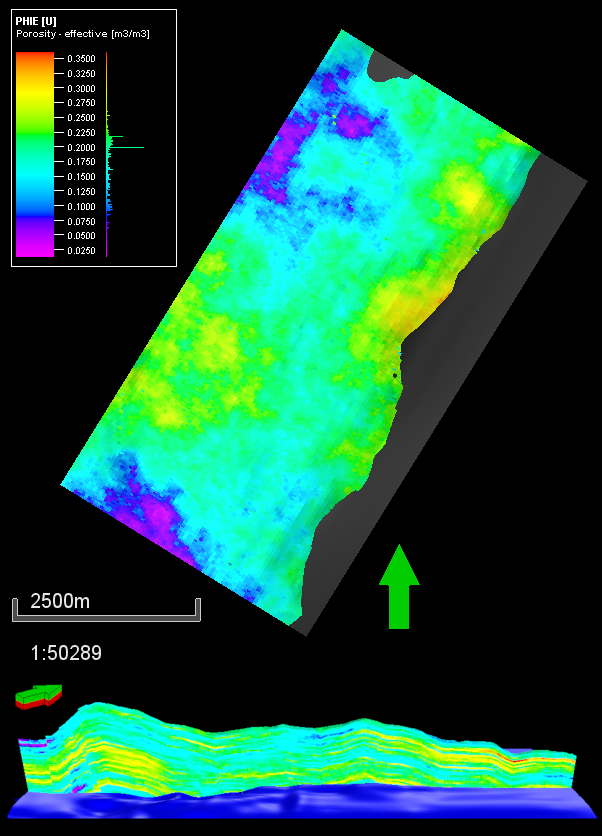


Adicionalmente, a observação do modelo de raio gama revela uma dependência significativa da profundidade com os valores modelados, com um aumento súbito de valores próximos ao topo do Quissamã indicando o início de uma possível transição de litologias.

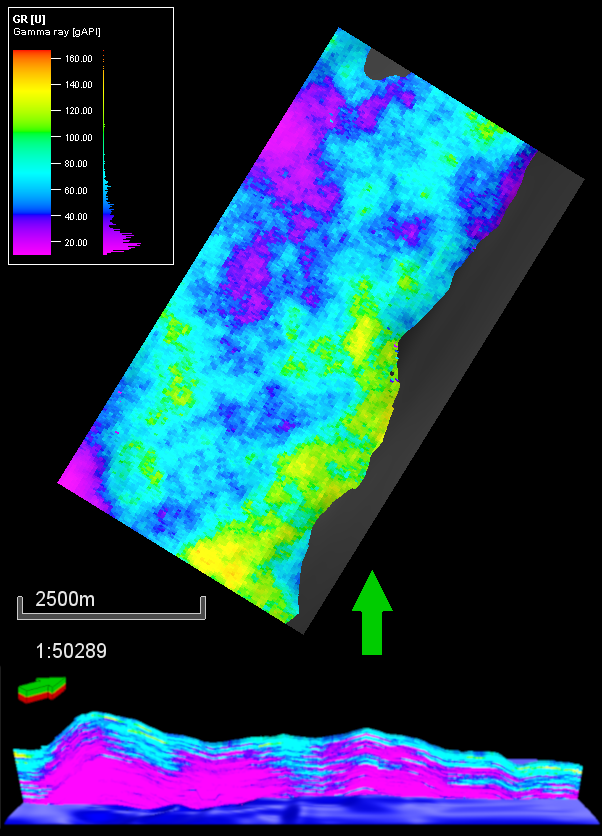
Quanto aos modelos de resistividade e densidade, o modelo de resistividade apresenta valores muito elevados em seu CV devido ao predomínio de baixas resistividades por toda a área do modelo exceto pelos altíssimos picos associados aos locais de baixa saturação de água. Já o modelo de densidade mostra valores aproximadamente constantes em torno dos valores médios.



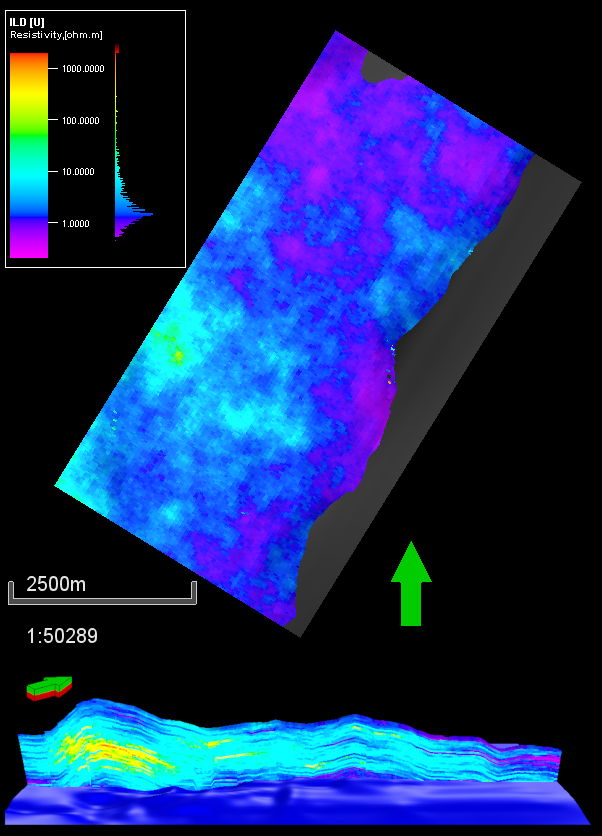
**Figura 28 – Modelo de saturação de água (SW). Valores distribuem-se de maneira similar para diferentes níveis de saturação de água. Algoritmo de KDE resulta em transições de valores suavizadas. Nota-se os maiores valores de saturação próximos ao topo e base do reservatório, evidenciando as regiões de fronteira do reservatório, assim como a presença de uma região na porção sudoeste do reservatório com valores muito baixos de saturação.**



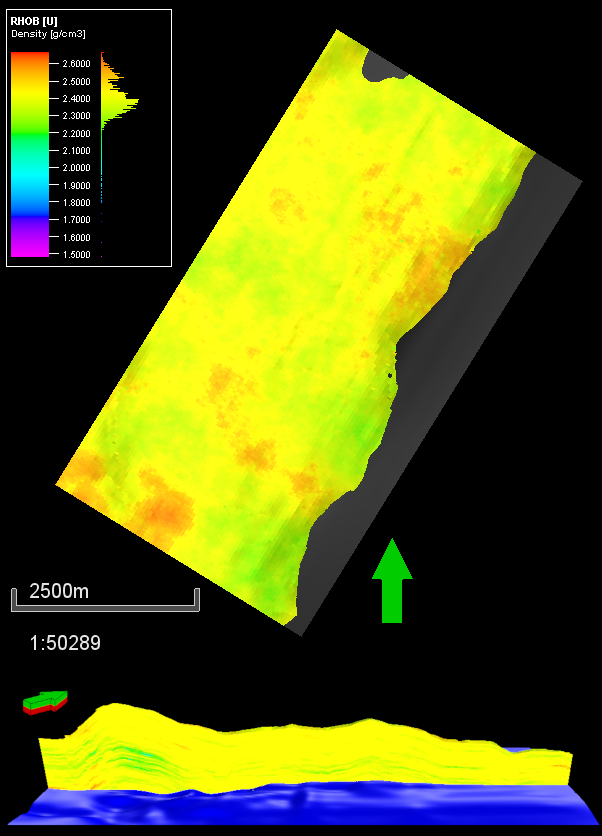
**Figura 29 - Modelo de porosidade efetiva (PHIE). Valores distribuídos em camadas horizontais contendo maiores valores. Observa-se associação entre maiores valores de porosidade nos mesmos locais onde ocorrem menores valores de saturação de água (SW).**



**Figura 30 - Modelo de raio gama (GR). Observa-se dependência entre os valores modelados e profundidade, com valores próximos ao topo da Formação Quissamã apresentando um aumento nos valores de raio gama observados.**

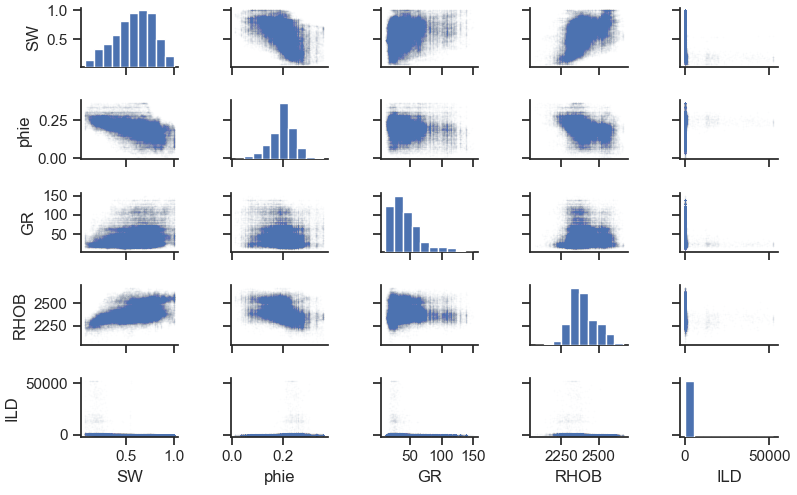


**Figura 31 – Modelo de resistividade (ILD). Observa-se predominância de valores baixos. A ocorrência de picos com valores bastante elevados na região sudoeste do modelo explica os altos valores de CV observados na Tabela 6.**



**Figura 32 – Modelo de Densidade (RHOB). Valores aproximadamente constantes em torno da média de aproximadamente 2400g/cm³ explicam baixos valores de CV observados na Tabela 6. Ao mesmo tempo, raras variações são observadas na região sudoeste, com quedas nos valores de Densidade associadas a valores de alta porosidade, alta resistividade e baixa saturação de água.**

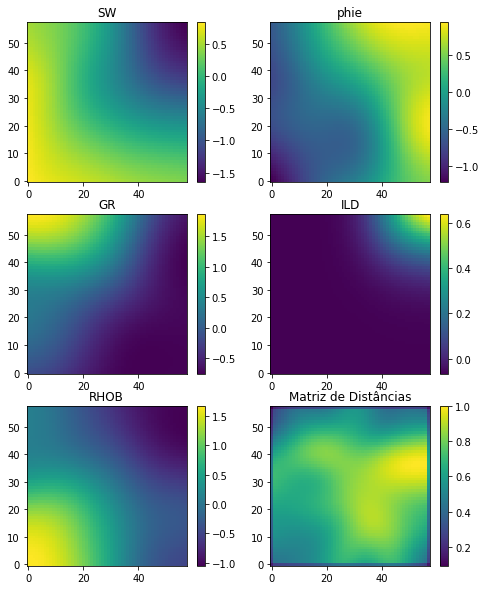
Refletindo os valores de correlação observados entre as propriedades modeladas de saturação de água e porosidade efetiva, a distribuição dos dados num plano mostra uma tendência entre os dois valores (Figura 33). O mesmo pode ser observado para os demais valores de correlação. Sendo possível notar que os valores de raio gama distribuem-se de maneira aproximadamente independente dos valores de porosidade e saturação, o que reflete os valores de correção relativamente baixos para os pares.



**Figura 33 – Distribuições par a par das propriedades modeladas. Histogramas de distribuição das propriedades em casos em que a propriedade é comparada a ela mesma. Valores intermediários de correlação entre SW-PHIE e SW-RHOB observados na Tabela 7 podem ser visualizados. Nota-se ainda a distribuição concentrada em baixos valores observada para a propriedade de resistividade (ILD).**

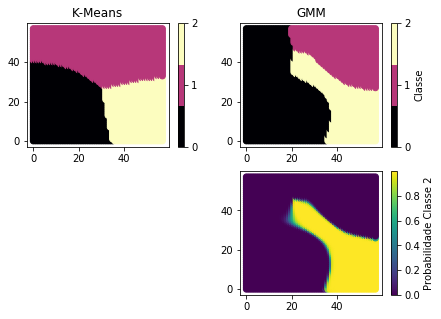
Os resultados do algoritmo de mapas auto organizáveis permitem reduzir a dimensionalidade do problema de 3 (PHIE, GR, RHOB) para 2 dimensões, e, principalmente, reduz o espaço amostral de mais de 400000 células para um mapa 2D composto por 3364 neurônios. Observa-se a resposta dos neurônios do mapa gerado para cada variável de entrada e também a matriz correspondente às distâncias entre cada uma das respostas (Figura 34).

Nota-se que os neurônios com pesos mais altos para medidas de porosidade também concentram respostas menos intensas para medidas de saturação de água e densidade. Comportamento observado anteriormente nas propriedades modeladas. Ao mesmo tempo, devido a distribuição concentrada dos picos de alta resistividade, observa-se que os neurônios com alta resposta de resistividade ocorrem também de maneira concentrada, mostrando relação inversa com as respostas de saturação de água, comportamento esperado em uma rocha reservatório. Ao mesmo tempo, a observação da matriz de distâncias não revela um agrupamento óbvio dos dados, ainda que os valores mais altos centrais cercados por diferentes zonas periféricas de baixos valores sugiram que existe um agrupamento das respostas da matriz de distâncias.



**Figura 34 – Matriz de Distâncias e respostas à cada variável do algoritmo de mapas auto organizáveis (SOM). Observa-se a relação entre altas respostas de porosidade e resistividade com baixas respostas de saturação de água e densidade.**

Observa-se a maior correspondência do algoritmo de GMM com a distribuição observada no mapa de distâncias, com a classe 0 correspondendo aos neurônios de maiores distâncias. Nota-se que o número da classe não é o mesmo entre os 2 algoritmos e depende da inicialização aleatória do algoritmo, não tendo qualquer significado além de separar os dados.



**Figura 36 – Classificações da matriz de distâncias. Valores entre 0 e 1 para o mapa de probabilidades de GMM são a probabilidade de cada neurônio pertencer a classe 2.**

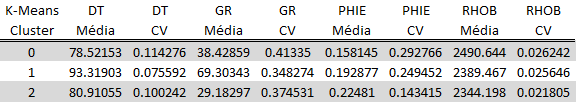
Atribuindo a cada célula do grid do modelo as classes e probabilidades geradas pelos algoritmos de K-Means e GMM correspondentes ao neurônio que atua como resposta desta célula na camada de saída do mapa auto organizável, é possível mapear a distribuição entre classes entre os valores das propriedades. Chama-se aqui a classe de menor saturação de água e/ou maior porosidade como classe alvo, enquanto a classe contrária com maiores saturações e menor porosidade é chamada aqui como classe de descarte. Chama-se a classe entre os dois valores de classe intermediária.

Observa-se que o algoritmo de GMM conseguiu individualizar melhor as células pertencentes a classe alvo (Classe 1), mostrando maiores valores de porosidade (24%) e menores de saturação (35%) simultaneamente (Tabela 8), enquanto o K-Means obteve duas classes com porosidades semelhantes em torno de 21% (Classes 1 e 2) e uma classe (Classe 1) com a menor saturação de água em 46% (Tabela 9), definida como a classe alvo. Ao mesmo tempo, para os dois algoritmos observa-se a atribuição dos valores de alta resistividade à classe alvo, justificando os altos valores de resistividade média e coeficiente de variação observados na ocorrência dos picos de resistividade em células de muito baixa saturação de água (Figura 33; Figura 35).

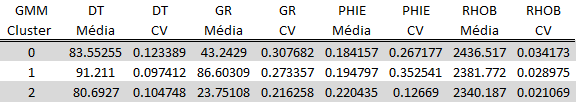
Ao mesmo tempo, nota-se que o GMM classificou como alvo apenas 123,262 células quando comparado às 191,565 células alvo do K-Means. Considerando os valores similares de porosidade e saturação de água observados para as classes intermediária e descarte nos dois modelos, é provável que o K-Means tenha associado altos valores de raio gama à classe alvo. Classificando um grande número de células com alta resposta de raio gama e altas saturações de água como alvo. Tal efeito pode ser observado na classificação ainda no mapa auto organizável, tal que os neurônios com altas respostas para raio gama ocorrem na área superior esquerda e os com baixa resposta de saturação de água na área superior direita (Figura 35). Enquanto o K-Means coloca os neurônios de alta resposta em raio gama na mesma classe que os neurônios com baixa resposta em saturação, o GMM classifica os neurônios de alta resposta em raio gama de forma separada das baixas respostas em saturação (Figura 36).

Os dados obtidos pelo algoritmo de

**Tabela 7 - Distribuição estatística das propriedades modeladas entre as classes obtidas com o algoritmo de K-Means.**



**Tabela 8 – Distribuição estatística das propriedades modeladas entre as classes obtidas com o algoritmo de Modelo de Mistura Gaussiana, GMM.**

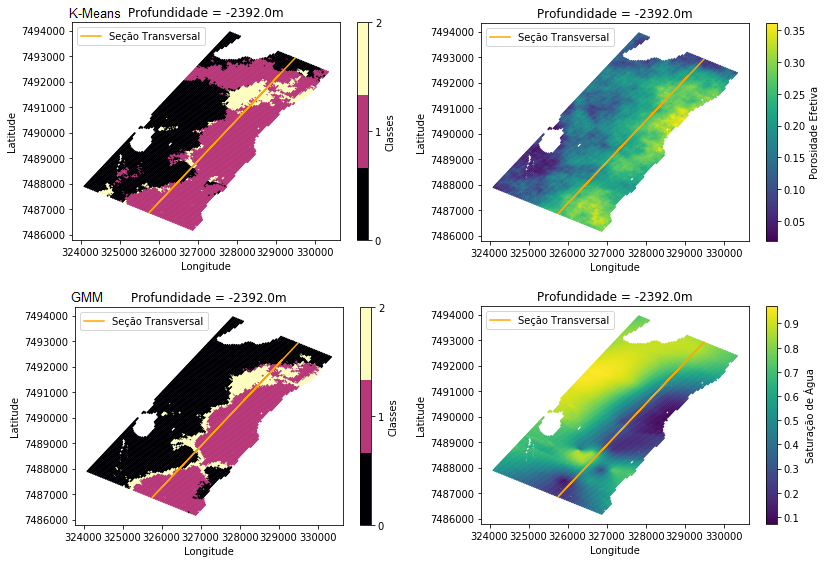


Observando a distribuição espacial da classificação para os dois algoritmos é possível observar que o K-Means classifica algumas regiões de porosidade intermediária e alta saturação de água como alvo (Figura 37). Este comportamento não é observado na classificação obtida do algoritmo de GMM. A classificação do GMM também resulta em uma transição melhor entre as classes alvo, intermediária e descarte, com células classificadas na classe intermediária geralmente circundando células classificadas como alvo. A distribuição observada no GMM é mais condizente com a provável distribuição de diferentes litologias no reservatório, ainda que a classificação leve em conta valores relativos ao conteúdo de fluidos do espaço poroso como resistividade e saturação de água.

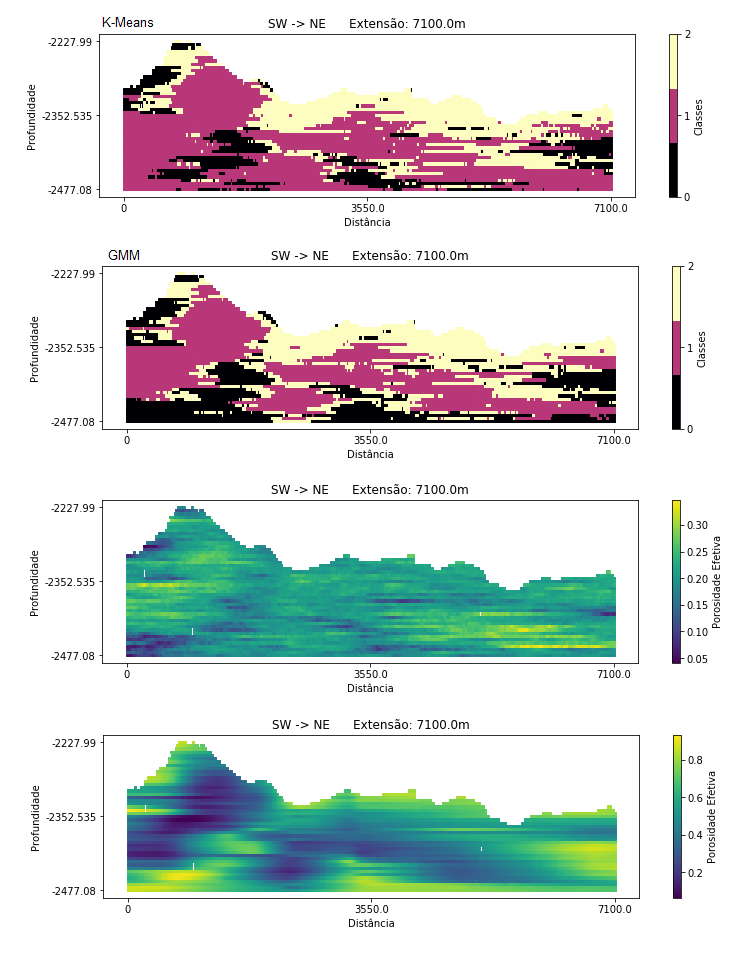
Seções transversais dos mapas da Figura 37 apresentam o mesmo comportamento, com K-Means classificando zonas de alta saturação de água como alvo (Figura 38). Células classificadas como alvo também podem ser observadas circundadas por células classificadas como intermediárias.

De forma adicional, os resultados da classificação obtida do algoritmo de GMM também tem respostas não discretas para a probabilidade de cada célula pertencer a cada classe obtidas da aplicação a cada célula das probabilidades correspondentes ao seu neurônio de resposta nos mapas auto organizáveis.

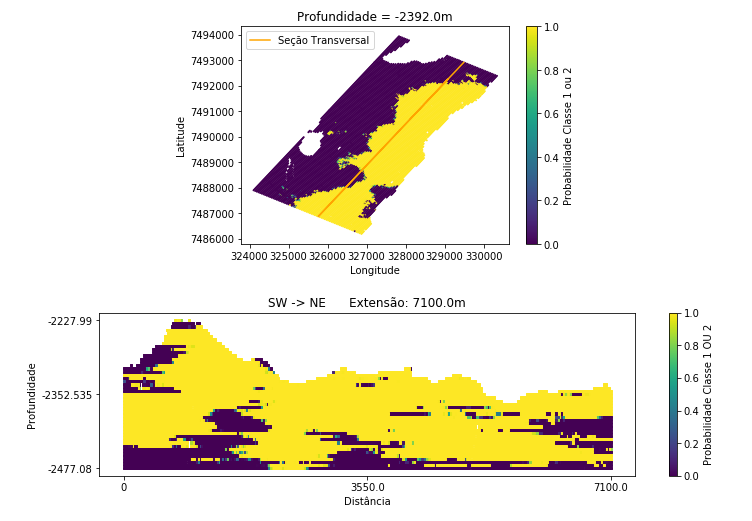
Assim, a observação da distribuição espacial das probabilidades de cada célula não pertencer a classe de descarte (ou, alternativamente, de cada célula pertencer a classe alvo ou intermediária) revela que o algoritmo de GMM classificou os neurônios da matriz de distâncias com um alto grau de certeza, visto que as probabilidades distribuem-se de maneira aproximadamente booleana em 1 ou 0 (Figura 39).



**Figura 37 – Mapas de distribuição das classificações para K-Means e GMM e propriedades de porosidade efetive e saturação de água para profundidade de 2392m. Linha laranja indica as seções transversais da Figura 39.**



**Figura 38 - Seções transversais para as distribuições das classificações obtidas de K-Means e GMM. Seções correspondem as linhas indicadas nos mapas da Figura 38.**



**Figura 39 – Mapa e seção transversal de distribuição de probabilidades de cada célula pertencer a Classe 0 ou 1 na classificação obtida do método GMM.**