

UNIVERSIDADE FEDERAL FLUMINENSE
INSTITUTO DE GEOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO DE DINÂMICA DOS OCEANOS E DA
TERRA
DEPARTAMENTO DE GEOLOGIA E GEOFÍSICA



Rodrigo Andrade Santos

Impacto das Estruturas de Correlação na Avaliação Geológica de
Prospectos na Indústria de Óleo e Gás.

Niterói-RJ
2019

Intencionalmente deixada em branco.

RODRIGO ANDRADE SANTOS

Impacto das Estruturas de Correlação na Avaliação Geológica de
Prospectos na Indústria de Óleo e Gás.

Dissertação apresentada, como requisito parcial para a obtenção do título de Mestre, ao Programa de Pós-Graduação de Dinâmica dos Oceanos e da Terra da Universidade Federal Fluminense. Áreas de Concentração: Geologia, Geofísica, Risco e Incerteza.

Orientador: Prof. Dr. Marco Antônio Cetale Santos

Co-Orientador: Prof. Dr. Luiz Alberto Santos

Niterói-RJ
2019

Ficha catalográfica automática - SDC/BIG
Gerada com informações fornecidas pelo autor

S237i Santos, Rodrigo Andrade
Impacto das Estruturas de Correlação na Avaliação
Geológica de Prospectos da Indústria de Óleo e Gás / Rodrigo
Andrade Santos ; Marco Antônio Cetale Santos, orientador ;
Luiz Alberto Santos, coorientador. Niterói, 2019.
189 f. : il.

Dissertação (mestrado) - Universidade Federal Fluminense,
Niterói, 2019.

DOI: <http://dx.doi.org/10.22409/PPGDOT.2019.m.02557267597>

1. Prospecto. 2. Correlação. 3. Simulação de Monte-
Carlo. 4. Cópulas. 5. Produção intelectual. I. Cetale
Santos, Marco Antônio, orientador. II. Santos, Luiz Alberto,
coorientador. III. Universidade Federal Fluminense. Instituto
de Geociências. IV. Título.

CDD -

RODRIGO ANDRADE SANTOS

IMPACTO DAS ESTRUTURAS DE CORRELAÇÃO NA AVALIAÇÃO GEOLÓGICA DE PROSPECTOS NA INDÚSTRIA DE ÓLEO E GÁS.


Dissertação apresentada, como requisito parcial para a obtenção do título de Mestre, ao Programa de Pós-Graduação em Ciências dos Oceanos e da Terra da Universidade Federal Fluminense. Área de Concentração: Geologia, Geofísica, Risco e Incerteza.

Aprovada em Outubro de 2019. Assinaturas colhidas separadamente devido à pandemia de COVID-19.

BANCA EXAMINADORA



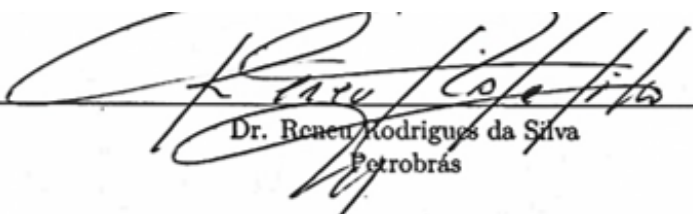
Prof. Dr. Marco Antônio Cetale Santos - Orientador
UFF/DOT/GISIS



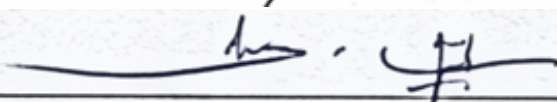
Prof. Dr. Luiz Alberto Santos - Co-Orientador
UFF/DOT/GISIS & Petrobrás



Prof. Dr. Rodrigo Bijani Santos
UFF/DOT



Dr. Renan Rodrigues da Silva
Petrobrás



Dr. Salvador José Chrispim
Petrobrás

Niterói-RJ
2019

Intencionalmente deixada em branco.

Dedicatória

Dedico este trabalho à minha família e amigos.

Agradecimentos

Agradeço a todos que me deram apoio não só na confecção e revisão desta dissertação, mas que me apoiaram ao longo do caminho que me trouxe até aqui.

Gostaria de agradecer a minha família, minha esposa e filha, a pequena princesa Luiza. Agradeço também a meus familiares (pai, mãe, e irmã) por terem me ajudado. E uma dedicação especial a minha filha

Agradeço aos meus orientadores Luiz Alberto Santos e Marco Antônio Cetale Santos, pela ajuda e paciência. Gostaria de agradecer à Banca examinadora pelas valiosas correções e pelo tempo dispendido na avaliação do trabalho.

Gostaria de agradecer a ajuda e os valiosos conselhos e ensinamentos do Reneu Rodrigues da Silva, Salvador Chrispim, Antônio Affonso, Luciano Arantes R. Costa, Carlos Lopo Varela e Laury Medeiros de Araújo.

Um agradecimento especial a Pedro O. V. Piaulino e Wagner Alixandrini, que além de me ajudar na dissertação, muitas vezes me apoiaram no trabalho e foram muito mais do que colegas, foram verdadeiros amigos.

Agradeço também aos meus colegas (e mentores) de sala, que tanto me ajudam pacientemente: Paulo Roberto de Sá e Benevides, Carlos Rogério Oliveira Rodrigues (Turbidito), Jair Rodrigues de Salles Soares Filho, Fátima Regina Blanco de Dios, Luiz Carlos Empinotti e Fausto Machado Coelho. Agradeço a eles por todo o conhecimento, carinho e apoio que me deram durante os últimos anos de convivência.

Destaco que o fim deste trabalho somente ocorreu graças à compreensão do meu atual gerente, Carlos R. Becker, que permitiu que eu me dedicasse quase que integralmente à conclusão da dissertação.

Não poderia deixar de agradecer à Petrobras, como instituição, que me acompanhou, até este momento por toda a minha vida. Desde o meu nascimento através do meu pai (petrolheiro), até os dias de hoje onde atuo como empregado, com muito orgulho. Por fim, mas não menos importante gostaria de agradecer aos profissionais desta grande empresa que propiciaram a minha pós-graduação na UFF: André Luiz Romanelli Rosa e Aloisio Pellon.

Lista de Figuras

1.1	Preços do barril de Petróleo desde 1986 até os dias atuais, intervalo de amostra: 1 mês. Fonte: FRED (2019)	3
2.1	Um leque de velocidades ao redor de uma função central (esquerda), <i>gathers</i> de velocidade variável que correspondem a esse leque (centro) e <i>gathers</i> corrigidos de NMO com velocidades exibidas (direita).	7
2.2	Exemplos de distribuições utilizadas em estudos de probabilidade e estatística. Distribuição uniforme (esquerda e acima), distribuição normal (direita acima), distribuição lognormal (abaixo e esquerda), distribuição triangular (abaixo e direita). Todos os gráficos acima foram gerados com a biblioteca <code>numpy.random</code>	8
2.3	Lançamento de quatro dados e distribuição resultante de valores possíveis baseados na soma independente de resultados, note que a distribuição se torna mais semelhante a uma forma de sino. (Modificado de Rose (2001))	10
2.4	Exemplos de distribuições lognormais verificadas em ocorrências de hidrocarbonetos na bacia de Perth (Oeste da Austrália). A esquerda Coluna das acumulações, e a direita a distribuição de campos ao longo da bacia (Ferdinando and Longley, 2016).	11
2.5	Lançamento de dois dados e distribuição resultante de valores possíveis baseados na multiplicação independente de resultados (esquerda) a distribuição resultante é bastante assimétrica com mais valores menores do que maiores. Lançamento de quatro dados e distribuição resultante de valores possíveis baseados na multiplicação independente de resultados (direita). Note que a distribuição agora se torna mais suave na seção região com menores valores e mais errática na seção com valores maiores (modificada de Rose (2001)).	11
2.6	Relação entre a distribuição normal e lognormal. Para esta figura foi gerada uma lognormal com $P_{10}=90$ e $P_{90}=10$, com truncamento em p_{01} e p_{99} . A distribuição normal subjacente (esquerda) foi obtida pelo operador logarítmico.	12
2.7	Exemplo de resultados previstos de volumes de acumulações e seus respectivos volumes verificados. Note que a maioria das acumulações está abaixo da reta de ajustes perfeitos, abaixo do previsto. Fonte: Shell (2014).	13
2.8	Exemplos de <i>flags</i> visualizados na sísmica, em canais meandrantos próximos a falhas.	14
2.9	Exemplos de distribuições lognormais truncadas em p_{99} e p_{01} . Foram utilizados valores típicos de saturação aplicados em prospectos.	15
2.10	Exemplos de distribuições de porosidade do poço 2-ANP-0001-RJS, disponibilizadas na apresentação técnica da ANP da primeira rodada de Partilha de produção do Pré-Sal. Fonte: Petersohn and Abelha (2013).	16
2.11	Caracterização hierárquica de elementos de um sistema petrolífero criada a partir dos conceitos de Magoon and Dow (1994) e Otis and Schneidermann (1997).	17

2.12	Um sistema petrolífero inclui uma geradora madura, caminho de migração, rocha reservatório, armadilha e selo em conjunção de tempo adequada. O tempo relativo de formação desses elementos e os processos de geração, migração e acumulação são críticos para a acumulação e preservação de hidrocarbonetos sendo estes abarcados pelo sincronismo. Fonte: Modificado de Railsback (2011).	18
2.13	Elementos do Sistema Petrolifero necessários para que ocorra uma acumulação de hidrocarbonetos. Os três fatores de risco primários seriam Carga, Armadilha e Reservatório, que podem ser desdobrados em outros elementos conforme ilustrado.	20
2.14	Matriz de adequação para avaliação da favorabilidade de cada elemento do sistema petrolífero, com base na quantidade e qualidade da informação disponível e no julgamento feito pelo avaliador sobre essas informações. Fonte: Costa et al. (2013).	21
2.15	Faixas de chance de sucesso consideradas razoáveis para tipos diferentes de prospectos. Fonte: Modificado de Otis and Schneidermann (1997)	22
2.16	Mudança na incerteza e método de avaliação num ciclo de vida típico de projeto de E&P.Fonte: (SPE, 2007).	23
2.17	Variações de dimensionalidade em relação a abordagem volumétrica. Do modelo de valores médios 1D utilizado tipicamente no método probabilístico até o modelo 3D normalmente utilizado no método determinístico. O modelo 2D é pouco explorado e normalmente é obtido a partir de extração de atributos do modelo 3D.	24
2.18	Correlação entre MMBtu de calor por barril de óleo <i>versus</i> o grau API do óleo e/ou condensado, a figura foi publicada com base em diversas fontes e talvez a não padronização das proporções de gás e/ou medidas tenha causado as descontinuidades observadas. Fonte: SPE (2007).	28
2.19	Figura ilustrativa do fechamento estrutural da estrutura Old Harry na Bacia de Newfoundland-Canadá, com área de aproximadamente 30kmx12km e 450m de coluna. Fonte: Corridor-Resources-Inc (2016).	29
2.20	Figura ilustrativa definindo diferentes volumes de Rocha para um prospecto no Golfo do México com base em diferentes modelos de velocidade. Fonte: Osypov et al. (2013).	30
2.21	Exemplo de definição de percentis de áreas baseados em curvas de contorno com p10, p90 e p50 do prospecto Odin na Austrália. Fonte: Metgasco (2018).	30
2.22	Exemplo de definição de estimativa de Gross Pay do prospecto Odin na Austrália. Fonte: Metgasco (2018).	31
2.23	Figura ilustrativa de um anticlinal com mergulho variável e espessura verdadeira constante, e com espessura aparente distorcida.	32
2.24	Figura ilustrativa de uma camada inclinada e poço inclinado.	32
2.25	Figura ilustrativa de um prospecto estratigráfico, que diferencia coluna de espessura.	33
2.26	Figura ilustrativa exemplificando os fatores geométricos note que quanto mais espesso e dobrado maior será o fator geométrico.	34
2.27	Ábaco utilizado para estimar o fator geométrico, note que a medida que a razão Espessura do Reservatório/ Altura do Fechamento aumenta o fator geométrico reduz e quanto menos achatada é a estrutura menor é o fator geométrico. Fonte: Modificada de White (1987)	34
2.28	Diagrama ilustrando as diferenças entre Gross Pay, Net Litologia, Net Reservatório e Net Pay. Acima os estágios do cálculo volumétrico representando estas diferentes etapas. Onde NTG = <i>Net-To-Gross</i> , e TF = <i>Trap-Fill</i>	36
2.29	Ilustração dos empacotamentos com grãos esféricos e suas respectivas porosidades exatas calculadas matematicamente. Fonte: modificado de Glover (2016)	38
2.30	Ilustração de <i>trend</i> de porosidade versus profundidade para os plays Mioceno Médio e Plioceno no GOM. Fonte : Amado (2013)	40

2.31	Seção Sísmica com arquitetura de falhas do campo de Tidepole (Noroeste da Austrália) e mapa estrutural. Fonte: Smith et al. (2018).	41
2.32	Modelo esquemático do campo de Tidepole ilustrando as variações de <i>Trap Fill</i> causado pelas falhas. Fonte: Smith et al. (2018).	41
2.33	Modelo esquemático de reservatório molhável a água (esquerda) e a óleo (direita) após a migração secundária de hidrocarbonetos. Note que em ambos os modelos ocorrem saturações residuais de água. Fonte: Crain (2019).	43
2.34	Modelo esquemático com as frações de saturação no espaço poroso.	43
2.35	Ábaco proposto por Standing and Katz (1942), para determinar o fator de compressibilidade (fator-z) de gases reais. Os valores de pressão e temperatura reduzida são obtidos matematicamente. Verifique que a escala superior de pseudo-pressão reduzida diz respeito aos valores das curvas acima e a escala inferior diz respeito às curvas de baixo.	47
2.36	Exemplo de fator de recuperação vs profundidade para diferentes campos no Golfo do México portadores de óleo (note que eles estão agrupados por idade). Fonte: Amado (2013).	56
2.37	Ilustração hipotética da relação entre a temperatura e a venda de gelo (esquerda), e temperatura e a venda de aquecedores em uma cidade (direita).	57
2.38	Crossplots de permeabilidade vs porosidade em reservatórios carbonáticos, a esquerda em carbonatos da Formação Quissamã, Bacia de Campos (Favreto et al., 2016), a direita em coquinas da idade Jiquiá também na Bacia de Campos (Muniz and Bosence, 2017). Em ambos os gráficos as diferenças nos pontos estão relacionadas a diferentes fácies.	58
2.39	Exmplos de coeficiente de correlação de Pearson. Vários conjuntos de pontos (x, y), com o coeficiente de correlação de Pearson entre x e y para cada conjunto. Observe que a correlação reflete o ruído e a direção de um relacionamento linear (linha superior). Ela não é sensível à inclinação desse relacionamento (meio). Este coeficiente não é bom para medir relacionamentos não-lineares (parte inferior). Note que a figura no centro tem uma inclinação de 0, mas nesse caso o coeficiente de correlação é indefinido porque a variância de Y é zero. Fonte : https://en.wikipedia.org/wiki/Correlation_and_dependence .	60
2.40	Gráfico de coeficientes de correlação Pearson (Cor_{AB}) versus o desvio padrão (σ_B) da normal subjacente da distribuição B (N_B). As curvas superiores representam as correlações máximas das distribuições lognormais. Elas foram obtidas a partir do coeficiente de correlação das distribuições subjacentes $\delta_{N_A, N_B} = 1$. As curvas inferiores representam correlações mínimas, obtidas a partir de $\delta_{N_A, N_B} = -1$. As cores variam para diversos desvios padrões σ_A das normais subjacentes N_A simuladas. Fonte: modificado de Demirtas and Hedeker (2011).	61
2.41	Diagrama esquemático do algoritmo <i>Generate Sort and Correlate</i> , (GSC) proposto por Demirtas and Hedeker (2011).	63
2.42	Exemplos de <i>crossplots</i> equivalentes a uma distribuição bivariada normal. Foram criadas duas distribuições normais com 50000 amostras, ambas com média 0 e ambas com desvio padrão de 1, a correlação das distribuições é zero (esquerda). Alternativamente (direita) temos duas distribuições também com média 0 e desvio padrão diferentes. A variável no eixo das abscissas tem desvio padrão 1 e a variável no eixo das ordenadas tem desvio padrão 2, a correlação das distribuições é zero.	67
2.43	Um diagrama com os passos para obter distribuições normais correlatas através da transformação linear de Cholesky em distribuições Normais.	69
2.44	Um diagrama com os passos para obter distribuições normais correlatas através da transformação linear de Cholesky em distribuições quaisquer.	70

2.45	Exemplos de <i>crossplot</i> equivalente a uma distribuição bivariada normal. Foram criadas duas distribuições normais com 50000 amostras, média 0, desvio padrão da variável no eixo das abscissas de 1 e desvio padrão no eixo das ordenadas de 2. Na figura da direita a correlação das distribuições é zero, na do centro a correlação é 0,5 e na da direita é 0,9.	70
2.46	Diferentes tipos de correlação obtidos por diferentes funções cópula. Os parâmetros dos <i>crossplots</i> foram ajustados de maneira a atingir um mesmo valor de correlação de Pearson de 0,8	72
2.47	Exemplos de cópula <i>t-student</i> com diferentes parâmetros de grau de liberdade ν e diferentes correlações.	75
2.48	Superfícies de CDFs e PDFs da cópula de Clayton, obtidos para diferentes valores de θ , e resultam em diferentes comportamentos das distribuições.	79
2.49	<i>Crossplot</i> de pontos gerados pela distribuição de Clayton com diferentes valores de θ . 80	
2.50	<i>Crossplot</i> de pontos gerados pela distribuição de Clayton com diferentes valores de θ negativos.	81
2.51	<i>Crossplot</i> de pontos gerados pela distribuição de Clayton com diferentes valores de θ negativos, com o algoritmo modificado.	81
2.52	Superfícies de CDFs e PDFs da cópula de Gumbel-Hougaard, obtidos para diferentes valores de θ , que resultam em diferentes comportamentos das distribuições. 83	
2.53	<i>Crossplot</i> de pontos gerados pela distribuição de Gumbel-Hougaard com diferentes valores de θ	84
2.54	<i>Crossplot</i> de pontos gerados pela distribuição de Gumbel-Hougaard com diferentes valores de θ negativos.	85
2.55	Superfícies de CDFs e PDFs da cópula de Frank, obtidos para diferentes valores de θ , que resultam em diferentes comportamentos das distribuições.	86
2.56	<i>Crossplot</i> de pontos gerados pela distribuição de Frank com diferentes valores de θ . 87	
2.57	<i>Crossplot</i> de pontos gerados pela cópula de Frank com diferentes valores negativos de θ	88
2.58	Séries Correlacionadas geradas pelos métodos propostos por Miller (1998) em seu artigo original (legendas foram traduzidas). Reamostragem da função normal analítica (A).O problema que abordaremos aqui é referente à figura na parte central (B). Amostragem estratificada (C). Todos os exemplos com as mesmas distribuições marginais e com correlação de Pearson aproximadamente 0.8.	89
2.59	Fluxograma de mudança de correlação numericamente por troca de pares de variáveis.	90
2.60	Fluxograma de mudança de correlação numericamente por troca de variáveis com o uso de janela móvel.	91
2.61	<i>Crossplot</i> de pontos com coeficiente para comparar as dispersões de pontos utilizando a a troca de pares e método GSSC. Ambas as figuras tem coeficiente de correlação de Pearson igual a 0,8	92
2.62	<i>Crossplot</i> de pontos com coeficiente para comparar as dispersões de pontos utilizando a cópula de Frank e GSSC. Ambas as figuras tem coeficiente de correlação de Pearson igual a 0,82	92
2.63	<i>Crossplot</i> de pontos de duas distribuições normais. Acima temos a distribuição completa com coeficiente de correlação de Pearson de 0.95, seguindo uma estrutura de correlação que segue a cópula de Clayton ($\theta=1$). Os retângulos nas figuras são as sub amostras selecionadas e apresentadas na parte inferior. Note na parte inferior esquerda que a subamostra preserva a alta correlação mas apresenta uma estrutura semelhante à cópula de Gumbel enquanto a figura da parte central (quadrado inferior direito) apresenta um baixíssimo coeficiente de correlação. . .	94

3.1	Modelo esquemático em seção de prospecto quaquaversal com espessura variável com aumento da área.	97
3.2	Topo do reservatório do Modelo esquemático (ES1) que constitui um prospecto quaquaversal radialmente simétrico (A), superfície com valor de z constante simulando um contato cortando a estrutura (B).	97
3.3	<i>Crossplot</i> de área vs espessura baseado em cortes do contato da estrutura do Modelo ES1 (representado na Figura 3.2), volume de rochas em cores.	98
3.4	Topo do reservatório do Modelo esquemático (ES2) de prospecto quaquaversal deformado (A), superfície simulando um contato cortando a estrutura (B).	98
3.5	<i>Crossplot</i> de área vs espessura baseado em cortes de contato do Modelo ES2 conforme representado na Figura 3.4, volume de rochas em cores. Note que a diferença da geometria da estrutura impacta a geometria de correlação mas ela continua com forte comportamento linear e preserva um alto coeficiente de correlação (98.6%).	99
3.6	Modelo esquemático em seção de prospecto quaquaversal simétrico com variação nos flancos da estrutura causada por uma possível variação de velocidade (20% a -20%).	99
3.7	<i>Crossplot</i> de área vs espessura baseado em cortes de contato da estrutura simétrica esquematizado na Figura 3.8, volume de rochas em cores. Note que há uma replicação do gráfico da Figura 3.3 onde a espessura cresce até atingir assintoticamente a diferença entre topo e base do reservatório (nesse caso definido como 10m).	100
3.8	Modelo esquemático em seção de prospecto quaquaversal com espessura de reservatório aproximadamente constante.	100
3.9	<i>Crossplot</i> de área versus espessura média baseado em cortes de contato do Modelo ES3, volume de rochas em cores. Note que a espessura cresce até atingir assintoticamente o valor de espessura da camada (nesse caso definido como 10m).	101
3.10	Modelo esquemático em seção de prospecto quaquaversal com espessura de reservatório aproximadamente constante definida em 10m (Modelo ES3), com variação nos flancos da estrutura de +20% à -20%	101
3.11	<i>Crossplot</i> de área vs espessura média baseado em cortes de contato da estrutura simétrica esquematizado na Figura 3.10, volume de rochas em cores. Note que espessura cresce até atingir assintoticamente a entre topo e base do reservatório (nesse caso definido como 10m).	102
3.12	Modelo esquemático em seção de prospecto quaquaversal com variação da espessura de reservatório em cada iteração seguida da variação da curvatura.	103
3.13	<i>Crossplot</i> de Área versus espessura média, baseado em cortes de contato da estrutura simétrica esquematizado pelos passos da Figura 3.12, variando espessura e curvatura. Volumes de rochas em cores.	103
3.14	Workflow Petrel 2016 utilizado para representar variações no topo do reservatório de um prospecto estrutural com espessuras que variam uniformemente em cada iteração. Juntamente implementamos a variação lateral nos flancos de -20% a +20% . O workflow está comentado de maneira a possibilitar a reprodução.	104
3.15	Exemplo de modelo de variação lateral de fácies (esquerda) e exemplo de incorporação de variações no modelo de reservatório (direita).	105
3.16	Mapa estrutural do topo do reservatório do pseudo-prospecto quaquaversal, o mesmo do modelo ES1	105
3.17	Ilustração do Modelo FR1, relevo do mapa estrutural do topo do reservatório do pseudo-prospecto quaquaversal superposto com feições estratigráficas propostas neste trabalho. As cores representam os diferentes valores de <i>net pay</i>	106

3.18	Modelo FR1 e ilustração do corte no contato máximo @ 2900m de profundidade. Superposto ao modelo estão as respectivas fácies reservatórios. O contato está representado pelo plano constante.	107
3.19	<i>Crossplot</i> de Área versus <i>net pay</i> médio baseado em cortes de contato considerando as fácies representadas no Modelo FR1, Figura 3.18. Profundidade do contato em cores.	107
3.20	Modelo FR2, Cenário em que temos fácies mais empobrecida no ápice da estrutura e canais mais ricos em areias nos flancos.	108
3.21	<i>Crossplot</i> de área versus <i>netpay</i> do Modelo FR2.	109
3.22	Modelo FR3, cenário em que temos uma fácies reservatórios concentradas nos flancos da estrutura.	109
3.23	<i>Crossplot</i> de área versus <i>net pay</i> , para os cenários em que os reservatórios estão nos flancos da estrutura (para o modelo FR2). Profundidade do contato em cores.	110
3.24	Relação de RGO com B_o dependente baseado nas leis de Standing e equação empírica de Kamari et al. (2016).	112
3.25	Buckles plot com a relação de S_{wir} e ϕ com decrescimento hiperbólico definido pelo método de Buckles para diferentes constantes k.	113
3.26	Buckles plot com a relação de S_{wir} e ϕ com decrescimento hiperbólico conforme extraído de Baker et al. (2015) p 272. Os dados foram extraídos de testemunhos de recifes carbonáticos com óleo obtidos de da região de Bonnie Glen D-3A Pool, Alberta, Canada.	114
3.27	Complemento do Buckles plot com a relação de S_{Hc} e ϕ com crescimento hiperbólico definido pelo método de Buckles para diferentes constantes κ	114
3.28	Gráfico de Fator de Recuperação vs Porosidade para reservatórios de águas profundas portadores de óleo. As acumulações estão localizadas no Golfo do México, compostos de canais e lobos no mundo Amado (2013).	115
3.29	Gráfico de Fator de Recuperação vs Porosidade para diversos tipos de reservatório portadores de óleo (incluindo os de águas profundas da Figura 3.28) no mundo.	116
3.30	Gráfico de Fator de Recuperação vs Porosidade para diversos campos portadores de óleo e gás abarcando diversos tipos de reservatório incluindo os de águas profundas da Figura.	116
3.31	Exemplo de gráfico tornado aplicado na análise de sensibilidade de um prospecto.	119
3.32	Exemplo de gráfico tornado aplicado na análise de sensibilidade de um prospecto.	120
3.33	Exemplo de <i>crossplot</i> de Bg vs ϕ . Note a estratificação vertical mostrando que a incerteza de Bg tem grande impacto no cálculo de volumes.	121
3.34	Exemplo de <i>crossplot</i> de NTG vs ϕ . Note que não há estratificação vertical em nenhum dos quadrantes mostrando que a incerteza de Bg e NTG tem pouco impacto no cálculo de volumes para esta oportunidade.	121
3.35	Exemplo de <i>crossplot</i> de área versus Gross pay. O tamanho e cor dos pontos refletem o volume de hidrocarbonetos recuperáveis. Note que a concentração de pontos de maior volume no 1.o quadrante indica que as duas variáveis são muito relevantes para o cálculo da oportunidade o que indica que a correlação entre elas terá bastante impacto nos volumes.	122
3.36	Exemplo de gráfico de barras para avaliar o impacto da correlação das duas distribuições.	122
3.37	Exemplo de gráfico de barras para avaliar o impacto da correlação das duas distribuições.	123
3.38	Exemplo de interface Gráfica do Software ReV BR.	125
3.39	Exemplo de interface Gráfica do Software ReV BR.	125
3.40	Exemplo de resultados de volumes calculados extraídos na forma de tabela.	126

3.41	Exemplo de resultados de volumes calculados extraídos na forma de função de curva de probabilidade cumulativa.	127
3.42	Exemplo de Tornando plot do Software ReV BR.	127
3.43	Exemplo de gráfico de avaliação de controle de impacto da correlação.	128
3.44	Exemplo de crossplot do par de correlação avaliado em relação com respeito ao volume de hidrocarbonetos calculado (hidrocarbonetos são a cor e o tamanho dos pontos).	128
3.45	Painel de parâmetros de entrada utilizados no cálculo das distribuições de entrada. O botão verde é utilizado para calcular as distribuições.	129
3.46	Exemplo de plotagem de histograma com distribuição.	130
3.47	Exemplo de parâmetros p10 e p90 mal definidos (onde p90>p10).	131
3.48	Painel referente a correlação entre pares de variáveis.	131
3.49	Exemplo de plotagem de <i>crossplot</i> com distribuições de Área (eixo x), <i>Gross Pay</i> (eixo y) e Saturação (cores).	132
3.50	Painel referente a determinação dos fatores de chance de sucesso geológico.	132
4.1	Seção geológica esquemática com correlação do play pré-sal na margem brasileira e angolana (fonte: http://www.gasandoil.com/news/2013/08/west-africa-deep-water-discovery-disappoints).	134
4.2	Parte da carta estratigráfica da Bacia de Santos, apenas com a seção pré-sal, o círculo verde indica a geradora Jiquiá, enquanto os reservatórios são os estromatólitos da seção sag (Formação barra Velha) e as coquinas da Formação Itapema. Fonte Moreira et al. (2007) <i>apud</i> . Petersohn and Abelha (2013)	134
4.3	Correlação de poços e respectivos ciclos de Karagodin da seção pré-sal carbonática entre poço das Bacias de Campos, Brasil e Kwanza, Angola (Ceraldi and Green (2016)).	135
4.4	Estrutura da base sal, topo do reservatório, para o prospecto Libra (atual campo de Mero). Fonte: (Petersohn and Abelha, 2013).	136
4.5	Seção sísmica passando pela estrutura de Libra (CGGVeritas, 2013). Note que a figura faz referência a um possível <i>flat spot</i> coerente com o que já foi publicado sobre o campo.	137
4.6	Tornado <i>plot</i> com os parâmetros do prospecto Libra.	138
4.7	Parâmetros de correlação aplicados ao prospecto Libra, positiva entre área e <i>gross pay</i> e negativa entre <i>gross pay</i> médio e NTG médio. Note que em um mesmo <i>crossplot</i> é possível visualizar os dois pares de correlação uma vez que o <i>gross pay</i> está presente em ambos os pares de variáveis.	139
4.8	Correlação gerada indiretamente entre Área e NTG para o prospecto Libra a partir da correlação positiva entre área e <i>gross pay</i> e negativa entre <i>gross pay</i> médio e NTG médio. As cores representam o volume de argila (V. Shale) com comportamento independente.	139
4.9	Gráfico com o impacto das correlações entre área e <i>gross pay</i> médio (gaussiana positiva : +0.8) e NTG vs <i>gross pay</i> médio, (gaussiana negativa : -0.8).	140
4.10	<i>Crossplots</i> das correlações entre área e <i>gross pay</i> (gaussiana positiva : +0.8) e NTG vs <i>gross pay</i> , (gaussiana negativa : -0.8). Veja em ambos os gráficos os maiores valores de volume se concentram nos maiores valores de área e GP	140
4.11	Verifique os impactos das diferentes estruturas de correlação para um valor de correlação Pearson de 0.8. São exibidas aqui a estrutura de correlação gaussiana (superior esquerda), T-student (superior direita), Frank (centro esquerda), Clayton (centro Direita), Gumbel (inferior direita), GSSC (inferior esquerda)	141
4.12	Batimetria em cores evidenciando a localização da Bacia do Golfo do México profundo. Fonte : BOEM (2017).	143

4.13	Mapas de drenagem expressando a grande área de drenagem que foi depositada no GoM (<i>Gulf of Mexico</i>) ao longo do Terciário-Neógeno. Fonte: Galloway (2009).	144
4.14	Carta Estratigráfica do Golfo do México com os principais reservatórios e geradoras testados. Fonte: Galloway (2009).	145
4.15	Mapas estruturais, em branco contorno p01 do prospecto Alfa.	146
4.16	Modelo estratigráfico de deposição dos lobos e canais em ambiente de água profunda em bacias ricas em sal (com o GoM). Fonte : Buckles (2000).	147
4.17	Seção sísmica <i>strike</i> sem interpretação ao longo do "Prospecto Alfa".	148
4.18	Seção sísmica <i>strike</i> interpretada, o horizonte em vermelho representa a base da seção alvo e em amarelo o topo da seção alvo do "Prospecto Alfa".	148
4.19	Mapa de amplitudes RMS extraído entre os horizontes do topo e base da seção alvo (aproximadamente 200ms).	149
4.20	Mapa de amplitudes RMS extraído entre os horizontes do topo e base da seção alvo (aproximadamente 200ms), o contorno em branco representa a área de amplitudes p01 ($360km^2$) e em vermelho a área de amplitudes P90 ($86km^2$).	149
4.21	Mapa estrutural do topo da seção objetivo, o contorno em branco representa a área máxima de ocorrência de areias p01 ($360km^2$) e em vermelho a área de amplitudes P90 ($86km^2$).	150
4.22	Seção sísmica com maior detalhe sobre o prospecto Alfa, note as feições de pequeno comprimento de onda típicas de canal e sobre elas feições alongadas típicas de leques.	150
4.23	Distribuições de entrada utilizadas para o cálculo de distribuições de volumes geradas para o prospecto Alfa.	151
4.24	Volumes gerados para o prospecto Alfa. Em azul, curvas da <i>cdf</i> heurística de distribuições de volumes <i>in place</i> e em laranja a <i>cdf</i> heurística de volumes recuperáveis.	152
4.25	Tabela com os volumes <i>in place</i> e recuperáveis do prospecto Alfa sem o uso de correlação. Esta tabela traduz os dados da Figura 4.24.	153
4.26	Na esquerda o Tornado plot com o impacto de cada variável no volume final (em MMBOE) e na direita as razões p10/p90 de cada variável.	153
4.27	Parâmetros utilizado para aplicar a correlação entre Bo e o RGO e ϕ e FrO. Na direita o crossplot para controle de qualidade dos parâmetros utilizando. Note que mesmo com o alto valor de correlação (0.75) ainda temos uma dispersão considerável.	154
4.28	Impacto nos volumes recuperáveis causados pela escolha dos parâmetros de correlação Bo e o RGO (0.75) e ϕ e FrO (0.75) com estrutura gaussiana.	154
4.29	Impacto nos volumes recuperáveis causados pela escolha dos parâmetros de correlação Bo e o RGO (0.75) e ϕ e FrO (0.75) com estrutura gaussiana.	155
4.30	Gráfico de controle de qualidade (QC plot) do par de correlação (ϕ vs FrO) com estrutura gaussiana e correlação com intensidade de 0.75. Note o progressivo aumento do volume com o valor dos pares de correlação.	155
4.31	Gráfico de controle de qualidade (QC plot) do par de correlação (RGO vs Bo) com estrutura gaussiana e correlação com intensidade de 0.75. Note que não há um padrão claro de aumento ou diminuição de volumes com o valor dos pares de correlação.	156
4.32	Impacto da correlação para diferentes estruturas de correlação disponível para os pares de correlação do prospecto Alfa (RGO vs Bo) e (ϕ vs FrO).	156
4.33	Distribuições de volumes P90 (acima), média (centro) e P10 (abaixo) para multiplicação de distribuições com diferentes correlações (eixo horizontal dos 3 gráficos).	157

Lista de Tabelas

2.1	Valores típicos de incerteza em relação ao seu valor médio para parâmetros de avaliação volumétrica de prospectos. Fonte: SPE (2007)	26
2.2	Classificação de Reservatórios de acordo com a porosidade conforme Tiab and Donaldson (2012).	37
2.3	Valores de porosidade crítica para as rochas conforme Nur et al. (1998) apud Dvorkin and Nur (2001).	38
2.4	Valores típicos do parâmetro m (expoente de cimentação) para a equação de Archie (1942).	44
2.5	Alguns valores de pressão e temperatura crítica de substâncias típicas em gases. .	46
2.6	Valores de massa molar para as frações gasosas típicas utilizadas para o cálculo de densidade específica.	50
2.7	Algumas propriedades físicas (PF= ponto de fusão, PE = Ponto de Ebulição, ρ = densidade e EF= Estado Físico) das principais substâncias típicas em acumulações de hidrocarbonetos Ahmed (2010).	51
2.8	Tabela de valores com as estatísticas de entrada utilizadas por Kamari et al. (2016) para ajustar uma equação empírica de R_s e os demais parâmetros.	53
2.9	Dados gerais médios de hidrocarbonetos. Fonte: www.petrowiki.com.	55
2.10	Alguns fatores que impactam no fator de recuperação de uma acumulação.	56
2.11	Tabela com escala de correlação Pearson e sua possível interpretação (atente que os valores estão em módulo.). Fonte: Whitford (2005).	60
2.12	Tabela de parâmetros para algumas Cópulas arquimedianas. Fonte: Grimaldi et al. (2011).	77
3.1	Parâmetros de distribuições normais utilizadas na construção da Figura 3.24. Os valores de γ_o foram obtidos com base nos dados de API.	112
3.2	Alguns fatores que impactam no fator de recuperação de uma acumulação.	115
3.3	Tabela com correlações entre elementos volumétricos.	118
3.4	Alguns fatores que impactam no fator de recuperação de uma acumulação.	130
4.1	Parâmetros de entrada para modelar prospecto Libra.	138
4.2	Volumes de óleo equivalente <i>in place</i> (em MMBoe) calculados para os diferentes cenários.	142
4.3	Parâmetros de entrada do prospecto Alfa	151
4.4	Parâmetros de entrada do prospecto Alfa. Tabela Estendida com base nas distribuições geradas.	152

Sumário

Agradecimentos	ix
Lista de Figuras	xvi
Lista de Tabelas	xvii
1 Introdução	1
1.1 Objetivos	2
1.2 Contexto do Trabalho	2
1.3 Histórico e Trabalhos Anteriores	3
2 Fundamentos	4
2.1 Incerteza, Risco & Distribuições	4
2.1.1 A Distribuição Normal	9
2.1.2 Da Distribuição Normal para a LogNormalidade	10
2.2 Avaliação Geológica de Prospectos	17
2.2.1 A Avaliação de Risco Geológico de um prospecto	19
2.2.2 A Avaliação Volumétrica de Prospectos	23
2.3 Associação de Variáveis	57
2.3.1 Coeficientes de Correlação	58
2.4 Métodos para Obter Séries Aleatórias Correlatas	65
2.4.1 Métodos de Reordenamento : Iman-Conover - Cholesky	68
2.4.2 Métodos de Reordenamento : Funções Cópula	71
2.4.3 Métodos de Reordenamento: Reordenamento Numérico	88
2.4.4 Limitações em Correlações Multivariadas	93
2.4.5 A questão da escala em Correlações Multivariadas	93
3 Metodologia	95
3.1 Correlação de Parâmetros	95
3.1.1 Área e <i>Gross pay</i>	96
3.1.2 Correlação B_o e <i>RGO</i> .	110
3.1.3 Correlação Porosidade (ϕ) vs Saturação de hidrocarbonetos (S_{hc}).	112
3.1.4 Correlações envolvendo o <i>Fator de Recuperação</i>	114
3.1.5 Correlações Verificadas	118
3.2 Metodologia de avaliação do impacto da correlação.	119
3.3 Análise de Sensibilidade	119
3.4 Programa e Interface - ReVBR (Risco e Volume)	124
4 Aplicações	133
4.1 Relação entre Área <i>Net Pay</i> em Prospectos do Play Pré-Sal - Santos Brasil	133
4.2 Avaliação de Prospecto no Dataset 3D no Golfo do México	143
4.2.1 Interpretação do Prospecto Alfa	145

5 Resultados e Conclusões

Intencionalmente deixada em branco.

Resumo

A exploração e produção (E&P) de petróleo é uma atividade de alto risco não só pela exposição de vidas às condições de perigo mas também pela alta exposição financeira das instituições envolvidas. O risco advém da incerteza sobre as variáveis tanto ambientais quanto antrópicas. A incerteza nesta atividade é fruto do desconhecimento sobre o passado geológico e, o presente e futuro das atividades antrópicas de operação (e.g. ocorrência e tamanho da acumulação, financiamento e implantação do projeto de exploração e exploração, preço de venda, dentre outros).

Ao identificar um prospecto, o objetivo inicial da equipe de exploracionistas constituída de geólogos, engenheiros e geofísicos é caracterizar a possível jazida e avaliar corretamente o valor econômico da oportunidade.

A avaliação deve levar em conta: (1) os cenários geológicos mais prováveis para aquela região; (2) um projeto de perfuração e exploração adequado e (3) premissas econômicas-contratuais correntes. Quando estes três elementos são bem caracterizados conseguimos estimar o risco e prêmio (volume de hidrocarbonetos) do projeto e conseqüentemente o risco financeiro da atividade exploratória de forma coerente.

Dentre as metodologias para tratar a incerteza volumétrica associada a ocorrência de hidrocarbonetos de uma oportunidade, uma das mais usadas pela indústria é a estimativa probabilística via simulação de monte-carlo. Na abordagem tradicional, são simulados milhares de cenários com valores médios possíveis das propriedades envolvidas nos cálculos volumétricos a fim de calcular heurísticamente as probabilidades dos cenários possíveis.

Atualmente, a maioria das empresas e dos trabalhos acadêmicos publicados assumem a premissa de que as variáveis envolvidas no cálculo volumétrico, e de risco interagem de forma independente e aleatória. Neste trabalho iremos apresentar a ocorrência de correlações das propriedades envolvidas no cálculo volumétrico a partir de modelos e equações empíricas e seu impacto no cálculo de volumes de hidrocarbonetos. Serão apresentadas diferentes formas de implementar a relação entre as variáveis e então verificaremos os impactos da correlação na percepção volumétrica do prospecto.

Foi feita uma extensa revisão sobre os parâmetros que permeiam a avaliação de risco e volumetria de oportunidades. Adicionalmente, implementamos uma interface gráfica funcional e gratuita que permitirá a replicação de todos os produtos aqui apresentados, em ambiente acadêmico.

Por fim, apresentaremos dois exemplos de aplicações, uma (re)avaliação do prospecto Libra, no pré-sal da Bacia de Santos, e a avaliação de um pseudo-prospecto na seção miocênica do Golfo do México. Verificamos o impacto das estruturas de correlação na estimativa de volume das pseudo-oportunidades. Os resultados apontam para mudanças que podem variar em mais de 5% sobre a percepção de petróleo recuperável em relação às avaliações feitas dos mesmo prospectos sem o uso da correlação.

Capítulo 1

Introdução

Desde o trabalho de Capen (1992), a avaliação de prospectos exploratórios a partir de uma abordagem probabilística se tornou padrão para tratar incertezas volumétricas de hidrocarbonetos na indústria de petróleo.

Dada a impossibilidade de obter dados de investigação geológica direta e contínua em subsuperfície, recorreremos ao uso da investigação geofísica para complementar os afloramentos e informações de poços, mais precisas, que temos pontualmente numa bacia.

Há incerteza permeando toda a cadeia de informações, desde os dados de poços até o processamento sísmico mais preciso. A incerteza é fruto do tipo, qualidade e representatividade dos dados utilizados, e alternativamente devido à multiplicidade de cenários geológicos com mesma resposta geofísica. A medida que o grau de incerteza aumenta uma abordagem probabilística se torna mais necessária para representá-la.

Para lidar com a ambiguidade geológica/geofísica por vezes trabalhamos com incerteza através de variáveis que se comportam como distribuições (variáveis estocásticas). As variáveis estocásticas são então incorporadas na modelagem dos métodos geológicos/geofísicos.

Neste trabalho trataremos da modelagem dos parâmetros estocásticos que definem o volume de petróleo em um prospecto exploratório. Para tratar o cálculo de volumes de prospectos, utilizamos a equação de balanço de volumes de hidrocarbonetos em condições de subsuperfície em relação ao seu volume em superfície. No caso de uma acumulação de óleo e gás associado, a equação de volume de hidrocarbonetos inicialmente *in place* (STHIIP - *Stock Tank Hydrocarbon Initially In Place*):

$$STHIIP = \frac{GRV.NTG.TF.\phi.Sh}{Bo}(1 + RGO) \quad (1.1)$$

Onde:

- GRV (*Gross Rock Volume*) - Volume de Rocha que contém a acumulação;
- NTG (*Net to Gross*) - Indica quanto do volume de rocha (GRV) está preenchido de reservatório;
- TF (*Trap Fill*) - Indica quanto do reservatório está preenchido por óleo;
- Porosidade (ϕ) - Indica o percentual de espaço poroso preenchido por uma mistura de hidrocarbonetos e água em relação ao volume total de reservatório;
- Saturação de hidrocarbonetos (Sh) - Indica a razão de hidrocarbonetos em relação a todo o espaço poroso;
- Bo (*Bulk Oil/Formation Factor*) - Indica a razão de volumes de óleo em subsuperfície em relação ao volume em superfície. Deve representar as mudanças de temperatura e pressão e a perda da fração gasosa

- RGO (Razão Gás-Óleo) - Indica a proporção de gás associado em solução que é liberada do óleo quando em condições de temperatura e pressão de superfície.

Na equação 1.1 todas as variáveis são consideradas aleatórias e independentes. Elas são modeladas como **distribuições lognormais** geradas por métodos de monte carlo separadamente e posteriormente aplicadas no produto da equação 1.1.

Neste trabalho mostraremos o impacto destas variáveis não serem necessariamente independentes. Mostraremos também que o uso de diferentes valores e estruturas de correlações tem impacto direto na estimativa de volumes de um prospecto exploratório.

Para atingir este objetivo, no Capítulo 2 serão abordados os tópicos: (i) a incerteza que motiva o uso da simulação, (ii) o significado físico de cada variável da equação 1.1 e, por fim, (iii) métodos já apresentados na literatura para simular a correlação das variáveis em simulações de monte carlo.

No Capítulo 3, ilustraremos modelos que exemplificam a correlação entre os parâmetros do cálculo de volumes. Apresentaremos também uma interface gráfica que contou com algoritmos implementados neste trabalho. Este produto permitirá a rápida avaliação em prospectos conceituais e *leads* identificados em dados reais. Desta forma, o leitor pode reproduzir o conceito aqui utilizado em outras condições habituais ao ambiente de interpretação exploratória. Por fim no capítulo 3 são apresentadas

No Capítulo 4 (Aplicações), faremos aplicações em dois prospectos um no pré-sal e um identificado em dados públicos licenciados como *Creative Commons* no Golfo do México dos EUA. Nos Capítulo 5, concluímos o texto citando as principais contribuições deste trabalho no processo de avaliação de prospectos exploratórios.

Esperamos que esta dissertação seja uma ferramenta de ensino para a Universidade, ainda não explorada em nenhum dos cursos de geologia e geofísica do país, tanto para o corpo docente quanto para o corpo discente.

1.1 Objetivos

Este trabalho tem por objetivo principal atuar na incerteza relacionada à variáveis geológicas e geofísicas, mais especificamente a correlação que eventualmente um ou mais parâmetros podem ter entre si, e qual impacto a correlação de variáveis aleatórias tem na estimativa volumétrica de um prospecto avaliado pelo método probabilístico 2D (Rose, 2001).

Tivemos especial preocupação em produzir algoritmos e uma interface gráfica em software gratuito e de fácil compreensão (linguagem Python) de forma que a universidade possa fazer uso em projetos multi-cliente. O código utilizado é aberto e pode ser livremente modificado em futuros projetos.

1.2 Contexto do Trabalho

Após a quebra do monopólio da indústria brasileira de óleo e gás (O&G) em 1997, houve um aumento no nível de investimentos no Brasil que se seguiu até 2014, quando uma queda global nos preços de petróleo (Figura 1.1 obtidos do FRED¹), causada pela desaceleração da economia dos países emergentes (China, Rússia e Brasil) e o aumento da produção das acumulações não-convencionais, que inundou o maior mercado mundial (os EUA), com óleo barato e em abundância².

Em face a este novo cenário desafiador, há uma necessidade de estimativas mais realistas da incerteza relacionada à predição de volumes na exploração. Com este trabalho esperamos

¹FRED = Federal Reserves Department - Departamento de Reservas Federal dos Estados Unidos.

²<https://www.investopedia.com/ask/answers/030315/why-did-oil-prices-drop-so-much-2014.asp>

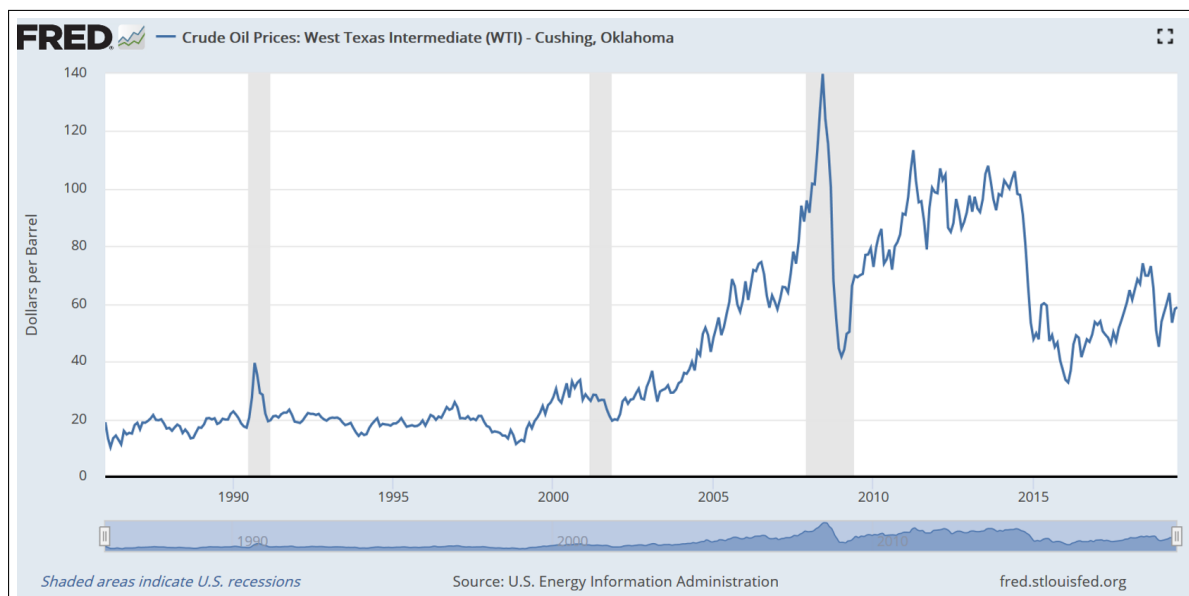


Figura 1.1: Preços do barril de Petróleo desde 1986 até os dias atuais, intervalo de amostra: 1 mês. Fonte: FRED (2019)

contribuir tanto para indústria quanto para universidade abordando o tema da correlação das variáveis aleatórias e seu impacto na incerteza volumétrica.

1.3 Histórico e Trabalhos Anteriores

A primeira aplicação do método de Monte-Carlo ter sido desenvolvido em 1777 com o problema das agulhas³, por George Louis Leclerc, conhecido como Conde Buffon Mas foi só no século 20, com o aumento da capacidade computacional em termos de processamento, memória e armazenamento, que houve a popularização do uso da ferramenta em diversos campos da ciência.

A correlação entre variáveis descritas por distribuições já era estudada no campo da estatística por Francis Galton no fim do século 19 e ganhou notoriedade com Pearson (1920). A aplicação da correlação em métodos de Monte-Carlo foi inicialmente rudimentar utilizando funções de ruído e posteriormente ganhou força com a publicação de Hastings (1970) que utilizou as Cadeias de Markov para modelar a dependência de variáveis em métodos de Monte-Carlo e Iman and Conover (1982) para simular estruturas de correlações gaussianas.

O método de Monte-Carlo em aplicações de geologia de petróleo era reportada como uma fronteira em Wolfe (1963) com aplicações em Jones (1969) e Bradley and Kaufman (1971). A utilização desta ferramenta no cálculo de volumes de hidrocarbonetos se popularizou nos anos 1990 com a publicação de Capen (1992). Desde então o tema foi padronizado nos livros de referência como em Rose (2001) e Harbaugh et al. (1995) se tornando um padrão de avaliação de oportunidades na indústria de óleo e gás.

Apesar de ser discutida em diversos artigos e livros como em Newendorp (1976), Smith and Jones (1992), Harbaugh et al. (1995) e Rose (2001), a aplicação da correlação na estimativa volumétrica probabilística 1D é relativamente pouco explorada na literatura, e por isso abordaremos este tema nesta dissertação. Como novidade verificaremos o impacto no cálculo de volumes não só da correlação mas das diferentes estruturas de correlação obtidas com as funções cópula (um tema pujante e atual que popula muitas dissertações de mestrado e teses de doutorado na área de estatística).

³Neste problema embrionário a disposição de múltiplas agulha jogadas de determinada altura em relação a lâminas de vidro gerou a ideia de distribuição de variáveis aleatórias geradas massivamente

Capítulo 2

Fundamentos

Neste Capítulo apresentamos os conceitos básicos necessários para o entendimento dos capítulos subsequentes. Cabe ressaltar que o objetivo aqui não é esgotar os assuntos abordados mas prover o leitor de uma visão geral sobre os temas apresentados. Fizemos uso de referências clássicas de cada assunto sempre que possível se remetendo ao tema central da dissertação que é a correlação de variáveis na avaliação de prospectos. O texto foi escrito com o objetivo de permitir uma leitura fluida e que, ao seu final, seja possível o uso consistente das ferramentas geradas no âmbito deste trabalho.

Inicialmente tratamos sobre como são definidas distribuições e incertezas que permearão os parâmetros de entrada da avaliação. Sempre que possível abordamos questões geológicas e geofísicas ou relacionadas à indústria do Petróleo. Em seguida faremos uma breve revisão das variáveis que serão objeto da correlação na avaliação geológica de prospectos. Por fim tratamos da estratégia para implementar a correlação na simulação de Monte Carlo.

2.1 Incerteza, Risco & Distribuições

O que motiva o uso de simulações no problema tratado neste trabalho é a incerteza sobre o valor real dos parâmetros de estudo. Por isso, discutiremos a incerteza a seguir.

Num sentido amplo "A incerteza advém de algo que não é conhecido"(Cambridge Dictionary). Ellsberg (1961) sugere uma definição mais precisa (ou diferente): "A incerteza seria um evento ou variável ambíguo ou de probabilidade desconhecida". Strand and Oughton (2009) fazem uma definição mais específica e segmentada e aqui adaptada para o caso de estudo, "A incerteza conota na linguagem cotidiana em três direções diferentes:

1. Relativas a acontecimentos futuros no mundo externo;
2. Relativas a realidades presentes e o nosso desconhecimento ou incapacidade de defini-las com precisão;
3. Relacionadas à mente.

Estamos interessados aqui nos primeiros fatores. Podemos dizer que o resultado de uma partida de futebol, uma eleição, ou as condições climáticas durante uma operação de perfuração são eventos com resultados incertos. Isto significa que o estado (futuro) de determinados eventos no mundo externo não é fixo ou determinado previamente.

Podemos dizer também que há uma incerteza considerável em uma previsão sobre a ocorrência da acumulação de hidrocarbonetos, ou no volume da mesma. Neste caso, a "incerteza" normalmente não é considerada como residindo no mundo em si, mas na qualidade imperfeita do nosso conhecimento sobre esse mundo: Existe uma quantidade determinada de óleo em uma acumulação, existe uma quantidade de moléculas em uma amostra, apenas não sabemos esses números.

Quando a incerteza está relacionada a chance de ocorrência de um evento, normalmente temos associada a ela o conceito quantitativo de probabilidade e portanto nos afastamos do conceito apresentado em Ellsberg (1961).

Quando a chance de ocorrência está associada a um evento negativo ou perigo lhe é conferido uma conotação negativa e por vezes é associada a palavra **risco** (Toma et al. (2012) *apud*. Rowe (1977)). No caso da exploração de óleo e gás (O & G), o risco está associado a chance de **não** encontrar pelo menos uma pequena acumulação de hidrocarbonetos e, portanto, também tem uma conotação negativa, enquanto a probabilidade de sucesso (PS) é dada pelo seu fator complementar.

$$PS = 100\% - RISCO \quad (2.1)$$

Atribuindo sentido financeiro ao conceito, de acordo com Smalley et al. (2008) risco é um evento (ou seja, um cenário específico pode ser descrita) que, caso ocorra, teria um impacto material sobre o negócio. Um exemplo seria o risco de prender coluna de perfuração em um poço sob alta pressão ou perfurar um poço seco, incorrendo em perda financeira.

Uma boa descrição de risco é aquela que descreve a causa, o risco do evento que poderia ocorrer como resultado e a consequência se ocorrer. Descrito desta forma, um risco terá dois valores quantitativos associados a ele: uma probabilidade de ocorrer e um impacto de negócios (valor) se isso ocorrer.

Knight (1921) fornece um conceito tipicamente utilizado por economistas para diferir risco de incerteza e que se aproxima daquele apresentado em Ellsberg (1961). O risco consistiria em eventos cuja probabilidade pode ser mensurada como por exemplo a chance de cair uma face ao lançar um dado. Por outro lado, a incerteza reúne eventos que não conseguem ser previstos e a probabilidade não pode ser definida. Isto significa que o risco poderia ser classificado e quantificado enquanto a incerteza não poderia.

A definição supracitada ainda é recorrentemente utilizada. Note, entretanto, que este conceito é pouco útil para o nosso problema uma vez que temos interesse de ao menos estimar (ainda que numa margem ampla) a faixa de volumes de uma oportunidade exploratória por isso deixamos de lado esta última definição.

Durante os anos 1960, a natureza subjetiva da geologia e a incerteza inerente aos dados de análise e avaliação de prospectos, criaram o entendimento no mercado que, de alguma forma, os geocientistas simplesmente não entendiam sobre economia prospectiva durante a década de 1980 (Rose, 2001). Outro fator que contribuía para esta percepção com os julgamentos intuitivos envolvidos no ranking prospectivo. Esperava-se que os geocientistas fossem otimistas, ao passo que os engenheiros se tornaram “porteiros” econômicos - ambos se tornaram profetas de estimativas não realizáveis que tiveram consequências negativas.

De acordo com Rose (2017), nos anos 1980 grandes empresas ¹ juntamente a empresas de menor porte trabalhavam internamente estabelecendo metodologias claras para avaliação de prospectos sob a luz da incerteza. A primeira publicação de impacto com toda a formulação matemática foi feita por Capen (1992) em seu celebre tratado *Business of Petroleum Exploration*, publicado pela AAPG. Daí em diante algumas publicações foram disponibilizadas na literatura e forneceram inclusive algoritmos como em Harbaugh et al. (1995), SPE (2007) e Rose (2001).

Mas quais fatores originariam a incerteza? Liner (2016) e Rose (2001), são categóricos no âmbito deste trabalho : "Não importa a quantidade de dados disponíveis nunca teremos 100% de certeza sobre a subsuperfície". Note que a resposta é abrangente mas pouco objetiva. Por isto, recorreremos a outros autores com o objetivo de ser um pouco mais preciso. Zio and Pedroni (2012), em seu trabalho sobre seguros compilaram (a partir de Strand and Oughton (2009), Zimmermann (2000) e Armacosta and Pet-Edwards (1999)) diferentes fatores que podem causar incerteza ao fazer uma análise de risco. Estes elementos podem ser utilizados também na indústria de O&G:

¹Das quais se incluem Shell, Elf-Total, Chevron-Gulf, Statoil (hoje Equinor), Exxon Mobil

- **Falta de informação (ou conhecimento)** : Falta de informação, conhecimento e/ou dados sobre um fenômeno, sistemas e eventos a serem analisados. Esta é a principal fonte de incerteza: pode ser de natureza quantitativa (por exemplo, o analista não sabe o valor preciso da probabilidade de um dado evento de interesse) ou qualitativa (por exemplo., o analista conhece as probabilidades do evento de interesse, mas as informações disponíveis não permitem descrição determinística do problema a ser analisado).
- **Aproximação:** Outra situação caracterizada pela falta de conhecimento é chamada aproximação. Ela ocorre quando o analista não tem informações suficientes para descrever exaustivamente o fenômeno de interesse ou quando ele/ela deliberadamente usa um nível mais baixo de detalhes do que o necessário. Em alguns casos, a aproximação é declarada explicitamente, enquanto em outros casos está oculta. Obviamente, esta causa de incerteza pode ser reduzida ganhando mais conceitos, informações e dados sobre o problema em questão.
- **Abundância de informação (ou conhecimento):** Esse tipo de incerteza é devido a incapacidade humana de assimilar e elaborar muitos dados e informações ao mesmo tempo. Nesta situação, o analista geralmente concentra sua atenção apenas naqueles parâmetros e os dados e informações que ele considera mais importantes, negligenciando os outros. A identificação de um rigoroso (e possivelmente automatizado) procedimento para selecionar os dados (entre centenas ou milhares), informações e parâmetros para a aplicação em questão é o fator mais crítica. O analista tem que enfrentar esse tipo de incerteza quando, por exemplo, ele tem que escolher entre diferentes modelos para simular um determinado fenômeno.
- **Natureza conflitante de informações/dados:** Pode acontecer que alguns pedaços de informações e dados disponíveis sugiram um determinado comportamento do sistema, enquanto outros possam sugerir um comportamento diferente. Neste caso, aumentando a quantidade de dados e informações disponíveis não diminuiriam a incerteza, mas aumentariam o conflito entre os diferentes pedaços de informação e dados. Este conflito pode ser devido ao fato de:
 - Algumas informações são afetadas por erros, mas o analista não consegue identificá-las, ou;
 - Alguns dos dados disponíveis não são relevantes para o problema em questão ou;
 - O modelo do sistema utilizado pelo analista não é correto (por exemplo, caracterizado por um *bias*). Mais uma vez, a fim de reduzir essa fonte de incerteza, o analista precisa fazer uma escolha entre as informações e dados disponíveis e, possivelmente, descartar alguns para reduzir o conflito.
- **Ambiguidade linguística:** Todas as línguas contêm palavras que têm diferentes significados dependendo do contexto de análise. Note que esta fonte de incerteza pode ser considerada devido à “falta de informação”, porque fornecer mais detalhes sobre o contexto de análise ajudaria a reduzir a incerteza associada.
- **Subjetividade das opiniões dos analistas:** A incerteza pode derivar da subjetividade da interpretação das informações e dados disponíveis pelo analista: diferentes analistas podem fornecer diferentes interpretações da mesma informação e dependência de dados com base em sua formação cultural e competência no campo da análise. Esta fonte incerteza pode ser reduzida recorrendo à avaliação de múltiplas opiniões de diferentes especialistas.

A incerteza pode se dar também, pelas limitações da natureza do dado de análise. Como por exemplo na Figura 2.1, retirado de <http://seismicreflections.globeclaritas.com/2013/10/>, podemos ver a incerteza relacionada a análise de velocidades em seções profundas dos dados sísmica onde o *semblance-VELAN* (*VELolity ANalysis*) perde resolução.

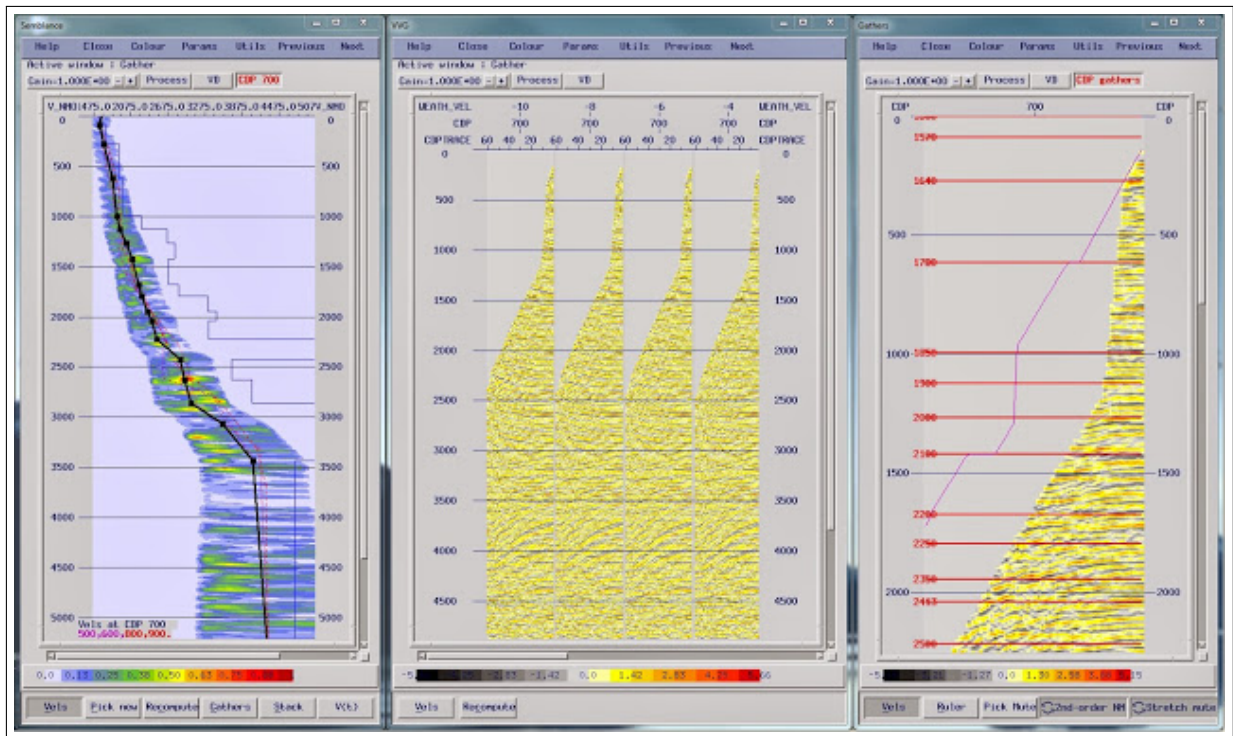


Figura 2.1: Um leque de velocidades ao redor de uma função central (esquerda), *gathers* de velocidade variável que correspondem a esse leque (centro) e *gathers* corrigidos de NMO com velocidades exibidas (direita).

Conforme Smalley et al. (2008), quando a incerteza incorre em risco associamos um valor estático baseado em dados e premissas. Entretanto, a probabilidade de ocorrência e os impactos são eles mesmos incertos e são comumente melhor tratados como intervalos ou distribuições de probabilidade. O risco pode ser gerenciado de duas formas principais (outras ações são subconjuntos destas):

- Mitigações: Onde as decisões são tomadas e os planos mudaram proativamente para reduzir a probabilidade do elemento de risco ocorrer e/ou reduzir seu impacto caso ocorra.
- Contingências: quando os planos são feitos de forma proativa, mas só são implementados quando o elemento de risco ocorre de modo a reduzir o seu impacto.

Do ponto de vista exploratório a mitigação de risco pode envolver atividades proativas como adquirir mais dados geológicos/geofísicos ou atividades passivas como aguardar o resultado de um poço que objetiva um prospecto análogo. Gerenciar risco usando contingencias envolve adicionar mais flexibilidade ao projeto (garantindo sucesso ou percepção do impacto para diversos cenários).

Para lidar com a incerteza, os profissionais de áreas diferentes atuam de formas diferentes. Aqueles que trabalham com medidas fornecem dados de dispersão ou confiança na forma de barras ou valores. Aqueles que trabalham com projeções, por vezes, simulam cenários com base em equações determinísticas definidas empiricamente ou por leis da natureza já provadas matematicamente. Neste último caso as variáveis desconhecidas assumem diversos valores e são simuladas dentro de um intervalo que irão interagir com as variáveis controladas de forma a gerar cenários.

A geração de cenários pode ser feita com um teste de sensibilidade simples utilizando poucos cenários baseados em variações percentuais. É possível também fazer uma simulação com o método *bootstrap* onde ajusta-se os modelos aos dados históricos, obtém-se os resíduos, reamostra-se os resíduos, injeta-se os resíduos reamostrados no modelo para obter os cenários.

Outra estratégia é estimar distribuição de múltiplas variações com variáveis que se comportam aleatoriamente. Neste último exercício, comumente a estratégia utilizada é a simulação de Monte Carlo.

Conforme descrito por Metropolis (1987), a simulação de Monte Carlo constitui uma classe de métodos estatísticos que se baseiam em uma massiva amostragem aleatória para obter resultados numéricos. Repetindo sucessivas simulações em elevado número de vezes (milhares), podemos calcular probabilidades heurísticamente, tal como se registrassem os resultados reais de realizações de jogos de cassino (daí o nome, vem da cidade de Monte Carlo em Mônaco, famosa por seus cassinos).

A simulação de Monte Carlo **moderna** foi desenvolvida para resolver problemas relacionados à difusão de nêutrons para o desenvolvimento da bomba atômica no projeto Metropolis no *Alamos Scientific Laboratory* (não por acaso um de seus autores foi N. Metropolis, em 1944).

Para a caracterização completa da simulação definida por variáveis aleatórias o primeiro passo é determinar o modelo ou lei de interação das variáveis. Em seguida é necessário determinar algumas propriedades das variáveis aleatórias, como seu comprimento (quantidade de amostras) e distribuição de frequências de ocorrência.

Existem diversos tipos de distribuições de frequências das quais podemos citar as distribuições normal, lognormal, uniforme, triangular, chi-quadrado, exponencial. Algumas delas representadas na Figura 2.2.

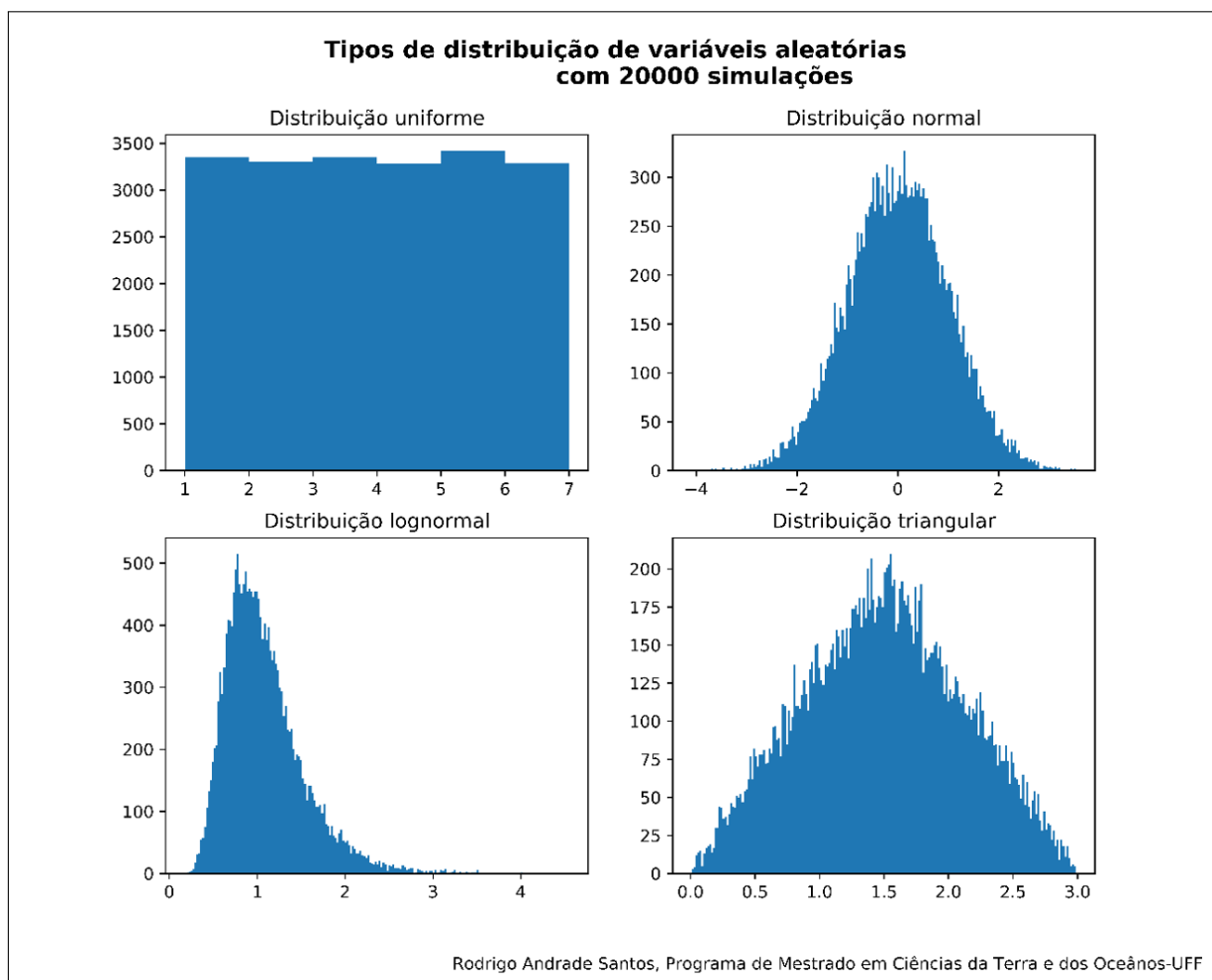


Figura 2.2: Exemplos de distribuições utilizadas em estudos de probabilidade e estatística. Distribuição uniforme (esquerda e acima), distribuição normal (direita acima), distribuição lognormal (abaixo e esquerda), distribuição triangular (abaixo e direita). Todos os gráficos acima foram gerados com a biblioteca `numpy.random`.

2.1.1 A Distribuição Normal

É comum utilizar como primeira aproximação ou primeira tentativa de modelagem de distribuições empíricas uma distribuição Gaussiana, ou distribuição normal. Essa distribuição aparece em diversos fenômenos independentes na natureza. A distribuição Gaussiana é caracterizada por dois parâmetros: média (μ) e o desvio-padrão (σ). A notação para variável x governada por uma distribuição Normal é $x \sim N(\mu, \sigma)$ que é dada por:

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (2.2)$$

Onde $g(x)$ é o valor da variável independente simulada.

Uma das razões da presença abrangente da distribuição Gaussiana se deve a ela emergir naturalmente como uma distribuição limite para processos aleatórios, como consequência do Teorema Central do Limite ou Teorema do Limite Central (TLC): que enuncia que a média da soma de infinitas variáveis aleatórias independentes com variância finita é descrita pela distribuição normal.

Tradicionalmente os lançamentos de moedas são os exemplos utilizados na literatura para o entendimento de distribuições normais e do TLC, utilizaremos aqui o exemplo do jogo de múltiplos dados, assim como em Rose (2001). Esta escolha foi feita pois um exemplo similar será usado na explicação da distribuição lognormal, comumente utilizada na análise volumetria de prospectos.

Ao lançar 1 dado todas as faces tem igual chance de serem sorteadas, ou seja, uma distribuição uniforme homogênea. Entretanto, ao somarmos os resultados dos resultados mais dados (variáveis) a distribuição resultante começa a tomar a forma de um sino que se assemelha uma distribuição normal ver a Figura 2.3. Este é o comportamento previsto pelo TLC, conforme enunciado anteriormente.

A distribuição normalizada para uma média 0 (zero) e desvio padrão igual a 1 (um) recebe o nome de distribuição normal padrão (*standard normal distribution*) esta definição é especialmente útil quando a distribuição normal é utilizada como base para geração de outras distribuições (e.g. a distribuição lognormal discutida a seguir).

Se utilizarmos a função da distribuição normal como uma função de densidade de probabilidades (*probability density function* - pdf), os valores menores que um desvio padrão da média representarão 68,27% do espaço de possibilidades, dois desvios padrão da média representam 95,45%, e três desvios-padrão representam 99,73% do espaço amostral.

Do ponto de vista teórico, a distribuição normal tem características interessantes. Ela é uma curva absolutamente contínua e suave, é simétrica e não atinge o valor zero em nenhum dos pontos do domínio real. Outra característica importante e especialmente útil em aplicações geofísicas, é a sua dualidade no domínio do tempo e frequência, a transformada de Fourier de uma curva gaussiana é também outra gaussiana com variância inversamente proporcional a original. Esta característica é extremamente útil na aplicação de filtros em séries temporais.

Ao utilizar a distribuição normal na análise de incerteza assume-se o conhecimento do valor mais provável de ocorrência (a média) a incerteza a ela associada (o desvio padrão). Estamos, portanto, desconsiderando totalmente o conceito de Knight (1921) e Ellsberg (1961) onde a incerteza é dada por uma variável cujas probabilidades são desconhecidas.

Note também, que a sua simetria garante que a distribuição de eventos será (em grandes distribuições) equilibradas em torno do valor mais provável. Este tipo de distribuição pode ser útil ao modelar um prospecto baseado em um *play* conhecido e bem calibrado, contudo, ao observar o comportamento das propriedades da natureza verifica-se que outra distribuição é observada com bastante frequência, a lognormal.

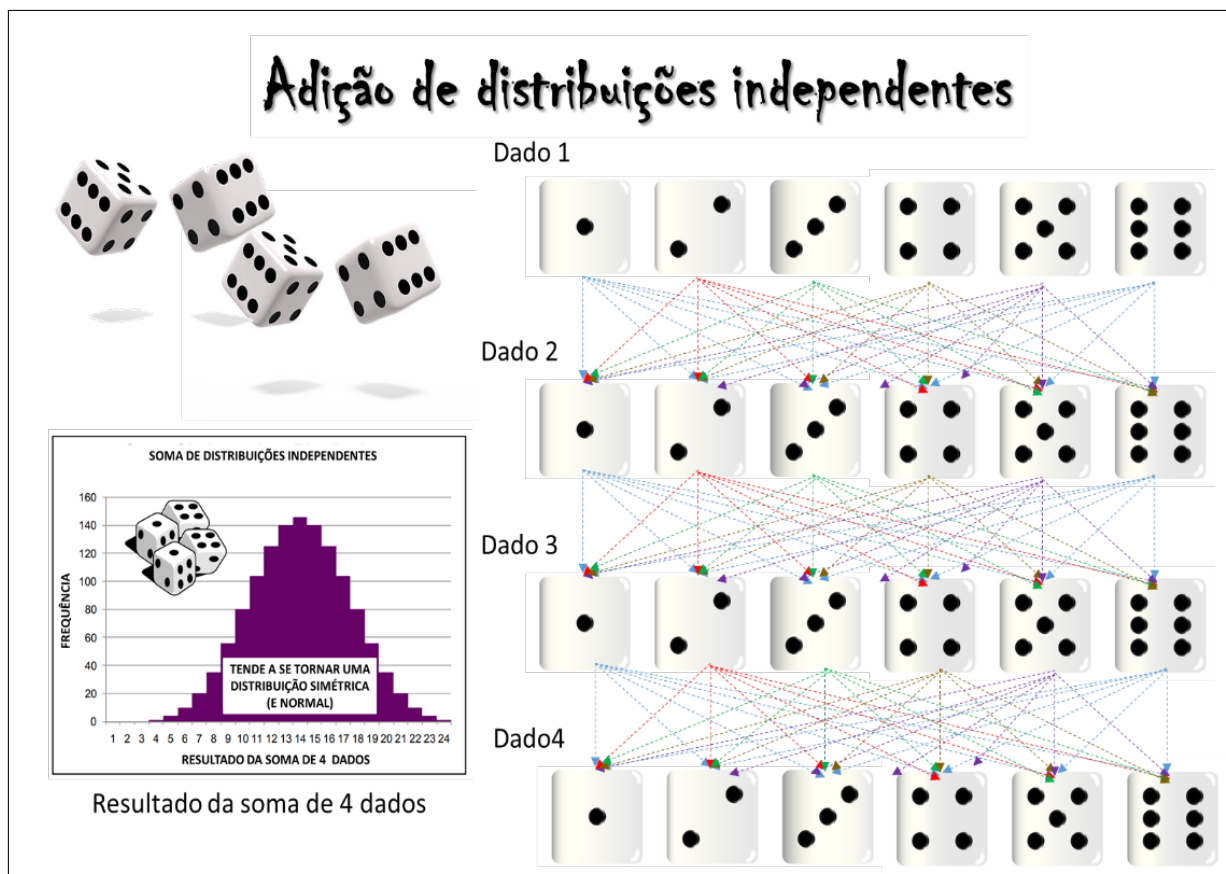


Figura 2.3: Lançamento de quatro dados e distribuição resultante de valores possíveis baseados na soma independente de resultados, note que a distribuição se torna mais semelhante a uma forma de sino. (Modificado de Rose (2001))

2.1.2 Da Distribuição Normal para a LogNormalidade

Apesar de comumente utilizada, a distribuição normal não é sempre a mais adequada para representar fenômenos da natureza (Gunver et al., 2018). Conforme Limpert et al. (2001) descreveram, a distribuição lognormal ocorre frequentemente em diversos campos da ciência e em especial nas ciências naturais. Nas áreas de Geologia e Geofísica (G&G) aplicadas à exploração de hidrocarbonetos as variáveis em escalas bastante diferentes tem boa aproximação na distribuição lognormal: desde a distribuição granulométrica de sedimentos num corpo arenoso, à coluna de hidrocarbonetos e tamanhos de campos em uma bacia com grau relativamente alto de maturidade exploratória (Figura 2.4).

O comportamento lognormal de uma distribuição aparece naturalmente da multiplicação de eventos independentes como podemos ver no exemplo do lançamento de dados ilustrado na Figura 2.5 (modificada de Rose (2001)). Este processo gera assimetria de ocorrência de pequenos valores em grande frequência em detrimento a eventos de grande dimensão ocorrendo em menor número. Segundo Limpert et al. (2001), o primeiro experimento para estudar este tipo de distribuição foi proposto por Galton (1879) utilizando máquinas de distribuição de ocorrências com separadores (triângulos assimétricos).

A extensão da multiplicação de elementos para a geologia de petróleo pode ser vista de forma simples. Dado que muitas distribuições resultam da superposição multiplicativa de processos independentes que se interligam, como a distribuição de porosidades em um reservatório que depende dentre outros fatores do ambiente de deposição, composição da matriz, variação granulométrica e processos digenéticos pós e sin-deposicionais que interagiriam entre si, de forma semelhante a um produto.

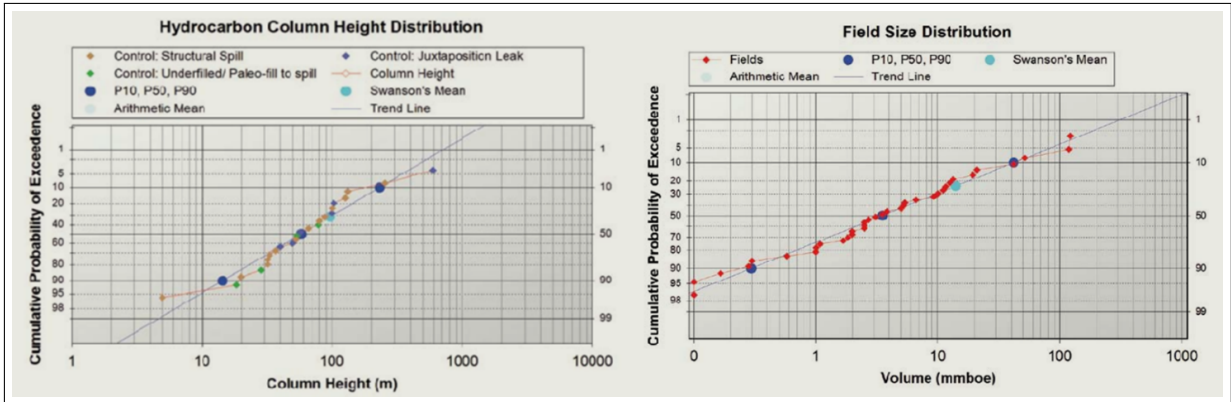


Figura 2.4: Exemplos de distribuições lognormais verificadas em ocorrências de hidrocarbonetos na bacia de Perth (Oeste da Austrália). A esquerda Coluna das acumulações, e a direita a distribuição de campos ao longo da bacia (Ferdinando and Longley, 2016).

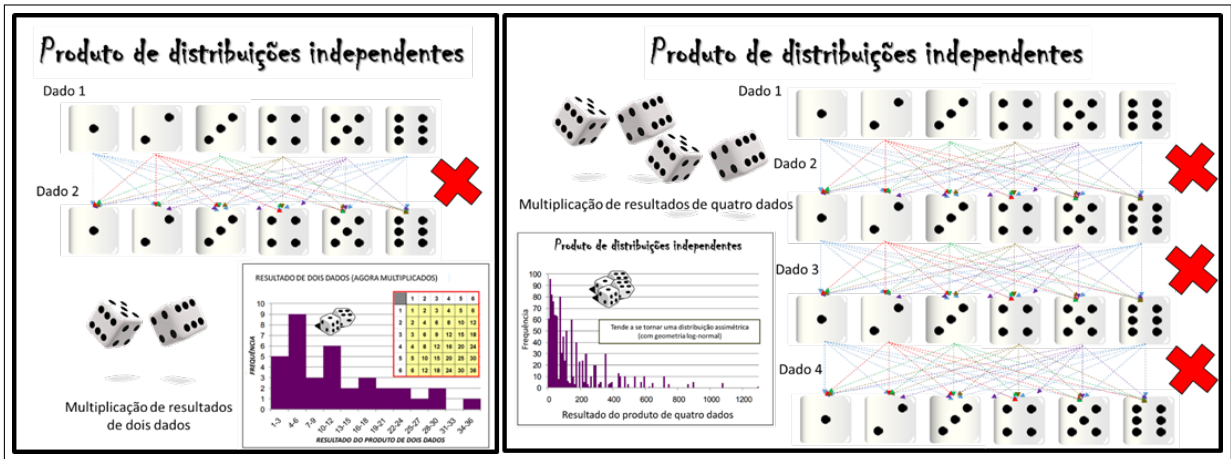


Figura 2.5: Lançamento de dois dados e distribuição resultante de valores possíveis baseados na multiplicação independente de resultados (esquerda) a distribuição resultante é bastante assimétrica com mais valores menores do que maiores. Lançamento de quatro dados e distribuição resultante de valores possíveis baseados na multiplicação independente de resultados (direita). Note que a distribuição agora se torna mais suave na seção região com menores valores e mais errática na seção com valores maiores (modificada de Rose (2001)).

A própria ocorrência de acumulação de petróleo depende que todos os elementos do sistema petrolífero ocorram em consonância em todas as partes do reservatório (como um produtório) o que influencia, por si só, o tamanho da acumulação, que é subsequentemente filtrada a cada fração de ocorrência.

A origem do termo "lognormal" advém do fato de que se aplicado o operador logarítmico às amostras desta distribuição temos uma distribuição com geometria aproximadamente normal. De forma equivalente, ao exponenciar os valores das amostras de uma distribuição normal teremos uma distribuição lognormal.

Este conceito é importante pois a geração de distribuições lognormais normalmente faz uso desta relação (Figura 2.6). Os algoritmos implementados em bibliotecas como excel ou numpy (utilizada neste trabalho) criam uma distribuição normal (dita subjacente) que é então exponenciada gerando uma distribuição lognormal.

Utilizando a equação 2.2 e a relação da Figura 2.6 é fácil verificar que, matematicamente, a curva que representa uma distribuição lognormal pode ser descrita por:

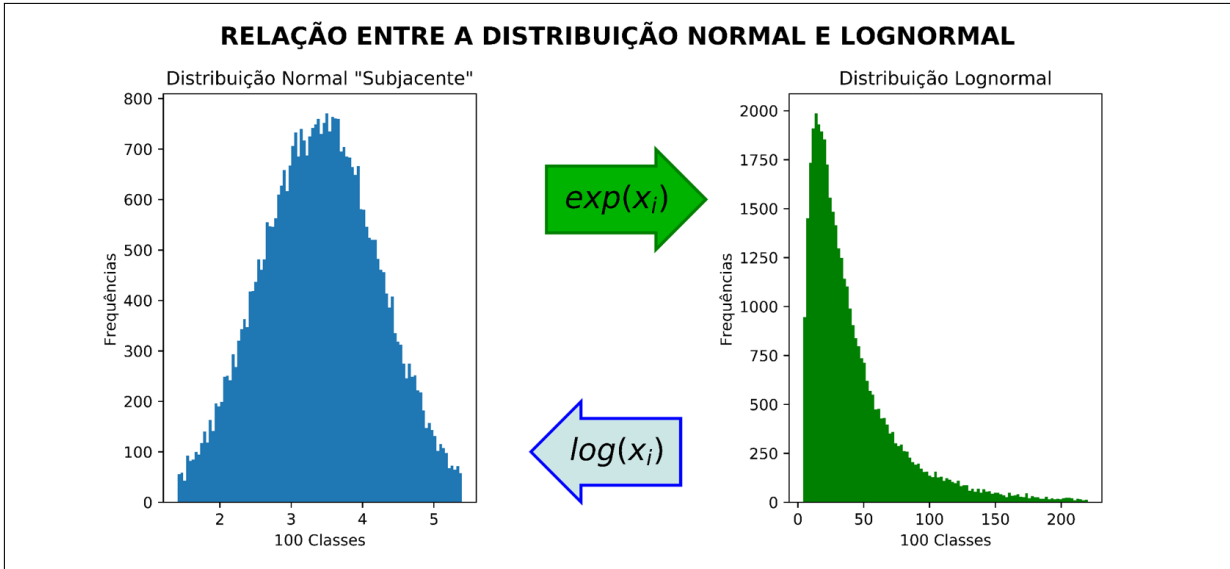


Figura 2.6: Relação entre a distribuição normal e lognormal. Para esta figura foi gerada uma lognormal com P10=90 e P90=10, com truncamento em p01 e p99. A distribuição normal subjacente (esquerda) foi obtida pelo operador logarítmico.

$$g(x) = \frac{1}{\sigma x \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{\ln(x) - \mu}{\sigma} \right)^2} \quad (2.3)$$

Ao trabalharmos com distribuições como funções de densidade de probabilidade (pdf - *probability density function*) é comum o uso de percentis de referência que indicam a probabilidade de que um evento ocorra. Na análise volumétrica utilizamos os percentis P10 para representar o caso otimista que apresenta grandes valores da população e pequena chance de ocorrência (menor ou igual a 10%) e P90 para o cenário conservador que abrange valores menores (90% de chance de ocorrência) dentro das amostras da distribuição. Em termos populacionais, numa escala crescente, o P10 significa que 10% da população é maior que "x" e P90 significa que 90% da população é maior que um outro valor "y".

A assimetria da distribuição lognormal, faz com que sua construção seja pouco intuitiva. Por isso, na indústria do petróleo, a sua construção se dá com base na estimativa dos valores de P10 e P90 (caso otimista e pessimista), Rose (2001). De posse destes valores é possível gerar a distribuição onde o caso mais provável ou cenário base representa a média das amostras (que **não** é o P50 devido à assimetria da distribuição).

Alguns eventos na natureza parecem atender a distribuições diferentes, mas que também podem ser aproximadas para lognormal. Por exemplo, se utilizarmos distribuições com P10 e P90 suficientemente próximos temos uma boa aproximação da distribuição normal com pequeno desvio. Enquanto no caso de distribuições aparentemente exponenciais, as mesmas são geralmente limitadas pela amostragem por efeito de escala e/ou limiares comerciais (Harbaugh et al., 1995; Rose, 2001).

Dessa forma mesmo essas distribuições são naturalmente truncadas pela abordagem de trabalho que fazemos na indústria e os dados relevantes disponíveis para nós são melhor aproximados por distribuições lognormais.

Um outro fator importante abordado por Capen (1992) e Rose (2001) é que os geocientistas são tradicionalmente mais otimistas na fase exploratória, etapa típica onde a modelagem puramente probabilística é utilizada. Isto é constatado inclusive em gráficos de resultados de volume predito versus verificado como na Figura 2.7 disponibilizada pela Shell sobre resultados na Nigéria. A distribuição lognormal por sua geometria, concentraria mais valores menores do que maiores, e dessa forma contrabalancearia o ímpeto otimista dos avaliadores.

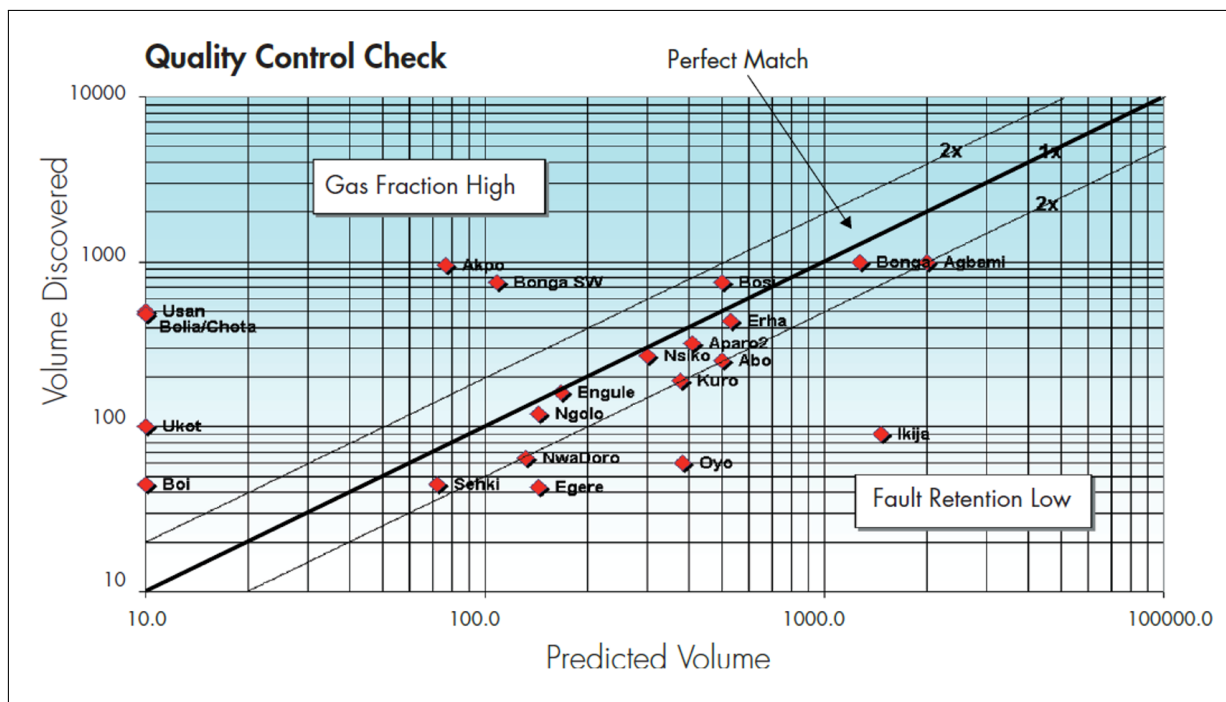


Figura 2.7: Exemplo de resultados previstos de volumes de acumulações e seus respectivos volumes verificados. Note que a maioria das acumulações está abaixo da reta de ajustes perfeitos, abaixo do previsto. Fonte: Shell (2014).

Um característica da distribuição lognormal é que ela pode apresentar valores excessivamente baixos ou altos em ambos os extremos, e, portanto, as variáveis geradas em seus extremos podem ser consideradas um problema conforme aponta Liner (2016).

Na direita da distribuição lognormal (maiores valores) a variável pode ser maior do que qualquer valor razoável. Exemplos clássicos destas distorções envolvem saturações e porosidades superiores a 100% - valores como estes devem ser tratados e evitados.

Do outro lado da escala, valores extremamente pequenos não são nem classificados como prospectos e assim são sub-amostrados. Para exemplificar este último caso, suponha possíveis acumulações com coluna e área extremamente pequenas associadas a falhas, essas ocorrências são classificadas como *flags*, e além de não serem perfuradas, são evitadas com o objetivo de tornar os projetos de poços mais baratos.

Para evitar erros de estimativa nos valores exageradamente altos ou baixos, algumas atitudes devem ser tomadas. A primeira delas é truncar a distribuição em seus valores P01 e P99. Note que estas amostras são relativamente escassas (Figura 2.9).

O passo seguinte consiste em ajustar os valores de P10 e P90 de forma que assumam valores realistas e, por fim, podem ser implementados filtros lógicos que descartam amostras superiores/inferiores a valores limítrofes considerados razoáveis pela equipe de interpretação.

Note que os valores absolutamente altos (o P01) da maioria das distribuições é provavelmente o melhor percentil a ser calibrado Rose (2001). As distribuições de área podem usar o contorno de fechamento em prospectos estruturais. A extensão do reservatório em armadilhas estratigráficas pode ser utilizada também como um critério de área. Valores de saturação de hidrocarbonetos podem considerar a menor quantidade de água residual para o reservatório e a porosidade pode assumir o seu valor crítico.

É importante ressaltar que ao criarmos uma distribuição estamos criando uma distribuição de valores médios para todo o prospecto. Isto significa que valores excessivamente altos e baixos na distribuição representam médias excessivamente altas e baixas.

Além da criteriosa análise dos valores absolutos de P10 e P90, a relação entre eles deve ser

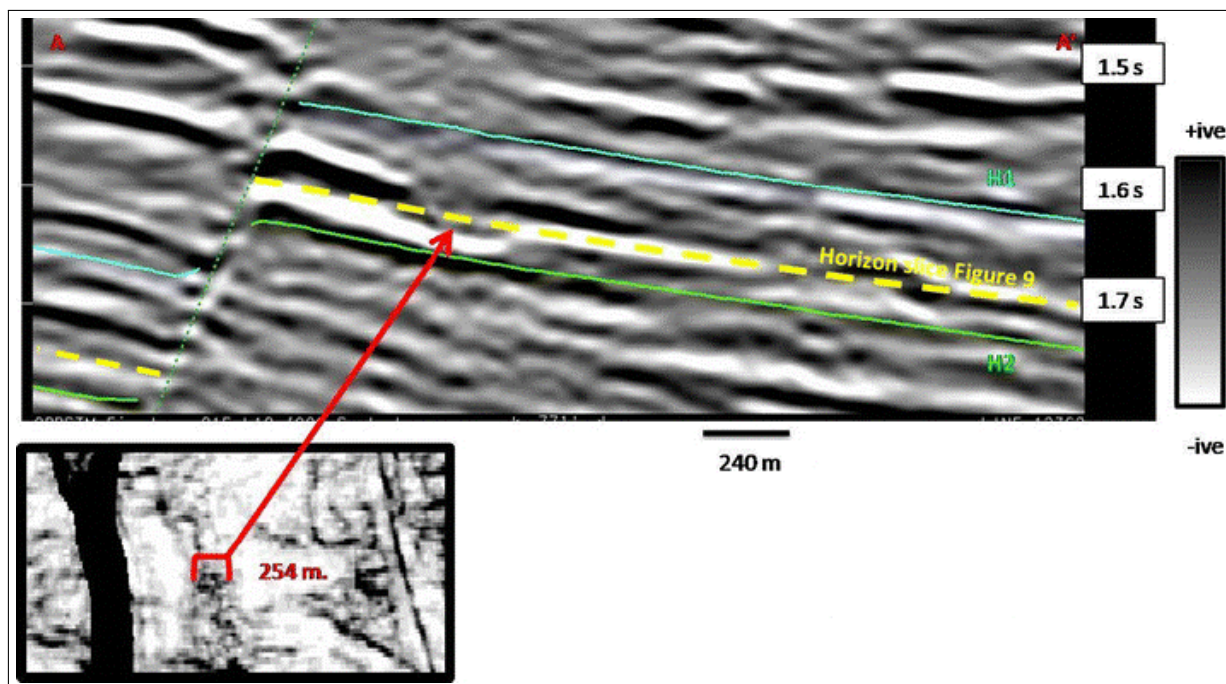


Figura 2.8: Exemplos de *flags* visualizados na sísmica, em canais meandrantos próximos a falhas.

coerente à disponibilidade de informações (e consequente incerteza) sobre o prospecto.

As razões de P10/P90 altas indicam grande incerteza e/ou escassez de dados e baixas razões de P10/P90 indicam pequena incerteza relacionados, por exemplo, a *plays* conhecidos.

Os valores das razões P10/P90 dependem também das variáveis envolvidas no processo. Tome como exemplo um reservatório fechado que produz por fraturas com um range de porosidade entre 2% e 8% $P_{10}/P_{90} = 4$ em detrimento da incerteza em turbiditos terciários em uma bacia conhecida com porosidades entre 18% e 24% teremos os mesmos 6% de variação, mas com P_{10}/P_{90} de 1,25.

A avaliação da incerteza sobre um parâmetro em termos de razão p_{10}/p_{90} permite de alguma forma quantificar a incerteza num prospecto e compara-la a outros prospectos em um portfólio exploratório.

Ao fazer uma normalização de razões P10/P90, permitimos que a incerteza seja consistente em prospectos que se encontrem em mesmas condições geológicas e exploratórias, do ponto de vista da qualidade e quantidade de dados.

Ao avaliar uma oportunidade estamos avaliando uma amostra (*prospecto/lead*) e estamos assumindo que ela é bem representada por características de uma população (*play*), conforme brevemente descrito em Capen (1992) e Liner (2016). Estamos assumindo, portanto que as distribuições e, portanto, as características de incerteza dos parâmetros são bem representadas por esta população.

Curiosamente, alguns parâmetros apresentam esta distribuição mesmo em escala menos representativa, conforme pode ser visto na distribuição de porosidades do poço 2-ANP-0001-RJS (Figura 2.10), o que por si só também não suporta o uso desta distribuição para representar a média de toda a área do prospecto mas ilustra a variedade de ocorrência em diferentes escalas mas ajuda a explicar também por que utilizar esta distribuição para representar a faixa de incerteza para a média deste parâmetro.

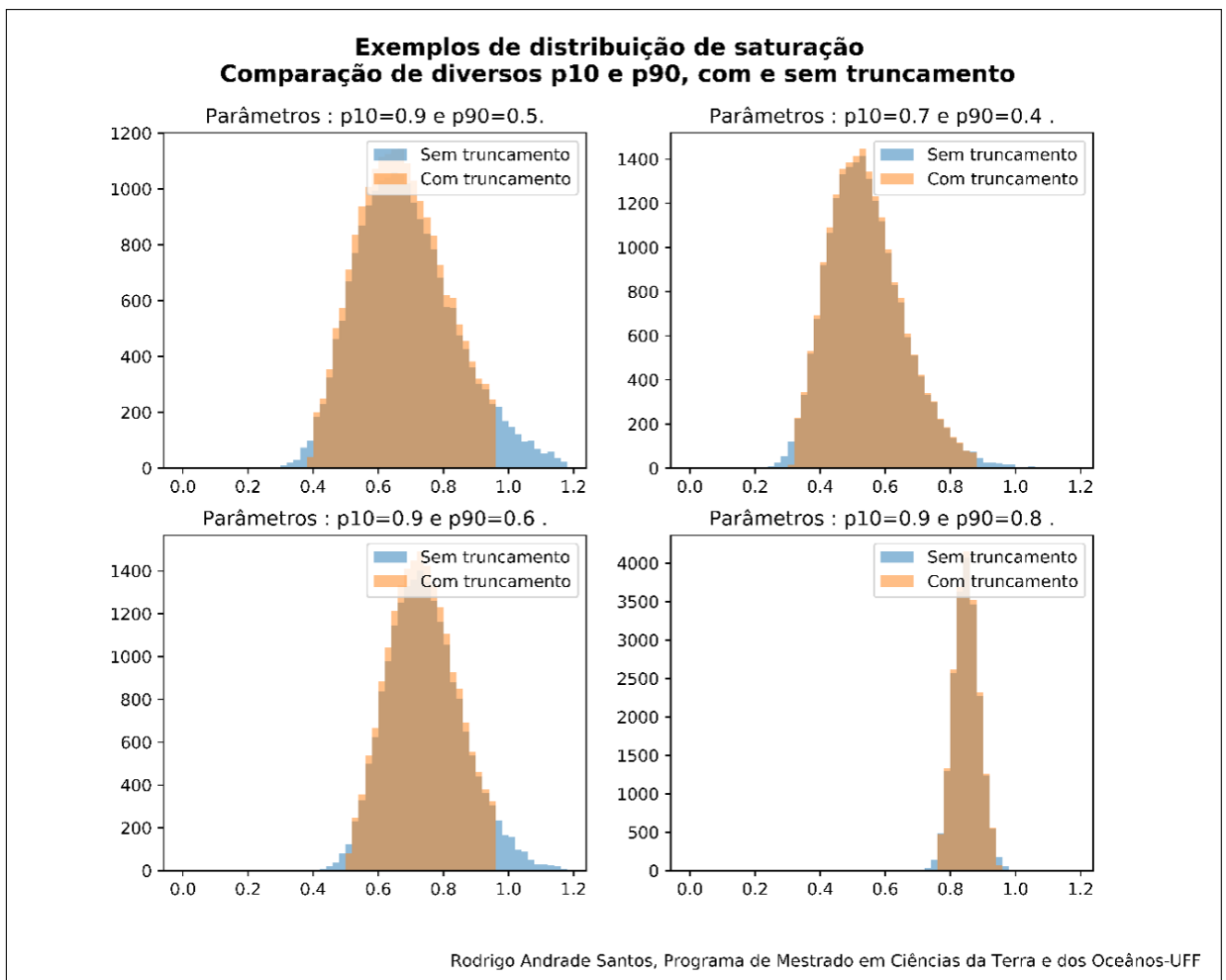


Figura 2.9: Exemplos de distribuições lognormais truncadas em p99 e p01. Foram utilizados valores típicos de saturação aplicados em prospectos.

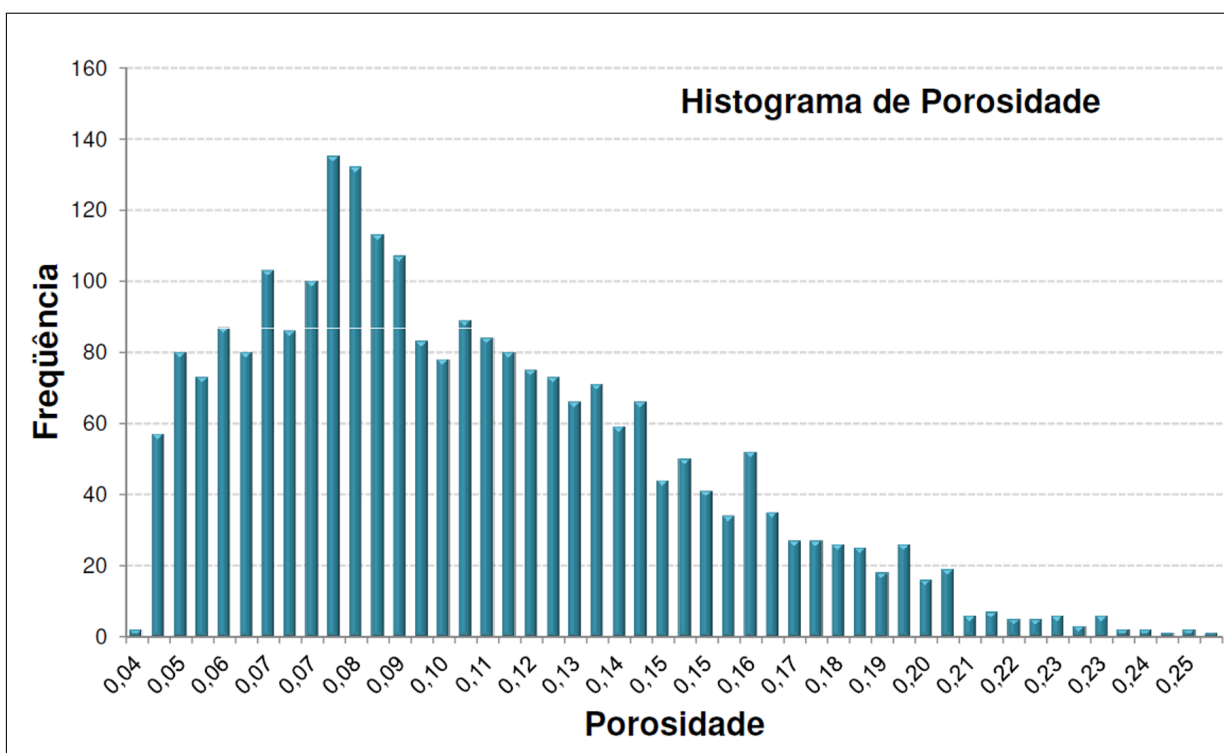


Figura 2.10: Exemplos de distribuições de porosidade do poço 2-ANP-0001-RJS, disponibilizadas na apresentação técnica da ANP da primeira rodada de Partilha de produção do Pré-Sal. Fonte: Petersohn and Abelha (2013).

2.2 Avaliação Geológica de Prospectos

Conforme comentado na seção anterior estimar os recursos de um prospecto é parte crucial do processo de exploração no setor de óleo e gás. A avaliação geológica normalmente é dividida em duas etapas (Costa et al., 2013). Estas são a avaliação de Risco Geológico (que consiste na possibilidade de se encontrar uma acumulação produzível), e a Avaliação Volumétrica (que consiste na estimativa de que faixa de volumes de hidrocarbonetos serão encontrados caso as premissas da avaliação sejam satisfeitas).

Apesar da análise econômica na indústria de petróleo ser feita em prospectos, os estudos de prospecção começam numa escala muito mais ampla (Otis and Schneidermann, 1997). A Figura 2.11 ilustra elementos na escala hierárquica de avaliação dos exploracionistas conforme descrito por Magoon and Dow (1994). Poderíamos ampliar ainda mais o escopo da Figura 2.11 para bacias sedimentares (múltiplos sistemas petrolíferos) e até mesmo para o estudo de tectônica de placas (múltiplas bacias sedimentares). Entretanto, isto tornariam a discussão desta seção muito longa e possivelmente infrutífera, por isso iniciaremos com o sistema petrolífero.

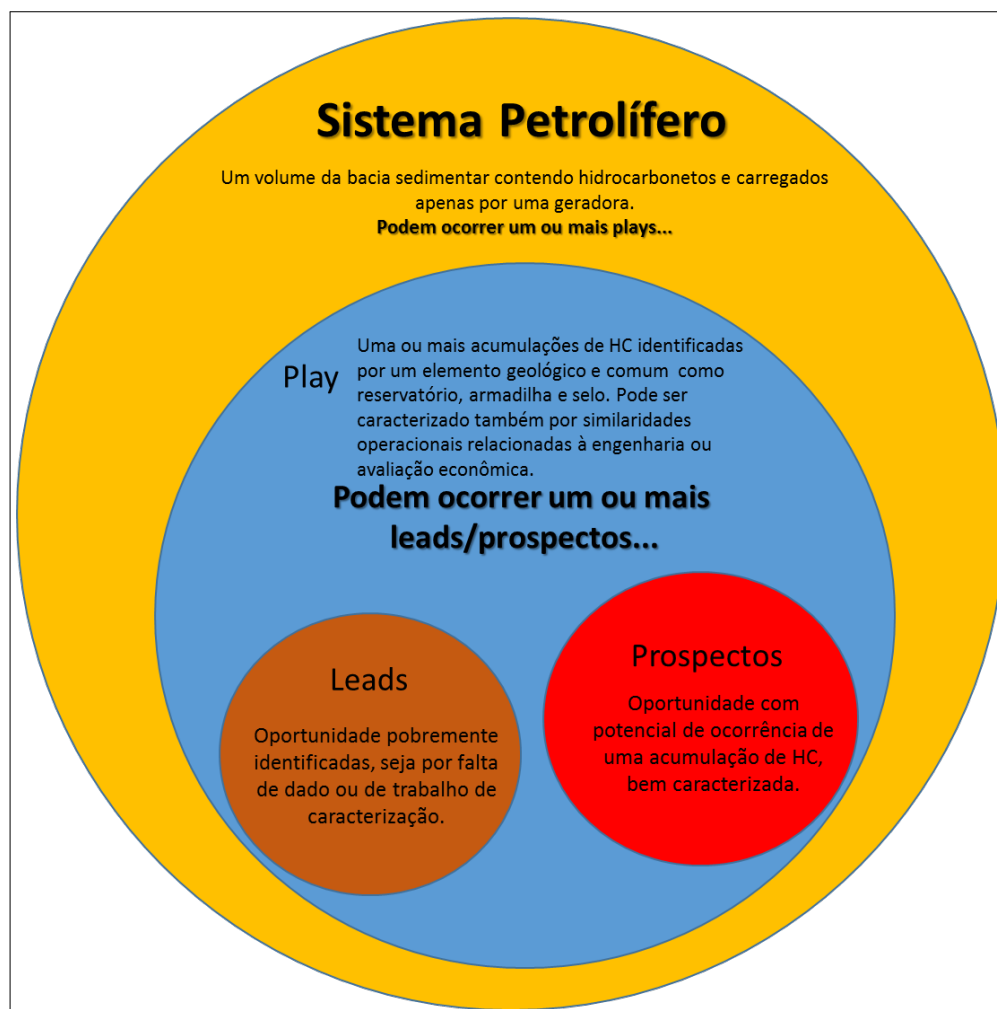


Figura 2.11: Caracterização hierárquica de elementos de um sistema petrolífero criada a partir dos conceitos de Magoon and Dow (1994) e Otis and Schneidermann (1997).

De acordo com Magoon and Dow (1994), um sistema petrolífero descreve as relações genéticas entre um *pod* de geração e as acumulações de óleo e gás resultantes. A primeira referência formal a este termo foi publicada em Dow (1978) baseado em relações óleo-óleo e óleo-geradora.

Diversas referências foram publicadas desde então. Sendo Magoon (1988), uma fonte básica para aqueles que gostariam de obter um trabalho pioneiro de compilação e resumo histórico

até 1988 (focado nos sistemas petrolíferos dos EUA). Magoon and Dow (1994) um trabalho de referência utilizado até hoje na indústria. Otis and Schneidermann (1997) seria indicado para uma abordagem mais concisa sobre o tema.

Conforme mencionado já em Magoon (1988), um sistema petrolífero inclui todos aqueles elementos e processos geológicos que são essenciais para uma acumulação de petróleo e/ou gás ocorra. Os elementos básicos de um sistema petrolífero são uma rocha geradora de petróleo, caminho de migração, rocha reservatório, selo e armadilha.

Os processos geológicos que relacionam cada um desses elementos básicos devem ser corretamente colocados no tempo e no espaço de modo que a matéria orgânica na rocha geradora possa ser convertida em uma acumulação de petróleo (Otis and Schneidermann, 1997). Um exemplo simples de acumulação em um sistema petrolífero pode ser encontrado na Figura 2.12.

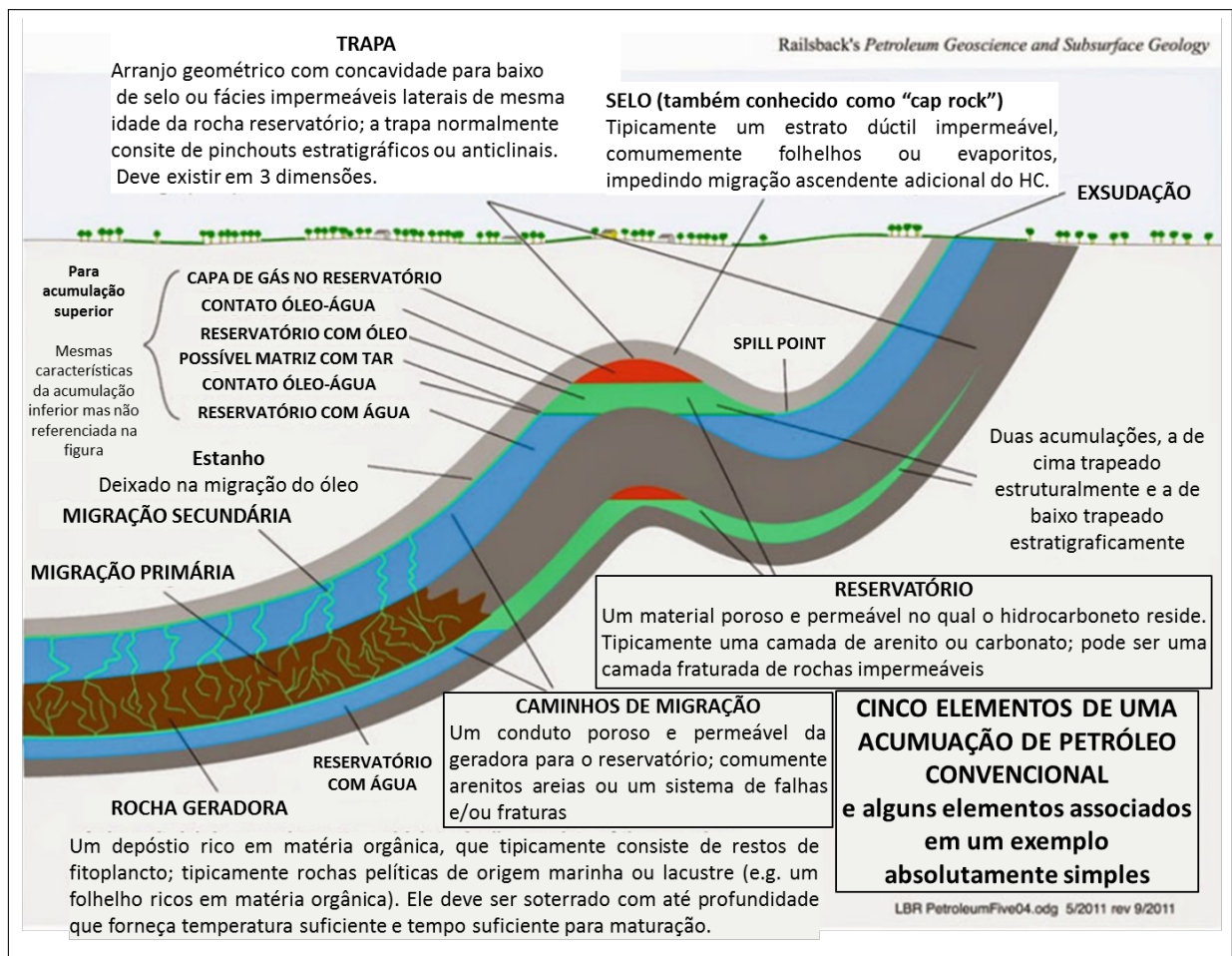


Figura 2.12: Um sistema petrolífero inclui uma geradora madura, caminho de migração, rocha reservatório, armadilha e selo em conjunção de tempo adequada. O tempo relativo de formação desses elementos e os processos de geração, migração e acumulação são críticos para a acumulação e preservação de hidrocarbonetos sendo estes abarcados pelo sincronismo. Fonte: Modificado de Railsback (2011).

Ao estudar potenciais armadilhas de hidrocarbonetos e acumulações com características semelhantes (tipo de estrutura, idade do reservatório, um selo regional etc) estamos estudando as características de um *play* (Magoon and Dow, 1994). *Plays* são limitados e definidos estratigraficamente pela mudança significativa de um de seus elementos do sistema petrolífero (e.g variação importante do selo, a variação no ambiente deposicional/fácies do reservatório, mudança da idade da rocha geradora etc).

Durante uma campanha exploratória, ao identificar uma possível estrutura, temos inicialmente um *Lead*. O *lead* nada mais é que uma possível trapa que não foi estudada adequadamente e

demanda mais dados, maior detalhamento da interpretação ou mais estudos. Após ser maturada, a oportunidade pode (ou não) se tornar um prospecto (Schlumberger, 2019).

Um prospecto é qualquer trapa perfurável em um *play* (Rose, 2001). Uma definição de prospecto um pouco mais específica pode ser adaptada de Schlumberger (2019): Uma área e profundidade exploratória não perfurada na qual foi predita ocorrência de hidrocarbonetos em quantidade econômica. Um prospecto é comumente uma anomalia, como uma estrutura geológica ou amplitude sísmica, que é recomendada por exploracionistas para perfurar um poço.

2.2.1 A Avaliação de Risco Geológico de um prospecto

Normalmente a primeira avaliação feita em uma oportunidade é a avaliação de risco. Utilizaremos aqui a definição de Otis and Schneidermann (1997) e Rose (2001). Eles consideram risco a possibilidade de não existir uma acumulação de hidrocarbonetos produzível, que traduzimos em risco geológico.

Uma acumulação produzível é aquela capaz de gerar um fluxo estabilizado de hidrocarbonetos em teste de formação Otis and Schneidermann (1997). Esta definição é interessante pois exclui o caso de reservatórios fechados com baixa permeabilidade, óleos extremamente biodegradados ou *flags* e *shows* de hidrocarbonetos (pequenas ocorrências muito restritas).

O risco geológico é avaliado considerando a probabilidade de que os elementos do sistema petrolífero ocorram em sintonia (Otis and Schneidermann, 1997; Rose, 2001). As instituições (empresas, universidades etc) variam em relação a quantidade de elementos que utilizam para a caracterização do sistema petrolífero, mas a maioria utiliza uma combinação daqueles descritos na Figura 2.13.

Diferentemente de Magoon (1988), para Otis and Schneidermann (1997) são utilizados quatro fatores de risco a serem avaliados no sistema petrolífero para que um *play*/prospecto tenha sucesso. (1) Presença de geradora madura com capacidade de migrar óleo para o reservatório (Carga), (2) Presença de rocha de reservatório, (Reservatório), (3) Presença de uma armadilha ("Trapa"), (4) Dinâmica do *play* (Sincronismo ou *Timing*).

Existem instituições que utilizam apenas três elementos (carga, reservatório e "Trapa"), enquanto outras utilizam 5 elementos (reservatório, trapa, geração, migração e *timing*) e outras ainda mais, destrinchando a trapa chegando a 6 elementos do sistema petrolífero (reservatório, retenção, geometria, geração, migração e *timing*).

Tome por exemplo o macro elemento carga que diz respeito a possibilidade de ocorrer a geração em uma cozinha, do óleo gerado migrar para o prospecto de forma a preencher (ainda que parcialmente) a trapa.

A carga é o elemento de caráter mais regional num sistema petrolífero. A simples ocorrência de acumulações ou presença de hidrocarbonetos próximo a uma área de avaliação reduz significativamente o risco de carga de um prospecto (Otis and Schneidermann, 1997). Estes 3 sub elementos, (geração, migração e sincronismo) podem ser analisados em conjunto para chegar a um valor de risco ou analisados separadamente para depois serem combinados.

É importante ressaltar que para aqueles que trabalham com modelagem numérica de sistema petrolífero carga pode significar o volume de hidrocarbonetos migrado para o reservatório modelado, uma vez que a premissa básica do modelo numérico é que há uma geradora.

A correta estimativa de cada parâmetro destrinchado ou analisado conjuntamente deveria ser a mesma, independentemente do número de parâmetros utilizados, conforme esquematicamente representado na Figura 2.13.

A faixa de incerteza (ou de certeza) razoável na avaliação de cada parâmetro deve ser balizada não só na percepção geológica (traduzida num modelo) dos intérpretes, mas também balizada na quantidade de dados disponíveis que permitiram chegar até aquelas conclusões. A Figura 2.14 extraída de Costa et al. (2013), fornece um excelente resumo para este guia de avaliação.

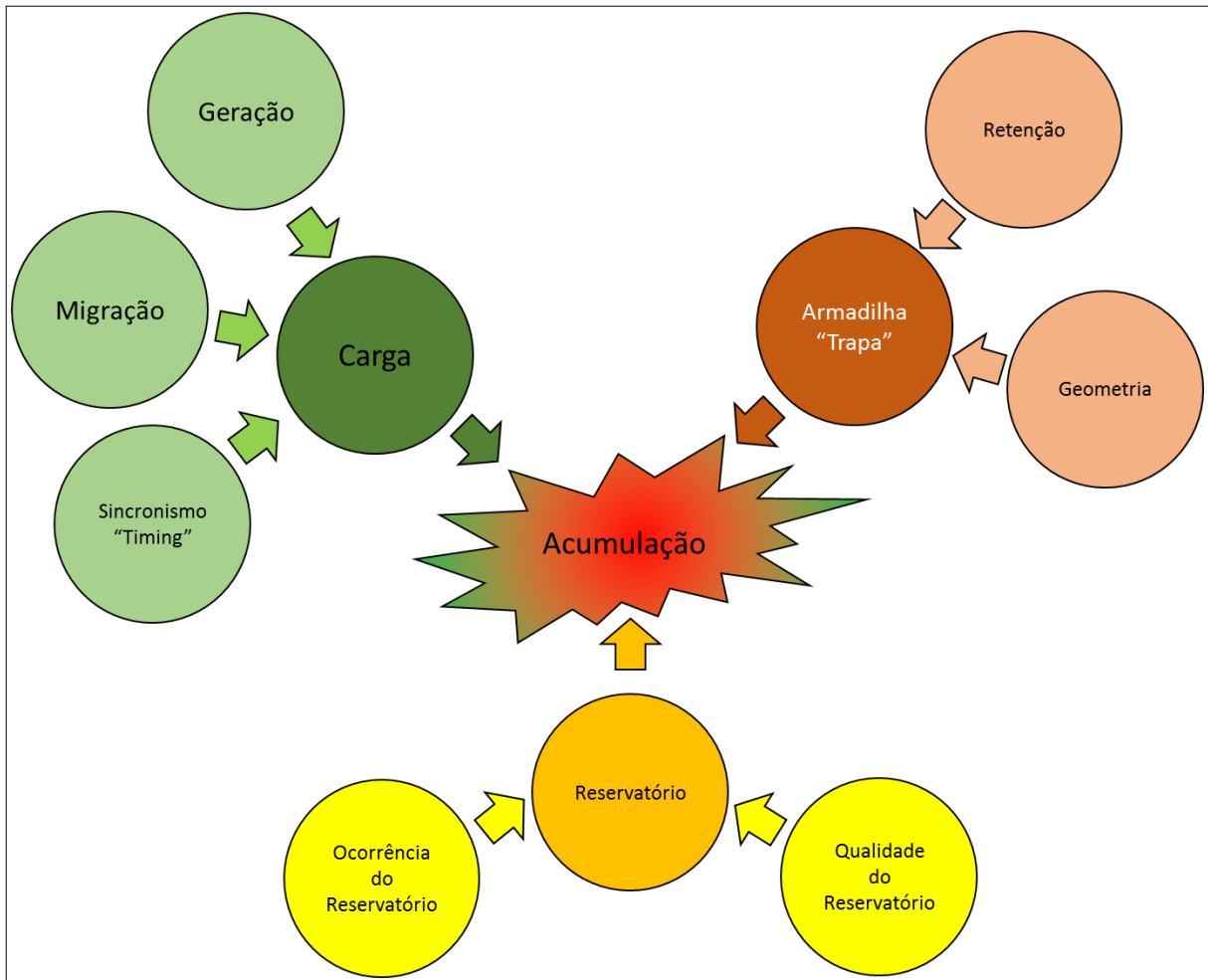


Figura 2.13: Elementos do Sistema Petrolífero necessários para que ocorra uma acumulação de hidrocarbonetos. Os três fatores de risco primários seriam Carga, Armadilha e Reservatório, que podem ser desdobrados em outros elementos conforme ilustrado.

A probabilidade de Sucesso

Para que uma acumulação ocorra, todos os elementos devem ocorrer conjuntamente (Magoon, 1988; Magoon and Dow, 1994; Otis and Schneidermann, 1997; Costa et al., 2013). Matematicamente isso significa que a probabilidade de sucesso dos prospectos é encontrada pela multiplicação dos seus elementos um a um (seja lá qual for a quantidade). Se um deles for zero não haverá uma acumulação.

$$PS_{Prospecto} = PS_{Reservatorio} * PS_{Migracao} * PS_{Geracao} * PS_{Geometria} * PS_{Retencao} * PS_{Sincronismo} \quad (2.4)$$

Ao atribuir a chance de sucesso (ou risco) a um prospecto estamos na verdade estimando uma métrica de comparação com outras oportunidades que só é verificada se testada estatisticamente. A interpretação por trás desta afirmação é que ao perfurar uma oportunidade estamos fazendo uma verificação binária sobre o resultado do prospecto entre sucesso e fracasso e, portanto, idealmente teríamos uma análise de risco polarizada em 0 para os prospectos fracassados e 1 para os prospectos com sucesso (Shell, 2014). Portanto, a avaliação do valor final não deve ser encarada como uma simples multiplicação, mas interpretada sob a luz de todo portfólio disponível e idealmente revisada sempre pelo mesmo grupo de avaliadores.

Contudo, teoricamente se perfurarmos muitas oportunidades semelhantes àquele prospecto,

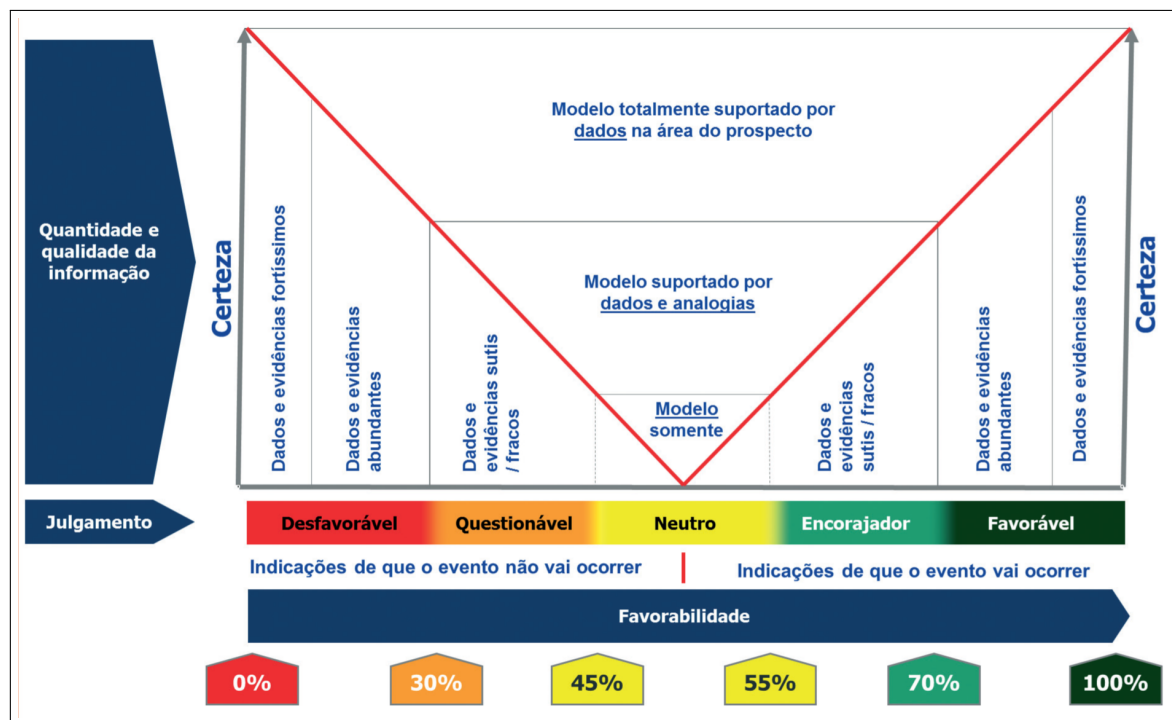


Figura 2.14: Matriz de adequação para avaliação da favorabilidade de cada elemento do sistema petrolífero, com base na quantidade e qualidade da informação disponível e no julgamento feito pelo avaliador sobre essas informações. Fonte: Costa et al. (2013).

com a mesma quantidade de informações, teríamos a taxa de sucesso igual ao fator de chance.

A PS tem forte impacto econômico na avaliação de um projeto. Isto significa que além de servir como elemento de comparação entre oportunidades, é também importantíssimo na avaliação de viabilidade econômica do projeto.

Otis and Schneidermann (1997) fornece também uma excelente discussão sobre as faixas de valores de chance de sucesso esperados para cada tipo de prospecto descrito a seguir:

- Um risco muito baixo (PS entre 0,5 e 0,99). Todos os fatores de risco são favoráveis. Esta categoria está associada a poços que testam *plays* comprovados em prospectos adjacentes (<5 km) a produção existente.
- Baixo risco (PS entre 0,25 e 0,5). Todos os fatores de risco são encorajadores a favoráveis. Esta categoria está associada a poços que testam *plays* comprovados perto de produção existente (5-10 km).
- Risco moderado (PS entre 0,125 e 0,25). Dois ou três fatores de risco são favoráveis - um ou dois fatores são encorajadore(s) ou neutro(s). Esta categoria está associada com poços testando novos *plays* em bacias com produção ou *plays* comprovados longe de (> 10 km) uma produção existente.
- Alto risco (PS entre 0,063 e 0,125). Um ou dois fatores de risco são encorajadores - dois ou três fatores são neutros ou encorajando a neutralidade. Esta categoria é frequentemente associada com poços testando novos *plays* em uma área não perfurada.
- Risco muito alto (PS entre 0,01 e 0,063). Dois a três fatores de risco não são melhores do que neutro, com um ou dois fatores questionáveis ou desfavoráveis. Esta categoria é geralmente associada com *plays* totalmente desconhecidos em áreas de fronteira.

Esta avaliação é representada na Figura 2.15. Note que Otis and Schneidermann (1997) utilizavam 4 elementos de risco, enquanto aqui descrevemos 6 (segundo Costa et al. (2013)). Entretanto, conforme mencionado no início do capítulo, desde que os elementos sejam corretamente normalizados é possível que se convirja a esta escala.

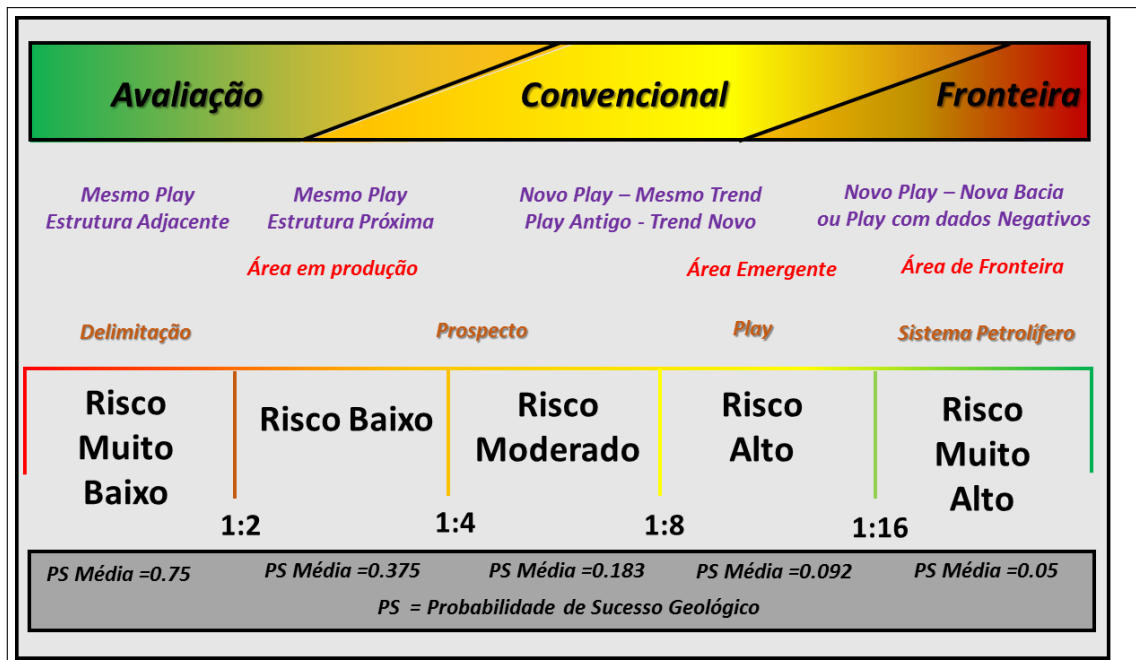


Figura 2.15: Faixas de chance de sucesso consideradas razoáveis para tipos diferentes de prospectos. Fonte: Modificado de Otis and Schneidermann (1997)

É importante ressaltar que a percepção de risco de um *play* ou prospecto pode mudar com o tempo devido novos modelos, novos poços etc. Um exemplo próximo a nossa realidade é o *play* Pré-Sal na Bacia de Santos, até o primeiro poço ser perfurado era uma área de fronteira. Entretanto, depois dos seus primeiros resultados obteve uma redução tão grande no risco que impulsionou uma variação contratual do governo (que não me cabe julgar se foi correta ou não) de concessão para partilha de produção.

A percepção de Risco sobre uma bacia pode mudar também com a aquisição de novos dados, isto significa que dados diferentes podem mostrar viabilidade (ou inviabilidade) de ocorrência de um ou mais elementos do sistema petrolífero (Costa et al., 2013). Por exemplo, uma sísmica 3D pode revelar anomalias de amplitude, arquiteturas que se assemelham a canais e leques siliciclásticos. Isto pode aumentar a chance de sucesso de reservatório. Contudo, a mesma sísmica pode mostrar também que o fechamento estrutural anteriormente interpretado não existia, piorando assim a chance de trapa.

Conforme ilustrado na apresentação da Shell (2014), o esforço mais importante para melhorar o desempenho exploratório é, com certeza, a reconstrução de uma abordagem exploratória baseada em *plays*. Isso poderia conduzir a uma maior identificação de oportunidades através de melhor compreensão dos sistemas petrolíferos, identificar áreas mais prospectivas, melhor risco, melhor polarização de POS e campanhas exploratórias mais eficientes.

A avaliação numa escala de *play* traz o benefício de avaliar prospectos de forma correlacionada e trazer sinergia a projetos. Por outro lado ao ter o correto entendimento sobre um determinado *play* é possível se posicionar rapidamente frente a mudanças relativas ao resultado de um poço em diferentes posições de uma bacia tomando vantagem da rápida variação de chance de sucesso e portanto percepção de risco e economicidade das oportunidades.

2.2.2 A Avaliação Volumétrica de Prospectos

As instituições variam entre a utilização de métodos de avaliação volumétrica de prospectos com abordagens determinística ou probabilística (sendo esta última utilizada neste trabalho). A decisão de que abordagem de cálculo uma equipe utilizará pode ser feita com base não só em seus padrões vigentes, mas na fase de avaliação da oportunidade/alvo em questão (Rose, 2001). Em geral a fase exploratória tem muita incerteza, por isto a volumetria normalmente é feita com métodos probabilísticos e à medida que o processo de delimitação e produção da acumulação evolui para o método determinístico, baseado em modelos de reservatório, ganha espaço no dia-dia de trabalho conforme ilustrado na Figura 2.16.

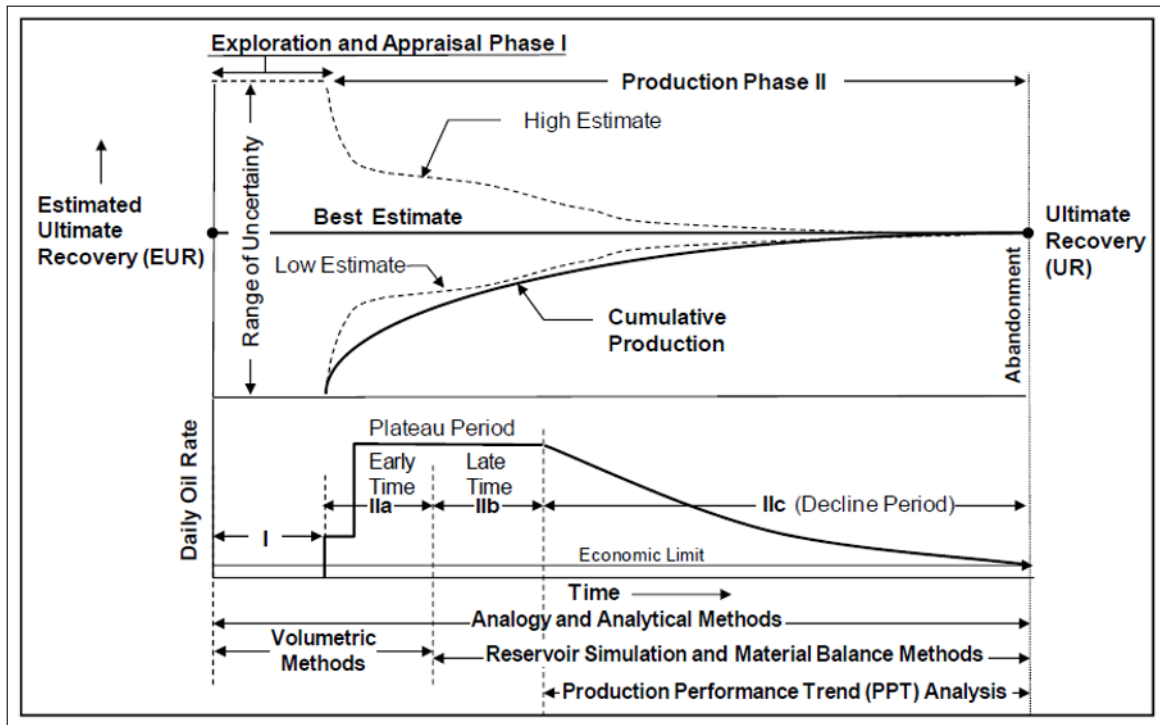


Figura 2.16: Mudança na incerteza e método de avaliação num ciclo de vida típico de projeto de E&P. Fonte: (SPE, 2007).

O uso de métodos determinísticos pressupõe conhecimento e precisão das variáveis envolvidas na estimativa dos recursos de uma oportunidade, de forma que um ou mais cenários estáticos, representem, de forma significativa, as possibilidades que a história geológica reservou à área particular do prospecto.

A abordagem determinística apresenta a vantagem do modelo ser facilmente visualizável, e as premissas de engenharia serem construídas com base num modelo em que se possa fazer algum controle de qualidade. Conforme mencionado em Rose (2007), existem alguns motivos para o método determinístico persistir mesmo na fase de exploração:

- O determinismo representa a característica ortodoxa do setor de óleo e gás e a mudança principalmente no corpo técnico mais experiente pode encontrar alguma resistência;
- O corpo gerencial que não seja familiarizado com o negócio pode se sentir mais confortável ao tomar decisões baseadas em um único valor e não uma faixa de probabilidades.
- Os processos de auditoria e contabilidade tem dificuldade de lidar com faixas probabilísticas de volumes e cenários;

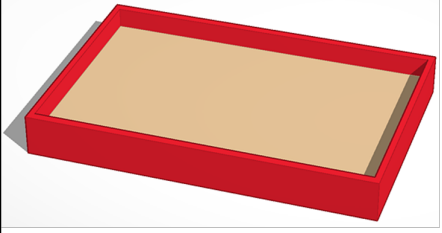
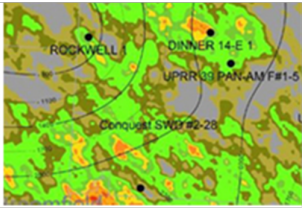
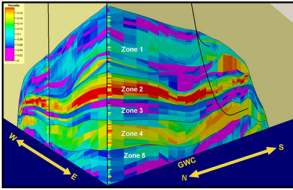
$k=f(x)$	$k=f(x,y)$	$k=f(x,y,z)$
1D Método probabilístico inicial 	2D 	3D Método determinístico 
<i>Valores Médios</i>	<i>Mapas de distribuição de Valor Médio vertical</i>	<i>Modelo de Reservatório 3D</i>

Figura 2.17: Variações de dimensionalidade em relação a abordagem volumétrica. Do modelo de valores médios 1D utilizado tipicamente no método probabilístico até o modelo 3D normalmente utilizado no método determinístico. O modelo 2D é pouco explorado e normalmente é obtido a partir de extração de atributos do modelo 3D.

- Os valores utilizados no caso base do método determinístico por vezes são correspondentes a um cenário otimista (e.g. último contorno estrutural fechado de um prospecto) se comparado à modelagem probabilística melhorando a percepção sobre um prospecto.

Apesar de ser uma ideia comum entre os geocientistas, a utilização de modelos geológicos estáticos 3D não significa necessariamente utilização do método determinístico. Mesmo em um modelo de reservatório estático é possível utilizar distribuições para algumas de suas variáveis, e neste ponto o modelo estático utilizado funciona apenas como uma 'semente' para um sub-método probabilístico (Azevedo and Soares, 2017).

Ao lidar com incerteza expressamos a realidade do processo de avaliação e suas limitações. Trabalhar com cenários é especialmente útil para empresas e investidores entenderem as consequências de encontrar cenários inferiores ao médio (comumente utilizado no método determinístico) e o potencial retorno e exposição financeira ao negócio nestes casos (Rose, 2007). A incerteza é inerente à amostragem e às variações geológicas em escala vertical e horizontal nos estágios iniciais da exploração.

Em face a incerteza, a utilização de métodos probabilísticos é aparentemente mais adequada, de forma que um somatório de milhares cenários médios abrangerá com alguma confiabilidade o modelo real conforme proposto por (Rose, 2001) e ilustrado na Figura 2.17.

A aplicação de métodos probabilísticos na volumetria de prospectos aplica, em geral, simulações de Monte-Carlo que são na verdade uma classe de métodos estatísticos que se baseiam em imensas amostragens de valores médios aleatórios para obter resultados numéricos. Isto é, repetindo sucessivas simulações um elevado número de vezes (milhares), podemos calcular probabilidades heurísticamente, tal como se de fato fossem registrados os resultados reais de realizações de jogos de cassino (daí o nome, vem da região de Monte Carlo em Mônaco, famosa por seus cassinos), Metropolis (1987).

Note que a utilização de valores médios expeditos (comumente chamado de modelo de caixa de areia) não significa necessariamente a utilização do método probabilístico, é possível ter um cenário estático determinístico simplificado baseado em valores médios, ainda que seja simplificado demais para tomar decisões.

Conforme discutido em SPE (2007), na teoria (e na prática) não existe diferença entre o método determinístico e o probabilístico quando o método probabilístico aborda apenas o valor esperado das distribuições, matematicamente definido por sua média. Os valores esperados se comportam como valores determinísticos considerando que o valor esperado de uma soma é igual à soma dos valores esperados, e da mesma forma para produtos de fatores **não-correlacionados**.

Esta definição se mostra especialmente importante neste trabalho, uma vez que estamos aqui analisando justamente a correlação entre as variáveis de entrada que serão aplicadas em um produto.

A incorporação da incerteza em projetos enfrenta uma dificuldade cultural pois ela normalmente tem conotação negativa (Rose, 2007). Isto nem sempre precisa ser verdade, a incerteza pode estar relacionada ao fenômeno de observação ou da nossa incapacidade de prever eventos com precisão e pode estar associada também a cenários positivos/otimistas. Normalmente ao avaliar uma oportunidade a margem de confiança é dada num intervalo de 80% e por isso as estimativas de volume utilizadas se encontram entre o percentil 90 (p90) e percentil 10 (p10) (Capen, 1992).

Ao lidar com distribuições a melhor estimativa é uma medida de tendência central (média), também conhecido como valor esperado, a estimativa conservadora é dada pelo p90 e a estimativa otimista é o valor de P10 (Rose, 2001).

Frequentemente a pdf (*probability density function*) de uma propriedade utilizada nestes cálculos é erroneamente interpretada. Por exemplo, a pdf que define a incerteza na estimativa da porosidade média, não é a mesma que a faixa real de valores de porosidade encontrados no reservatório.

A distribuição completa, neste trabalho truncada, de P99 (mínimo) a P01 (máximo) deve ser sempre visualizada para garantir que ela represente um intervalo válido de incerteza para os valores médios desse parâmetro (Capen, 1992). É importante salientar que devido à assimetria da curva lognormal que utilizamos nas estimativas dos parâmetros, o P50 não é equivalente a média e sim à mediana.

Ao avaliar a incerteza de um parâmetro devemos considerar a qualidade, abundância e representatividade dos dados com os quais é feita a avaliação e considerar também o modelo geológico em avaliação (Rose, 2001). Imagine por exemplo um *plug*, amostrado em um poço que não teve por objetivo testar as melhores fácies de um *upside* (e.g. visando um objetivo mais profundo), obtendo valores de porosidade de 2%, não podemos utilizar estes valores para representar todo o objetivo devido a sua falta de representatividade.

O aumento na abundância dos dados, normalmente, reduz a incerteza em relação a um objetivo mas não significa redução de incerteza em relação a outros plays/prospectos. O nível de incerteza para canais anastomosados perfurados por 2 poços possivelmente serão maiores do que a incerteza de arenitos eólicos perfurados por apenas um poço (dada a menor complexidade deste último sistema) o que significa que cada prospecto deve ser analisado de forma bastante particular. SPE (2007) fornece alguns valores de incertezas para parâmetros de avaliação de volumes podem ser encontrados na Tabela 2.1, mas ressalto que esta tabela deve ser utilizada com parcimônia.

O problema proposto aqui é o cálculo volumétrico (que tem equações determinísticas), a estas equações aplicaremos as variáveis estocásticas. Conforme SPE (2007), a equação 2.5 é utilizada para cálculo volumétrico de óleo *in place*, *Stock Tank Oil Initially In Place* (STOIIP)²:

$$STOIIP = \frac{VR * NTG * \phi * TF * (1 - S_w)}{Bo} \quad (2.5)$$

Onde:

- VR = O Volume de rocha (também conhecido como GRV *Gross Rock Volume*) normalmente é calculado com base na área média e espessura média (Gross Pay) equação (2.6), mas pode ser calculado também com base na soma de *voxels*³ de um volume sísmico em uma determinada faixa de amplitudes ou com base em um modelo análogo.

$$VR = Area \times GrossPay \times FG \quad (2.6)$$

² É comum encontrar também o termo *Stock Tank Oil Originally In Place* (STOOIP), ambas as siglas tem o mesmo significado

³Voxel: Elemento de volume, utilizado normalmente em dados 3D

Parâmetro	Faixa de Incerteza	Dados
Gross Rock Volume	+/-30%	Sísmica 3D
	+/-40%	Sísmica 2D
	+/-5%	Baseado em anomalias
Net-To-Gross	+/-20%	Perfil de Poço
Porosidade	+/-15%	Perfil de Poço
	+/-10%	Testemunho + Perfis
Saturação de HC	+/-20%	Perfil de Poço
Fator de Formação de óleo	+/-5%	Dados de PVT

Tabela 2.1: Valores típicos de incerteza em relação ao seu valor médio para parâmetros de avaliação volumétrica de prospectos. Fonte: SPE (2007)

Sendo que:

- Área = Área do prospecto em metros quadrados (m^2)
- *Gross Pay* = É a espessura de rocha reservatório(m).
- FG = É o fator geométrico utilizado como um escalar para descontar o efeito dos *pinch-outs*⁴ nos flancos da estrutura (adimensional).
- Porosidade (ϕ) = É o espaço poroso do preenchido por fluido nas seções consideradas reservatórios. A unidade é a fração percentual adimensional (%).
- *TrapFill* (TF) = É a quantidade de reservatório preenchido de hidrocarbonetos (%);
- *Net to Gross* (NTG) = É a proporção de reservatório na espessura total de rocha;
- Saturação (S_w) = Saturação de água. A unidade é a fração percentual de água no espaço poroso, e $1-S_w$ por vezes é definido como saturação de hidrocarbonetos S_{hc} , é adimensional (%);
- B_o = É o Fator de Formação de Óleo, gerado pela perda do volume de gás, aliado à diferenças de pressão e temperatura do óleo em condições de reservatório. Este, quando em condições de superfície, sofre redução significativa em seu volume.

Os valores obtidos na equação (2.5) dizem respeito a fração líquida de uma acumulação de óleo, a fração de gás associado, *Stock Tank Gas Initially In Place* (STGIIP). Esta fração é obtida no separador em superfície e é estimada pela equação 2.7:

$$STGIIP = STOIP \times RGO \quad (2.7)$$

Onde RGO é a razão Gás-Óleo. Os valores de RGO devem responder às conversões volumétricas de modo a expressar os volumes de gás finais separados do óleo em condições de superfície. Quando avaliamos uma acumulação de condensado e gás não-associado obtemos os volumes de gás a partir da equação (2.8):

$$STGIIP_{NA} = \frac{VR * NTG * \phi * TF * (1 - S_w)}{B_g} \quad (2.8)$$

⁴É o afinamento de uma camada em uma direção específica.

Onde os valores de B_g respondem pela expansão do gás (conhecido também como fator de formação de gás) devido a sua variação de condições de pressão e temperatura do reservatório para a superfície. Os valores obtidos na equação 2.8 dizem respeito à fração gasosa da acumulação, a fração de líquidos condensados, obtida no separador em superfície é estimada pela equação (2.9):

$$STOIIP_{cond} = RLG \times STGIIP_{NA} \quad (2.9)$$

Onde RLG é a razão líquido-gás. A partir dos valores de hidrocarbonetos *in place* é possível também calcular o volume recuperável estimado a partir do fator de recuperação (Fr) estimado normalmente pela equipe de engenharia para cada fluido:

$$Recuperavel_{oleo} = STOIIP \times Fr_{oleo} \quad (2.10)$$

$$Recuperavel_{cond} = STOIIP_{cond} \times Fr_{cond} \quad (2.11)$$

$$Recuperavel_{gas} = STGIIP \times Fr_{gas} \quad (2.12)$$

Ao trabalhar com fator de recuperação médio para uma acumulação inteira, mais variáveis aparecem no sistema como o grau de compartimentação da oportunidade e a análise da engenharia que definirá quantidade ótima de poços para o máximo aproveitamento econômico (Rose, 2001). Esta abordagem pode deixar pequenos compartimentos para traz e, portanto, abaixo do fator de recuperação médio "geológico" do prospecto.

As equações (2.10, 2.11 e 2.12) concedem valores volumétricos em termos de metro cúbico. Entretanto na indústria de petróleo utilizamos tradicionalmente as unidades barril para óleo e Tcf (*Trillion Cubic Feet*) para gás, ou Milhões de barris de óleo equivalente para representar acumulações de óleo e gás (em linha com SPE (2007)). As conversões são parte importante para comparação e *benchmarking* por isso salientamos os valores utilizados:

$$1m^3 = 6.28981bbl \quad (2.13)$$

$$1Tcf = 28316846592m^3 \quad (2.14)$$

$$1Tcf = 178.107MMBOE \quad (2.15)$$

A conversão da equação (2.15) utiliza valores de conversão de energia (calor) do óleo para o gás e portanto universaliza as conversões e possibilita a comparação de prospectos em diferentes partes do mundo. A avaliação de uma acumulação de gás sob a ótica da energia utiliza o seu potencial calorífero expresso em MMBtu (milhões de Btu- *British thermal unit*) que em última análise depende da composição do gás e seu volume.

Uma regra geral para esta conversão seria de que 1 Tcf (10^{12}) de gás equivaleria a 1 quadrilhão (10^{16}) de MMBtu (Rose, 2001). A conversão também pode ser feita de barril de óleo para também pode ser feita para MMBtu, conforme mencionado em SPE (2007) 1 barril de óleo de 35°API equivalem a aproximadamente 5.8MMBtu (veja a Figura 2.18 para a estimativa de conversão de outros API).

Estas conversões são na verdade uma simplificação porque mantida uma equivalência volumétrica a equivalência energética dependerá da composição do óleo e do gás em análise o que leva à variações (tipicamente próximas de 175MMBOE para 1Tcf).

É importante destacar entretanto, que o trabalho do geocientista normalmente tem fins econômicos. Por isto, o valor de conversão de gás para óleo equivalente pode (e deve) utilizar o valor do preço do óleo e do preço do gás a fim de representar o real valor econômico das acumulações em termos de óleo equivalente.

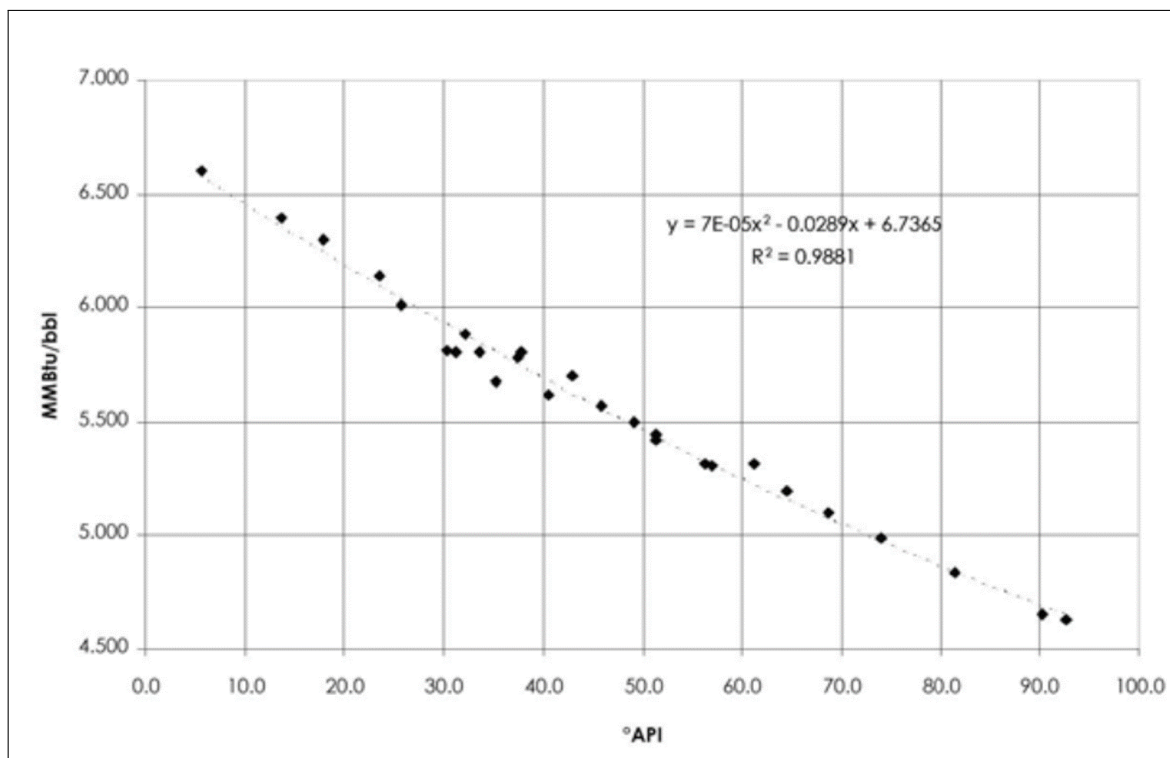


Figura 2.18: Correlação entre MMBtu de calor por barril de óleo *versus* o grau API do óleo e/ou condensado, a figura foi publicada com base em diversas fontes e talvez a não padronização das proporções de gás e/ou medidas tenha causado as descontinuidades observadas. Fonte: SPE (2007).

Neste momento por exemplo (02/03/2019), o preço do gás assume valores aproximados de US\$3/Mbtu nos EUA, US\$8/Mbtu no Brasil e US\$7/Mbtu na Europa. Esta variação torna evidente que ao contrário do óleo, o gás não é uma commodity global (ou seja seu preço é alvo de larga variação no mundo). Enquanto o preço do óleo assume valores que variam de US\$50/bbl a US\$55 por barril a depender da sua composição (e muito pouco em relação ao seu local de produção).

A conversão econômica pode ser ainda mais complicada. Se considerarmos que o hidrocarboneto produzido na acumulação só será efetivamente explorado após todos os investimentos de delimitação e instalação de infraestrutura de produção serem feitos. Uma avaliação realista requer uma projeção de preços para ambos os produtos (gás e óleo).

Neste trabalho não abordaremos a análise econômica do negócio. Como as propriedades dos prospectos são objeto da correlação de variáveis, faremos uma breve descrição de cada uma das propriedades envolvidas das equações 2.5 e 2.8. Primeiro trataremos das variáveis relacionadas à rocha e posteriormente as variáveis relacionadas a fluidos.

O Volume de Rochas (GRV *Gross Rock Volume*)

O GRV é o volume de rocha entre o topo do reservatório e base determinada por um contato entre hidrocarboneto e água ou, terminação do reservatório conhecido/postulado em uma trapa geológica (SPE, 2007). Ele é, frequentemente, o parâmetro mais influente na determinação da magnitude dos volumes de HC contidos ou potencialmente contidos numa armadilha (Capen, 1992). Em qualquer análise volumétrica de prospectos, é essencial calcular a melhor estimativa e a faixa de incerteza para a GRV com a precisão apropriada.

Existem pelo menos duas formas de calcular o volume de rocha (Rose, 2001): podemos definir o topo e base da oportunidade através de superfícies e obter o volume entre estas feições. Outra forma é definir a distribuição de área e distribuição de espessura do intervalo objetivo, de forma a

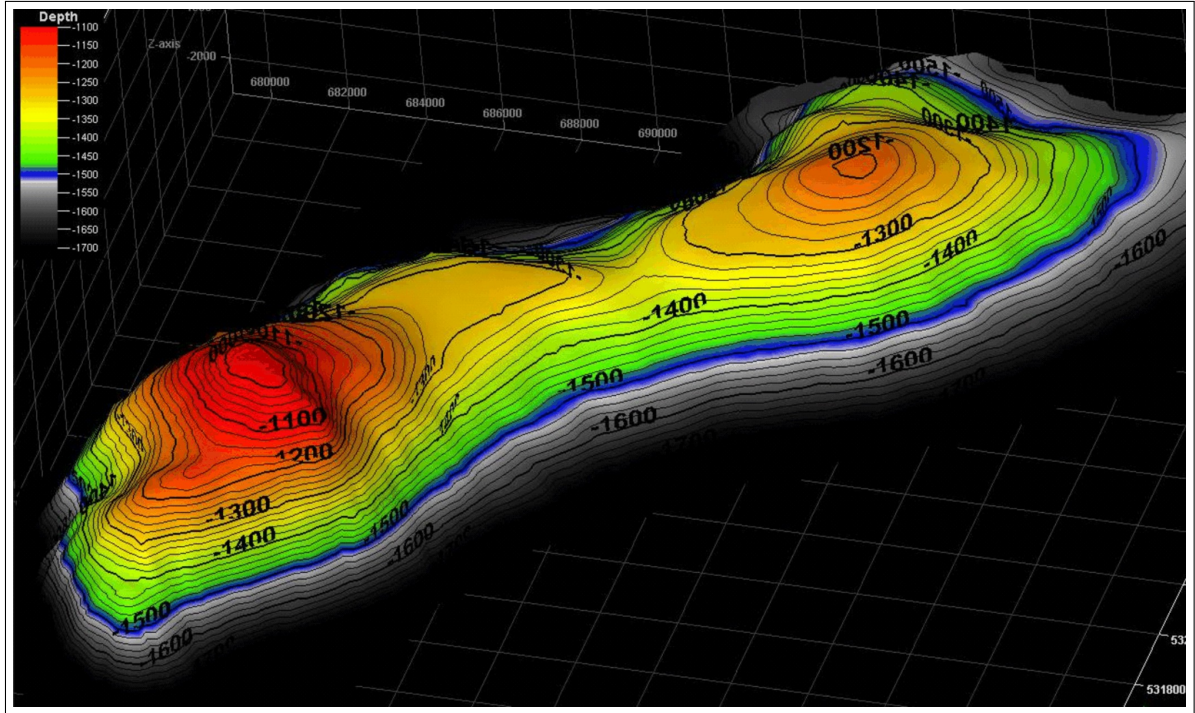


Figura 2.19: Figura ilustrativa do fechamento estrutural da estrutura Old Harry na Bacia de Newfoundland-Canadá, com área de aproximadamente 30kmx12km e 450m de coluna. Fonte: Corridor-Resources-Inc (2016).

determinar o volume de rocha pela interação das duas variáveis. No caso da metodologia de área e espessura é necessário por vezes também descontar o fator geométrico causado pelo *pinch-out* da coluna de óleo nos flancos da acumulação.

Em linhas gerais, quando temos maior riqueza de dados (grande número de poços e/ou dados sísmicos 3D por exemplo), a abordagem volumétrica é mais adequada. Entretanto, quando os dados são de uma malha esparsa de linhas, a utilização de distribuições de área e espessura parece ser mais adequada.

Na abordagem volumétrica baseada nas superfícies de topo e base, são mapeados o topo e a base do reservatório, ao longo do prospecto e então é feita uma integração numérica conforme a equação (2.16).

$$GRV = \sum_{i=xmin}^{xmax} \sum_{j=ymin}^{ymax} zT_{i,j} - zB_{i,j} \quad (2.16)$$

Onde zT e zB são as profundidades do topo e da base do reservatório.

É importante destacar que mesmo quando os dados são de boa qualidade, a incerteza inerente aos métodos geofísicos deixa espaço para uma análise de incerteza (Sheriff and Geldart, 1995). As equipes de interpretação tendem a utilizar variações no volume de rocha com base em variações no contato. Entretanto, é possível fazer esta avaliação também com base em diferentes modelos de velocidade, conforme ilustrado na Figura 2.20.

Por falta de melhor metodologia melhor, os parâmetros de área e espessura são estimados com base na sísmica, na interpolação de propriedades de poços e/ou feições de ocorrência que definem o prospecto.

A definição da área deve ter um critério claro, como uma curva de contorno representando um contato em prospectos estruturais (e.g. Figura 2.21) ou limite de ocorrência de reservatórios. Este limite pode ser dado por isócoras ou valores de anomalias sísmicas de amplitudes (Rose, 2001).

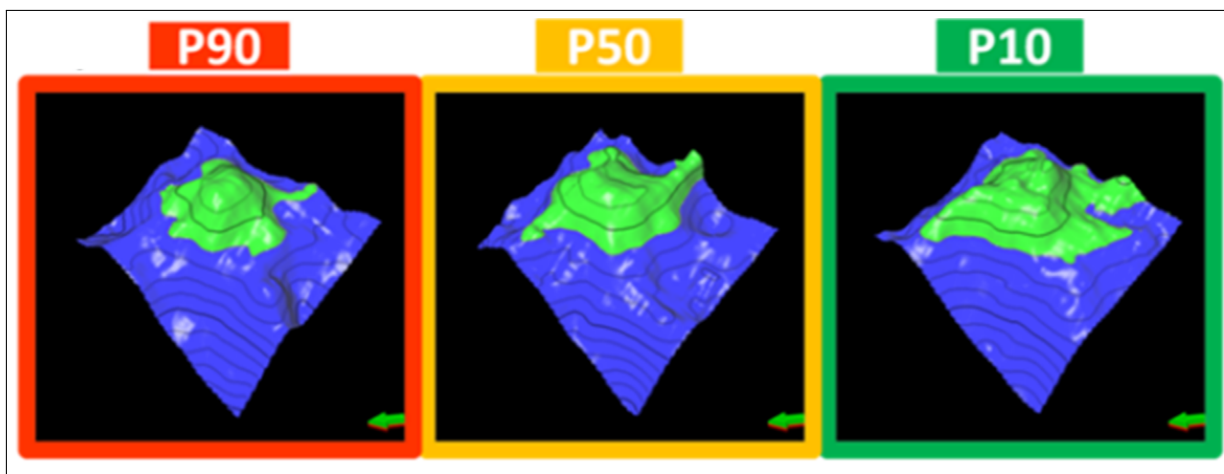


Figura 2.20: Figura ilustrativa definindo diferentes volumes de Rocha para um prospecto no Golfo do México com base em diferentes modelos de velocidade. Fonte: Osypov et al. (2013).

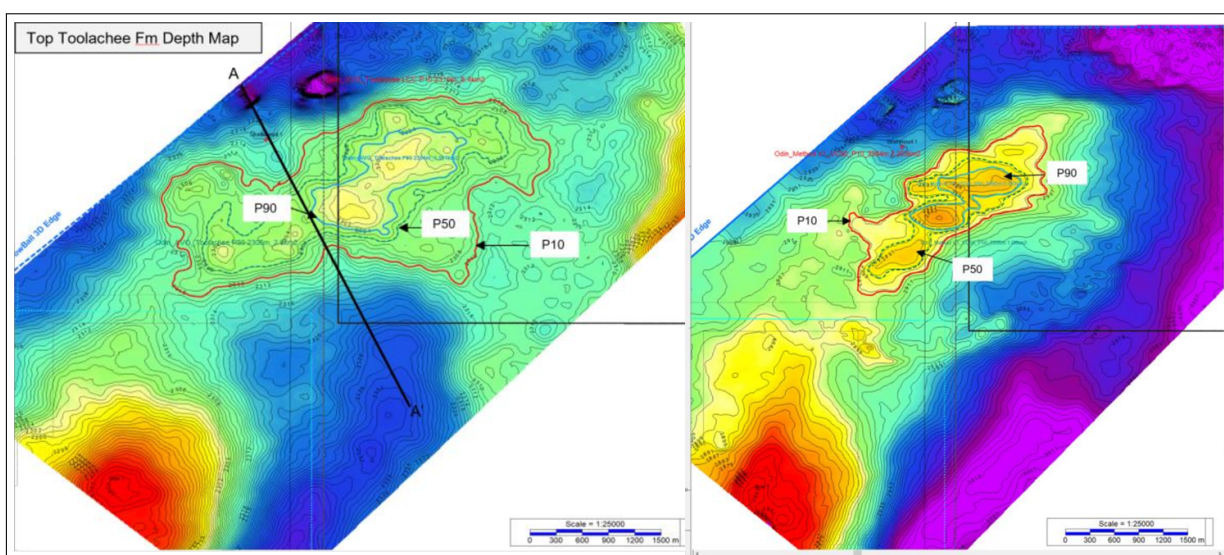


Figura 2.21: Exemplo de definição de percentis de áreas baseados em curvas de contorno com p10, p90 e p50 do prospecto Odin na Austrália. Fonte: Metgasco (2018).

De acordo com Otis and Schneidermann (1997), a área deve ser grande o suficiente a ponto de permitir que tenha óleo suficiente para fluir para o poço com um fluxo estável. Na outra ponta a área deve ter limites razoáveis, a ponto de englobar estruturas já perfuradas ou estruturas com fechamentos totalmente independentes.

De forma geral a área é o fator de maior dominância na incerteza volumétrica no GRV e consequentemente no prospecto. Uma exceção a esta regra é o caso onde ocorrem DHI (*Direct Hydrocarbon Indicator*). Neste caso o controle estrutural do prospecto passa a ser definido por variações de amplitude ou geometrias características como *flat spots*, e dessa forma a incerteza passa a ser consideravelmente menor.

A estimativa de *Gross Pay* (GP) deve contemplar todo o intervalo alvo, incluindo as feições não-reservatório (que serão posteriormente filtradas pelo parâmetro NTG). O valor de Gross-pay pode ser medido diretamente da sísmica utilizando os valores de topo e base do reservatório em (ou convertidos em) profundidade.

Alternativamente o GP pode ser estimado com base em poços de correlação que visaram o mesmo play. Apesar destas duas ferramentas poderem ser utilizadas de forma separada, devemos sempre utilizá-las de forma conjunta. Veja por exemplo o caso do prospecto Odin (da Figura

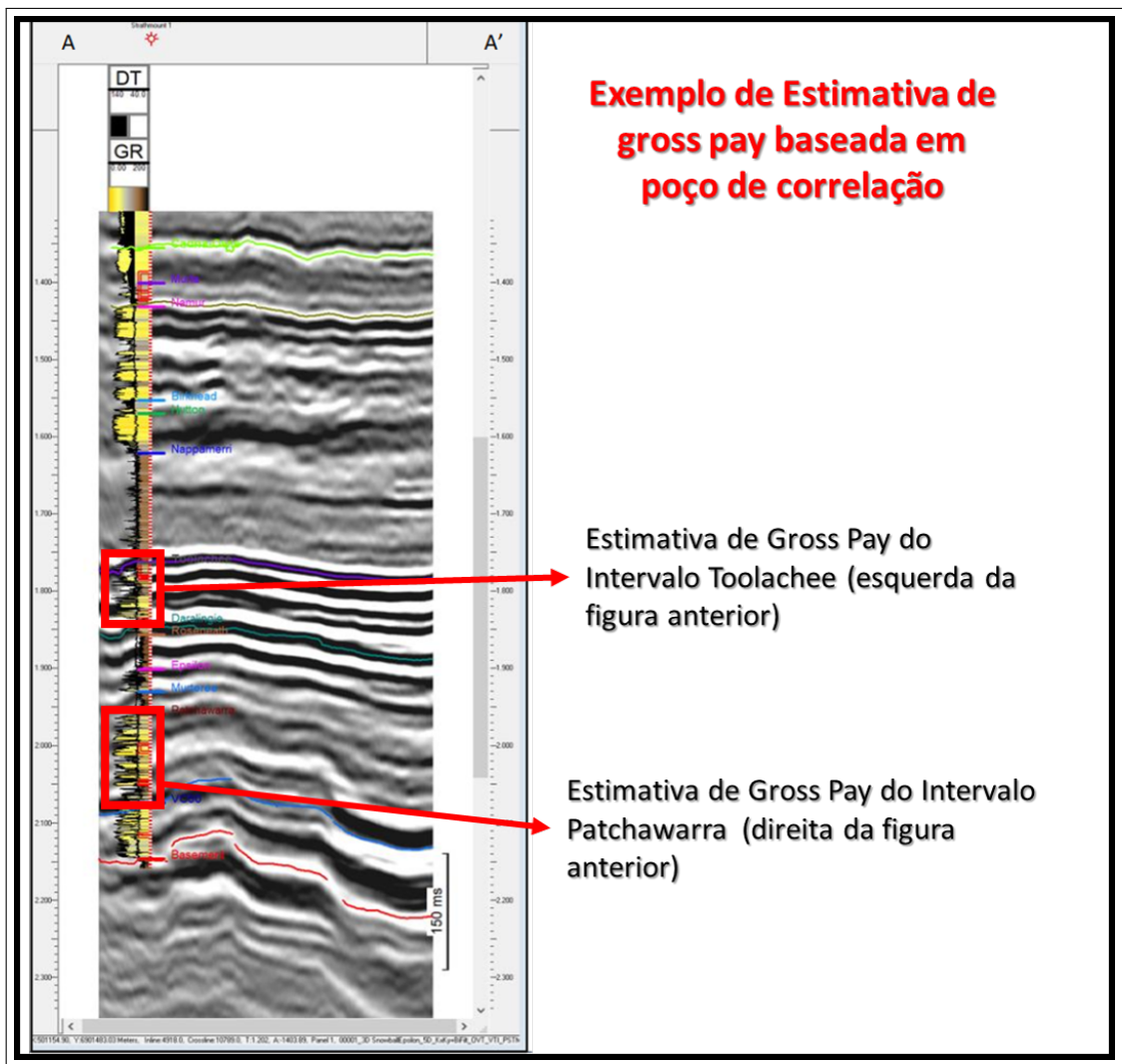


Figura 2.22: Exemplo de definição de estimativa de Gross Pay do prospecto Odin na Austrália. Fonte: Metgasco (2018).

2.21), naquela oportunidade foi utilizado o poço de correlação Strathmount-1 (Figura 2.22) que testou com sucesso gás nos Arenitos da Formação Patchawarra (160m de GP) e Tirrawarra (250m de GP) e a sísmica devidamente relacionada ao poço.

A incerteza na espessura pode ser oriunda de diversos fatores como: a confiança na velocidade sísmica de conversão (quando o GP é obtido diretamente da sísmica); a ausência de elementos que permitam boa definição de fácies sísmicas ou ; a incerteza na interpretação lito-estratigráfica dos poços. A equipe de interpretação deve tomar cuidado para utilizar valores de GP grandes o suficiente a ponto de contemplar todo o sistema de reservatório do objetivo e pequeno o suficiente a ponto de não englobar mais um intervalo que necessite de desenvolvimento da produção independente.

Um erro comum na estimativa de GP se dá quando trabalhamos com poços desviados e/ou prospecto com alto relevo. Deve-se lembrar que a estimativa de GP deve acompanhar o topo e base do objetivo conforme esquematicamente ilustrado na Figura 2.23, e não deve necessariamente seguir os valores medidos nos poços (ainda que eles sejam verticais).

Neste caso a espessura pode ser obtida com base na equação (2.17):

$$t = L \times \cos(\theta) \quad (2.17)$$

Onde L é a espessura vertical medida e θ o ângulo entre a vertical e a normal da camada. O

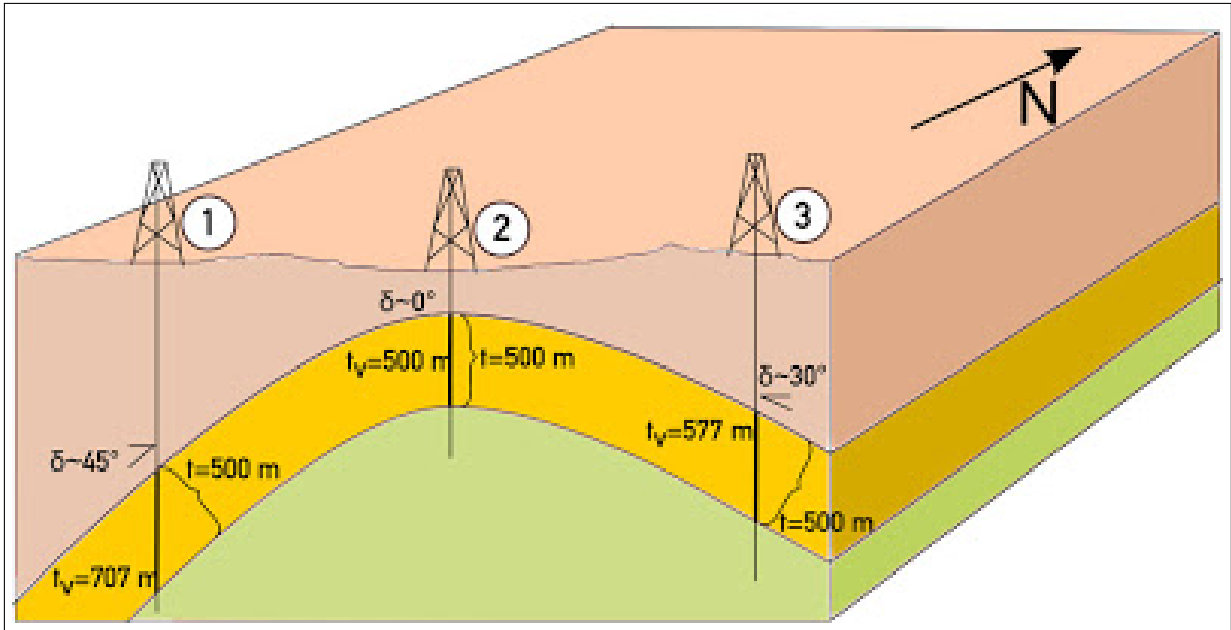


Figura 2.23: Figura ilustrativa de um anticlinal com mergulho variável e espessura verdadeira constante, e com espessura aparente distorcida.

exemplo pode ser ainda mais complicado se o poço for desviado e o intervalo objetivo também for desviado (conforme ilustrado na Figura 2.24), em que um passo adicional de verticalização do poço (não trivial) deve ser feito para a correção.

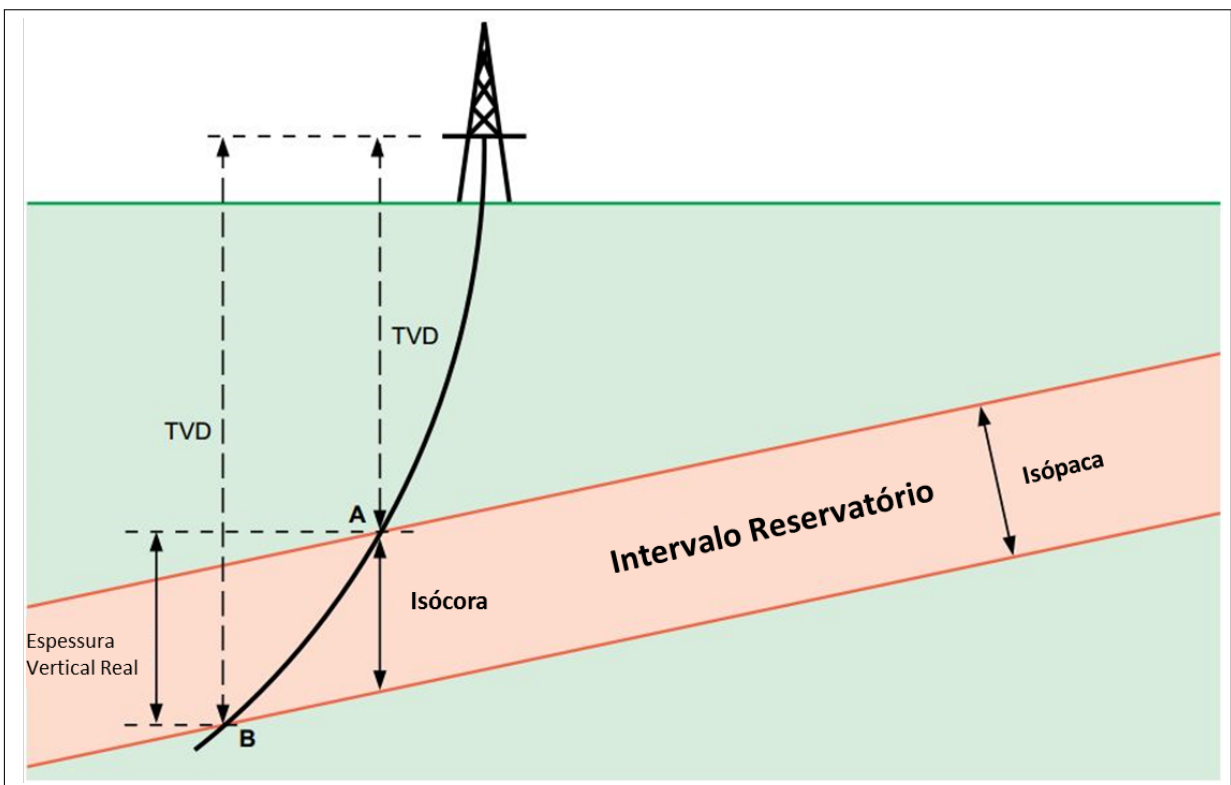


Figura 2.24: Figura ilustrativa de uma camada inclinada e poço inclinado.

Uma confusão principalmente entre os mais inexperientes se dá entre o conceito de coluna de óleo e *gross pay*. A coluna de óleo é a diferença vertical entre o ápice da estrutura preenchida por

determinado hidrocarboneto e o seu contato independentemente de seu mergulho (ver Figura 2.25). Ao passo que o GP diz respeito a espessura do reservatório

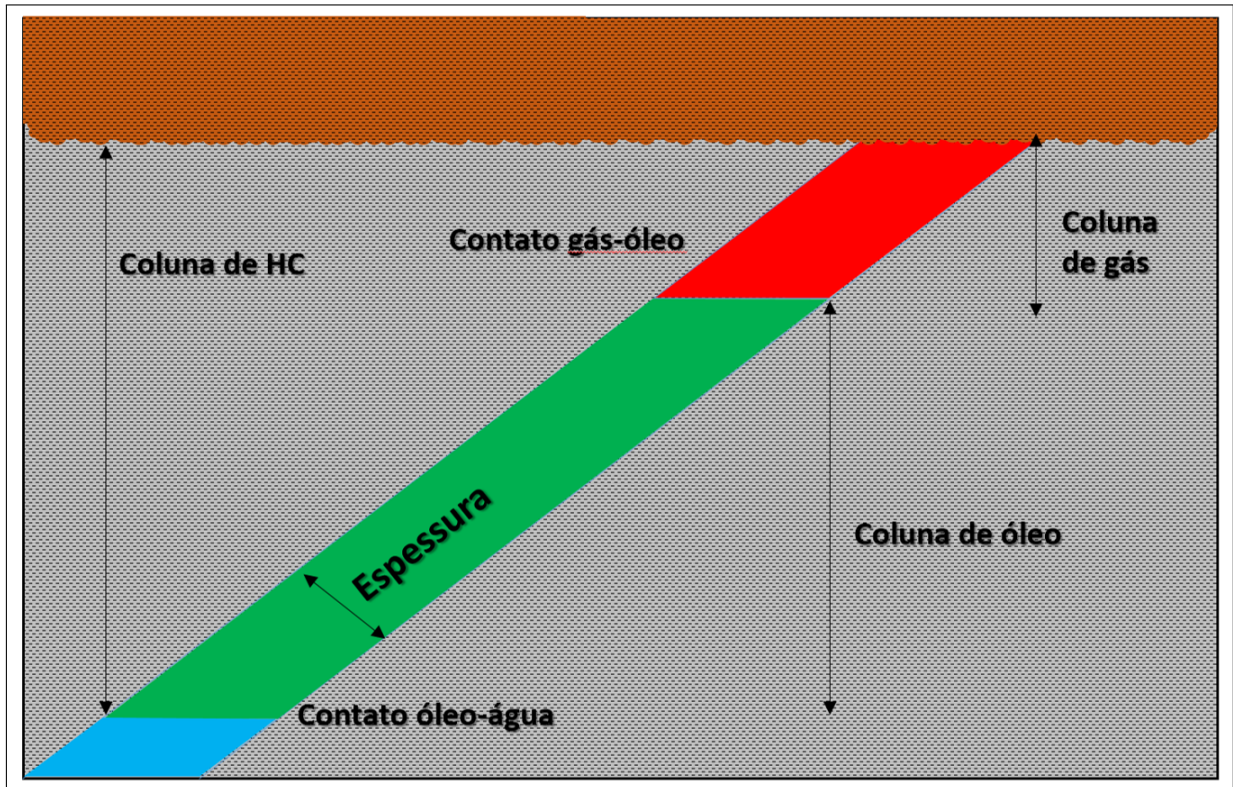


Figura 2.25: Figura ilustrativa de um prospecto estratigráfico, que diferencia coluna de espessura.

Colunas de aproximadamente 1000m podem ser consideradas máximas (p01) sendo um bom ponto de amarração para as estimativas de GRV.

O Fator Geométrico (FG)

Ao utilizar a abordagem de área e espessura devemos reduzir o GRV com o fator geométrico, que é um multiplicador aplicado para descontar a seção abaixo do contato nas bordas do prospecto onde ocorre o *pinch-out* da coluna de petróleo ou gás (White, 1987). A variável FG será fortemente afetada pela curvatura do prospecto (ver Figura 2.26).

Para guiar os intérpretes White (1987) fornece um ábaco útil para estimar a faixa de valores de fator geométrico (ver figura 2.27). Baseado no ábaco, temos uma visão clara de que a medida que a razão Espessura do Reservatório/ Altura do Fechamento aumenta o fator geométrico reduz, e quanto menos achatada é a estrutura menor é o fator geométrico.

Verificamos também que se as bordas são uma feição arredondada (como a metade de uma bola de futebol), será fortemente descontada com um fator de cerca de 0.5. Se o prospecto se assemelha a uma placa, o GRV não sofre alterações, com um fator geométrico de 1. Num caso intermediário, há uma geometria com topo plano - como um prato de sopa invertido.

Note que, apesar de ser extremamente incomum, pode ser necessário aplicar o fator geométrico mesmo em prospecto cuja abordagem foi definida pelo volume de rocha baseado em superfícies. Considere, por exemplo, um cenário em que o volume de rocha foi calculado com base na replicação do topo do reservatório verticalmente (de modo a refletir o Gross Pay) sem considerar o possível contato óleo água. Neste cenário o efeito do *pinch out* nas bordas persistirá mesmo utilizando a abordagem volumétrica.

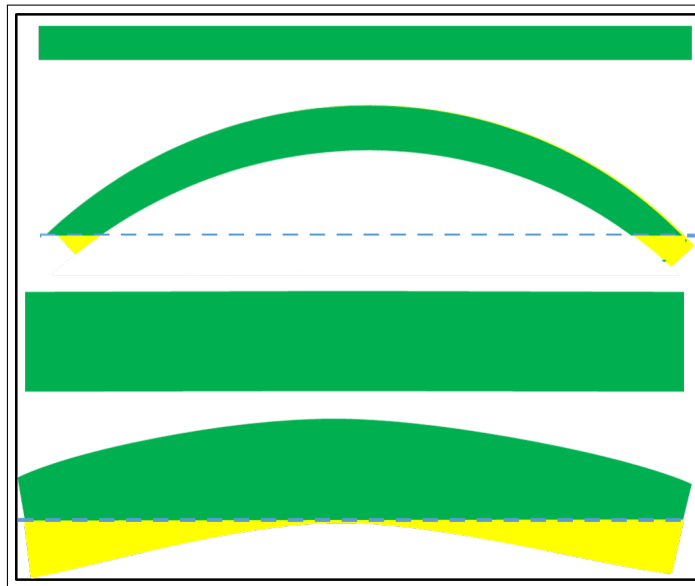
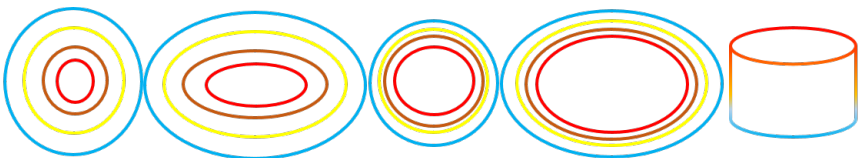


Figura 2.26: Figura ilustrativa exemplificando os fatores geométricos note que quanto mais espesso e dobrado maior será o fator geométrico.



<i>Geometrias do Prospecto</i>		Domo, Cone, Piramide	Anticlinal, Prisma, Cilindro			Domo Achatado	Anticlinal, Prisma, Cilindro		Bloco Vertical ou Cilindro
	Largura/Espessura	1	2	10	1	2	10	Qualquer	
Espessura do Reservatório / Altura do Fechamento	> 1	0,34	0,42	0,49	0,59	0,68	0,74	1,00	
	0,9	0,37	0,46	0,54	0,63	0,71	0,77	1,00	
	0,8	0,41	0,50	0,60	0,67	0,74	0,80	1,00	
	0,7	0,47	0,54	0,66	0,71	0,74	0,83	1,00	
	0,6	0,53	0,61	0,71	0,76	0,81	0,86	1,00	
	0,5	0,59	0,67	0,76	0,80	0,84	0,89	1,00	
	0,4	0,66	0,73	0,81	0,84	0,88	0,91	1,00	
	0,3	0,74	0,80	0,86	0,88	0,91	0,93	1,00	
	0,2	0,83	0,87	0,91	0,93	0,95	0,96	1,00	
	0,1	0,92	0,94	0,96	0,97	0,98	0,99	1,00	
	0	1,00	1,00	1,00	1,00	1,00	1,00	1,00	

Figura 2.27: Ábaco utilizado para estimar o fator geométrico, note que a medida que a razão Espessura do Reservatório/ Altura do Fechamento aumenta o fator geométrico reduz e quanto menos achatada é a estrutura menor é o fator geométrico. Fonte: Modificada de White (1987)

Net-to-Gross (NTG)

O *NTG* é a porção de todo o volume de rocha escolhido em que ocorrem reservatórios e, portanto sua faixa de valores varia de zero (ausência de rocha reservatório) a 1 (todo o volume de rocha composto de reservatório), Ellis and Singer (2007). Isto significa que a depender de como é limitado o reservatório são utilizados valores de *NTG* diferentes. Isto torna o *NTG* uma variável dependente dos parâmetros de espessura e excepcionalmente de área (se a variação lateral de fácies também é considerada neste parâmetro).

O *NTG* é definido normalmente com base em perfis de poços perfurados em prospectos análogos. Neles são aplicados valores de *cut-off*⁵ de porosidade e permeabilidade, de forma a representar apenas as rochas **capazes** de armazenar e fluir hidrocarbonetos (Nery, 2013). Existem diferentes definições para este parâmetro principalmente devido à confusão dos termos *net-pay* e *net-reservatório*. A Figura 2.28 sintetiza a diferença entre estes termos de forma ilustrativa.

O *NTG* descreve a relação entre GrossPay e Net-Reservatório. A relação de Net-Reservatório e Net-pay se dará pelo *Trap Fill* (discutido a seguir), nos sistemas onde toda o reservatório presente na trapa está preenchido por hidrocarbonetos o *NTG* será a relação entre o Gross Pay e Net-Pay.

Quando estimamos o *net-to-gross* com base em perfis, devemos ter em mente que temos dois pontos interpretativos, que são: (1) como é definido o *Gross Pay* e; (2) como são definidos os *cut-offs* de caracterização litológica e de porosidade.

Considerando por exemplo, uma espessura de 20m de net-reservatório num dado intervalo alvo, há uma relação inversa entre *gross-pay* e *NTG*. Se o Gross pay utilizado é de 100m teremos um *NTG* de 0.2, enquanto se o Gross pay utilizado é de 40m o *NTG* é de 0.5 . Em todos os cenários o net continua de 20m e este deve ser o parâmetro a ser observado ao avaliar o *NTG*.

Os *cut-offs* de porosidade serão discutidos nas seções de porosidade e saturação a seguir. Mas, por hora, é razoável utilizar regras práticas aplicadas na indústria. Valores típicos de *cut-off*, que definem uma rocha como reservatório ou não-reservatório de óleo, variam de aproximadamente 5% para reservatórios carbonáticos e 8% para reservatórios siliciclásticos, enquanto, reservatórios portadores de gás admitem porosidades menores com valores que chegam a 3-4%, (Ellis and Singer, 2007).

Note que ao assumir que os valores de perfis são representativos para média da acumulação, estamos assumindo que a acumulação teria aquele comportamento constante lateralmente. Na prática entretanto verificamos que as variações laterais de fácies ocorrem de forma abrupta e estas variações frequentemente são representadas nas amplitudes (Brown, 1999).

A porosidade (ϕ)

Na avaliação de prospectos a porosidade é definida como a porção da rocha reservatório preenchida ou potencialmente preenchida por fluido. Matematicamente a porosidade pode ser definida como uma proporcionalidade de volumes:

$$\phi = \frac{V_{poros}}{V_{amostra}} \quad (2.18)$$

A porosidade desempenha um papel muito importante na avaliação das acumulações pois ela controla o armazenamento de fluidos (água, óleo e gás) e influencia diversas propriedades. Dentre elas podemos citar, sistemas geotérmicos, a extensão e conectividade da estrutura porosa, fluxo de fluido através de formações geológicas, pressão de poros. Além destas, a porosidade também influencia sobretudo propriedades petrosísmicas da rocha como por exemplo densidade e velocidade compressional (Avseth et al., 2010).

⁵Valores de corte.

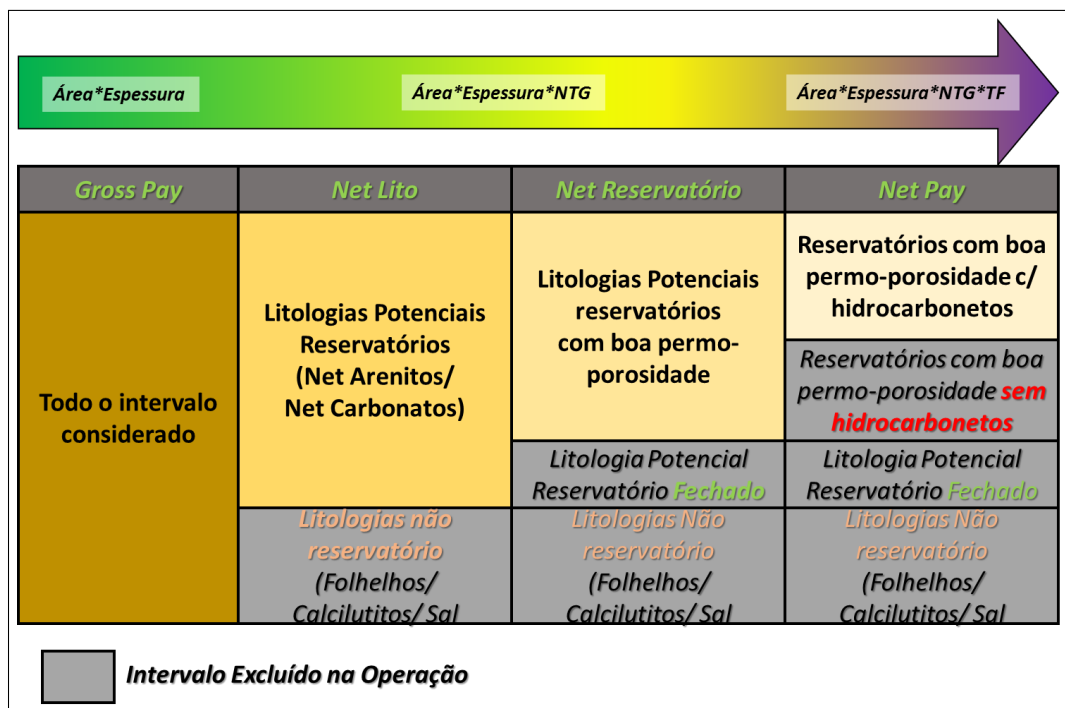


Figura 2.28: Diagrama ilustrando as diferenças entre Gross Pay, Net Litologia, Net Reservatório e Net Pay. Acima os estágios do cálculo volumétrico representando estas diferentes etapas. Onde NTG = *Net-To-Gross*, e TF = *Trap-Fill*.

A porosidade pode ser dividida em primária e secundária. A porosidade primária compreende os espaços porosos da rocha em configuração semelhante àquela da deposição, também chamada de porosidade pré-diagênese (Ellis and Singer, 2007). Ela é afetada por quatro parâmetros microestruturais principais. Estes são:

- O tamanho do grão;
- A forma dos grãos;
- O acondicionamento dos grãos e;
- A distribuição dos tamanhos dos grãos;

A porosidade inicial é raramente encontrada em rochas reais. Isto ocorre, pois, as camadas de rochas são subsequentemente afetadas por controles secundários de porosidade, como compactação e processos geoquímicos diagenéticos (Crain, 2019). A porosidade secundária é o resultado de processos geológicos após a deposição de sedimentos, chamada também de porosidade induzida, pode ser subdividida em três grupos. Estes grupos podem ser baseados no processo geológico mais dominante:

- Dolomitização, processo em que carbonatos calcários se tornam dolomíticos;
- Porosidade de fratura: Poros criados por na rocha reservatório sob tensão causada por atividades estruturais, como dobramento e falhamento.
- Porosidade por dissolução: Poros criados pela dissolução de uma fração da rocha como grãos, cimentos ou fração orgânica.

Outros processos também podem gerar variações de porosidade como por exemplo a erosão e redeposição dos sedimentos de um reservatório previamente erodido (Walker, 1984).

Em reservatórios carbonáticos, a porosidade secundária é muito mais importante do que porosidade primária, devido aos processos químicos. Uma vez que as rochas carbonáticas tendem a ter forte cimentação e baixa compressibilidade estas rochas tendem a sofrer menos com o soterramento se comparadas as rochas siliciclásticas (Lucia, 2007).

A porosidade primária é dominante em reservatórios clásticos, como arenitos, conglomerados, e certos carbonatos oolíticos (Walker, 1984). Estes serão muito mais influenciados por processos de compactação. É importante enfatizar que ambos os tipos de porosidade (primária e secundária) frequentemente ocorrem num mesmo intervalo reservatório em diferentes proporções ao longo de sua extensão (Ellis and Singer, 2007).

Em subsuperfície, o espaço poroso será normalmente preenchido por água, gás (incluindo CO_2) ou óleo. Para se ter uma ideia de valores, um reservatório arenito típico comumente irá ter valores de porosidade na faixa de 10% à 30% (Tiab and Donaldson, 2012). Alguns poucos reservatórios, como areias inconsolidadas sobrepressurizadas, carbonatos altamente carstificados ou corpos ígneos ricos em vesículas podem ter porosidades médias que superam os 30% em pequenas porções, mas a distribuição destas feições tende a ser errática e não ser representativa dos recursos prospectivos de uma bacia.

A tabela 2.2 resume as classificações de qualidade de reservatório baseado em valores típicos de porosidade conforme proposto por Tiab and Donaldson (2012).

Faixa de Porosidade	Qualidade
0-5%	Fechado (não-reservatório)
5-10%	Pobre
10-15%	Razoável
15-20%	Bom
>25%	Muito Bom

Tabela 2.2: Classificação de Reservatórios de acordo com a porosidade conforme Tiab and Donaldson (2012).

O valor mínimo de porosidade de uma rocha seria 0%, onde há ausência absoluta dos poros, mas normalmente rochas com porosidades tão baixas são consideradas não-reservatório. Conforme anteriormente citado, valores típicos de *cut-off*, que definem uma rocha como reservatório ou não reservatório, variam de aproximadamente 5% para reservatórios carbonáticos e 8% para reservatórios siliciclásticos enquanto reservatórios portadores de gás admitem porosidades menores com valores que chegam a 3-4% (Tiab and Donaldson, 2012).

No outro extremo valores de porosidade assumem um valor máximo teórico de 1 (100%), onde temos apenas fluidos e ausência total de grãos, de fato, não teríamos uma rocha neste extremo. Para considerar os casos encontrados na natureza em faixa prospectiva, uma forma visual de verificar as limitações de valores de porosidade é avaliar as porosidades considerando empacotamentos perfeitos numa rocha composta por grãos esféricos. Conforme ilustrado em Glover (2016), os diferentes empacotamentos respondem a diferentes porosidades matematicamente relacionadas (Figura 2.29).

Estes valores, entretanto, são raramente encontrados, uma vez que os grãos normalmente não são esféricos e apresentam tamanhos heterogêneos.

Mesmo que os sedimentos sejam depositados com um empacotamento cúbico e seja uni-modal, eles tendem a assumir um empacotamento romboédrico em função da compactação reduzindo assim a porosidade da rocha. Em um sistema siliciclástico bem soterrado (>2km de cobertura sedimentar), porosidades excessivamente altas (maiores que 36-40%) indicam que os grãos não estão em contato entre si e são suportados por fluido (Nur et al., 1998). Esta faixa é conhecida como porosidade crítica (Nur et al. (1998) e Dvorkin and Nur (2001)). O valor de porosidade crítica pode ser mais facilmente aplicado em siliciclásticos, onde o empacotamento fornece um

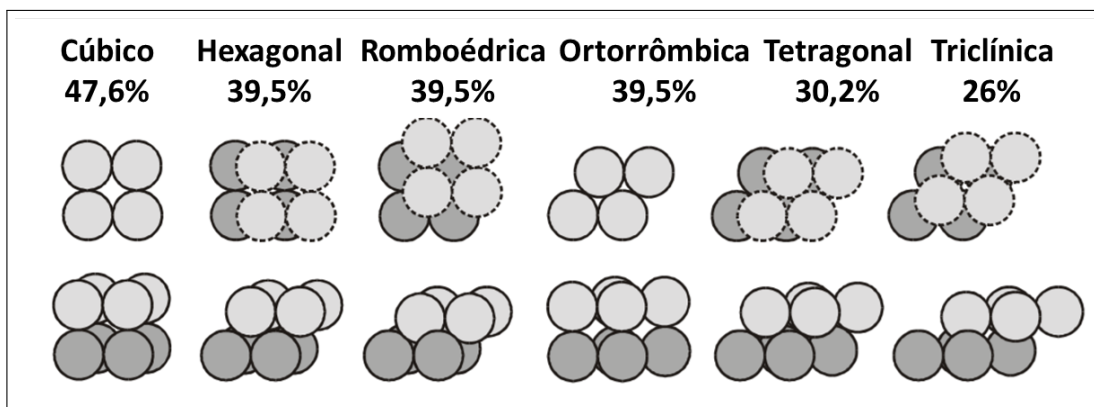


Figura 2.29: Ilustração dos empacotamentos com grãos esféricos e suas respectivas porosidades exatas calculadas matematicamente. Fonte: modificado de Glover (2016)

excelente entendimento visual do fenômeno. Contudo, o conceito de porosidade crítica também pode estar associado a outras rochas (carbonatos, ígneas fraturadas etc). Na Tabela 2.3 é possível encontrar valores de porosidade crítica normalmente encontrados na natureza (Nur et al. (1998) e Dvorkin and Nur (2001)).

Rocha	Porosidade crítica
Arenitos	40%
Carbonatos Calcários	40%
Carbonatos Dolomíticos	40%
Púmice	70%
Chalk	65%
Sal	40%
Ígneas Fraturadas	5%
Basaltos Oceânicos	20%

Tabela 2.3: Valores de porosidade crítica para as rochas conforme Nur et al. (1998) apud Dvorkin and Nur (2001).

Note que os valores acima são médios e, portanto, variarão de rocha para rocha para uma mesma classe. O entendimento de porosidade crítica é importante pois ao avaliar prospectos estamos trabalhando com faixas de valores razoáveis (Rose, 2001). Ao atribuir valores de porosidade anormalmente altos para uma oportunidade, temos que ter em mente o seu impacto na resposta esperada nos dados de estudo. Isto é, devemos atribuir uma relação de causalidade desta propriedade aos dados avaliados.

Durante a sedimentação e litificação, alguns dos espaços porosos inicialmente desenvolvidos tornam-se isolados dos outros espaços porosos por vários processos diagenéticos e catagenéticos, como cimentação e compactação (Dvorkin and Nur, 2001). Assim, muitos dos poros serão interconectados, enquanto outros serão completamente isolados. Isto leva a duas categorias distintas de porosidade, a total (absoluta) e a efetiva.

A diferença entre as porosidades total e efetiva é basicamente a porosidade isolada ou não efetiva. A porosidade absoluta é a razão entre o espaço poroso total na amostra e o volume da amostra, independentemente de esses espaços porosos estarem ou não interligados Lucia (2007). Uma rocha pode ter porosidade absoluta considerável e ainda não ter condutividade de fluidos por falta de interconexões de poros. Exemplos disso são lava, pedra-pomes e outras rochas com porosidade vesicular ou carbonatos com porosidade *vuggy*.

A porosidade efetiva é afetada por vários fatores litológicos:

- Tipo, conteúdo e hidratação das argilas presentes na rocha;
- A heterogeneidade de tamanhos de grãos;
- O empacotamento e cimentação dos grãos;
- Qualquer intemperismo e lixiviação que podem afetar a rocha;

Muitos dos poros podem ser isolados com apenas uma entrada para o sistema principal de canais de poros. Dependendo da molhabilidade dos grãos, esses poros podem ser preenchidos com água ou óleo, estes fluidos estarão presos, até que algum elemento externo permita um fluxo contínuo Crain (2019).

Para recuperar petróleo e gás dos reservatórios, os hidrocarbonetos devem fluir várias centenas de metros através dos poros da rocha antes que eles atinjam o poço produtor. Se o petróleo ocupa poros não conectados, pode não ser produzido e é de pouco interesse para a indústria. Assim sendo, porosidade efetiva é o valor utilizado em todos os cálculos de engenharia do reservatório, enquanto as estimativas de porosidade na etapa exploratória tendem a considerar a porosidade total, posteriormente descontada no fator de recuperação ou risco tecnológico. Estes parâmetros devem refletir a porosidade efetiva e permeabilidade, se não anteriormente descontada no próprio cálculo de volume de hidrocarbonetos *in place*.

A estimativa de porosidade normalmente é feita com base em perfis de poços e/ou de ensaios laboratoriais (em amostras laterais e testemunhos), Nery (2013). Alternativamente, ela pode ser definida e extrapolada por dados obtidos com a sísmica.

Trends Regionais

Trends de porosidade com a profundidade ou soterramento geralmente fornecem uma boa fonte de informações para estimar as faixas de porosidade para diferentes profundidades em uma bacia, formação ou play específico.

A Figura 2.30 mostra duas tendências de porosidade com profundidade para dois plays importantes (Mioceno Médio e Plioceno) nas águas profundas do Golfo do México (GOM), Amado (2013). Como pode ser visto, este gráfico pode ser usado (ou derivado para outras bacias ou *plays*) para obter intervalos de porosidade para um prospecto situado em uma profundidade particular.

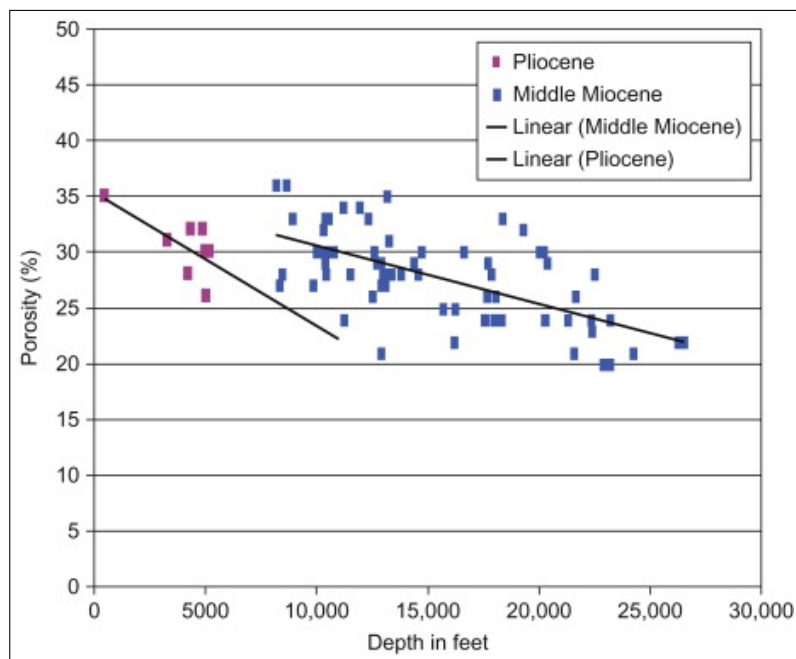


Figura 2.30: Ilustração de *trend* de porosidade versus profundidade para os plays Mioceno Médio e Plioceno no GOM. Fonte : Amado (2013)

Trap Fill (*TF*)

Não é incomum que os intervalos reservatórios das estruturas geológicas testadas não estejam completamente preenchidos de petróleo ou gás. Existem diversas explicações para isso, como incapacidade do selo para reter grandes colunas, falta de migração para determinados compartimentos ou mesmo uma geração insipiente. O fator que descontinua o efeito do não preenchimento de intervalos reservatório é o *Trap Fill*.

Conforme mencionado por Peacock (2007), os poços exploratórios muitas vezes são direcionados para posições próximas à crista dos altos estruturais e são objeto de premissas excessivamente otimistas sobre preenchimento da trapa. A suposição de preenchimento até o *spill point* uniformemente, foi (e em algumas empresas ainda são) uma das principais causas de falha na estimativa de incerteza volumétrica.

É importante ressaltar, entretanto, que nos casos em que o *Trap Fill* leva em conta o não preenchimento da estrutura em termos de contato ao longo de toda a trapa, esta variação não deve ser considerada também no volume de rocha.

Assim como discutido nas seções de geração e migração, grande parte dos hidrocarbonetos gerados são retidos na rocha geradora sem migrar, aproximadamente 50% para uma rocha geradora que ainda não entrou na janela de gás (Law, 1999). Estes volumes permanecem como alvos para o desenvolvimento das acumulações não-convencionais. Uma vez gerados e expulsos das rochas geradoras, os hidrocarbonetos fluem vertical e lateralmente nos sedimentos que circundam a rocha geradora.

Problemas de migração, quando há geração de volumes consideráveis de óleo e gás na cozinha, normalmente são decorrentes do fato das armadilhas geológicas poderem exibir geometrias altamente variáveis e muitas vezes compartimentadas (Bishop, 2012).

Nestes casos, em que os reservatórios são altamente segmentados, é possível que a migração ou retenção não seja efetiva para alguns compartimentos. Veja por exemplo o campo de Tidepole (Figuras 2.31 e 2.32) em que os contatos variam dentro de um mesmo compartimento.

Como regra geral ao utilizar o volume de rochas calculado diretamente com base no modelo (considerando pouca variação do mesmo). Normalmente utiliza-se o *Trap Fill* para representar as seções de reservatório sem hidrocarbonetos.

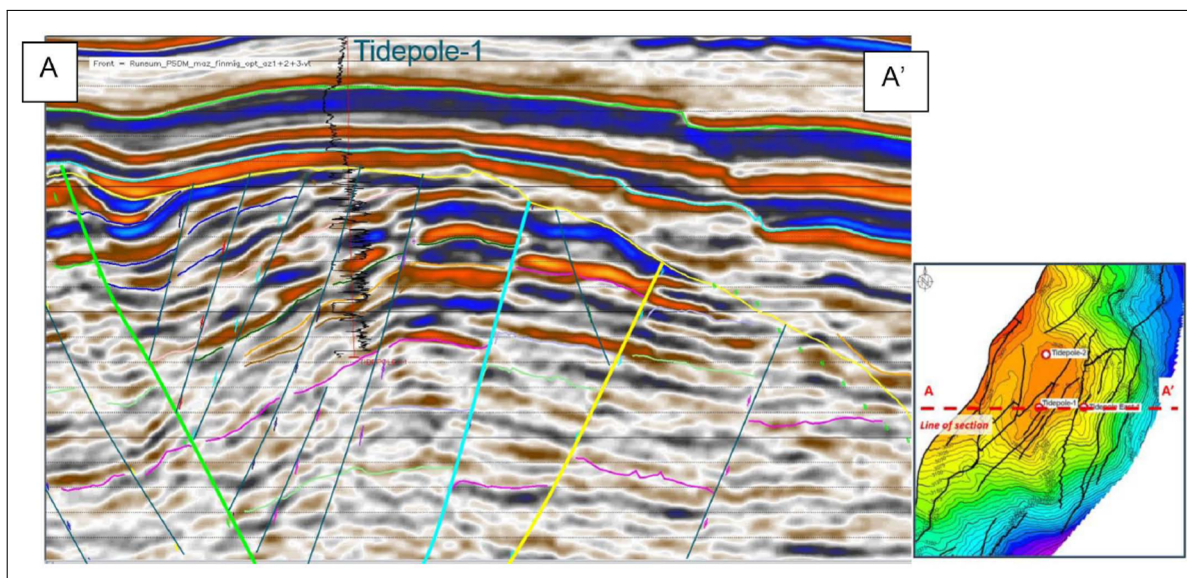


Figura 2.31: Seção Sísmica com arquitetura de falhas do campo de Tidepole (Nordeste da Austrália) e mapa estrutural. Fonte: Smith et al. (2018).

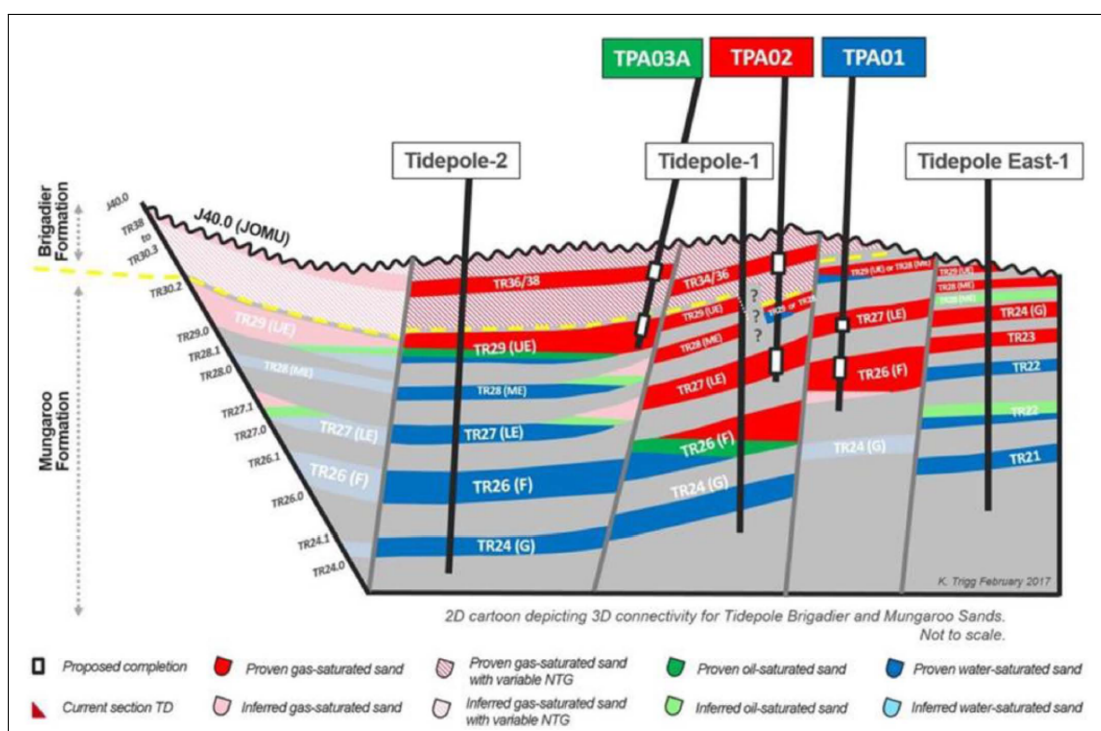


Figura 2.32: Modelo esquemático do campo de Tidepole ilustrando as variações de *Trap Fill* causado pelas falhas. Fonte: Smith et al. (2018).

Quando usamos área e espessura ou volume de rocha, já delimitando apenas o reservatório com hidrocarboneto não faz sentido utilizar este parâmetro.

Devemos descontar do reservatório as seções subsaturadas (o parâmetro saturação será discutido a seguir). Como regra geral saturações inferiores a 50% não são caracterizados como intervalo *pay* (Rose, 2001).

Em avaliações onde toda a estrutura é considerada como preenchida em termos de volume de rocha (o GRV apresenta pouca dispersão) sem fatores concretos para isto, (e.g. ausência de anomalias ou estruturas cheias até a *spill point* próximos ao alvo). O *Trap Fill* deve ser utilizado

com valores razoáveis girando em torno de 0,3 a 1.

Saturação de água / Saturação de HC

A saturação é a proporção relativa com que os poros da rocha são preenchidos com fluidos específicos. Essa propriedade é chamada de saturação de fluido e é expressa como a fração ou porcentagem do volume total de poros ocupado pelo óleo, gás ou água. Assim, por exemplo, a saturação do óleo é igual a:

$$S_o = \frac{V_{oleo}}{V_{poros}} \quad (2.19)$$

Todos os fluidos terão seu percentual definido por saturações de forma que o somatório de todos eles totalize 100%:

$$\sum_{i=1}^{n_{fluidos}} S_i = 100\% \quad (2.20)$$

Idealmente, devido à diferença nas densidades de fluidos e pressão capilar, um reservatório de petróleo é formado de tal forma que, de cima para baixo, haverá gás, óleo e água (uma exceção a este padrão são os gases de centro de bacia não discutidos neste trabalho).

Para avaliar a saturação, precisamos entender o processo de entrada dos hidrocarbonetos na trapa e sua respectiva drenagem. Os reservatórios são inicialmente preenchidos por água, seja por sua deposição sub-aquosa, seja fato de, em algum momento, chegarem abaixo do nível do lençol freático.

O reservatório é posteriormente preenchido por hidrocarbonetos como resultado da migração secundária. Neste processo, a maior parte da água de formação será removida restando o que chamamos de saturação residual de água (Nery, 2013). A saturação residual é uma fração que se formará nas paredes dos grãos nos reservatórios molháveis a água e/ou no interior dos poros nos reservatórios não molháveis a óleo (ver Figura 2.33). Esta fração de água é chamada de água conata ou água irreduzível e desde que a rocha esteja totalmente saturada será a fração que permanecerá no reservatório (Ellis and Singer, 2007).

Conforme Crain (2019), normalmente os reservatórios são molháveis a água, assim a área de superfície define a saturação de água irreduzível. Dessa forma, poros pequenos têm maior área de superfície em relação ao seu volume poroso, de modo que a saturação de água irreduzível assume maior percentual (Amado, 2013). Se os poros forem pequenos o suficiente, a saturação de água irreduzível pode ser de 1, não deixando espaço para o acúmulo de óleo ou gás.

A saturação inicial de água em um reservatório preenchido por hidrocarbonetos tende a ser a saturação irreduzível em uma zona sem vazamento do selo e/ou produção prévia de qualquer poço anterior, ou seja, o processo de saturação de hidrocarbonetos tende a ser eficiente na expulsão da água móvel.

Assim como ocorre para água, no processo de drenagem da produção parte dos hidrocarbonetos permanecem no reservatório o que nos leva uma informação: Os valores de saturação de hidrocarbonetos nunca atingirão o valor de 100%. Similarmente, o fator de recuperação sempre será reduzido pelos hidrocarbonetos que nunca serão recuperáveis.

Este fenômeno se dá pelo fato da permeabilidade relativa dos fluidos atingirem o valor zero antes de sua saturação atingir o valor zero e, portanto, a fase não-contínua dos fluidos fica aprisionada no reservatório.

Para fazer uma avaliação coerente da saturação devemos considerar os tipos de saturação que coexistem no meio poroso (ver Figura 2.34), Crain (2019). Onde apenas os hidrocarbonetos móveis constituirão a fase pay do reservatório.

A avaliação baseada em perfis de poços deve levar em consideração alguns processos artificiais (Nery, 2013). Um exemplo para isto é a entrada da parte filtrada do fluido de perfuração na

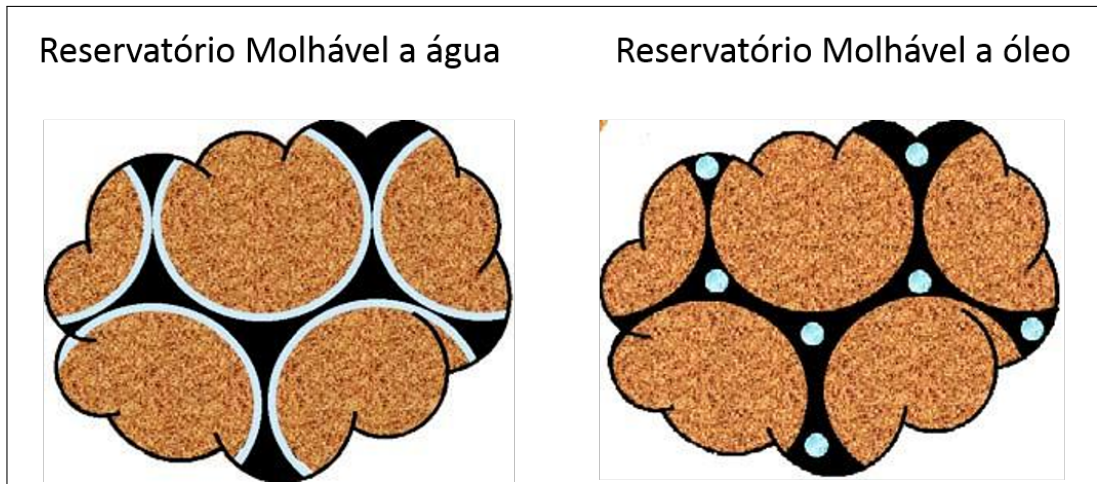


Figura 2.33: Modelo esquemático de reservatório molhável a água (esquerda) e a óleo (direita) após a migração secundária de hidrocarbonetos. Note que em ambos os modelos ocorrem saturações residuais de água. Fonte: Crain (2019).

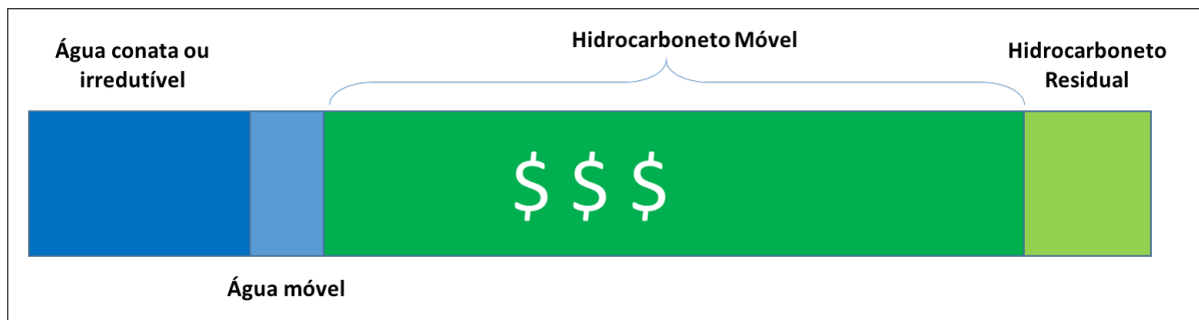


Figura 2.34: Modelo esquemático com as frações de saturação no espaço poroso.

região próxima ao poço (bem como sua composição que constituirá mais uma fase no reservatório) e o seu efeito nas ferramentas que terão diferentes profundidades de investigação.

Os diferentes valores de saturação podem ser bem estimados em dados de laboratório, através do processo de injeção dos diferentes fluidos. Com a ausência de amostras é mais comum obtermos os valores de saturação de cada fluido baseado em perfis de resistividade e porosidade, limitando-se a saturação total por tipo de fluido.

Archie (1942) determinou uma expressão empírica entre a resistividade, a porosidade e diferentes valores de saturação com água salgada. Ele fez isto a partir de testemunhos e amostras com porosidade e permeabilidades conhecidas, expressa pela seguinte relação:

$$S_w = \left(\frac{a \cdot R_o}{\phi^m R_t} \right)^{1/n} \quad (2.21)$$

Onde:

- n= expoente de saturação
- m= expoente de cimentação
- a= coeficiente litológico (ou coeficiente de tortuosidade)

Esta equação é especialmente importante neste trabalho pois evidência explicitamente a relação entre porosidade e saturação de água (dois parâmetros que serão discutidos a seguir).

Embora a formulação original de Archie não contenha o parâmetro "a", ele passou a ser amplamente utilizado na indústria. Ainda hoje o seu uso divide opiniões principalmente sobre o seu significado físico (ou a falta dele). Para aqueles que atribuem um significado físico para o coeficiente, é muitas vezes conhecido como o "índice de tortuosidade". Outros autores, como Doveton (1986), vêem este coeficiente simplesmente como elemento de acomodação de mudanças no expoente de cimentação (m), na equação original. Este é certamente o caso da equação de Winsauer et al. (1952) onde a=0.62 e m=2.15.

O fato é que, fisicamente, considerar $a \neq 1$ é considerar que $R_w \neq R_w$ o que torna o uso do valor de "a" diferente de 1, um ajuste útil porém mal posto. O parâmetro "m" diz respeito ao grau de conexão entre os poros que em última análise tem forte relação com a permeabilidade e cimentação da rocha Nery (2013). Isto significa que areias não consolidadas com tamanhos de grãos razoavelmente uniformes terão um valor baixo de m (1.2-1.3). No outro extremo, arenitos fortemente cimentados, ou mal classificados, onde os poros tem baixa interconexão são caracterizados por altos valores de m (2-2.2). A tabela 2.4 abaixo, resume bem os valores típicos de m:

Rocha	Valor de m típico
Cimentados	1.9-2.2
Moderadamente cimentados	1.8-1.9
Fracamente cimentados	1.6-1.7
Inconsolidado	1.3-1.5
Arenitos Limpos	1.2-1.3

Tabela 2.4: Valores típicos do parâmetro m (expoente de cimentação) para a equação de Archie (1942).

Para calcários fechados, que são normalmente rochas altamente cimentadas, o valor de m pode ser tão alto quanto 3. É importante enfatizar que estes resultados são empíricos e os valores de m muitas vezes diferem de poço a poço nas mesmas formações de um reservatório e portanto seu uso deve ser utilizado pelo que ele é : **uma aproximação**.

A partir da equação de Archie, é possível determinar um fator de resistividade de formação F_R relacionado com a porosidade, uma relação entre essas duas variáveis é especialmente útil. A forma geral desta expressão é:

$$F_R = \frac{a}{\phi^m} \quad (2.22)$$

Onde "a" e "m" devem idealmente ser obtidos de medidas de laboratório. Quando o valor destes coeficientes não podem ser determinados e o uso de a=1 e m=2 atinge um bom ajuste.

Fator de Formação de Gás (B_g)

O fator de formação de gás é usado para relacionar o volume de gás medido nas condições do reservatório, com o volume do gás medido em condições padrão, isto é, 60°F e 14,7psia. Matematicamente podemos descrever B_g como:

$$B_g = \frac{V_{TP,res}}{V_{TP,padrao}} \quad (2.23)$$

Onde:

- $V_{TP,res}$ - Volume de gás temperatura e pressão nas condições de reservatório;
- $V_{TP,padrao}$ - Volume de gás temperatura e pressão nas condições atmosféricas;

Idealmente podemos utilizar a equação geral dos gases para determinar os valores de volume em ambas as condições, e assim estimar o valor de B_g .

$$p.V = n.R.T \quad (2.24)$$

Onde:

- p = pressão (dada em psia ou atm)
- V = Volume (dada em L, m^3 ou pés³)
- n = número de mols lb-mole
- R = constante geral dos gases 0,082 atm.L/mol.k (ou 10.730 psia pés³/lb-mol)

Ao lidar com gases a uma pressão muito baixa, a aproximação de gás ideal e real é uma ferramenta conveniente e geralmente satisfatória. Em pressões mais altas, como aquelas que encontramos em condições de reservatórios, o uso da equação de estado do gás ideal pode levar a erros da ordem de 500%, em comparação com erros de 2 a 3% à pressão atmosférica Hall and Yarborough (1973).

Basicamente, a magnitude dos desvios dos gases reais em relação à lei dos gases ideais aumenta com o aumento da pressão e temperatura e varia muito com a composição do gás (Ahmed, 2010). A razão para isso é que a lei para gases perfeitos foi derivada sob o pressuposto de que o volume de moléculas é insignificante e que nenhuma atração molecular ou repulsão existe entre eles. Este não é o caso de gases reais a alta pressão (Standing and Katz, 1942).

Numerosas equações de estado foram desenvolvidas na tentativa de correlacionar as variáveis pressão-volume-temperatura para gases reais com dados experimentais. Para expressar uma relação mais exata entre as variáveis p , V e T , um fator de correção chamado de fator de desvio de compressibilidade do gás, ou simplesmente o fator z , deve ser introduzido para corrigir a saída das condições ideais dos gases e, portanto, a equação tem a seguinte forma (Hall and Yarborough, 1973):

$$P.V = z.n.R.T \quad (2.25)$$

Onde o fator z é adimensional e é definido como uma razão entre o volume real dos gases em relação ao ideal (Hall and Yarborough, 1973). Estudos de compressibilidade de gás para gases naturais de várias composições mostraram que o fator z pode ser bem aproximado quando ele é expresso em termos de duas propriedades adimensionais:

- Pressão Pseudo-Reduzida (p_{pr}) e;
- Temperatura Pseudo-Reduzida (T_{pr}).

Estes termos podem ser definidos como:

$$p_{pr} = \frac{p}{p_{pc}} \quad (2.26)$$

$$T_{pr} = \frac{T}{T_{pc}} \quad (2.27)$$

Onde T

- T = temperatura do reservatório;
- p = pressão do reservatório;

- T_{pr} = temperatura pseudo-reduzida.
- P_{pr} = pressão pseudo-reduzida.
- T_{pc} = temperatura pseudo crítica
- p_{pc} = pressão pseudo crítica

Os valores de pressão (P) e temperatura (T) podem ser obtidos com dados de PVT próximos à locação e/ou estimado a partir de *trends* regionais.

$$T_{pc} = \sum_{i=1} y_i \cdot T_{ci} \quad (2.28)$$

$$p_{pc} = \sum_{i=1} y_i \cdot p_{ci} \quad (2.29)$$

Onde os valores de p_{ci} e T_{ci} são os valores de pressão e temperatura críticas para cada substância e y_i a sua respectiva fração molar (Hall and Yarborough, 1973). A tabela 2.5 apresenta alguns valores de pressão e temperatura crítica para frações gasosas típicas presentes em acumulações.

Gás	T_c ($^{\circ}R$)	p_{ci} (psia)
Nitrogênio(N ₂)	227.16	493.1
Dióxido de Carbono(CO ₂)	547.58	1071
Sulfeto de Hidrogênio (H ₂ S)	672.35	1306
Metano (C ₁ H ₄)	343.04	666.4
Etano (C ₂ H ₆)	549.59	706.5
Propano (C ₃ H ₈)	665.73	616.0
i - Butano (C ₄ H ₁₀)	734.13	527.9
n - Butano (C ₄ H ₁₀)	765.29	550.6
i - Pentano (C ₅ H ₁₂)	828.77	490.4
n - Pentano (C ₅ H ₁₂)	845.47	488.6
pseudo C ₆ H ₁₄	913.27	436.9
pseudo C ₇ H ₁₆	972.37	396.8
pseudo C ₈ H ₁₈	1023.89	360.7
pseudo C ₉ H ₂₀	1070.35	331.8
pseudo C ₁₀ H ₂₂	1111.67	305.2

Tabela 2.5: Alguns valores de pressão e temperatura crítica de substâncias típicas em gases.

É possível calcular os valores de p_{pc} e T_{pc} a partir de aproximações empíricas a partir da gravidade específica dos gases (γ_g):

$$T_{pc} = A + B\gamma_g + C\gamma_g^2 \quad (2.30)$$

$$p_{pc} = D + E\gamma_g + F\gamma_g^2 \quad (2.31)$$

Para sistemas com gás natural (A=168,B=325, C=-12,5, D=677, E=15 e F=-37.5) e para sistemas com gás-condensados (A=187,B=330, C=-71.5, D=706, E=51.7 e F=11.1). A vantagem destas equações é que não é necessário saber a composição exata dos gases, apenas uma estimativa da sua gravidade específica (Standing and Katz, 1942).

De posse dos valores de temperatura e pressão pseudo-reduzidas utiliza-se o ábaco de (Standing and Katz, 1942) (ver Figura 2.35) para determinar o fator de compressibilidade.

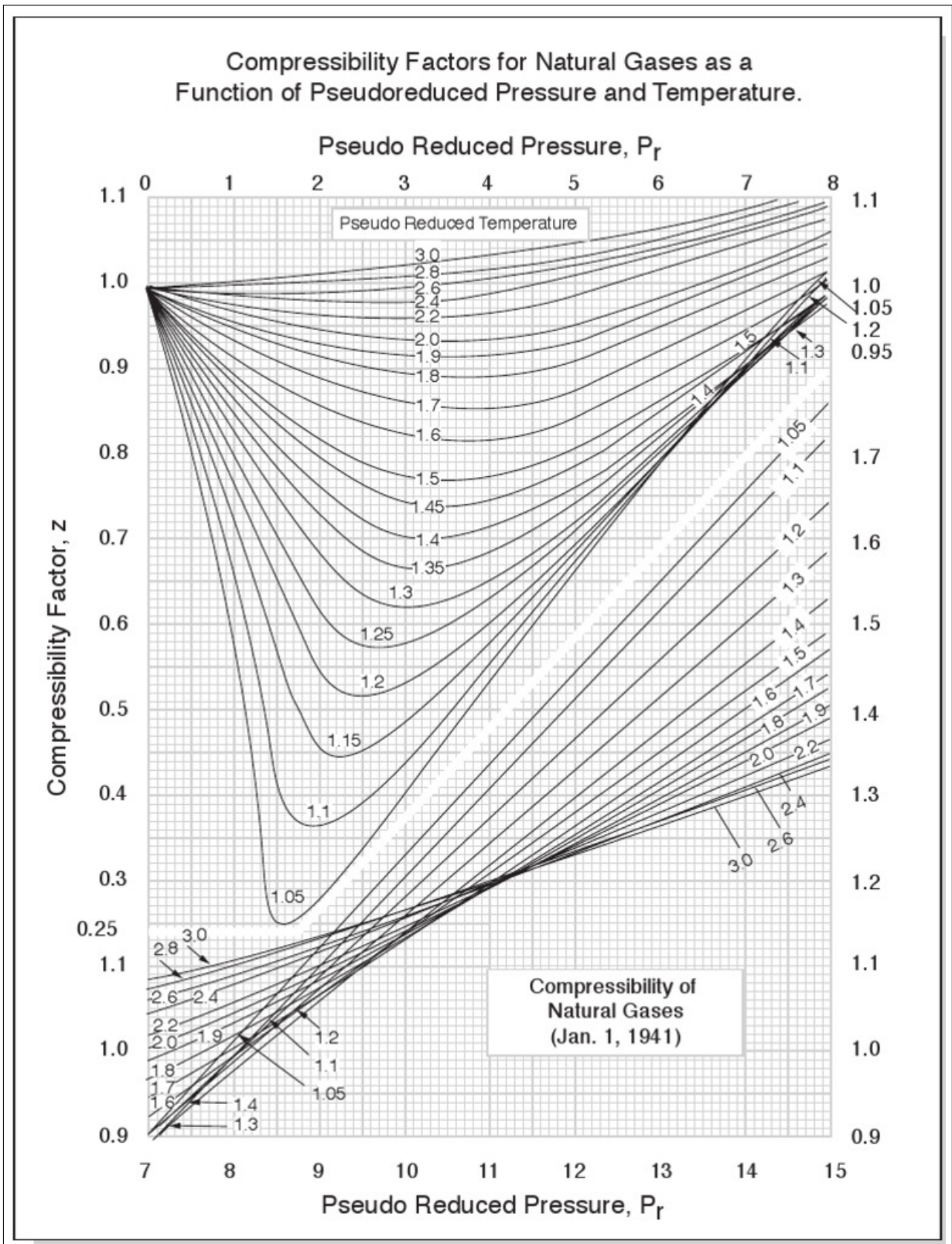


Figura 2.35: Ábaco proposto por Standing and Katz (1942), para determinar o fator de compressibilidade (fator- z) de gases reais. Os valores de pressão e temperatura reduzida são obtidos matematicamente. Verifique que a escala superior de pseudo-pressão reduzida diz respeito aos valores das curvas acima e a escala inferior diz respeito às curvas de baixo.

Este ábaco foi aproximado por Hall and Yarborough (1973), apresentando equações de estado que representam com alguma precisão o fator de compressibilidade (z). Eles propuseram equações não-lineares baseados no ajuste de pontos do ábaco gerando uma equação não-linear conforme abaixo:

$$z = \left[\frac{A1 \cdot p_{pr}}{Y} \right] \quad (2.32)$$

Onde:

- p_{pr} = pressão pseudo-reduzida
- Y = densidade reduzida, que pode ser obtida a partir da raiz da equação (2.33) abaixo.

$$- A1 \cdot p_{pr} + \frac{Y + Y^2 + Y^3 + Y^4}{(1 - Y)^3} - A2 \cdot Y^2 + A3 \cdot Y^{A4} = 0 \quad (2.33)$$

e

- $A1 = 0.06125 \cdot t \cdot \exp 1.2(1 - t)^2$
- $A2 = 14.76t - 9.76t^2 + 4.58t^3$
- $A3 = 90.7t - 242.2t^2 + 42.4t^3$
- $A4 = 2.18 + 2.82t$
- $t = 1/T_{pr}$

Note que a equação (2.33) é não linear e pode ser resolvida iterativamente para Y utilizando métodos como Newton-Raphson ou Gauss-Newton com base em um chute inicial de Y . Dessa forma Y deve ser incrementado com valores negativos ou positivos de forma a fazer a equação assumir o valor 0.

Kareem et al. (2014) fornece aproximações alternativas que não requerem métodos iterativos de solução, entretanto, em face à capacidade computacional disponível hoje em dia, isto já não é mais um problema.

Os gases naturais frequentemente contêm outros materiais além de componentes hidrocarbonetos, como nitrogênio, dióxido de carbono e sulfeto de hidrogênio (Kareem et al., 2014). As equações de estado acima apresentadas normalmente têm bom ajuste quando componentes que não hidrocarbonetos estão presentes em pequenas frações (inferiores a 5% da fração molar total).

Os gases de hidrocarbonetos são classificados como doces ou azedos dependendo teor de sulfeto de hidrogênio H_2S (os que tem teor menor do que 0.5% são classificados como doce), Kareem et al. (2014). Ambos os gases doces e azedos podem conter nitrogênio, dióxido de carbono, ou ambos. Em condições em que estes não-hidrocarbonetos apresentam grandes proporções a estimativa do fator z tem erros superiores a 10% (Kareem et al., 2014).

Para corrigir o efeito das proporções das outras frações não compostas de HC utiliza-se normalmente dois métodos (Carr-Kobayashi-Burrows e Wichert-Aziz) ambos os métodos se baseiam na correção das temperaturas e pressões pseudo-críticas de forma a tornar o ábaco de Standing e Katz (Figura 2.35) eficaz novamente (Ahmed, 2010).

Wichert-Aziz desenvolveram um método para ajustar os valores de temperatura e pressão pseudo-crítica.

$$T'_{pc} = T_{pc} - \epsilon \quad (2.34)$$

$$p'_{pc} = \frac{p_{pc} T'_{pc}}{T_{pc} + B(1 - B)\epsilon} \quad (2.35)$$

Onde:

- T_{pc} = Temperatura pseudo-crítica, °R
- p_{pc} = pressão pseudo-crítica, psia
- T'_{pc} = Temperatura pseudo-crítica corrigida, °R
- p'_{pc} = pressão pseudo-crítica corrigida, psia
- B = fração Molar de H_2S da mistura de gás.
- ϵ = fator de ajuste temperatura pseudo-critica dado pela equação (2.36).

$$\epsilon = 120[A^{0.9} - A^{1.6}] + 15(B^{0.5} - B^{4.0}) \quad (2.36)$$

Onde A é a soma das frações molares de H_2S e CO_2 , ($A = y_{H_2S} + y_{CO_2}$).

O método de Carr-Kobayashi-Burrows utiliza equações simplificadas que também permitem esta correção, as equações foram originalmente desenvolvidas para calcular a viscosidade gases ricos em H_2S e CO_2 em concentrações até 15% e em temperaturas que variam de 32 a 400F e pressões superiores a 12kpsia (Ahmed, 2010).

$$T'_{pc} = T_{pc} - 80y_{CO_2} + 130y_{H_2S} - 250y_{N_2} \quad (2.37)$$

$$p'_{pc} = p_{pc} + 440y_{CO_2} + 600y_{H_2S} - 170y_{N_2} \quad (2.38)$$

Para o cálculo da gravidade específica do gás (γ_g), ou seja sua densidade em relação a substância de referência (normalmente utilizamos o ar). Por se tratar de um gás utiliza-se normalmente as frações molares para fazer esta estimativa.

$$\gamma_g = \frac{M_a}{28.96} \quad (2.39)$$

Onde Ma_{pc} pode ser obtido com a equação (2.40), abaixo:

$$Ma_{pc} = \sum_{i=1} y_i \cdot MM_i \quad (2.40)$$

As frações molares (y_i) normalmente obtidas por ensaios de PVT ou estimada a partir de outro fluido esperado para a formação (Kareem et al., 2014). Os valores de massas molares dos gases típicos encontrados na natureza podem ser encontrados na tabela 2.6:

Dessa forma, com os valores de z, T e P é possível determinar o B_g , utilizando a equação geral dos gases corrigida pelo fator z. Assumindo as condições padrões de pressão (14.696 psia ou 101.325 kPa) e temperatura (60°F ou 519.67°R ou 288.71°K) substituímos as variáveis na equação (2.23) e obtemos a equação 2.41.

$$B_g = 0.0282793 \cdot \frac{z \cdot T_{res}}{p_{res}} (m^3/m^3) \quad (2.41)$$

Onde a temperatura é dada em Kelvin e a pressão em kPa. Alternativamente, utilizando a pressão em psia e a temperatura em °R obtemos:

$$B_g = 0.350959 \cdot \frac{z \cdot T_{res}}{p_{res}} (pes^3/pes^3) \quad (2.42)$$

Valores típicos de B_g giram em torno de 0.004 para gases extremamente ricos em metano, chegando a valores que giram em torno de 0.07 para gases ricos em CO_2 e outros gases pesados.

Gás	<i>Massa Molar(MM)</i>
Nitrogênio(N ₂)	28.01
Dióxido de Carbono(CO ₂)	44.01
Sulfeto de Hidrógeno(H ₂ S)	34.08
Metano (C ₁)	16.04
Etano (C ₂)	30.07
Propano (C ₃)	44.1
i - Butano (C ₄)	58.12
n - Butano (C ₄)	58.12
i - Pentano (C ₅)	72.15
n - Pentano (C ₅)	72.15
pseudo C ₆ H ₁₄	84.0
M-C-Pentano	84.16
Benzeno	78.11
Ciclohexano	84.16
pseudo C ₇ H ₁₆	96.0
M-C-Hexano	98.19
Tolueno	92.14
pseudo C ₈ H ₁₈	107.0
C ₂ -Benzeno	106.17
M/P-Xileno	106.17
O-Xileno	106.17
pseudo C ₉ H ₂₀	121.0
pseudo C ₁₀ H ₂₂	134.0

Tabela 2.6: Valores de massa molar para as frações gasosas típicas utilizadas para o cálculo de densidade específica.

Razão Gás-Óleo (RGO), a Razão Líquido-Gás (RLG)

A razão gás óleo (RGO) diz respeito a quantidade de gás produzida por unidade de óleo em um determinado momento. Quando lidamos com compostos ricos em gases com pequena fração líquida utilizamos a variável razão líquido-gás (RLG) ou razão condensado-gás (RCG), que matematicamente corresponde ao inverso do RGO ($RLG=1/RGO$), Rose (2001).

Ao trabalhar com razões de fases líquidas em relação a gasosa (ou vice-versa). Estamos lidando principalmente com eventos coordenados pelas mudanças de fases e solubilidades de líquidos e gases que darão origem aos principais tipos de fluidos hidrocarbonetos: óleo típico, óleo volátil, condensado (gás retrógrado), gás úmido e gás seco. A obtenção dos parâmetros para determinar e quantificar o tipo de fluido e as suas proporções governadas pelos parâmetros RGO, RLG e RCG normalmente são feitas com base nas análises de laboratório de fluidos obtidos em condições semelhantes ao do prospecto (Ahmed, 2010).

Para fazer a classificação dos compostos e determinar o que é gás e o que é líquido devemos ter em mente algumas propriedades físicas dos compostos mais comuns em acumulações (ver a tabela 2.7).

Em outras palavras, o que chamamos de gases consiste basicamente de compostos como N₂, CO₂, H₂S e a fração (C1-C4) enquanto o que convencionalmente chamamos de líquido são as frações de maiores que C5. Note, entretanto, que eventualmente algumas frações de C5-C7 ainda ocorrem em solução nos gases, principalmente a medida que aumentamos a pressão e temperatura e por isso convencionalmente chamamos de frações líquidas ou pesadas de C7+. Os tipos de hidrocarbonetos variarão de composição proporção e estado físico para chegar às características

Nome	PF °C	PE °C	ρ g/L @ 25°C	EF @ 25°C
Nitrogênio(N ₂)	-210	-195	1.251	gás
Dióxido de Carbono(CO ₂)	-55.6	-78.5	1.977	gás
Sulfeto de Hidrógeno(H ₂ S)	-122	-74.9	0.668	gás
Metano (C ₁ H ₄)	-138	-164	0.668	gás
Etano (C ₂ H ₆)	-183	-89	1.265	gás
Propano (C ₃ H ₈)	-190	-42	1.867	gás
n - Butano (C ₄ H ₁₀)	-138	-1	2.493	gás
n - Pentano (C ₅ H ₁₂)	-130	36	626	Liquido
pseudo C ₆ H ₁₄	-95	69	659	Liquido
pseudo C ₈ H ₁₈	-57	125	703	Liquido
pseudo C ₁₀ H ₂₂	-30	174	730	Liquido

Tabela 2.7: Algumas propriedades físicas (PF= ponto de fusão, PE = Ponto de Ebulição, ρ = densidade e EF= Estado Físico) das principais substâncias típicas em acumulações de hidrocarbonetos Ahmed (2010).

que conhecemos.

Como as variações de pressão e temperatura ocorrem principalmente ao produzir petróleo, os grupos de química e engenharia voltados ao desenvolvimentos de acumulações são os *experts* na avaliação da composição dos hidrocarbonetos, por isso utilizamos aqui algumas das equações publicadas principalmente em *Handbooks* de Engenharia de Reservatório como por exemplo Ahmed (2010).

Trataremos inicialmente de acumulações líquidas com frações de gás dissolvido representado pela constante *RGO*.

Óleo Típico, Óleo volátil e a Razão Gás-Óleo (RGO)

Como a produção de óleo altera a composição do remanescente de hidrocarbonetos no reservatório, o RGO muda com a produção. Por isso ele é referenciado como RGO instantâneo (ou inicial) nos grupos de engenharia de reservatório (hua Han, 2001). Estas variações ocorrem em resposta ao efeito da produção que atua diretamente na composição dos fluidos, pressão e temperatura do reservatório, no nosso trabalho estamos estudando o caso de avaliação estático atual (pré-produção) que evoluiu ao longo do tempo geológico dinamicamente.

Conforme descrito em Ahmed (2016), podemos descrever o RGO matematicamente de acordo com a equação (2.43):

$$RGO = R_s + \left(\frac{k_{rg}\mu_o B_o}{k_{ro}\mu_g B_g} \right) \quad (2.43)$$

Onde:

- RGO é a razão gás óleo, dada em scf/STB;
- R_s é a solubilidade do gás, dada em scf/STB;
- k_{rg} é a permeabilidade relativa do gás (em milidarcy);
- k_{ro} é a permeabilidade relativa do óleo (em milidarcy);
- μ_g é a viscosidade do gás (em centipoise);
- μ_o é a viscosidade do óleo (em centipoise);
- B_g é o fator de formação de gás;

- B_o é o fator de formação de óleo;

Caso o reservatório esteja em uma pressão acima da pressão de bolha, todo o gás estará dissolvido no óleo, portanto $k_{rg}=0$ e $RGO=R_s$. Dessa forma, caso a acumulação esteja razoavelmente soterrada ($<3\text{km}$) o RGO inicial da acumulação será dado pela solubilidade do gás em óleo. Uma vez que a pressão do reservatório declina abaixo do ponto de bolha devido à extração de fluido, o gás é liberado dentro do reservatório (formando uma capa de gás) e o GOR, medido na superfície, excede R_s , então mais gás será produzido devido a maiores mobilidades de gás. O gás que foi liberado da solução é conhecido como gás livre.

A solubilidade do gás em óleo (R_s)

A solubilidade do gás mede a capacidade do óleo de manter o gás em solução a uma determinada pressão e temperatura. Por isso esta propriedade dependerá dos parâmetros ambientais, das características do óleo (e.g. API) e gás (γ_g), Abdul-Majeed (2016).

Para um dado óleo a uma dada temperatura, a solubilidade irá aumentar até que atinja a pressão de saturação (conhecida como pressão de bolha). A esta pressão todos os gases disponíveis na solução estão dissolvidos no óleo e a solubilidade chega ao seu maior valor (Abdul-Majeed, 2016). Para pressões inferiores ao ponto de bolha o gás forma uma fase separada e o ponto de bolha reduz gradualmente. Existem várias aproximações na literatura para aproximar a solubilidade do gás em óleo como por exemplo a aproximação de Glaso (1980):

$$R_s = \gamma_g \left[\frac{API^{0.989}}{(T - 460)^{0.172}} (10^u) \right]^{1.2255} \quad (2.44)$$

Onde:

$$u = 2.8869 - [14.1811 - 3.3093 \log(p)]^{0.5} \quad (2.45)$$

Glaso desenvolveu a equação com base em 45 amostras de óleo no Mar do Norte (uma pequena representatividade estatística). A solubilidade do gás em óleo pode ser calculada rigorosamente a partir de dados de PVT medidos na pressão e temperatura especificadas para o reservatório. A seguinte expressão relaciona a solubilidade do gás (R_s) com a densidade do óleo (ρ_o), gravidade específica do óleo (γ_o), a gravidade específica do gás (γ_g) e o fator de formação de óleo (B_o):

$$R_s = \frac{B_o \rho_o - 62.4 \gamma_o}{0.0136 \gamma_g} \quad (2.46)$$

Como a equação (2.46) é analiticamente rigorosa, o seu erro dependerá apenas da precisão dos dados disponíveis. Esta equação é de especial interesse para este trabalho porque ela relaciona explicitamente duas variáveis da equação de STOIP o B_o e o RGO (uma vez que a $R_s = GOR$ para $p < p_{bolha}$).

A pressão de bolha (p_b)

A pressão de bolha (p_b) de uma mistura de óleo e gás é definida como a maior pressão na qual uma bolha de gás é liberada pela primeira vez do óleo. Esta importante propriedade pode ser medida experimentalmente através da realização de um teste de expansão de composição constante (Ahmed, 2010).

Na ausência da pressão do ponto de bolha medida experimentalmente, é necessário fazer uma estimativa dessa propriedade do petróleo bruto a partir dos parâmetros de análogos. Existem vários gráficos e relações matemáticas para determinação de p_b foram propostas durante as últimas décadas. Estas correlações são essencialmente baseadas na suposição de que a pressão do ponto de bolha tem uma forte ligação com a solubilidade do gás R_s , gravidade do gás γ_g , grau API do óleo e temperatura T.

Standing and Katz (1942) propôs a utilização de um ábaco baseado em medidas de óleos de campos na Califórnia (EUA) que mais tarde foi aproximada matematicamente por Standing (1977) pela expressão empírica abaixo:

$$p_b = 18.2[(R_s/\gamma_g)^{0.83}(10)^{0.00091(T-460)-0.0125(API)} - 1.4] \quad (2.47)$$

É evidente na equação (2.46) a relação entre B_o e R_s . Entretanto, esta equação depende de parâmetros como permeabilidade, viscosidade e fator de formação de gás (B_g). Por isso utilizaremos a relação empírica 2.48 proposta por Kamari et al. (2016), que obteve boa correlação para um amplo espectro de óleos (tabela 2.8) completamente saturados, acima do ponto de bolha:

$$R_s = A + B - 15.849 \quad (2.48)$$

Onde :

$$B = 0.0064332 * (p + API) * (API * \gamma_g - 14.811)$$

$$A = 0.14624 * (p - API) + 802.44/p + ((2.727 * p - API * t)/(API - 995.53))^2$$

Propriedade	Unidade	Mínimo	Média	Máximo
R_s	SCF/STB	7,08	515,32	3298,66
γ_g	-	0,52	1	3,44
T_r	$^{\circ}F$	54,9	173,30	360,93
API	-	6,00	33,41	56,8
P_b	psi	58,01	1755,58	7127,01

Tabela 2.8: Tabela de valores com as estatísticas de entrada utilizadas por Kamari et al. (2016) para ajustar uma equação empírica de R_s e os demais parâmetros.

Condensados e Gás Úmido - Razão Líquido Gás (RLG)

Os condensados e gases úmidos são compostos principalmente de frações gasosas, (metano e a faixa C1-C4) mas incluem também frações líquidas de baixo ponto de ebulição (derivados de hidrocarbonetos C5 a C8) e, menos frequentemente, constituintes com um peso molecular mais elevado (tipicamente derivados de hidrocarbonetos C9 a C12). Em função da redução isotérmica da pressão de formação, parte dos constituintes condensados irá condensar na forma de condensado de gás (hua Han, 2001). A dissolução do condensado é chamada de condensação retrógrada porque é contrária ao comportamento de substâncias puras, que vaporizam quando a pressão cai abaixo da pressão de saturação sob condições isotérmicas (temperatura constante), Schlumberger (2019). Quando o gás se comporta desta forma em subsuperfície ele é chamado de gás retrogrado.

Reservatórios com uma quantidade relativa de 5 a 10 g de condensado por metro cúbico são geralmente referidos como reservatórios de condensado (hua Han, 2001). Por isso a faixa de produção líquida típica em reservatórios de gás condensado é 1-10 bbl/MMm³ de líquido (barris de líquido por milhão de metros cúbicos de gás).

As acumulações de gás e condensado estão geralmente localizadas nos estratos mais profundos sob alta temperatura e alta pressão (Ungerer et al., 1995; Sun et al., 2012) estas condições estão tipicamente em faixas de profundidade que permitem temperaturas de 200 a 400°F e 3000 e 8000psi, na faixa de 3km (que variará com o gradiente geotérmico da bacia).

Fator de Formação de Óleo (B_o)

O fator de formação de óleo, B_o , é definido como a razão do volume de óleo (mais o volume de gás em solução) nas condições de temperatura e pressão do reservatório e o volume de óleo em condições padrões (equação 2.49). O B_o é sempre maior ou igual a 1, uma vez que o óleo é sempre visto em condições de superfície (*seeps*) ou soterrado onde as pressões e temperaturas são alterados por efeito da compactação e gradiente geotérmico.

$$B_o = \frac{V_{T,P-Reservatorio}}{V_{T,P-padrao}} \quad (2.49)$$

Onde o B_o é dado em unidade de volume por unidade volume ($bbl/STB, m^3/m^3$). O fator de formação do óleo normalmente aumenta com a pressão até a pressão de bolha, para pressões abaixo da pressão de bolha o gás em solução é liberado e o B_o é reduzido gradativamente até que a pressão atinja o valor da pressão atmosférica onde o B_o assume o valor de 1.

Matematicamente o B_o normalmente é descrito como uma equação do tipo:

$$B_o = f(R_s, \gamma_o, \gamma_g, \rho_o, P, T) \quad (2.50)$$

Onde :

- γ_o é a gravidade específica ⁶ do óleo em relação à água;
- γ_g é a gravidade específica média dos gases (em relação ao ar);
- R_s é a solubilidade do gás em scf/STB nas condições de reservatório (*standard cubic feet / stock tank barrel*);
- ρ_o é a densidade do óleo nas condições do reservatório;
- P é a pressão nas condições de reservatório e;
- T é a temperatura nas condições de reservatório.

Por vezes os valores de P e T são implícitos na equação, uma vez que os mesmos influenciam os valores de densidade e solubilidade dos gases (Ahmed, 2010). Os valores de γ_o podem ser obtidos a partir da massa molar e proporções dos constituintes do óleo e gás em condições padrões comparada com a densidade da água e ar respectivamente. Onde a densidade do ar é dada por 28.96 e a densidade da água é de aproximadamente 62.4 lb/ft³ o que nos leva às seguintes equações:

$$\gamma_o = \frac{\rho_o}{62.4}, 60^\circ/60^\circ \quad (2.51)$$

A gravidade específica do gás é discutida na seção que discute o cálculo de B_g , a solubilidade do gás foi discutida na seção referente à razão gás óleo (RGO).

A gravidade específica do óleo pode ser estimada alternativamente pelo grau API ou vice-versa. Na verdade, o grau API é a medida padrão na indústria para classificação de óleos. O API pode ser calculado pela equação (2.52), e inversamente a gravidade específica pode ser estimada caso o API esteja disponível.

$$API = \frac{141.5}{\gamma_o} - 131.5 \quad (2.52)$$

Os valores típicos de grau API vão de 10^o para óleos pesados, a 50^o API para óleos leves/condensados.

Ahmed (2010) fornece uma excelente revisão sobre a modelagem de fluidos em reservatórios incluindo diversas aproximações úteis de B_o para acumulações em pressões menores do que a pressão de bolha. Neste texto citarei apenas a aproximação utilizando balanço de materiais do fluido podemos à equação (2.53). Há diversas aproximações disponíveis na literatura dadas por correlações distintas: *Standing's correlation*, *The Vasquez-Beggs correlation*, *Glaso's correlation*, *Marhoun's correlation* e *The Petrosky-Farshad correlation*.

⁶A gravidade específica é a densidade de uma substância em relação a outra

$$B_o = \frac{62.4\gamma_o + 0.0136R_s\gamma_g}{\rho_o} \quad (2.53)$$

Assim como já discutido anteriormente, a precisão da equação (2.53) depende apenas da acurácia em determinar as constantes envolvidas, sendo ela a que tem maior rigor analítico dentre as aproximações disponíveis.

Para uma abordagem mais simplificada pode-se utilizar correlação de Arp reportada em Frick (1962) para estimar o fator de volume de formação de óleo quando as propriedades solubilidade do gás são conhecidas.

$$B_o = 1,05 + 0,0005.R_s \quad (2.54)$$

Outra opção estimada com base em dados de óleo leve no Egito também bastante simplificada, é a expressão empírica de Hanafy et al. (1997):

$$B_o = 1.079 + 0.0006.R_s \quad (2.55)$$

Note que os valores das constantes e variáveis utilizadas aqui estão em unidades de medidas tipicamente aplicadas na indústria americana, a temperatura em °F e a pressão em psi, caso seja conveniente a conversão do B_o de bbl/STB para m_3/m_3 possuem a mesma razão de proporcionalidade não precisando de conversão, porém não é incomum medidas em SCF/STB que devem ser calculadas utilizando a constante de proporcionalidade:

$$B_o(m_3/m_3) = 5.61B_o(SCF/STB) \quad (2.56)$$

Valores típicos de B_o giram em torno de 1 a 1.1 para óleo extra pesados, 1.2-1.5 para óleos típicos e valores que chegam a superar 3 em condensados (veja na tabela 2.9).

Algumas características típicas de óleo podem ser vistas na tabela 2.9, que podem ser utilizadas caso nenhuma informação sobre análogos esteja disponível.

Fluidos de Petróleo e Suas Características					
Características	Óleo			Gás	
	Óleo Pesado e Tar	Black Oil	Óleo Volátil	Gás Condensado	Gás Úmido e Gás Seco
Cor do Óleo	Preto	Marrom a Verde Claro	Esverdeado a Laranja	Laranja a Branco	Branco a Incolor
Peso Molecular de Fluido Inicial	210+	70 a 210	40 a 70	23 a 40	<23
Stock Tank Oil Gravity API	5 a 15	15 a 45	42 a 55	45-60	45+
Fração Maior que C7 mol%	>50	35 a 50	10 a 30	1 a 6	0 a 1
GOR inicial scf/STB	0-200	200-900	900-3500	3500-30000	30000+
Boi Inicial (FVF) RB/STB	1 a 1,1	1,1 a 1,5	1,5 a 3,0	3 a 20	20+
Temperatura Típica de Reservatório F	90 a 200	100 a 200	150 a 300	150 a 300	150 a 300
Pressão de Saturação em psia	0 a 500	300 a 5000	3000 a 7500	1500 a 9000	-
Óleo Volatil/Razão de gás STB/MMscf	0	0 a 10	10 a 200	50 a 300	0 a 50
OOIP STB/acre-ft (bulk)	1130 a 1240	850 a 1130	400 a 850	60 a 400	0 a 60
OGIP, Mscf/acre-ft (bulk)	0 a 200	200 a 700	300 a 1000	500 a 2000	1000 a 2200

Tabela 2.9: Dados gerais médios de hidrocarbonetos. Fonte: www.petrowiki.com.

Fator de Recuperação

O fator de recuperação é o elemento que deverá representar a razão entre a quantidade de hidrocarbonetos que pode ser extraído em subsuperfície e o valor total de hidrocarbonetos presentes originalmente na acumulação. Ele será, portanto, o elemento de ligação entre os recursos *in place* e as reservas ou volume recuperável potencial que serão efetivamente produzidos.

Um fato interessante é que normalmente mais óleo é deixado para trás nos campos de petróleo do que será recuperado até o final da exploração da acumulação, ou seja, o fator de recuperação é

em geral menor do que 50% (Shepherd, 2009). Já nos campos de gás a eficiência de recuperação será normalmente maior.

O fator de recuperação (F_r) é determinado por fatores geológicos, fatores relacionados ao conceito de desenvolvimento do campo e fatores relacionados à operação. Os fatores relacionados à engenharia (conceito de desenvolvimento e operação) são guiados pelo problema geológico as tecnologias disponíveis e as premissas econômicas que em última análise definirão a viabilidade do projeto. Alguns dos pontos chaves dos fatores supracitados são ilustrados na Tabela 2.10:

Fatores Geológicos	Desenvolvimento	Operação
Porosidade	N de Poços	Eficiência de Injeção
Permeabilidade	Conceito de desenvolvimento	Otimização da planta
Conectividade	Razão de Injetores/Produtores	Intervenção nos poços
Propriedade dos fluidos	Tamanho da facilidade de produção	Paradas de Manutenção
Saturação de fluidos	Tempo de exploração	Gerenciamento do campo

Tabela 2.10: Alguns fatores que impactam no fator de recuperação de uma acumulação.

Em *plays* maduros o fator de recuperação pode ser estimado com base em *trends* de fator de recuperação feitos para descobertas com mesmas características (como idade e/ou soterramento) de forma a dar representatividade Amado (2013). Um exemplo pode ser visto na Figura 2.36, com exemplos de diferentes *plays* no GOM.

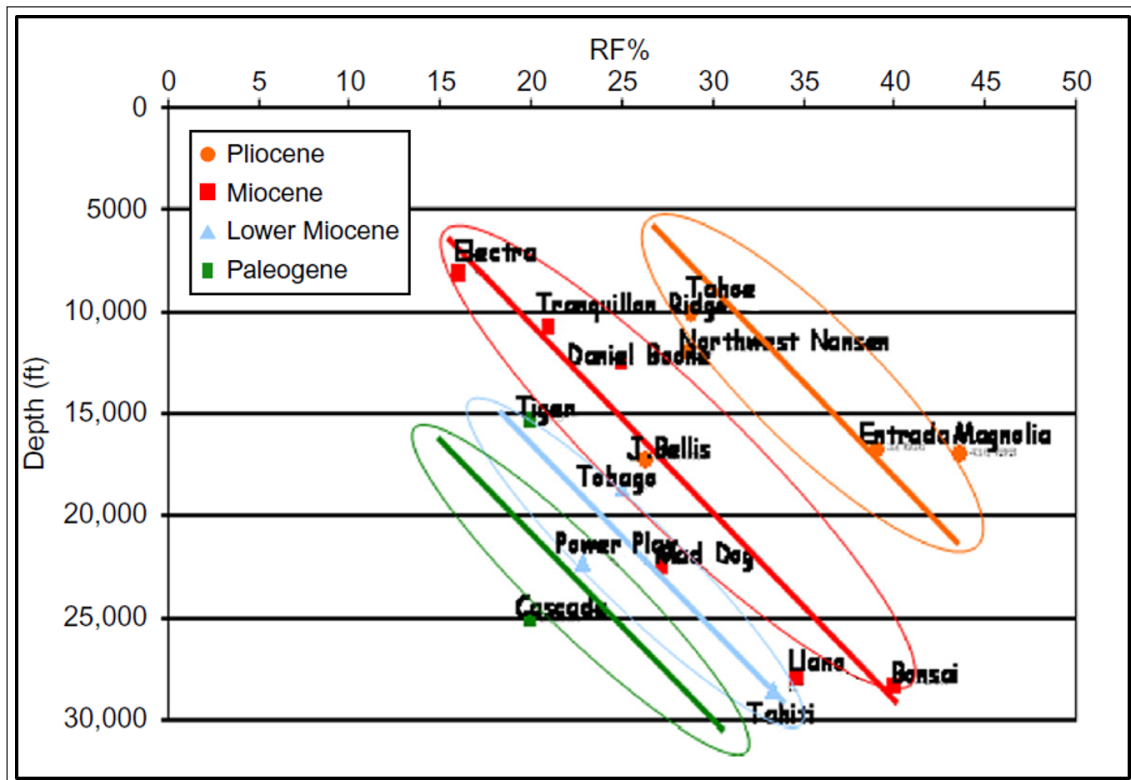


Figura 2.36: Exemplo de fator de recuperação vs profundidade para diferentes campos no Golfo do México portadores de óleo (note que eles estão agrupados por idade). Fonte: Amado (2013).

2.3 Associação de Variáveis

De acordo com Rodgers and Nicewander (1988) e Beaumont (2012), sir Francis Galton (1822-1911) propôs pela primeira vez o termo "co-relação" entre duas variáveis em 1885, quando estudava a relação entre altura e comprimento do antebraço de 348 homens.

A associação de variáveis começou muito antes de 1885, poderíamos dizer que esta relação é tão antiga quanto o comércio onde há uma relação entre preço e produto. Ao estabelecer uma relação de dependência entre duas variáveis estamos criando uma associação entre elas.

Muitas publicações da estatística nas primeiras décadas do século 20 se preocuparam em medir a correlação de variáveis dentro de grandes *datasets*, com o objetivo de entender os fatores que controlam a ocorrência de determinado evento (de maneira bastante descritiva). O famoso artigo "Notas sobre a história da Correlação", (Pearson, 1920) fornece uma excelente histórica revisão sobre o assunto.

Quando avaliamos dois eventos possivelmente associados tentamos verificar uma simetria em suas tendências que pode ser um indicativo de que a relação entre elas pode ser positiva (eventos que ocorrem com aumento ou diminuição conjuntos, ou negativos com comportamento oposto). Utilizando um gráfico hipotético (Figura 2.37) a relação entre temperatura e a venda de gelo, e temperatura e a venda de aquecedores em uma cidade podemos ilustrar esta tendência.

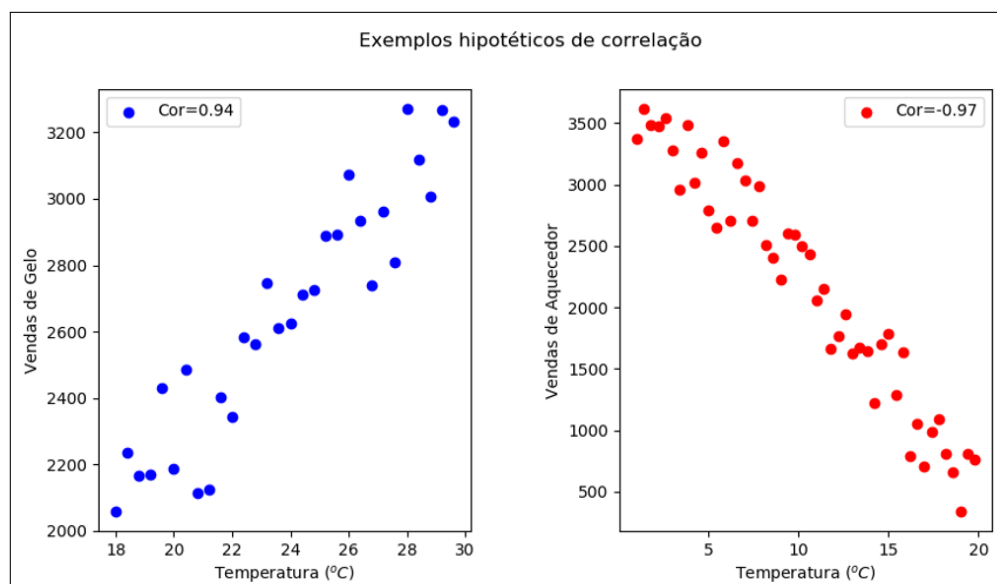


Figura 2.37: Ilustração hipotética da relação entre a temperatura e a venda de gelo (esquerda), e temperatura e a venda de aquecedores em uma cidade (direita).

Espera-se que com o aumento da temperatura as vendas de gelo aumentem e ocorra uma diminuição na venda de aquecedores. Para expressar relações lineares como esta de forma numérica utilizamos os coeficientes de correlação.

A correlação pode ser definida como uma família de métodos para resumir a intensidade e direção da relação entre duas variáveis (Whitford, 2005). A correlação como conceito serve como base para declarações sobre relacionamentos e associações estatísticas simples. Ela pode ser usada em testes de hipóteses e métodos mais avançados para avaliar separadamente múltiplos parâmetros num problema de alta complexidade. Estas variáveis podem ser puramente matemáticas, relacionadas a ações antrópicas, descrever fenômenos físicos etc.

As correlações também ocorrem na geologia do petróleo com frequência. Se fizermos uma análise regional dos campos descobertos em uma bacia, utilizando valores médios de propriedades de acumulações, verificaríamos que algumas propriedades por vezes apresentam comportamento proporcional.

Isto pode ser visto, por exemplo, na relação porosidade (ϕ) vs permeabilidade (k) em reservatórios uma mesma idade, Fator de Formação de óleo (B_o) vs Razão gás-óleo (RGO).

Estas relações não são coincidentes, verifica-se que um aumento no espaço poroso tende a tornar os poros conectados e, portanto, a relação entre porosidade e permeabilidade aparece naturalmente nestas condições, de maneira análoga um aumento no volume de frações leves dissolvidas no óleo (RGO) permite que ele seja mais comprimido e dessa forma, tem impacto direto no fator de formação de óleo (B_o) estimado para aquela acumulação.

Mesmo em reservatórios complexos como os carbonáticos estas relações são rotineiramente reportadas na literatura. Isto é reportado em Favoreto et al. (2016) e Muniz and Bosence (2017) (Figura 2.38) que ilustram a relação de porosidade com a permeabilidade em carbonatos albianos e coquinas, respectivamente, na Bacia de Campos.

Por outro lado, o uso destas relações deve ser feito com cuidado uma vez que a relação entre variáveis pode não ser uniforme ao longo de uma bacia. Exemplos clássicos em que estas premissas falham são por exemplo em reservatórios fraturados onde mesmo com porosidades relativamente baixas é possível ter permeabilidades relativamente altas. Também ocorrem reservatórios carbonáticos com porosidade *vuggy* ou arenitos cimentados classificados como fechados (*tight sands*), onde o espaço poroso existe, mas são fracamente conectados e portanto tem pouca relação com a permeabilidade.

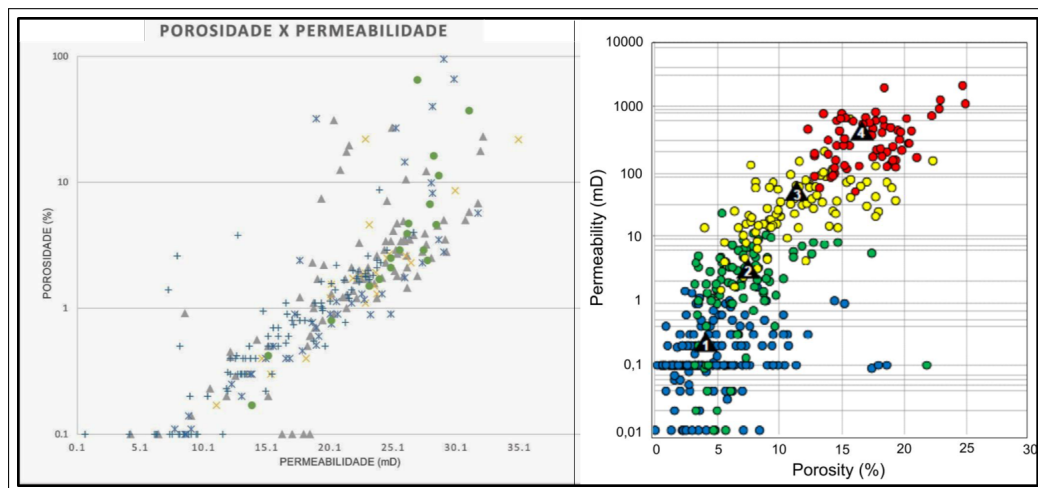


Figura 2.38: Crossplots de permeabilidade vs porosidade em reservatórios carbonáticos, a esquerda em carbonatos da Formação Quissamã, Bacia de Campos (Favoreto et al., 2016), a direita em coquinas da idade Jiquiá também na Bacia de Campos (Muniz and Bosence, 2017). Em ambos os gráficos as diferenças nos pontos estão relacionadas a diferentes fácies.

O cenário pode ser ainda mais complexo quando mais de duas variáveis se relacionam entre si. Um exemplo muito claro disso é o fator de recuperação de óleo em um compartimento do reservatório em produção, que depende da permeabilidade (que por vezes traz uma relação com a porosidade), ao fluido a ser escoado ao longo do reservatório.

Ao trabalhar com fator de recuperação médio para uma acumulação inteira, mais variáveis aparecem no sistema como o grau de compartimentação da estrutura e a análise da engenharia que definirá quantidade ótima de poços para o máximo aproveitamento econômico. Estes elementos serão superpostos e dessa forma a relação de linearidade das variáveis será reduzida à medida que o número de variáveis entram no sistema.

2.3.1 Coeficientes de Correlação

A covariância e o coeficiente de correlação de Pearson são as versões mais populares para fazer estimativa de relação entre duas variáveis (Whitford, 2005). Mas, só são apropriados dadas

suposições específicas sobre a relação entre as variáveis envolvidas e as escolhas feitas sobre a sua medição. Matematicamente, a covariância é dada pela expressão:

$$Cov_{AB} = \sum_{i=1}^n \{E[(A_i - \mu_A)(B_i - \mu_B)]\} \quad (2.57)$$

onde $E[\]$ é o operador do valor esperado da variável, **normalmente** dado por sua média:

$$E[A] = \frac{1}{n} \sum_{i=1}^n E[A_i] = \mu_A \quad (2.58)$$

Considerando que $E[A] = \mu_A$ e $E[B] = \mu_B$, temos:

$$Cov_{AB} = \sum_{i=1}^n \{E[(A_i - E[A])(B_i - E[B])]\} \quad (2.59)$$

Aplicando a propriedade distributiva na equação (2.59), obtemos:

$$Cov_{AB} = \sum_{i=1}^n \{E[A.B - A.E[B] - B.E[A] + E[A].E[B]]\} \quad (2.60)$$

Aplicando o operador esperança matemática nos elementos da equação 2.60, chegamos a:

$$Cov_{AB} = \sum_{i=1}^n \{E[A.B] - E[A].E[B] - E[B].E[A] + E[A].E[B]\} \quad (2.61)$$

Simplificando obtemos:

$$Cov_{AB} = \sum_{i=1}^n \{E[A.B] - E[A].E[B]\} \quad (2.62)$$

Ao reordenarmos de maneira consistente distribuições suficientemente amostradas, com mesma geometria, estamos adicionando linearidade a relação A e B. Portanto, mudamos diretamente o termo $E(A.B)$. Por outro lado, no caso em que A e B são independentes:

$$\sum_{i=1}^n \{E[A.B]\} = \sum_{i=1}^n \{E[A].E[B]\} \quad (2.63)$$

E assim:

$$Cov_{AB} = 0 \quad (2.64)$$

Note que o valor de Cov_{AB} não é afetado pelo número de amostras, mas pela escala de valores de ambas as distribuições. Para normalizar os valores de covariância, dividimos o valor calculado pela variâncias (Var) das distribuições com dispersões diferentes. Este coeficiente é o coeficiente de correlação de Pearson dado por:

$$Cor_{AB} = \frac{Cov_{AB}}{\sqrt{Var(A).Var(B)}} \quad (2.65)$$

e a variância é dada por:

$$Var(A) = E[(A_i - \mu_A)^2] \quad (2.66)$$

Note que a partir da equação (2.66) verificamos que a variância é um caso particular da covariância (Cov(A,B)) em que A=B. Ao analisarmos as equações 2.57 à 2.65, verificamos que a correlação não depende diretamente da ordem das amostras em uma das distribuições mas sim da ordem de ocorrência conjunta das mesmas. Em outras palavras, estabelecidos os pares de

variáveis é possível embaralhar ou organizar as distribuições mantendo os pares e assim manter o mesmo nível de correlação. Por outro lado, ao mudar a ordem de apenas uma das distribuições possivelmente a correlação será afetada.

Observe na Figura 2.39 que o coeficiente de correlação de Pearson é limitado aos valores 1 e -1. Este coeficiente reflete a intensidade (ou inversamente o grau de ruído) e a direção de um relacionamento linear (linha superior da Figura 2.39). Note que, o grau de inclinação desse relacionamento (linha do meio da Figura), tem uma relação com a faixa de valores das variáveis mas não com a correlação. Outros aspectos do relacionamento entre as variáveis como a não-linearidade ou curvatura (parte inferior) também não são corretamente representados por este coeficiente.

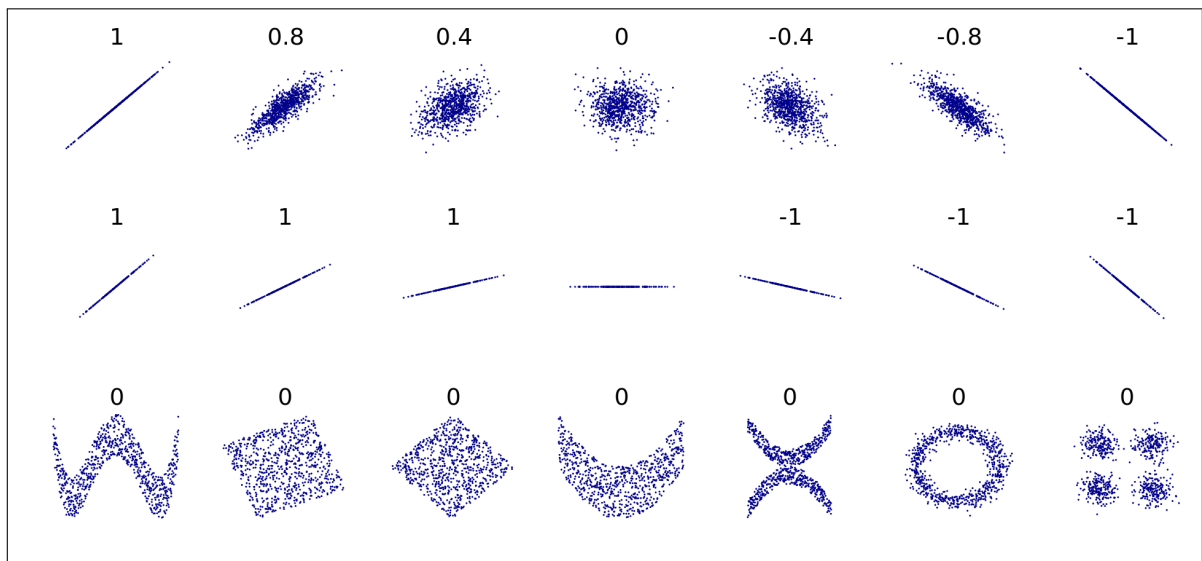


Figura 2.39: Exmplos de coeficiente de correlação de Pearson. Vários conjuntos de pontos (x, y), com o coeficiente de correlação de Pearson entre x e y para cada conjunto. Observe que a correlação reflete o ruído e a direção de um relacionamento linear (linha superior). Ela não é sensível à inclinação desse relacionamento (meio). Este coeficiente não é bom para medir relacionamentos não-lineares (parte inferior). Note que a figura no centro tem uma inclinação de 0, mas nesse caso o coeficiente de correlação é indefinido porque a variância de Y é zero. Fonte : https://en.wikipedia.org/wiki/Correlation_and_dependence.

Se a relação entre duas variáveis A e B for descrita por uma equação de segundo grau em forma de "U" ainda que sem nenhuma dispersão o resultado será uma correlação próxima a zero.

Ao lidar com distribuições com relações lineares podemos utilizar o coeficiente de correlação de Pearson como a intensidade da relação entre duas variáveis. Considerando este cenário, podemos utilizar a escala proposta por Whitford (2005) para estimar quão boa é a correlação entre duas variáveis (Tabela 2.11):

$0.0 \leq Cor_{A,B} < 0.3$	não há correlação entre as variáveis
$0.3 \leq Cor_{A,B} < 0.5$	correlação baixa
$0.5 \leq Cor_{A,B} < 0.7$	correlação moderada
$0.7 \leq Cor_{A,B} < 0.9$	correlação alta
$0.9 \leq Cor_{A,B} < 1$	correlação muito alta

Tabela 2.11: Tabela com escala de correlação Pearson e sua possível interpretação (atente que os valores estão em módulo). Fonte: Whitford (2005).

Em distribuições lognormais, os valores limítrofes do coef. de correlação de Pearson já não são mais necessariamente -1 e 1. Este desequilíbrio acontece, pois, a função é altamente assimétrica

em frequências e valores. A dispersão dos poucos valores de P10 fazem com que os valores de $E(A.B)$ fiquem bastante elevados e limitam principalmente os valores da parte negativa da correlação.

É possível verificar este desequilíbrio com desenvolvimento matemático analítico da expressão do coeficiente de correlação Pearson (2.65) conforme verificado em (Demirtas and Hedeker, 2011) (com pequeno erro no numerador, corrigido aqui).

Sejam $N_A \sim N(\mu_A, \sigma_A^2)$ e $N_B \sim N(\mu_B, \sigma_B^2)$ duas distribuições normais de média μ e variância σ^2 subjacentes às distribuições lognormais A e B , funções lognormais geradas por $A = \exp(N_A)$ e $B = \exp(N_B)$. Substituindo A e B na equação 2.65 obtemos:

$$Cor_{AB} = \frac{e^{\delta_{AB} \cdot \sigma_A \cdot \sigma_B} - 1}{\sqrt{(e^{\sigma_A^2} - 1) \cdot (e^{\sigma_B^2} - 1)}} \quad (2.67)$$

Onde Cor_{AB} é o coeficiente de correlação de Pearson entre as funções lognormais, δ_{N_A, N_B} é o coeficiente de correlação entre as distribuições normais N_A e N_B de B e A . Demirtas and Hedeker (2011), também foram acertos ao produzir o gráfico de variação Cor_{AB} em função σ_B utilizando pequenos incrementos de σ_A mostrando que quanto maior forem as diferenças dos desvios padrões menor será a faixa de correlações possíveis. O gráfico foi reproduzido aqui na Figura (2.40).

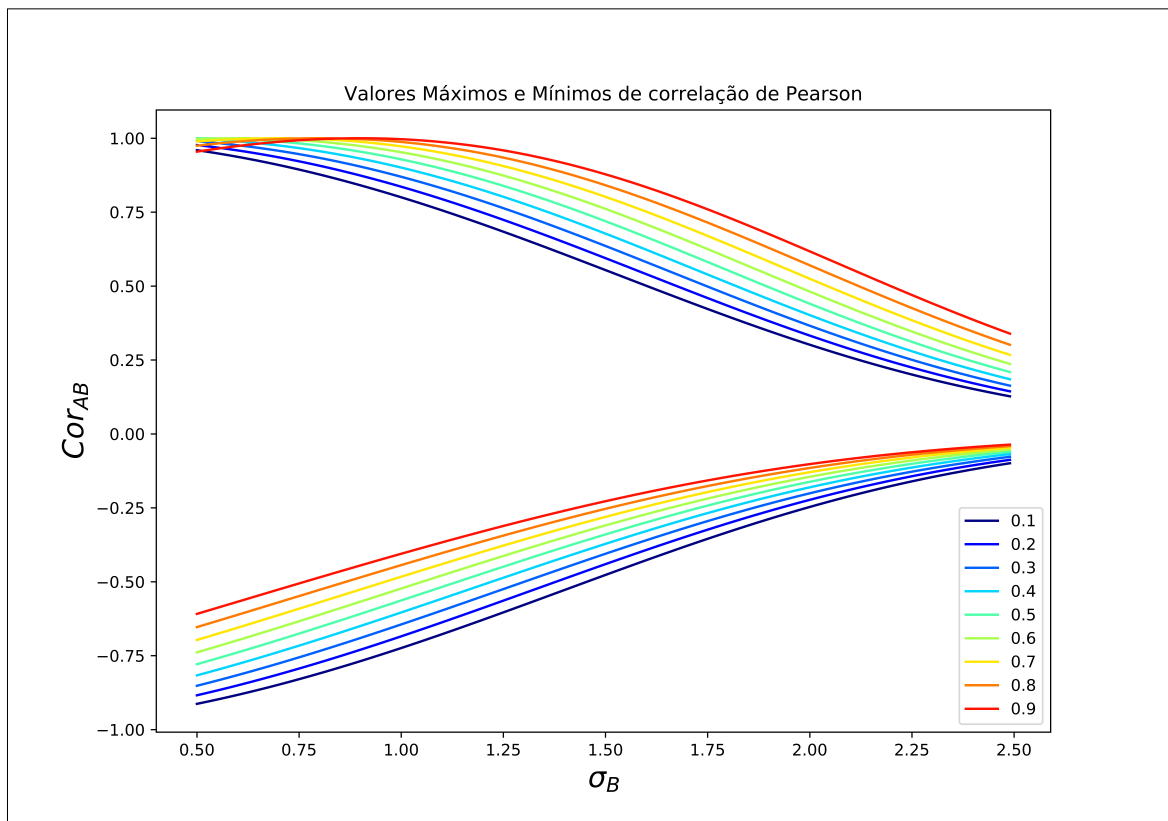


Figura 2.40: Gráfico de coeficientes de correlação Pearson (Cor_{AB}) versus o desvio padrão (σ_B) da normal subjacente da distribuição B (N_B). As curvas superiores representam as correlações máximas das distribuições lognormais. Elas foram obtidas a partir do coeficiente de correlação das distribuições subjacentes $\delta_{N_A, N_B} = 1$. As curvas inferiores representam correlações mínimas, obtidas a partir de $\delta_{N_A, N_B} = -1$. As cores variam para diversos desvios padrões σ_A das normais subjacentes N_A simuladas. Fonte: modificado de Demirtas and Hedeker (2011).

O valor de correlação é uma medida indireta do grau de linearidade e efetividade da relação entre as variáveis. A incerteza decorrente da ausência de dados para um determinado *play* gera altos valores da razão P10/P90 que tem impacto direto nos valores de desvio dispersão

de valores e portanto inclinação da reta de correlação. Esta característica deverá ser comum principalmente nas etapas exploratórias iniciais de prospectos em áreas de fronteira e em áreas onde a resposta sísmica em termos de DHI não é conclusiva. Nestas áreas as oportunidades são baseadas eminentemente em modelo conceitual.

Como a medida de incerteza no método probabilístico de avaliação de prospectos é dado pela razão (P10/P90), e os valores de σ são calculados com base na normal padrão, os valores absolutos das variáveis não exercem nenhum papel importante nos limites de correlação de Pearson e incerteza.

Tome, por exemplo, um prospecto com porosidade média de 22,3% ($P10 = 27\%$, $P90 = 18\%$ e $\sigma = 0.158$) tem uma razão $P10/P90 = 1.5$ e desvio padrão subjacente $\sigma = 0.158$. Analogamente tomemos o parâmetro área com média de 124km ($P10 = 150km$ e $P90 = 100km$) também tem uma razão $P10/P90 = 1.5$ e desvio padrão subjacente $\sigma = 0.158$. Duas variáveis com médias em escala de grandeza bastante diferentes, mas ao ter os mesmos valores de $P10/P90$ terão desvio padrão das normais subjacentes (σ) iguais.

A prova matemática desta propriedade é simples e feita a seguir:

$$\sigma = \frac{\ln(P10) - \ln(P90)}{2 * 1.282} \quad (2.68)$$

Se o P10 pode ser reescrito como n.P90 temos:

$$\sigma = \frac{\ln(n.P90) - \ln(P90)}{2 * 1.282} \quad (2.69)$$

Utilizando a propriedade multiplicativa dos logaritmos:

$$\sigma = \frac{(\ln(n) + \ln(90)) - \ln(P90)}{2 * 1.282} \quad (2.70)$$

Logo em casos em que P10 é um múltiplo qualquer n de P90, temos que:

$$\sigma = \frac{\ln(n)}{2 * 1.282} \quad (2.71)$$

A consequência prática desta relação é que conhecida a razão $p10/p90$ e um percentil é possível definir toda a distribuição lognormal teórica. Para trabalhar com a limitação entre valores máximos e mínimos algumas abordagens podem ser utilizadas:

- Utilizar valores de correlação entre os limites definidos.
- Utilizar valores de correlação das distribuições normais subjacentes que serão posteriormente transformados analiticamente para as distribuições de estudo.
- Utilizar valores de correlação não lineares.

O algoritmo GSC - *Generate Sort and Correlate*

Conforme comentado anteriormente, os valores de correlação para distribuições assimétricas (como a lognormal) com desvios diferentes tem valores máximos e mínimos de coeficiente de correlação de Pearson diferentes de 1.

Para tratar este problema podemos usar o algoritmo *Generate, Sort and Correlate* (GSC) que foi proposto inicialmente em Demirtas and Hedeker (2011) com o objetivo de estimar os valores limítrofes para distribuição em análise. O algoritmo funciona em três passos (Figura (2.41) conforme seu nome:

- *Generate* - Nesta etapa é possível utilizar amostras já adquiridas de campo ou distribuições geradas por simulações;

- *Sort* - As distribuições são ordenadas de maneira crescente ou decrescente em mesma direção ou em direções opostas (para obter o valor mínimo de correlação);
- *Correlate* - Faça o cálculo da correlação entre as distribuições em mesma direção (para obter o valor máximo de correlação) e em direções opostas (para obter o valor mínimo de correlação);

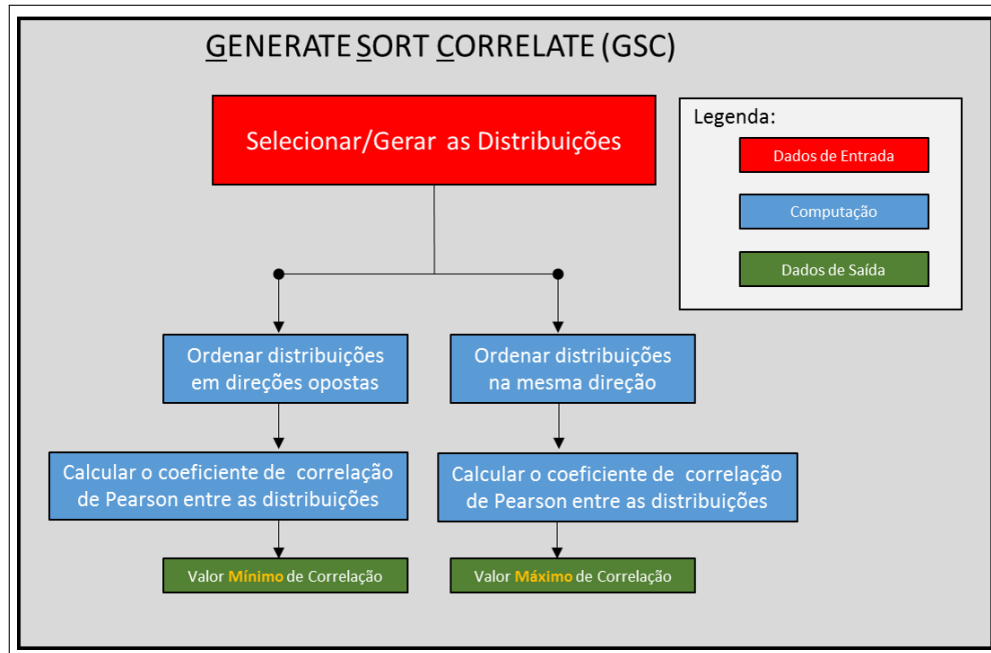


Figura 2.41: Diagrama esquemático do algoritmo *Generate Sort and Correlate*, (GSC) proposto por Demirtas and Hedeker (2011).

Note que, neste algoritmo estamos utilizando um esquema dependente da correlação de Spearman (ou rank correlation) para determinar os valores máximos e mínimos de correlação de Pearson.

Algumas das vantagens advogadas pelos autores é que esta abordagem funciona com vários tipos de distribuições que inclusive não precisam ser do mesmo tipo entre si, portanto é flexível.

Contanto que a distribuição seja bem amostrada, a solução analítica das distribuições e de sua correlação não precisa ser conhecida ou aproximada. Ainda assim, é possível obter bons valores máximos e mínimos de correlação para as distribuições.

Outra vantagem é que consiste em uma solução rápida pois depende de operações triviais (ordenação e correlação), Demirtas and Hedeker (2011) .

Medidas de Correlação não-lineares

O coeficiente de correlação de Spearman (r) é um caso especial do coeficiente de correlação Pearson para quando ambas as variáveis são medidas quantitativamente como escalas ordinais (Whitford, 2005). Neste caso um critério determinará as posições definindo um *rank* e por este motivo este coeficiente é chamado de *Spearman Rank Correlation Coefficient* ou simplesmente *rank correlation*.

Este último ponto pode ser utilizado para avaliar a correlação entre duas distribuições e também para avaliar o grau de organização entre suas amostras. Matematicamente o coeficiente de Spearman, tradicionalmente atribuído à letra ρ , é dado por:

$$\rho_{AB} = \frac{Cov_{R_A, R_B}}{\sqrt{Var(R_A) \cdot Var(R_B)}} \quad (2.72)$$

Onde R_A é a série de posições de amostras ordenadas da distribuição A, e R_B a série de amostras ordenadas da distribuição B.

Ao utilizar o coeficiente de correlação de Spearman, ou *rank correlation*, o que fazemos é normalizar a unidade de cada amostra para 1 e o incremento também para 1, esta operação lineariza qualquer distribuição independente da sua forma e por este motivo este coeficiente de correlação é chamado de "não-paramétrico" (Whitford, 2005).

Note que, estaremos de alguma forma, dando o mesmo peso para amostras de valores pequenos (que tem grande frequência nas distribuições log-normais) e valores grandes que tem frequência bastante reduzida.

Um conceito importante que permeia a relação de posições é o de comportamento comonotônico. Ao lidar com distribuições univariadas a comonotonicidade diz respeito à ordenação crescente dos elementos da distribuição (Cherubini et al., 2004). Tome por exemplo, uma distribuição X dispostos de maneira que os elementos da distribuição $X = [x_1 < x_2, \dots, x_{n-1} < x_n]$. No caso das distribuições bivariadas as duas distribuições são organizadas de maneira igualmente crescente e ordenada (maximizando o coeficiente de correlação de Spearman $\rho = 1$).

No caso de distribuições bivariadas, há também o conceito de distribuições contra-comonotônica (Cherubini et al., 2004). Neste caso, as distribuições são organizadas de maneira estritamente inversas, em que uma distribuição é organizada de maneira crescente e a outra de maneira decrescente (minimizando o coeficiente de correlação de Spearman $\rho = -1$).

Outra medida de correlação não-paramétrica é o coeficiente de correlação de Kendall, este coeficiente avalia o grau de similaridade entre dois conjuntos de posições (assim como o coeficiente de correlação de Spearman) dados a um mesmo conjunto de dados (Whitford, 2005). A sua medida é feita com base na diferença de probabilidade de concordância e discordância entre os pares de duas variáveis aleatórias (e.g. X e Y). Matematicamente o coeficiente de correlação de Kendall pode ser descrita pela equação 2.73.

$$\tau = \frac{2(C - D)}{n(n - 1)} \quad (2.73)$$

Onde n é o número de amostras em cada um dos conjunto de dados, C é o número de pares concordantes e D é o número de pares discordantes. O coeficiente de correlação de Kendal τ pode ser interpretado como a diferença entre a probabilidade de dois pares de objetos estarem na mesma posição [denotada P(igual)] e a probabilidade desses objetos estarem em uma posição diferente [denotado P(diferente)] (Whitford, 2005). Matematicamente esta relação pode ser interpretada como:

$$\tau = [(X_1 - X_2).(Y_1 - Y_2) > 0] - P[(X_1 - X_2).(Y_1 - Y_2) < 0] \quad (2.74)$$

Onde (X_1, Y_1) e (X_2, Y_2) são vetores aleatórios independentes e identicamente distribuídos com a mesma CDF (*Cumulative Density Function*) ou simplificadamente:

$$\tau = P(\text{igual}) - P(\text{diferente}) \quad (2.75)$$

De acordo com McNeil et al. (2005), é um pouco melhor usar essas medidas do que o coeficiente correlação linear. Em sua opinião, no entanto, deve-se escolher um coeficiente para a estrutura de dependência que reflete um conhecimento mais detalhado do problema em questão, em vez de resumir a dependência com um único número como correlação linear ou correlação de posições. O coeficiente de correlação de Kendall também é uma medida utilizada para modelar a dependência através das cópulas, discutida a seguir.

2.4 Métodos para Obter Séries Aleatórias Correlatas

O uso de séries aleatórias é parte importante na configuração de algoritmos que tentam reproduzir cenários, para fenômenos da natureza que podem ser considerados aleatórios.

A primeira questão quando tratamos a modelagem de eventos aleatórios numericamente é: o que é/são número(s) aleatório(s)? Números aleatórios essencialmente não existem (e.g 1254 é um número aleatório? não). Aleatoriedade advém da nossa incapacidade em prever o(s) próximo(s) número(s) de uma série. Portanto, quando tratamos com a aleatoriedade estamos falando de uma série de números não previsíveis e aí sim, 1254 pode ser o próximo número aleatório mas, o seu sorteio não o classifica como tal.

Os computadores como conhecemos hoje, são máquinas por si só determinísticas, eles seguem uma série de instruções que se repetidas deverão dar o mesmo resultado e, portanto, uma máquina não é capaz de gerar números aleatórios sozinha.

Um gerador de números aleatórios real é um sistema que gera números a partir de uma fonte verdadeira de aleatoriedade. Estas fontes podem ser algo físico, como um contador Geiger, onde os resultados são transformados em números aleatórios. Entretanto, este método necessitaria de um hardware adicional. Ele não seria reproduzível e, não seria rápido para que pudéssemos realizar 40.000 iterações para vários parâmetros (como fazemos na modelagem de prospectos) em tempo factível.

Note que em muitos problemas que enfrentamos nós não precisamos necessariamente de aleatoriedade real. Este é o caso ao fazer avaliações probabilísticas de prospectos, principalmente por lidar com distribuições definidas. Em vez disso, podemos usar a pseudo-aleatoriedade.

A pseudo-aleatoriedade consiste na amostragem de números que "parecem aleatórios", mas foram gerados usando um processo determinístico (Rukhin et al., 2001). Para simular a aleatoriedade em computadores utilizam-se 2 elementos: (1) Uma semente que é o ponto de partida do algoritmo e; (2) Um algoritmo que gere uma sequência de números chamado de gerador de números pseudo-aleatórios.

Se a semente não for declarada explicitamente, o gerador de números pseudo-aleatórios poderá usar a hora atual do sistema em segundos ou milissegundos como a semente combinada com outros elementos como temperatura da CPU etc. De posse da semente, os números são gerados em uma sequência determinística que dependerá do algoritmo específico implementado.

Métodos baseados em algoritmos pseudo-aleatórios têm vantagens importantes, uma delas é que são reproduzíveis (característica advinda do caráter determinístico do processo de geração de amostras). Isto significa que ao usar a mesma semente devemos ter uma mesma sequência de números permitindo a repetição de experimentos e, portanto, contanto que o algoritmo e o tamanho do vetor sejam conhecidos só necessitamos da semente para gerar uma sequência de milhares de números. Outra vantagem é que estes métodos são também rápidos para gerar amostras e (normalmente) não precisam de nenhum *hardware* adicional.

A desvantagem óbvia é a ausência de aleatoriedade real, mas conforme mencionado em Rukhin et al. (2001): "ironicamente algoritmos pseudo-aleatórios bem construídos fornecem séries aparentemente mais aleatórias do que muitas fontes verdadeiramente aleatórias" uma vez que alguns algoritmos introduzem não-linearidades sobre não-linearidades destruindo qualquer relação de entrada e saída esperada.

A biblioteca Python *numpy.random* usa um gerador de números pseudo-aleatórios popular e robusto chamado *Mersenne Twister* (MT) proposto inicialmente por Matsumoto and Nishimura (1988). Este gerador possui as seguintes vantagens: (1) sua velocidade de geração de amostras é muito rápida; (2) tem um grande período de recorrência $2^{19937-1}$; (3) tem propriedade de equidistribuição de alta dimensão (623 dimensões equidistribuídas).

O algoritmo se baseia num método chamado de *Linear Congruential Generator* que usa a expressão 2.76 para a geração de amostras.

$$x_{n+1} = \text{resto}[(x_n * a + c)/d] \quad (2.76)$$

Onde \mathbf{a} é um multiplicador, \mathbf{c} um incremento, \mathbf{d} um divisor e x_0 é a semente que irá gerar outro número que será a próxima semente e assim por diante ($\mathbf{a}, \mathbf{c}, \mathbf{d}$ podem variar dinamicamente ou ser previamente definida). O algoritmo faz uso de métodos adicionais baseados em álgebra linear para ter uma sequência com característica multidimensional.

Mas o que faz um algoritmo gerar sequências de números que "parecem aleatória"?

1. Incapacidade de fazer previsões sobre o próximo termo. A sequência deve seguir um padrão altamente complexo a ponto de ser definida como "sem padrão", que não pode ser predito seja por uma expressão matemática ou por repetição da própria série, note que essa característica é de fato o que traz aleatoriedade ao algoritmo;
2. Ter a abrangência e igualdade de probabilidade sobre todo o espectro de números possíveis para a série (a não ser que seja definida para ocorrer de maneira diferente). Um exemplo simples de algoritmo que não atende a esta característica é um código que simule um dado de seis faces, mas só consiga retornar uma variedade de cinco números;
3. A capacidade de rápida divergência. Se mudarmos a amostra inicial de 2 para 2,001 deveríamos ter séries bastante diferentes;

Para estudos de análise de incerteza que envolvem amostragem de uma única distribuição (gerando 'n' observações), não há significado para a ordem em que essas observações são obtidas e a aleatoriedade tem pouca importância.

O que importa ao amostrar uma única distribuição é que todo o conjunto de observações preserve fielmente as propriedades da distribuição da qual foi amostrada. Entretanto, ao avaliar equações que envolvem mais de uma distribuição, a ordem na qual as variáveis são geradas e consequentemente combinadas passa a ser um fator importante, e, portanto, estudos de correlação são necessários.

Conforme mencionado por Iman and Conover (1982), foi devido ao avanço do poder computacional nos anos 1980 que a simulação de Monte Carlo se popularizou nas empresas. As companhias passaram a empregar muitas técnicas de simulação na avaliação de risco. Ainda de acordo com os autores, a maioria das implementações sobre técnicas de amostragem era univariada ou multivariada independente.

Apesar da disponibilidade da teoria sobre a correlação na simulação numérica já tivessem sido produzidas previamente, a implementação destas teorias de maneira conjunta ainda era pouco incorporada. Bibliotecas gráficas de fácil uso foram também importantes na visualização de *crossplots* entre as variáveis, o que propiciou a percepção da relação entre elas e da necessidade do uso de correlações para um público maior.

A visualização de distribuições bivariadas (com duas variáveis) é comumente feita através de um histograma 2D ou de uma nuvem de pontos acompanhada dos respectivos histogramas. Neste tipo de gráfico a geometria da nuvem de pontos e histogramas dependerão da forma das distribuições. Caso sejam gaussianas com desvios padrão iguais, gerarão idealmente uma geometria circular. No caso de desvios padrões diferentes resultarão em geometria elíptica centrada nas médias das distribuições que será deformada pela correlação⁷.

A distribuição bivariada normal correlacionada possui *pdf* com equação analítica conhecida desde os anos 1920 (Owen, 1956). A equação (2.77) descreve esta distribuição:

$$p(x) = \frac{1}{2\pi\sqrt{\Sigma}} \exp -\frac{1}{2}\mathbf{x}^T \Sigma^{-1}\mathbf{x} \quad (2.77)$$

Onde duas distribuições quaisquer (X_1 e X_2) estão contidas no vetor $\mathbf{x}^T = [X_1, X_2]$, Σ é a matriz de covariância e $||$ representa o módulo da matriz.

⁷ caso da Figura 2.42 a correlação foi definida como zero em ambos os exemplos.

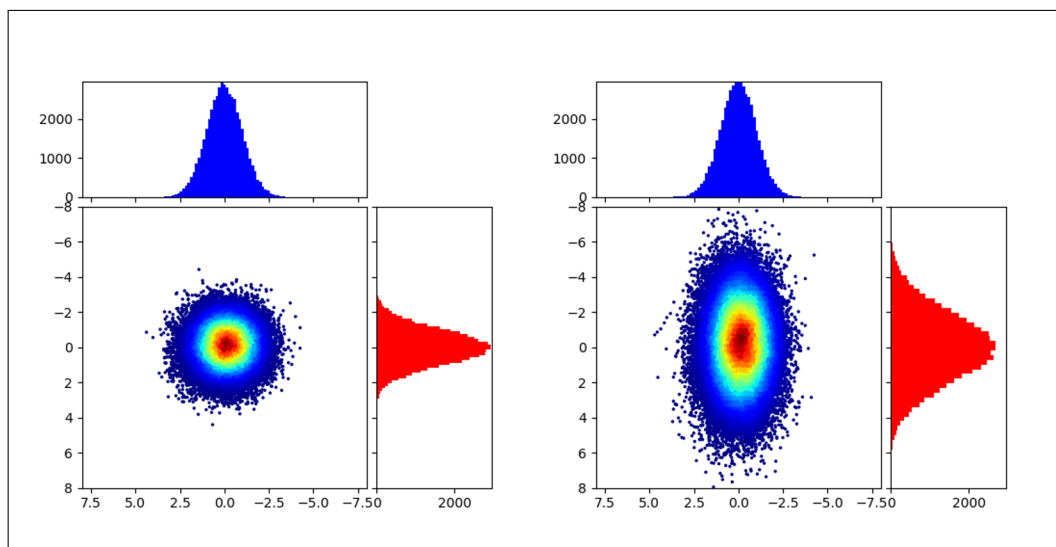


Figura 2.42: Exemplos de *crossplots* equivalentes a uma distribuição bivariada normal. Foram criadas duas distribuições normais com 50000 amostras, ambas com média 0 e ambas com desvio padrão de 1, a correlação das distribuições é zero (esquerda). Alternativamente (direita) temos duas distribuições também com média 0 e desvio padrão diferentes. A variável no eixo das abscissas tem desvio padrão 1 e a variável no eixo das ordenadas tem desvio padrão 2, a correlação das distribuições é zero.

Considerando a relação exponencial entre normal e lognormal, se exponenciarmos as distribuições normais geradas com base na reamostragem da equação (2.77) teríamos distribuições lognormais de alguma forma correlatas, mas seus valores de correlação Pearson não seriam os especificados em função da operação de exponenciação.

O problema pode se tornar um pouco mais complexo se as distribuições têm geometrias não-normais desconhecidas ou são determinadas por métodos numéricos.

A geração de amostras correlacionadas é baseada em dois processos: A geração de amostras sob um critério de restrição prévio que gere amostras de distribuições correlatas (e.g. Cadeias de Markov, *Bootstrapping*) ou; A geração de distribuições pseudo-aleatórias independentes (e.g. SRS - *Standard Random Simulation*, *Latin Hypercube Sampling*) seguidas de rearranjo de maneira a atingir o nível de correlação desejado (e.g. Iman Conover-Cholesky, Cópula e GSSC).

Não iremos discorrer sobre estratégias de geração de amostras, nos concentraremos nos métodos de reordenamento.

Apesar de apresentar as expressões matemáticas quando conveniente, a não ser que seja estritamente necessário, não abordaremos o desenvolvimento dos teoremas e definições com o formalismo matemático das referências utilizadas neste trabalho, privilegiaremos em contrapartida, tópicos relacionados aos algoritmos. Espera-se com isso que o leitor não-familiarizado com os tópicos de estatística e matemática, ao nível de pós-graduação, possa ler o conteúdo de maneira fluida e tenha acesso à proposta das ferramentas acompanhada de rápida implementação.

2.4.1 Métodos de Reordenamento : Iman-Conover - Cholesky

A origem de modelos lineares em operações relacionadas a matrizes de covariância tem seus primórdios nos trabalhos de Yule (1907), Yule (1927) e Gabriel (1962). Nestes trabalhos há relatos também sobre a parametrização implícita de distribuições normais multivariadas em termos de entradas de qualquer covariância Σ ou seu inverso. A manipulação da correlação através da álgebra linear é tradicionalmente feita através da decomposição de Cholesky (que recebeu este nome devido a seu descobridor, o cartógrafo e matemático francês André-Louis Cholesky 1875-1918).

A decomposição de Cholesky foi popularizada em estudos de simulação na área de geociências pela publicação de Iman and Conover (1982). Em sua forma original, a decomposição de Cholesky se baseia no seguinte produto:

$$\Sigma = U^T U \quad (2.78)$$

Onde U é uma matriz triangular superior com diagonais positivas e o sobrescrito T denota transposição de matriz (por isso as vezes é conhecida como transformação LU *Lower-Upper*) onde *Lower* seria representada por U^T . Alguns autores preferem trabalhar com matrizes triangulares inferiores, obtidas se definimos $L = U^T$, conseqüentemente teríamos $\Sigma = L.L^T$. As matrizes U ou L são conhecidas como matriz Cholesky (ou matriz "raiz quadrada").

Como não trabalhamos com valores complexos de correlação, todas as matrizes de covariância são hermitianas, a suposição de ser positiva definida depende de algumas condições como a verificação do teste:

$$\Sigma = k \Sigma k^T > 0 \quad (2.79)$$

Onde k é um vetor qualquer com tamanho definido de maneira a atender a dimensionalidade de Σ e permitir o produto. Σ é positivo definido também se e somente se, todos os seus autovalores forem positivos. Para decompor uma matriz 3x3 com Cholesky teríamos o seguinte sistema de equações:

$$\Sigma = L.L^T \quad (2.80)$$

$$\begin{bmatrix} \Sigma_{1,1} & \Sigma_{1,2} & \Sigma_{1,3} \\ \Sigma_{2,1} & \Sigma_{2,2} & \Sigma_{2,3} \\ \Sigma_{3,1} & \Sigma_{3,2} & \Sigma_{3,3} \end{bmatrix} = \begin{bmatrix} l_{1,1} & 0 & 0 \\ l_{2,1} & l_{2,2} & 0 \\ l_{3,1} & l_{3,2} & l_{3,3} \end{bmatrix} \times \begin{bmatrix} l_{1,1} & l_{2,1} & l_{3,1} \\ 0 & l_{2,2} & l_{3,2} \\ 0 & 0 & l_{3,3} \end{bmatrix} \quad (2.81)$$

Que se desenvolvida dá origem à:

$$\begin{bmatrix} \Sigma_{1,1} & \Sigma_{1,2} & \Sigma_{1,3} \\ \Sigma_{2,1} & \Sigma_{2,2} & \Sigma_{2,3} \\ \Sigma_{3,1} & \Sigma_{3,2} & \Sigma_{3,3} \end{bmatrix} = \begin{bmatrix} l_{1,1}^2 & l_{2,1}.l_{1,1} & l_{3,1}.l_{1,1} \\ l_{2,1}.l_{1,1} & l_{2,1}^2 + l_{2,2}^2 & l_{3,1}l_{2,1} + l_{3,2}l_{2,2} \\ l_{3,1}.l_{1,1} & l_{3,1}l_{2,1} + l_{3,2}l_{2,2} & l_{3,1}^2 + l_{3,2}^2 + l_{3,3} \end{bmatrix} \quad (2.82)$$

Este sistema de equações pode ser resolvido recursivamente a partir do termo $\Sigma_{1,1} = \sqrt{L_{1,1}}$. A função `numpy.linalg.cholesky` resolve as iterações de maneira otimizada utilizando a linguagem Python (preferencial nesta dissertação)⁸.

É importante destacar que esta transformação é perfeita apenas para distribuições normais padrão (média =0 e desvio padrão=1) o que significa que temos que transformar as nossas distribuições em distribuições normais padrão e realizar a transformação inversa ao final das operações.

⁸Note que ser positiva definida é um elemento necessário à decomposição. A própria função retornará um erro de álgebra se a função não for positiva definida

A transformação de distribuições não-normais (gamma, lognormal, bimodal etc) em distribuições normais para viabilizar o uso da decomposição de Cholesky em transformações relacionadas à correlação foi inicialmente proposta por Jonhson and Ramberg (1977) utilizando as equações analíticas inversas para algumas distribuições lognormais e seno hiperbólico inverso, mas a matemática se tornou intratável para distribuições complexas ou cuja solução analítica é desconhecida.

Iman and Conover (1982) propôs tratar o problema através da transformação das distribuições em distribuições de posições e assim utilizar a *rank correlation* ao invés do coeficiente de correlação de Pearson (ρ) com a premissa de que os dois coeficientes são semelhantes.

Conforme descrito em Nelsen (2006), apesar do coeficiente de correlação de Pearson ser amplamente utilizado suas limitações são raramente enfatizadas: (1) ρ tende ser altamente afetado por *outliers* e, portanto, não é adequado para análise de distribuições com valores extremos (e.g. o caso de curvas lognormais); (2) o valor de ρ pode mudar se X e/ou Y são transformados monotonicamente (como com uma exponenciação) enquanto suas correlações de Spearman permanecem as mesmas; (3) ρ é adequado apenas adequado para distribuições cuja correlação apresenta alto grau de linearidade (e.g. gaussianas ou elípticas).

Atualmente é possível fazer computacionalmente a conversão de distribuições uniformes em distribuição normais através da função inversa de densidade de probabilidade acumulada (*icdf - inverse cumulative density function*) para distribuições discretas baseada na conversão em percentis (*percentile probability function*) utilizada no mapeamento de amostras em funções uniforme (de posições) e posteriormente transformada em distribuição normal. A função `scipy.stats.norm.ppf` implementada em python pode ser utilizada nesta atividade.

De posse destes vetores mapeados para distribuições normais padrões podemos realizar a transformação linear pelo método de Cholesky. Para efetuar a transformação de distribuições normais padrão independentes em distribuições normais padrão correlacionadas. Para isso somente precisamos efetuar a multiplicação de \mathbf{NX} (a matriz transformada em normal padrão) e L (a matriz triangular inferior obtida pela decomposição de Cholesky).

$$NX_{icdf_{Cor}} = L_{cholesky} \times NX_{icdf_{independente}} \quad (2.83)$$

De posse da matriz transformada correlacionada podemos fazer a operação inversa da distribuição transformada para a distribuição original correlacionada (também utilizando o pacote `scipy.stats.norm`).

A Figura 2.43 ilustra os passos do algoritmo utilizado para tornar as variáveis aleatórias dependentes através do método de Cholesky. No caso de operações utilizando outras distribuições adicionamos mais alguns passos no algoritmo conforme ilustrado na Figura 2.44 (note que o fluxograma da distribuição normal é na verdade um caso particular daquele utilizado para outras distribuições).

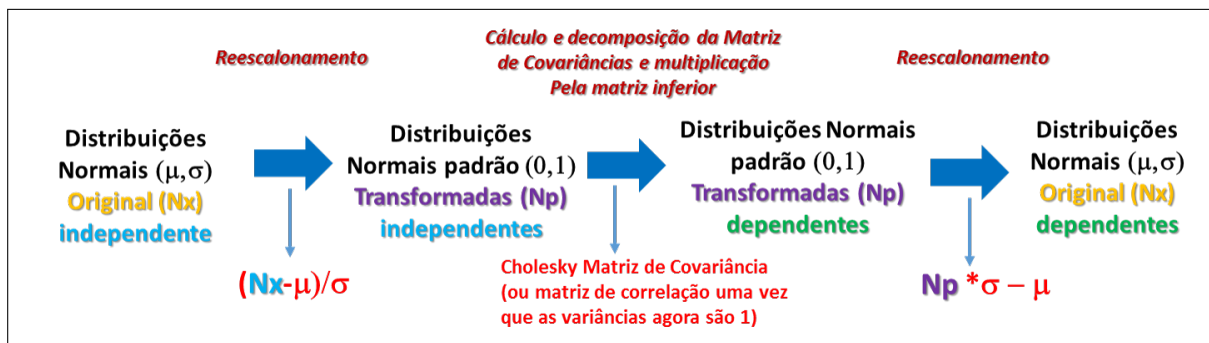


Figura 2.43: Um diagrama com os passos para obter distribuições normais correlatas através da transformação linear de Cholesky em distribuições Normais.

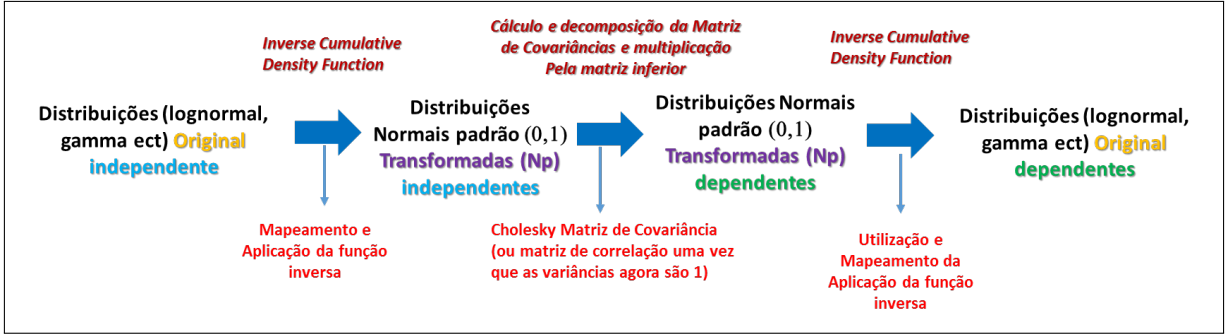


Figura 2.44: Um diagrama com os passos para obter distribuições normais correlatas através da transformação linear de Cholesky em distribuições quaisquer.

De maneira equivalente podemos transformar uma matriz correlacionada em uma matriz de distribuições não-correlacionadas multiplicando \mathbf{X} pela inversa da matriz de cholesky (L^{-1}). Podemos obter a inversa da matriz L numericamente com a função `numpy.linalg.inv` (também utilizando a linguagem Python):

$$NX_{icdf_{independente}} = L_{cholesky}^{-1} \times NX_{icdf_{Cor}} \quad (2.84)$$

Note que a matriz L não faz apenas uma operação de rotação (esta operação é resultado da mínima correlação e dos valores de $\Delta P_{ordenadas}/\Delta P_{abscissas}$ valores das distribuições de entrada). Temos aqui também uma operação de deformação longitudinal nos pontos controlado pelo coeficiente de correlação. Um exemplo claro é dado na Figura 2.45:

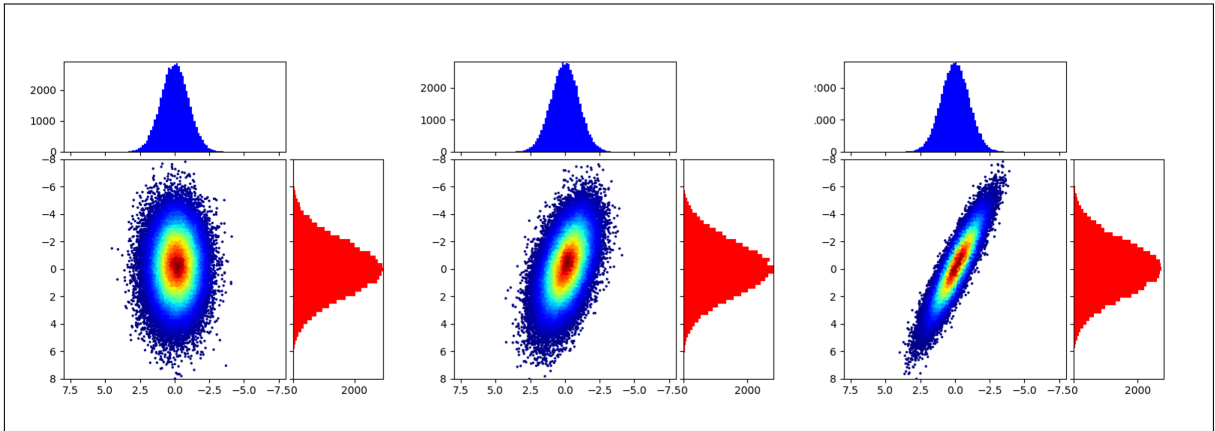


Figura 2.45: Exemplos de *crossplot* equivalente a uma distribuição bivariada normal. Foram criadas duas distribuições normais com 50000 amostras, média 0, desvio padrão da variável no eixo das abscissas de 1 e desvio padrão no eixo das ordenadas de 2. Na figura da direita a correlação das distribuições é zero, na do centro a correlação é 0,5 e na da direita é 0,9.

É importante destacar que a totalidade dos elementos positivos da matriz não garantem uma matriz positiva definida e alternativamente possuir correlações negativas não impedem a matriz de ser positiva e definida (e.g. a matriz Σ abaixo que é positiva definida):

$$\Sigma = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \quad (2.85)$$

A seguir deixo os passos do algoritmo utilizados nesta dissertação de maneira que este possa ser implementado em qualquer plataforma (a função em python está em anexo):

- Simular d variáveis com n amostras $\mathbf{X} = [X_1, \dots, X_d]$;
- Converter as d variáveis em variáveis uniforme com a distribuição de posições;
- Misturar as amostras em cada variável para garantir correlação próxima de zero entre elas;
- Converter as variáveis uniforme em normais com a iCDF;
- Definir a matriz de correlações desejada σ .
- Verificar se a matriz é positiva definida;
- Realizar a decomposição de Cholesky $\sigma = L.L^T$;
- Fazer o produto $X_{cor} = L.X_{indep}$;
- Transformar a matriz de volta com a CDF e transformar de volta para a distribuição original a partir dos postos.

Muitas modificações da decomposição de Cholesky foram estudadas no sentido de estimar a propriedades da matriz de covariância a partir de um *dataset* qualquer, com o objetivo óbvio de extrair padrões de correlação e dispersão dos dados. Algumas das implementações envolvem otimização e transformações que tornam a matriz de correlação positiva definida e a decomposição de matrizes de covariância com o uso de Cholesky para dividi-la em uma matriz de variâncias e uma matriz de correlação (até então misturadas na matriz de covariância).

Conforme descrito em Pourahmadi (2007), a decomposição da covariância em correlação-variância é o método mais simples e direto, cujos componentes são fáceis de interpretar em termos das variáveis originais (Styan (1968); Liang and Zeger (1986); J.C.W. and Rayner (1987); Barnard et al. (2000)). Entretanto, ela tem sido mais utilizada quando enfrentamos problemas inversos em que conhecemos os dados e ambos os elementos (variância e correlação) não são conhecidos, como nosso problema se concentra no método direto onde definimos os valores de correlação e variância para gerar os dados, não nos aprofundaremos neste assunto.

2.4.2 Métodos de Reordenamento : Funções Cópula

O uso do método de Iman and Conover (1982) pressupõe que a estrutura de dependência será linear e simétrica gaussiana, uma vez que as amostras se dispersarão de maneira uniforme ao longo do eixo de correlação máxima. Entretanto, esta geometria de correlação não é a única para um determinado valor de correlação e muitas vezes não é a mais indicada se comparada com a geometria verificada nos dados.

As funções cópulas oferecem uma maneira flexível e geral de descrever separadamente a estrutura de dependência e as distribuições de probabilidade marginais (as variáveis no cálculo volumétrico), e servem como uma ferramenta poderosa para modelagem de dados multivariada possibilitando obter distribuições correlatas em que os pontos assumem dispersões diferentes (e.g. Figura 2.46).

De acordo com Durante and Sempi (2009), o uso de cópulas tem sua origem preliminar nos estudos de Fréchet (1956); Féron (1956); Gumbel (1958) mas foi no trabalho de Sklar (1959) que o nome cópula foi cunhado e resultados mais substanciais foram obtidos culminando no **Teorema de Sklar**, enunciando que:

Seja (X_1, X_2, \dots, X_k) vetores aleatório em \mathbb{R}^k com função de distribuição conjunta F e marginais $F_i, i = 1, 2, \dots, k$. Então existe uma cópula C tal que $\forall x (x_1, \dots, x_k) \in a \mathbb{R}^k$:

$$F(x_1, \dots, x_k) = C(F_1(x_1), \dots, F_k(x_k)) \quad (2.86)$$

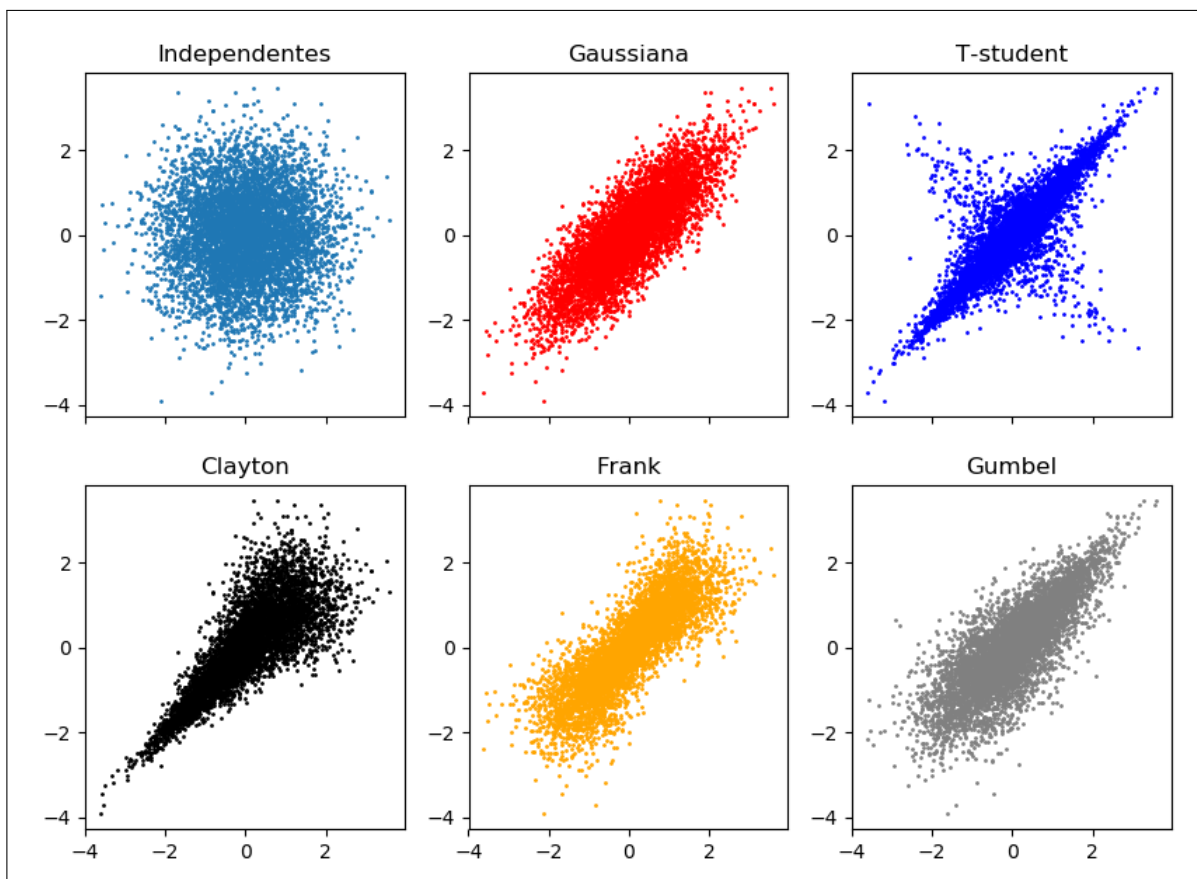


Figura 2.46: Diferentes tipos de correlação obtidos por diferentes funções cópula. Os parâmetros dos *crossplots* foram ajustados de maneira a atingir um mesmo valor de correlação de Pearson de 0,8 .

Onde $F_n(x_n) = u_n$ em que u_n representa uma amostra de uma função padrão uniforme U_n . Se F_1, \dots, F_k são contínuas, então C é única, e pode ser determinada pela equação inversa das funções:

$$C(u_1, \dots, u_k) = F(F_1^{-1}(u_1), \dots, F_k^{-1}(u_k)) \quad (2.87)$$

As equações 2.86 e 2.87 nos remetem ao livro de referência de Nelsen (2006) onde destaca-se que cópulas podem ser definidas por dois pontos de vista.

O primeiro ponto de vista resume cópulas como sendo funções que juntam ou acoplam distribuições multivariadas às suas distribuições marginais (equação 2.86). Ou seja, uma função de ligação entre as distribuições univariadas que tem como resultado distribuições multivariadas.

A outra forma de interpretação afirma que cópulas são distribuições multivariadas cujas distribuições marginais são uniformes no intervalo $]0,1[$ (equação 2.87).

Resumidamente, uma definição vê a cópula como uma função e outra como a própria distribuição multivariada. Os livros de Rachev (2003), Nelsen (2006) e Piotr Jaworski (2009) são excelentes referências para os leitores não familiarizados ao tema e que querem se aprofundar.

Note que a descrição da cópula das distribuições $C(u_1, \dots, u_k)$ representa a cdf para esta cópula composta por cdf's de outras distribuição, dessa forma para obter a função densidade de probabilidade é necessário obter:

$$f(u_1, \dots, u_n) = f_{u_1}(u_1) \times \dots \times f_{u_n}(u_n) = \frac{\partial^n C(F_1(u_1), \dots, F_n(u_n))}{\partial F_1(u_1) \dots \partial F_n(u_n)} \quad (2.88)$$

Que no caso bivariado seria:

$$f(u, v) = c(u, v) \cdot f_u(u) \cdot f_v(v) = \frac{\partial^2 C(F_u(u)F_v(v))}{\partial F(u)\partial F(v)} \quad (2.89)$$

Note que a cdf da função cópula ($C(u, v)$ no caso bivariado) deve apresentar algumas características:

- $C(u, 1) = u$ para $u \in [0, 1]$. Esta propriedade significa que para cópulas que unem uma distribuição unidimensional (u) à distribuições constantes **unitárias** (compostas apenas pelo valor 1). O resultado da cópula será a distribuição de probabilidade da distribuição u .
- $C(u, 0) = 0$ para $u \in [0, 1]$. Esta propriedade indica que para cópulas que unem uma distribuição unidimensional (u) à distribuições constantes **nulas** (compostas apenas pelo valor 0). O resultado da cópula será a distribuição de probabilidade constantes nula;
- $C(U, V)$ é crescente - esta propriedade advém do fato das distribuições de entrada serem estritamente positivas e por isso a cdf que as une não pode ser decrescente em nenhum ponto;
- Os limites da cópula é dado pelos limites de Fréchet: $\max(\sum_{i=1}^d u_i + 1 - d, 0) \leq C(u, v) \leq \min(u, v)$ Onde os extremos são os limites de Fréchet-Hoeffding inferior e superior.

Como as cópulas são medidas de dependência, elas nos permitem distinguir a dependência e independência de maneira direta. A independência entre variáveis é dada por:

$$C(F_u(u), F_v(v)) = F_u(u) \cdot F_v(v) \quad (2.90)$$

No que se refere às aplicações, apesar de ter sido definida já no fim dos anos 1950, até os anos 1990 as funções cópula eram um tema pouco discutido nas ciências aplicadas e até então se resumiam a aplicações em métricas estatísticas (Cherubini et al., 2004).

Atualmente o uso de cópulas está crescendo em outras áreas de conhecimento nas ciências aplicadas como por exemplo na engenharia (e.g Goda and Tesfamariam (2015)), hidrogeologia e.g (Grimaldi et al. (2011)) e na exploração e produção de petróleo (e.g Armstrong et al. (2004); Al-Harthya et al. (2007); Mendoza-Torres et al. (2017)).

A popularização desta ferramenta se deu pela sua versatilidade de modelagem de distribuições não-normais aliada a sua formulação matemática relativamente simples de entender se comparado ao que se propõe a tratar.

Existem diferentes famílias de cópulas que originarão diferentes agrupamentos para uma mesma correlação, neste trabalho discutiremos apenas as mais utilizadas, a escolha de que família utilizar deve considerar as limitações impostas por suas premissas e o bom ajuste da geometria de correlação produzida com a geometria esperada para o modelo.

Existem dezenas de famílias de cópulas, discutiremos as mais populares: as elípticas e as arquimedianas.

Uma cópula Elíptica é basicamente a cópula de uma distribuição elíptica específica (*T-Student*, Normal/Gaussiana, Chi-quadrada). Uma característica importante é que todas as distribuições elípticas são radialmente simétricas e se baseiam na transformação linear obtida pelo método de cholesky.

A cópula gaussiana é talvez uma das mais conhecidas, visto que ela é uma repetição da metodologia de Iman and Conover (1982), amplamente conhecida e por isso não será extensivamente rerepresentada.

A utilização de Cópulas já tem funções bem documentadas em linguagem R e algumas iniciativas preliminares em Python. Deixarei aqui os passos do algoritmo para aqueles que sintam a necessidade de por alguma razão implementar estas distribuições em outras linguagens. Adicionalmente as funções em python implementadas por nós e utilizadas neste trabalho serão deixadas em anexo.

Cópula Elípticas - T-Student

As cópulas *T-student* tem seu nome com base na distribuição univariada *T-student*. Esta distribuição foi definida por William Sealy Gosset, químico e estatístico inglês, que utilizava o pseudônimo *Student*⁹. A distribuição univariada de *T-student* pode ser obtida de maneira analítica com a equação 2.91 (Shaw and Lee, 2008):

$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu \cdot \pi} \Gamma(\frac{\nu}{2})} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \quad (2.91)$$

Onde Γ é a função gamma e ν se refere aos graus de liberdade da distribuição, discutida a seguir. A distribuição *t-Student* tem aplicações em problemas de inferência estatística associados à média de uma distribuição normal. Estas aplicações são especialmente úteis quando a média e a variância da distribuição é desconhecida e a quantidade de amostras é pequena.

Sua amostragem é dada da seguinte forma: Z e χ são variáveis aleatórias independentes, com Z tendo uma distribuição normal padrão, e χ tendo uma distribuição Chi-quadrado (com ν graus de liberdade). A distribuição chi-quadrado é gerada pela soma de ν normais independentes com seus elementos elevados ao quadrado. O grau de liberdade ν nada mais é do que a quantidade de amostras utilizadas no processo, então a variável aleatória que comporá distribuição *t-Student* será:

$$x_i = \nu \frac{Z_i}{\chi_i} \quad (2.92)$$

A simulação de distribuições multivariadas que seguem a cópula *t-Student* é bem semelhante ao da cópula gaussiana, entretanto ao gerar as distribuições gaussianas uniformes é aplicada a transformação da equação 2.92. Este processo é traduzido pelo seguinte algoritmo:

- Simular d variáveis com n amostras $\mathbf{U} = [U_1, \dots, U_d]$;
- Converter as d variáveis em variáveis uniforme com a distribuição de posições;
- Misturar as amostras em cada variável para garantir correlação próxima de zero entre elas;
- Converter as variáveis uniforme em normais com a iCDF;
- Definir a matriz de correlações desejada σ .
- Verificar se a matriz é positiva definida;
- Realizar a decomposição de Cholesky $\sigma = L.L^T$;
- Gerar uma distribuição chi-quadrada (χ);
- Fazer o produto $U_{cor} = L.U_{ind}$;
- Faça $\mathbf{x} = Y \cdot \sqrt{\frac{n}{\chi}}$;
- Transformar a as distribuições de volta com a CDF de T-student e transformar de volta para distribuição original a partir das posições.

Na Figura 2.47, seguem alguns exemplos da cópula *T-Student* em distribuições normais padrão com diferentes parametrizações de ρ e ν :

Como com a cópula Gaussiana, mudando o sinal do valor de correlação de Pearson (ρ) na matriz de correlações, alteramos a direção e grau de dispersão das variáveis. Cópulas com valores

⁹Conforme descrito em <https://leandro-est.webnode.com.br/products/o-curioso-caso-de-william-sealy-gosset/>: Ele não podia publicar os resultados de seus trabalhos com seu nome enquanto trabalhava na cervejaria *Guinness*

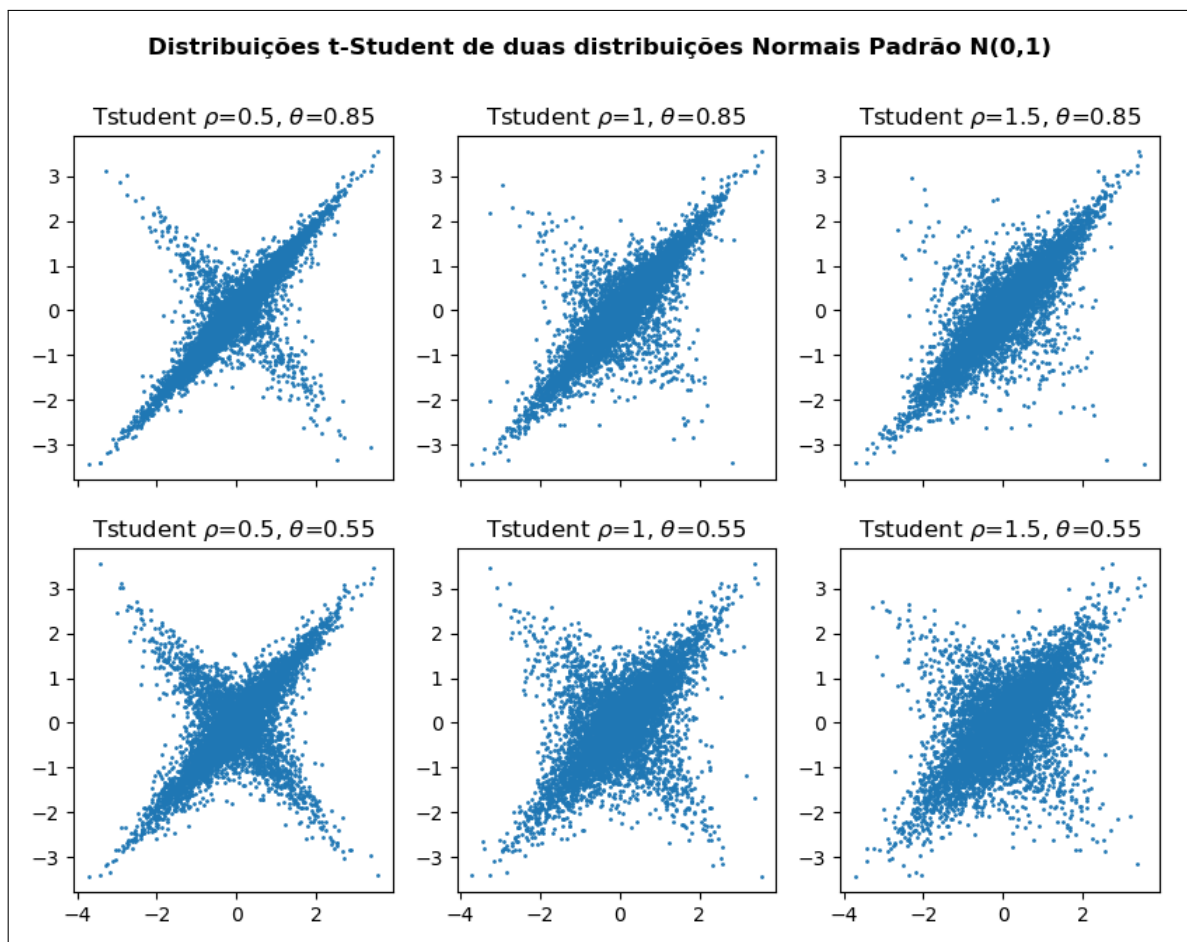


Figura 2.47: Exemplos de cópula *t-student* com diferentes parâmetros de grau de liberdade ν e diferentes correlações.

mais próximos dos extremos de correlação (1,-1) tendem a ter dispersões menores e se concentram no eixo máximo de correlação Nelsen (2006). Contudo, a cópula *T-Student* tem mais pontos concentrados em seus extremos do que a gaussiana e são capazes de produzir observações em direções opostas gerando uma geometria semelhante ao de uma estrela. Estas observações na direção ortogonal são mais frequentes com a redução do valor dos graus de liberdade. Para um valor grande de graus de liberdade (n), a distribuição *t-Student* é bem aproximada pela distribuição normal padrão (Lehoczky, 2015).

Este comportamento é pertinente para simular situações em que anti-correlações episódicas podem acontecer. No caso da geologia, poderíamos relacionar por exemplo a porosidade com a permeabilidade num eixo principal de correlação mas que por vezes aparecem *trends* perpendiculares devido um possível aumento de densidades fraturas ou ocorrência de porosidade modica.

Cópuas Arquimedianas- Conceitos

Em muitas aplicações parece razoável que haja uma dependência mais forte nas caudas das distribuições ou que a concentração de pontos assuma simetria em toda a sua extensão. É neste cenário que as cópuas arquimedianas tem valor.

Segundo Nelsen (2006), cópuas arquimedianas não apareceram originalmente em estatística, mas sim em estudos probabilísticos de espaços métricos, onde foram estudados como parte do desenvolvimento de um versão probabilística da desigualdade triangular. Para mais detalhes desta história, veja Schweizer (1991).

O exemplo clássico nos casos mais estudados se concentram no seguimento de mercado financeiro e de seguros. Ao avaliar uma carteira de investimentos espera-se haver uma grande correlação entre grandes perdas (por exemplo, uma quebra do mercado de ações) do que entre grandes ganhos Cherubini et al. (2004). Tais assimetrias não podem ser modeladas com relações de cópulas elípticas (como a gaussiana e a *T-Student*).

Para estudar estas dependências diferentes recorreremos a famílias de cópulas que simulem esta assimetria, das quais a mais comumente utilizada são as cópulas da classe arquimediana. As cópulas desta família são caracterizadas por uma função geradora ϕ a qual normalmente é associada a variável t , esta função tem as seguintes características:

- ϕ tem Domínio e Imagem dado por $[0, 1] \rightarrow [0, \infty]$;
- Os seguintes pontos devem ser definidos: $\phi(1) = 0$ e $\phi(0) = \infty$;
- Para $\forall t \in (0,1)$ $\partial\phi/\partial t(t) < 0$, ou seja $\phi(t)$ é estritamente decrescente;
- Para $\forall t \in (0,1)$ $\partial^2\phi/\partial t^2(t) \geq 0$ (ϕ é convexa);
- $\phi(t)$ tem uma inversa tal que $\phi^{-1}(t)$ tem Domínio e Imagem dado por $[0, \infty] \rightarrow [0, 1]$ e analogamente $\phi(0) = 1$ e $\phi(\infty) = 0$;
- A cdf da distribuição conjunta das distribuições uniformes padrão u e v , $C(u,v)$, é chamada Arquimediana se ela pode ser escrita como: $C(U, V) = \phi^{-1}(\phi(U) + \phi(V))$;

Uma importante fonte de funções geradoras para cópulas arquimedianas consiste da inversa da Transformada de Laplace para expressões conhecidas (Feller, 1971).

Um dos problemas das cópulas é a particularidade do parâmetro de correlação (normalmente designado pela variável θ) Cherubini et al. (2004). Em face a esta necessidade Genest and MacKay (1986) provaram que a relação entre a cópula e o coeficiente de correlação de Kendall (τ)¹⁰ no caso bivariado pode ser dada por:

$$\tau_K = 1 + 4 \int_0^1 \frac{\phi(t)}{\phi'(t)} dt \quad (2.93)$$

Várias propriedades estatísticas podem ser simplesmente expressas em termos da função geradora $\phi(t)$, como a função de distribuição paramétrica univariada de Kendall ($K_C(t)$) de cópulas.

$$K_c(t) = t - \frac{\phi(t)}{\phi'(t)} \quad (2.94)$$

Precisamos de condições específicas para afirmar que extensões multivariadas de cópulas bivariadas Arquimedianas são adequadas para aplicações multivariadas (n-variáveis). Outra desvantagem é que a extensão multivariadas de cópulas de Arquimedes em geral sofrem de falta (ou excesso) de escolha de parâmetros livres. Por este motivo utilizaremos apenas cópulas bidimensionais.

Note que neste trabalho nos propomos a gerar distribuições correlacionadas. Para implementar funções Cópula Arquimedianas para gerar distribuições correlacionadas podemos seguir o seguinte passo a passo proposto por Nelsen (2006) (**utilizado neste trabalho**):

- Simule duas variáveis aleatórias independentes (u,v) que se comportem como distribuições uniformes padrão U(0,1) :

¹⁰conforme já discutido pode ser definido como $\tau = [P((X_1 - X_2) \cdot (Y_1 - Y_2) > 0) - P((X_1 - X_2) \cdot (Y_1 - Y_2) < 0)]$, onde (X_1, Y_1) e (X_2, Y_2) são vetores aleatórios independentes e identicamente distribuídos com a mesma CDF conjunta de $C(X, Y)$

- Defina $B(u, v) = \frac{\partial C(u, v_{dep})}{\partial u}$;
- Definir $v_{dep} = B(u, v)^{-1}$, onde $^{-1}$ denota função inversa.

Alternativamente, Rachev (2003) sugere os seguintes passos:

- Simule duas variáveis aleatórias independentes (s,q) que se comportem como distribuições uniformes padrão U(0,1) :
- Defina $t = K_C^{-1}$;
- Definir $u_{dep} = \phi^{-1}(s\phi(t))ev_{dep} = \phi^{-1}((1-s)\phi(t))$.

Conforme comentado anteriormente as funções cópula utilizam normalmente distribuições uniforme como distribuições marginais de entrada. O que significa que toda a modelagem parte do princípio que transformamos as distribuições de entrada em funções uniforme padrão (ou pela cdf). O processo de mapeamento e transformação é bastante semelhante ao da transformação para normal mencionada no método de Iman e Conover e os passos do algoritmo seguem abaixo:

1. Criar uma distribuição de posições baseada na distribuição original (tal como usamos no rank correlation);
2. Criar uma a série de posições normalizadas tal que seu valor máximo seja igual a $1/(N+1)$. Esta distribuição será a utilizada na cópula;

Deve-se notar que, nem toda família de cópulas arquimedianas consegue acomodar toda a gama de valores de dependências (de dependências positivas à dependências perfeitamente negativas) por isso a escolha da família de cópulas a ser utilizada deve levar este fator em conta Nelsen (2006).

Para contornar este problema fizemos alterações nos algoritmos de maneira a acomodar as operações com dependências negativas.

A tabela a seguir resume as funções das cópulas utilizadas neste trabalho (Clayton, Frank e Gumbel) parametrizadas em função de t (parâmetro utilizado na função geradora) e θ (o parâmetro que controla a correlação.)

Família	$\phi(t)$	Int. de θ	$K_c(t)$	$\tau(\mathbf{t})$
Clayton	$\frac{1}{\theta}(t^{-\theta} - 1)$	$[-1, \infty)-[0]$	$t(1 + \frac{1}{\theta}) - \frac{t^{\theta+1}}{\theta}$	$\frac{\theta}{\theta+2}$
Frank	$-\ln \frac{\exp(-\theta t)-1}{\exp(-\theta)-1}$	$(-\infty, \infty)-[0]$	$t + (\frac{\exp(-\theta t)-1}{\theta}) \cdot \ln \frac{\exp(-\theta)-1}{\exp(-\theta t)-1}$	$1 + \frac{4}{\theta}[D_1(\theta) - 1]$
Gumbel	$(-\ln t)^\theta$	$[1, \infty)$	$t(1 - \frac{\ln t}{\theta})$	$\frac{\theta-1}{\theta}$

Tabela 2.12: Tabela de parâmetros para algumas Cópulas arquimedianas. Fonte: Grimaldi et al. (2011).

Note o uso da função Debye¹¹ de primeira ordem(D_1).

Uma distinção entre estas três cópulas pode ser feita no que diz respeito à capacidade de capturar a dependência da cauda: As cópulas de Gumbel e Clayton capturam a dependência da cauda superior e dependência da cauda inferior, respectivamente, enquanto a cópula Frank não mostra dependência de cauda. Para as três cópulas, o parâmetro θ pode ser estimado diretamente via $\tau(u, v)$.

Para avaliar a dependência que ocorrem nas caudas, Nelsen (2006) propôs a utilização do parâmetro assintótico da dependência de cauda inferior, λ_{inf} . Este parâmetro é a probabilidade condicional no limite em que uma variável assume valor muito baixo, dado que a outra também

¹¹Onde D_1 é a função de Debye de ordem 1 definida por: $D_1(\theta) = \int_0^\theta \frac{t}{\theta(\exp(t)-1)} dt$

assume um valor muito baixo. Analogamente também propôs, a utilização do parâmetro assintótico da dependência de cauda inferior, λ_{sup} que é a probabilidade condicional no limite que uma variável assume muito valor alto (próximo a 1), dado que a outra também assume um valor muito alto. Estes parâmetros são matematicamente definidos por:

$$\lambda_{inf} = \lim_{t \rightarrow [0^+]} \frac{C(t, t)}{t} \quad (2.95)$$

$$\lambda_{sup} = 2 - \lim_{t \rightarrow [1^-]} \frac{1 - C(t, t)}{1 - t} \quad (2.96)$$

Uma propriedade interessante das medidas de dependência de cauda é que elas são independentes das distribuições marginais das variáveis. Elas são invariantes sob transformações estritamente monótonicas de u e v . Diz-se que as variáveis u e v são assintoticamente independentes se $\lambda_{inf} = \lambda_{sup} = 0$ mas o contrário não é verdade (veja a distribuição de Frank a seguir).

Antes de usar qualquer resultado, o modelo escolhido deve ser avaliado. Existem vários métodos na literatura para testar a qualidade do ajuste de um modelo de cópula aos dados (por exemplo, Genest et al. (2006); Wang and Wells (2000)). De maneira mais próxima ao tema deste trabalho, a simulação de Monte Carlo pode também ser usada (Nikolaidis et al. (2011)). Neste método, o usuário compara os pares simulados de variáveis aleatórias extraídos do modelo de cópula com os pares de valores observados das variáveis reais.

Cópuas Arquimedianas- Cópula de Clayton

A cópula de Clayton foi introduzida inicialmente em Clayton (1978). Ela foi inicialmente utilizada para estudos de correlação de riscos por causa de sua capacidade de capturar a forte dependência na cauda inferior de eventos (e.g. cenário de um *crash* financeiro). A função geradora $\phi(t)$ é dada por:

$$\phi_{\theta}(t) = \frac{1}{\theta}(t^{-\theta} - 1) \quad (2.97)$$

Onde θ é o parâmetro da cópula restrito no intervalo $(-1, +\infty) - [0]$. Se $\theta = 0$ as distribuições são tratadas como independentes.

Enquanto sua inversa $(\phi(t)^{-1})$ é dada por:

$$\phi_{\theta}^{-1}(t) = (t.\theta + 1)^{-1/\theta} \quad (2.98)$$

A expressão bivariada da CDF desta cópula é dada por:

$$C_{\theta}^{Cl}(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta} \quad (2.99)$$

Note que para obter a distribuição da pdf temos que fazer $\frac{\partial^2 C(u, v)}{\partial u \partial v}$.

$$c_{\theta}^{Cl}(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v} = (\theta + 1)(u.v)^{-(\theta+1)}(u^{-\theta} + v^{-\theta} - 1)^{-\frac{-2\theta+1}{\theta}} \quad (2.100)$$

A Figura 2.48 ilustra o comportamento descrito pelos itens acima através das *cdf's* e *pdf's* para cópuas de Clayton com diferentes valores de θ .

Para obter os valores de v dependentes, necessitamos obter as derivadas parciais em relação às variáveis não avaliadas (no caso apenas u):

$$\frac{\partial C(u, v)}{\partial u} = v = (1 + u^{\theta} . (v_{dep}^{-\theta} - 1))^{\frac{1+\theta}{\theta}} \quad (2.101)$$

Resolvendo v_{dep} obtemos:

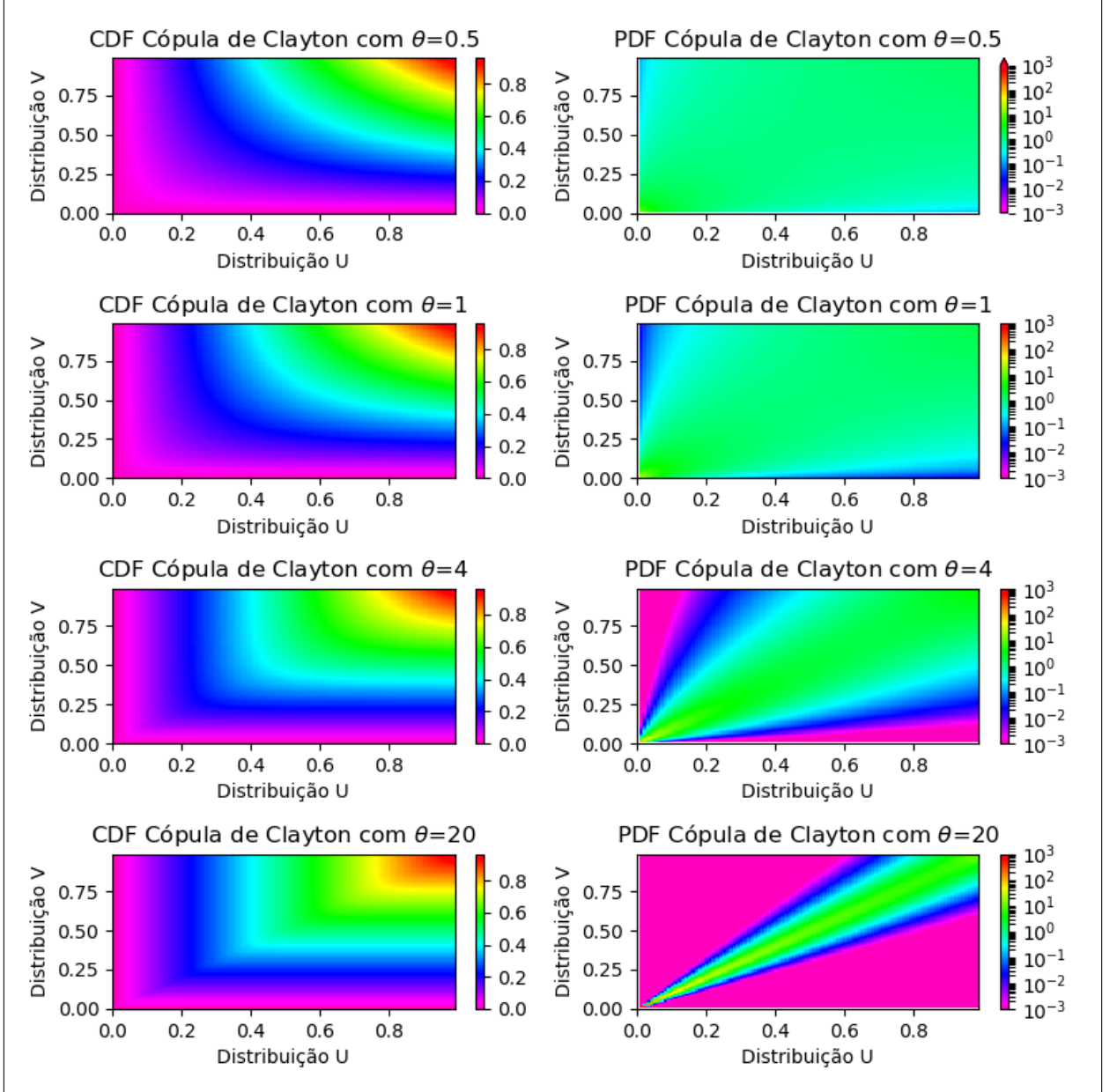


Figura 2.48: Superfícies de CDFs e PDFs da cópula de Clayton, obtidos para diferentes valores de θ , e resultam em diferentes comportamentos das distribuições.

$$v_{dep} = ((v^{\frac{-\theta}{1+\theta}} - 1)u^{-\theta} + 1)^{-1/\theta} \quad (2.102)$$

Dessa forma para gerar amostras de distribuições bivariadas com estrutura de dependência da cópula de Clayton podemos utilizar a equação (2.102) conforme o algoritmo a seguir:

1. Simular as duas variáveis com n amostras (u, v) ;
2. Converter as d variáveis em variáveis uniforme com a distribuição de posições normalizadas (ru, rv) ;
3. Misturar as amostras em cada variável para garantir correlação próxima de zero entre elas;
4. Definir o vetor $ru_{dep} = ru_{indep}$.
5. Definir o o vetor $rv_{dep} = (rv_{ind}^{-\theta/(1+\theta)} - 1).u_{indep}^{-\theta} + 1)^{-1/\theta}$

6. Converter as distribuições uniforme (ru_{dep}, rv_{dep}) nas distribuições originais (u, v) .

A Figura 2.49 ilustra alguns exemplos de correlações positivas obtidas com a função que segue o algoritmo descrito acima (deixada em anexo deste trabalho).

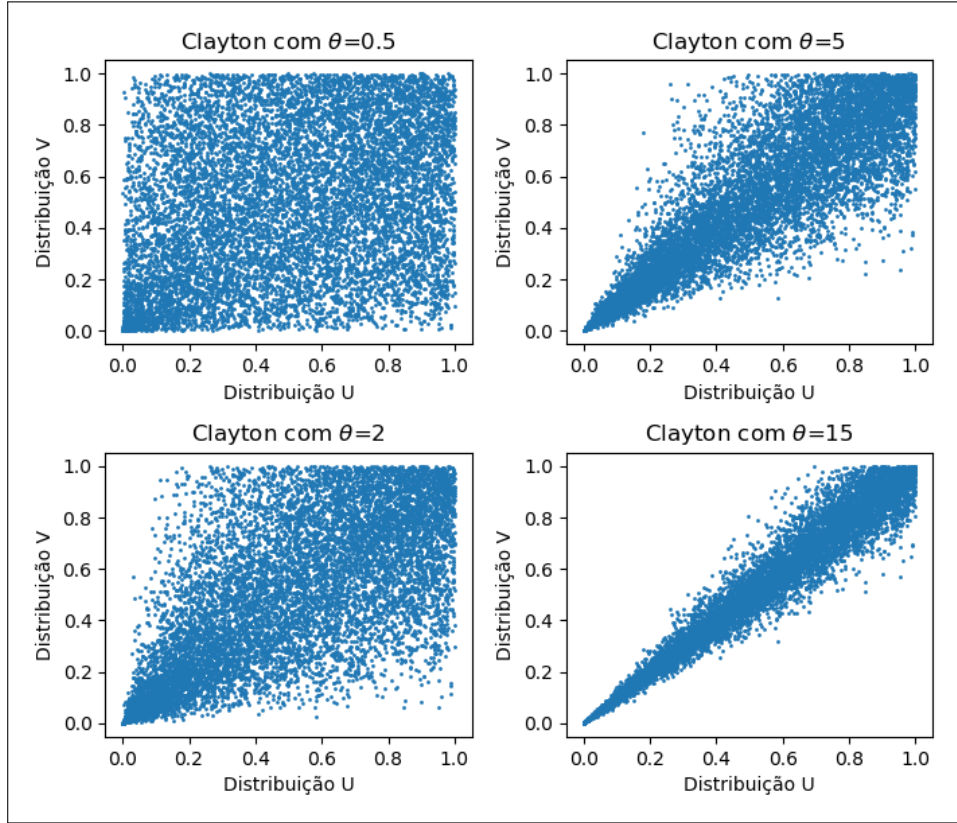


Figura 2.49: *Crossplot* de pontos gerados pela distribuição de Clayton com diferentes valores de θ .

Devido à indeterminação causada pelo parâmetro de dependência, a cópula de Clayton não pode assumir valores de dependência excessivamente negativos ($\theta < -1$). E mesmo no cenário de obter valores na faixa $-1 < \theta < 0$ causam um comportamento degenerado da correlação (Figuras 2.50 e 2.51).

Para contornar a limitação imposta pelo limite inferior de theta $(-1, +\infty)-[0]$, sugerimos aqui fazer uma rotação gráfica no resultado obtido com parâmetro θ positivo, de maneira a atingir o efeito desejado.

Esta rotação é obtida com os seguintes passos:

1. Se θ for negativo, defina $\theta_2 = \text{abs}(\theta)$;
2. Realize as operações com o os valores de θ_2 ;
3. Faça a rotação gráfica: $C_\theta(u, v) = -C_{\theta_2}(u, v) + 1$.

Note que o parâmetro θ é bastante particular à problemas relacionados a cópulas Arquimedianas, dessa forma normalmente utilizamos correlação de Kendall como uma medida mais "comum". A dependência entre o parâmetro θ e o coeficiente de correlação de Kendall é dada por Mahfoud (2012):

$$\tau = \frac{\theta}{\theta + 2} \quad (2.103)$$

E a dependência da cauda inferior é dada por $\lambda_{inf} = 2^{-1/\theta}$ (Cherubini et al. (2004)).

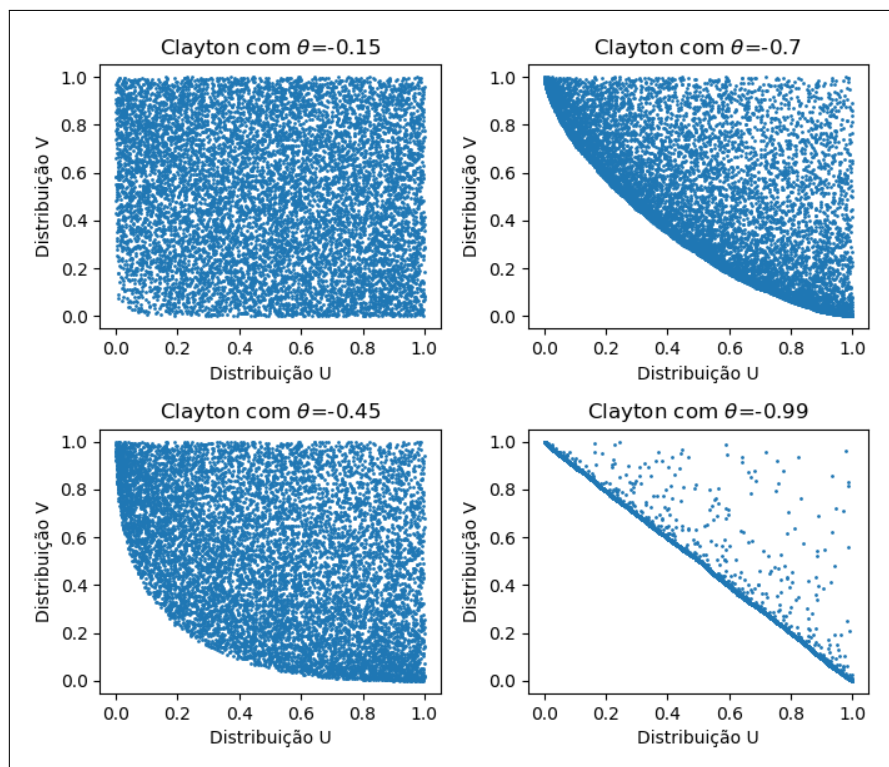


Figura 2.50: *Crossplot* de pontos gerados pela distribuição de Clayton com diferentes valores de θ negativos.

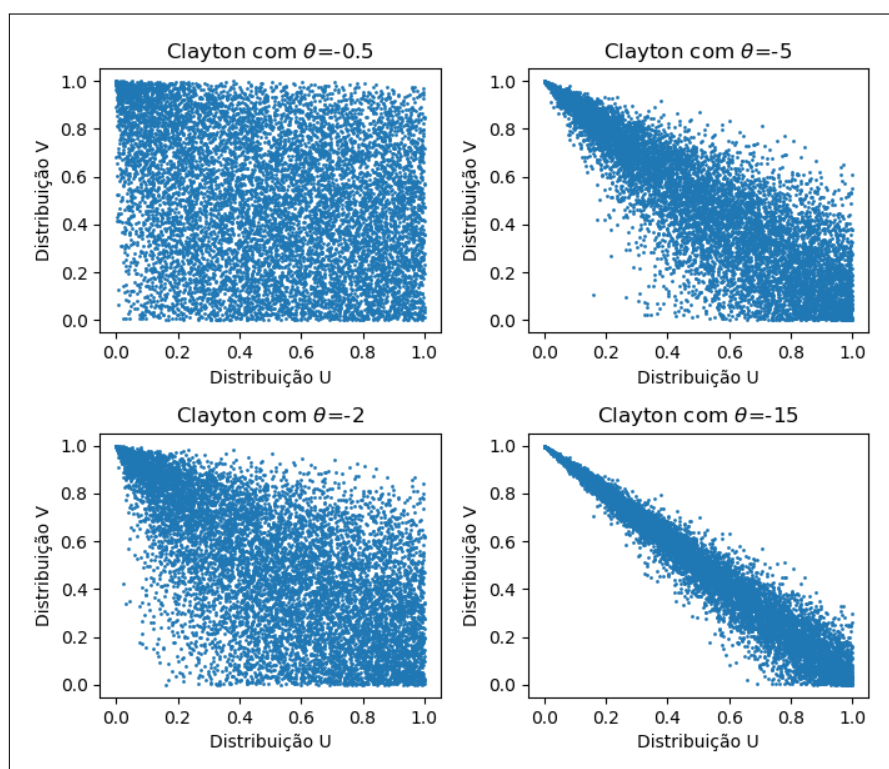


Figura 2.51: *Crossplot* de pontos gerados pela distribuição de Clayton com diferentes valores de θ negativos, com o algoritmo modificado.

Cópuas Arquimedianas- Cópula de Gumbel-Hougaard

A cópula de Gumbel-Hougaard foi introduzida inicialmente em Gumbel (1960) para modelar a dependência assimétrica quando há necessidade de capturar forte dependência da cauda

superior e fraca dependência da cauda inferior. Se esperamos que os resultados sejam fortemente correlacionados em valores altos, mas menos correlacionados em valores baixos, então a cópula de Gumbel é uma escolha apropriada. A função geradora $\phi(t)$ é dada por:

$$\phi_{\theta}(t) = (-\ln t)^{\theta} \quad (2.104)$$

Onde θ é o parâmetro da cópula restrito no intervalo $(0, +\infty) - [0]$. Se $\theta = 0$ as distribuições são tratadas como independentes.

Enquanto sua inversa $(\phi(t)^{-1})$ é dada por:

$$\phi_{\theta}^{-1}(t) = \exp(-t^{1/\theta}) \quad (2.105)$$

A expressão bivariada da CDF desta cópula é dada por:

$$C_{\theta}^{Cl}(u, v) = \exp[-[(-\ln(u))^{\theta} + (-\ln(v))^{\theta}]^{1/\theta}] \quad (2.106)$$

Esta cópula tem algumas características importantes como:

- Alta dispersão nos extremos inferiores das variáveis;
- Alta concentração no extremo superior das variáveis.

Assim como fizemos para a cópula de Clayton, para o obter a pdf de Gumbel-Hougaard temos que fazer $\frac{\partial^2 C(u, v)}{\partial u \partial v}$. E assim obtemos a equação 2.107:

$$c_{\theta}^{GH}(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v} = \frac{C_{\theta}(u, v)(\ln(u) \cdot \ln(v))^{\theta-1}}{u \cdot v [(-\ln(u))^{1/\theta} + (-\ln(v))^{1/\theta}]^{2-1/\theta}} ([(\ln(u))^{\theta} + (\ln(v))^{\theta}]^{\theta} + \theta - 1) \quad (2.107)$$

Note que a equação acima é minimamente extensa. A partir das equações (2.106) e (2.107) é possível gerar pdf's e cdf's (Figura 2.52) que ilustram os comportamentos já expostos:

Para obter os valores de v dependentes de u necessitamos obter as derivadas parciais em relação às variáveis e posteriormente realizar a operação inversa.

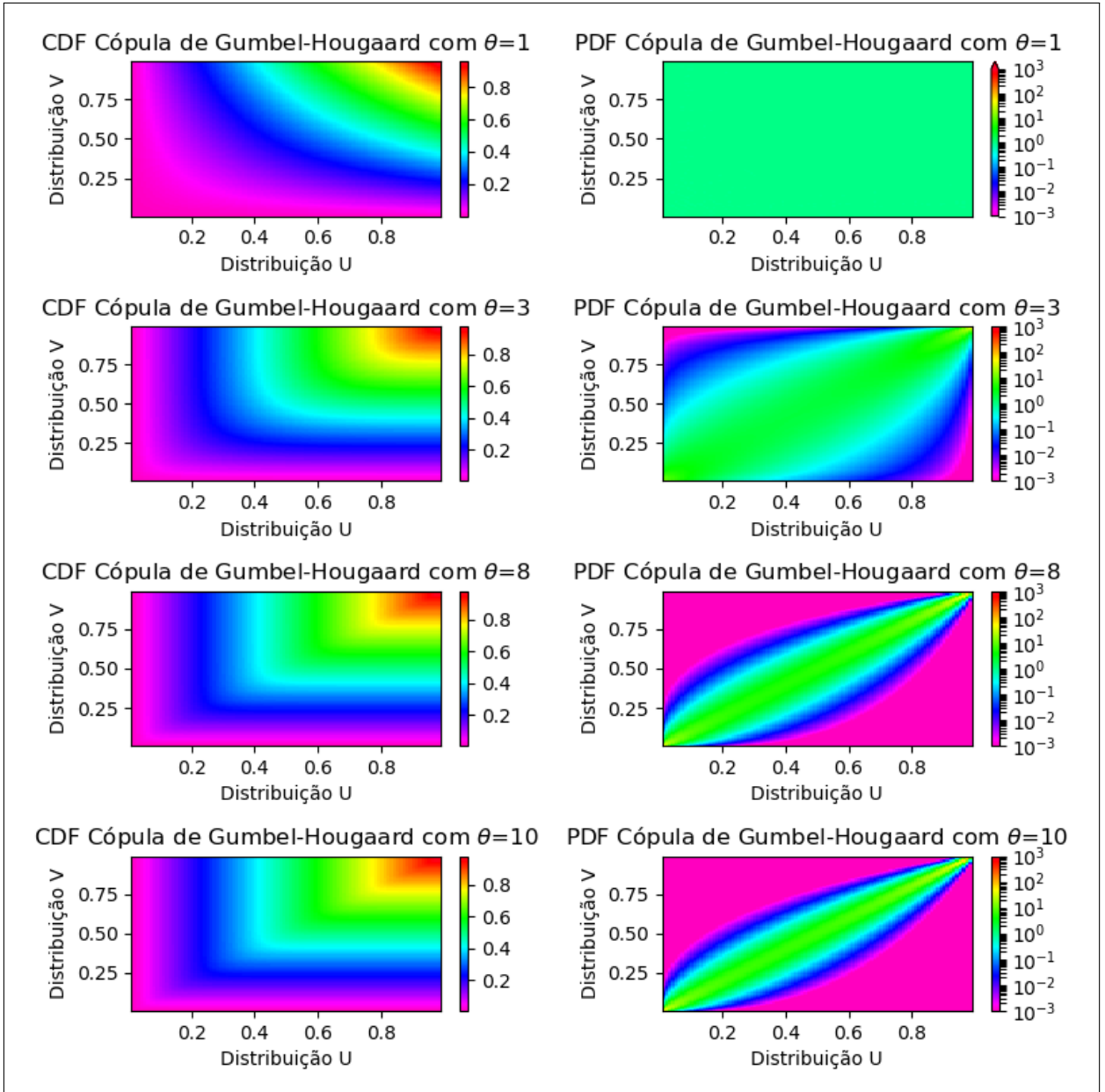


Figura 2.52: Superfícies de CDFs e PDFs da cópula de Gumbel-Hougaard, obtidos para diferentes valores de θ , que resultam em diferentes comportamentos das distribuições.

Conforme discutido em Embrechts et al. (2010), a cópula Gumbel-Hougaard não tem derivadas parciais inversíveis e por isso necessita de um procedimento especial para simular distribuições uniformes. Por isso, utilizaremos o procedimento sugerido por Embrechts et al. (2010) que modifica as variáveis conforme o algoritmo a seguir:

1. Simular as duas variáveis com n amostras (u, v) ;
2. Converter as d variáveis em variáveis uniforme com a distribuição de posições normalizadas (ru, rv) ;
3. Misturar as amostras em cada variável para garantir correlação próxima de zero entre elas;
4. Simular a distribuição uniforme $\kappa_1 = [-\pi/2, \pi/2]$ e calcular $\kappa_2 = \kappa_1 + \pi/2$;
5. Simular uma distribuições uniformes $\kappa_3 = [0, 1]$

6. Calcular $\alpha_1 = -\ln(\kappa_3)$
7. Calcular $\alpha_2 = \cos \kappa_1 - \kappa_1/\theta/\alpha_1$
8. Calcular $\alpha_3 = (\frac{\sin \kappa_2}{\theta})^{1/\theta} \cdot \frac{\alpha_2}{\cos \kappa_1}$
9. as variáveis dependentes serão dadas por:

$$\begin{aligned} ru_{dep} &= \exp(-\ln(ru)^{1/\theta}/\alpha_3) \\ rv_{dep} &= \exp(-\ln(rv)^{1/\theta}/\alpha_3) \end{aligned} \quad (2.108)$$

10. Converter as distribuições uniforme (ru_{dep}, rv_{dep}) nas distribuições originais (u, v) .

De posse deste algoritmo implementado no python (verificar o anexo). Pudemos gerar a Figura 2.53 que ilustra a geração de variáveis aleatórias para diferentes valores de θ :

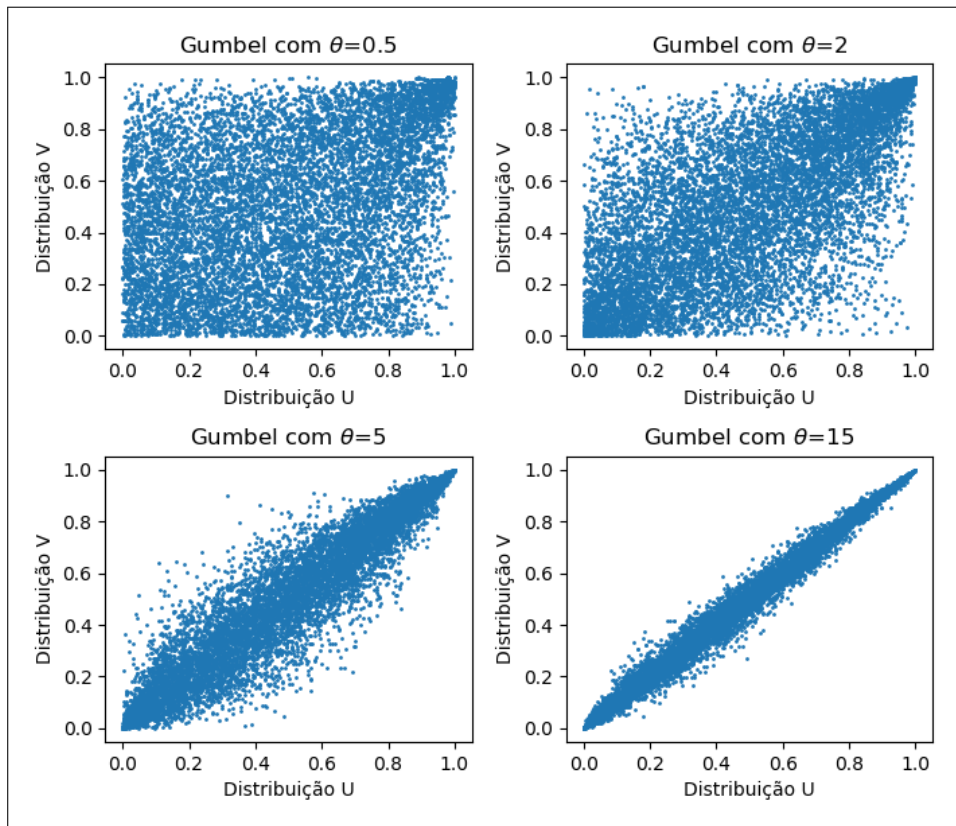


Figura 2.53: *Crossplot* de pontos gerados pela distribuição de Gumbel-Hougaard com diferentes valores de θ .

A distribuição de Gumbel não permite valores de θ menores do que 1 (excluindo assim os valores negativos). Por isso, implementamos a rotação gráfica assim como fizemos para a cópula de Clayton. Com esta operação "ampliamos" o domínio da simulação destas cópulas para os intervalos de $theta$ $(-\infty, -1) \cup (1, \infty)$.

A Figura 2.54, ilustra alguns exemplos de valores negativos:

Conforme (Mahfoud, 2012), a dependência entre o parâmetro $theta$ e o coeficiente de correlação de Kendall pode ser dada por :

$$\tau = 1 - \frac{1}{\theta} \quad (2.109)$$

E a dependência da cauda inferior é dada por $\lambda_{inf} = 0$ e $\lambda_{subp} = 2 - 2^{-1/\theta}$ (Cherubini et al. (2004)).

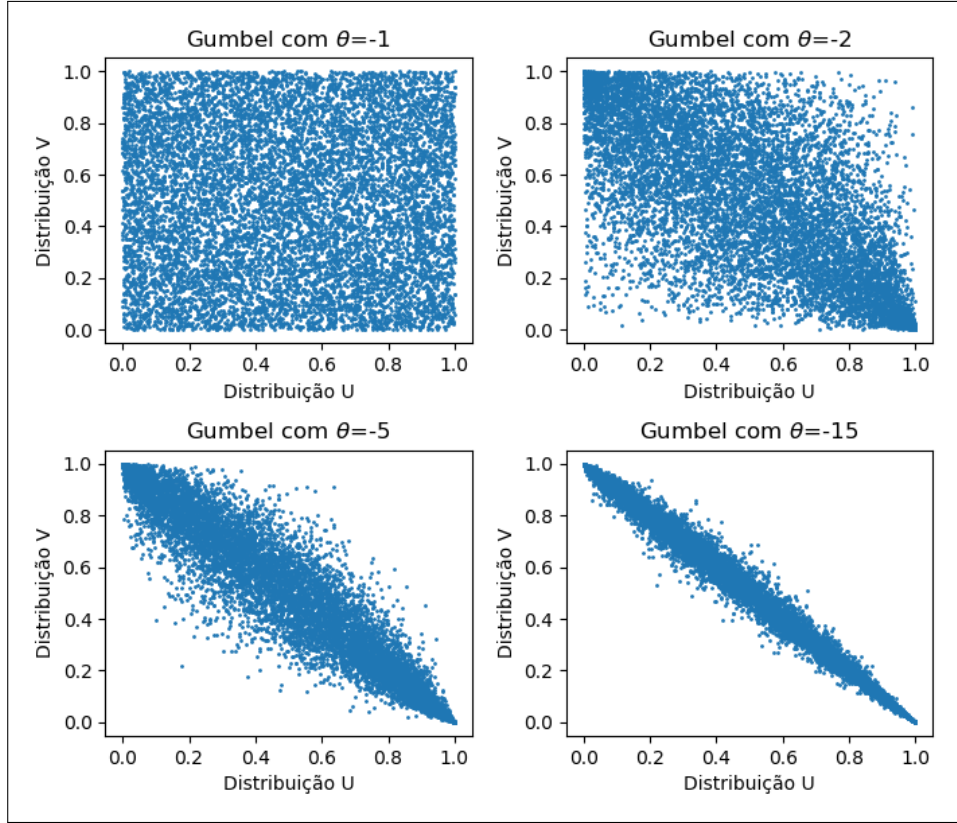


Figura 2.54: *Crossplot* de pontos gerados pela distribuição de Gumbel-Hougaard com diferentes valores de θ negativos.

Cópuas Arquimedianas- Cópula de Frank

A cópula de Frank foi introduzida inicialmente em Frank (1979) para modelar a dependências altamente simétricas em toda a sua extensão. O caso de independência será alcançado quando θ se aproxima de zero. Note também que a cópula de Frank não tem dependência em suas caudas inferior ou superior ($\lambda_i = \lambda_i = 0$). A cópula de Frank é, portanto, adequada para modelar dados caracterizados por fraca dependência da cauda. A função geradora ($\phi_\theta(t)$) da cópula de Frank é dada por:

$$\phi_\theta(t) = -\ln\left(\frac{\exp(-\theta t) - 1}{\exp(-\theta) - 1}\right) \quad (2.110)$$

Onde θ é o parâmetro da cópula restrito que abrange todo o intervalo $(-\infty, +\infty) - [0]$. Se $\theta = 0$ as distribuições são tratadas como independentes. Enquanto sua inversa ($\phi_\theta^{-1}(t)$) é dada por:

$$\phi_\theta^{-1}(t) = -\theta \ln(1 + \exp(-t)(\exp(-\theta) - 1)) \quad (2.111)$$

A expressão bivariada da CDF desta cópula é dada por:

$$C_\theta^{Fr}(u, v) = -\frac{1}{\theta} \ln\left(1 + \frac{(\exp(-\theta u) - 1)(\exp(-\theta v) - 1)}{(\exp(-\theta) - 1)}\right) \quad (2.112)$$

Algumas características desta cópula são:

- A ausência de uma cauda em cada extremidade da dispersão;
- Uniformidade da nuvem de pontos ao longo do eixo de correlação máxima de um extremo ao outro;

- Concentrações limitadas;
- A correlação é relativamente fraca devido ao grau amplo e uniforme de dispersão.

Para obter a distribuição da pdf temos que fazer $\frac{\partial^2 C(u,v)}{\partial u \partial v}$.

$$c_{\theta}^{Fr}(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v} = \frac{\theta[1 - \exp(-\theta)] \exp(-\theta(u+v))}{([1 - \exp(-\theta)] - [1 - \exp(-\theta.u)][1 - \exp(-\theta.v)])^2} \quad (2.113)$$

A partir das equações (2.112) e (2.113) é possível gerar pdf's e cdf's (Figura 2.55) que ilustram os comportamentos já expostos:

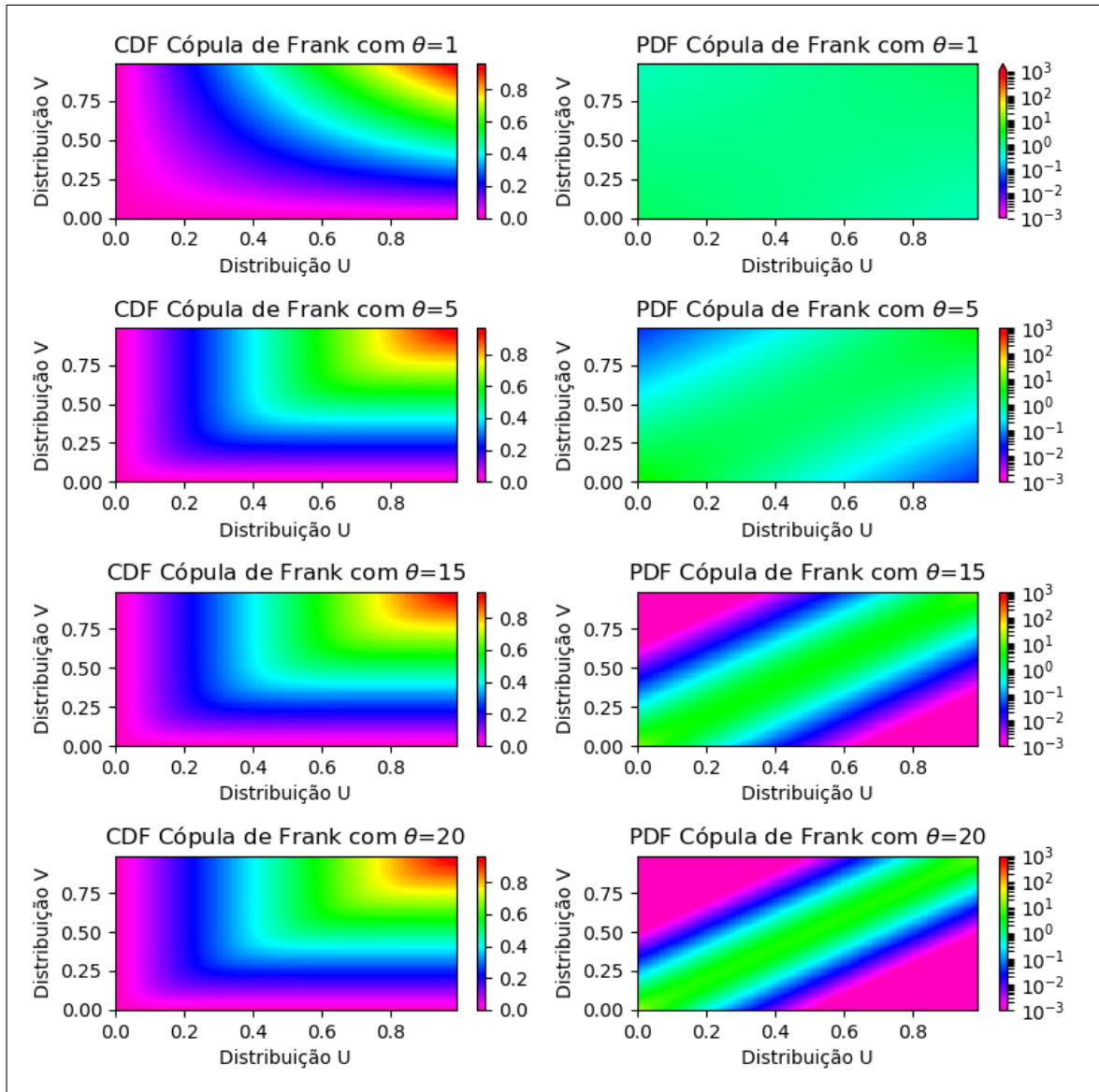


Figura 2.55: Superfícies de CDFs e PDFs da cópula de Frank, obtidos para diferentes valores de θ , que resultam em diferentes comportamentos das distribuições.

Para obter os valores de v dependentes de u necessitamos obter as derivadas parciais em relação às variáveis não avaliadas (no caso apenas u). E em seguida precisamos obter a inversa de v , dessa forma chegamos a seguinte equação:

$$v_{dep} = (1 - \theta) \cdot \ln \left(1 + \left(\frac{v(1 - \exp(\theta))}{-\exp(-\theta.u) + (-1 + \exp(-\theta.u)).v} \right) \right) \quad (2.114)$$

A partir da equação (2.114) podemos gerar amostras de distribuições bivariadas com estrutura de dependência da cópula de Frank conforme o algoritmo a seguir:

1. Simular as duas variáveis com n amostras (u, v) ;
2. Converter as d variáveis em variáveis uniforme com a distribuição de posições normalizadas (ru, rv) ;
3. Misturar as amostras em cada variável para garantir correlação próxima de zero entre elas;
4. Definir o vetor $ru_{dep} = ru_{indep}$.
5. Definir o o vetor $rv_{dep} = (-1/\theta) \cdot \ln(1 + num/den)$, onde:
 - $num = v \cdot (1 - \exp(-\theta))$
 - $den = -\exp(-\theta \cdot u) + (-1 + \exp(-\theta * u)) \cdot v$
6. Converter as distribuições uniforme (ru_{dep}, rv_{dep}) nas distribuições originais (u, v) .

De posse deste algoritmo implementado no python (verificar o anexo). Pudemos gerar a Figura 2.56 que ilustra a geração de variáveis aleatórias para diferentes valores de θ :

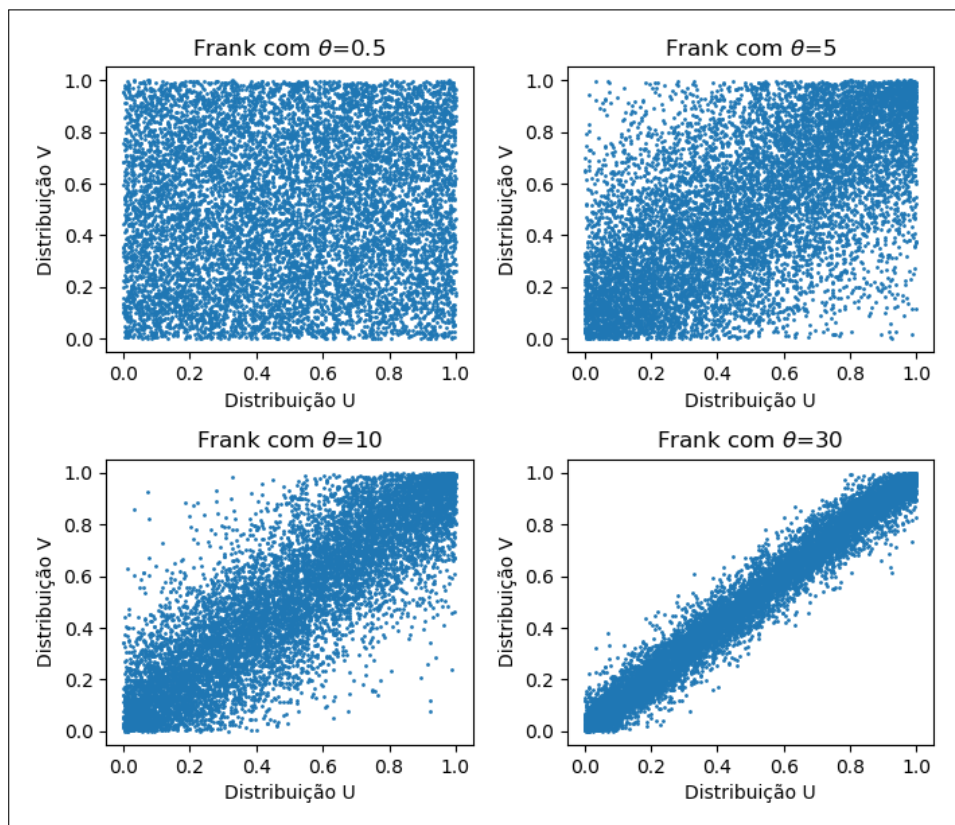


Figura 2.56: *Crossplot* de pontos gerados pela distribuição de Frank com diferentes valores de θ .

Ao contrário da cópula de Clayton e Gumbel, a cópula de Frank permite um amplo espectro de valores de dependência. Isto significa que o parâmetro θ da cópula de Frank permite a modelagem com dependência positiva ou negativa nos dados (Figura 2.57).

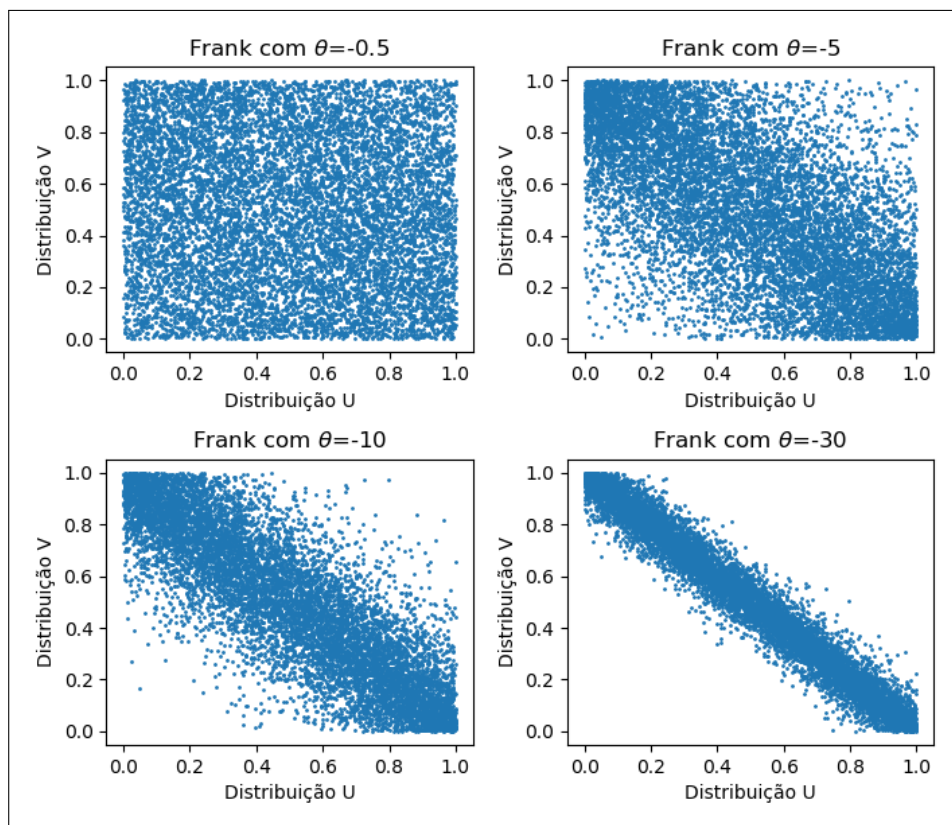


Figura 2.57: *Crossplot* de pontos gerados pela cópula de Frank com diferentes valores negativos de θ .

2.4.3 Métodos de Reordenamento: Reordenamento Numérico

A alteração da ordem das variáveis em uma série multivariada é o princípio básico para mudar a estrutura de correlação entre duas variáveis. Este processo preserva as características de média, assimetria, variância e curtose das distribuições marginais.

A operação de troca é normalmente feita por métodos numéricos recursivos. Esta técnica é utilizada em diversos campos de estudo como na própria simulação de Monte-Carlo (Miller, 1998; Vorechovsky, 2002). Em pesquisas estatísticas, com a designação de *data-swapping*¹² Dalenius and Reiss (1982); Moor (1996) e também na análise de correlação de séries temporais de aplicações financeira, com a designação de *empirical shuffling method* (e.g. Kenett et al. (2015)).

A ideia de permutar elementos de uma distribuição ou entre distribuições é, entretanto, mais antiga do que as referências supracitadas. A teoria evoluiu a partir dos trabalhos sobre teste de significância estatística de Fisher (1926) e Pitman (1937).

Embora Fisher tenha introduzido a aleatorização (isto é, a associação aleatória de tratamento de objetos) no planejamento experimental. Foi Pitman quem demonstrou teoricamente a correspondência entre aleatoriedade e significância. A partir de então a aleatoriedade se tornou uma das premissas para os principais testes estatísticos como o teste-t e o teste-F.

Miller (1998) resume três formas diferentes para lidar com a correlação de variáveis numericamente (duas delas já mencionadas neste trabalho). As técnicas apresentadas pelo autor são rapidamente descritas a seguir:

1. O primeiro se baseia em reamostrar a pdf gerada pelas distribuições normais gerada pela equação analítica bivarianda normal (similar ao *bootstrapping* mencionado no trabalho).

¹²Com o objetivo de tornar dados de pesquisas não-relacionados à população pesquisada, mantendo assim sua privacidade

2. O segundo se baseia em tornar todos os pares alinhados perfeitamente e progressivamente desalinhar alguns pares da distribuição (degradando a covariância e correlação) até chegar à correlação desejada;
3. O terceiro se baseia em dividir as duas distribuições em bandas de percentis e adicionar pontos nas duas variáveis em percentis equivalentes. A quantidade de divisões das distribuições determinará a correlação (esquema de monte carlo estratificado semelhante ao Hiper-cubo Latino McKay et al. (1979) proposto 20 anos antes da publicação de Miller).

A Figura 2.58 ilustra os resultados publicados por Miller (1998):

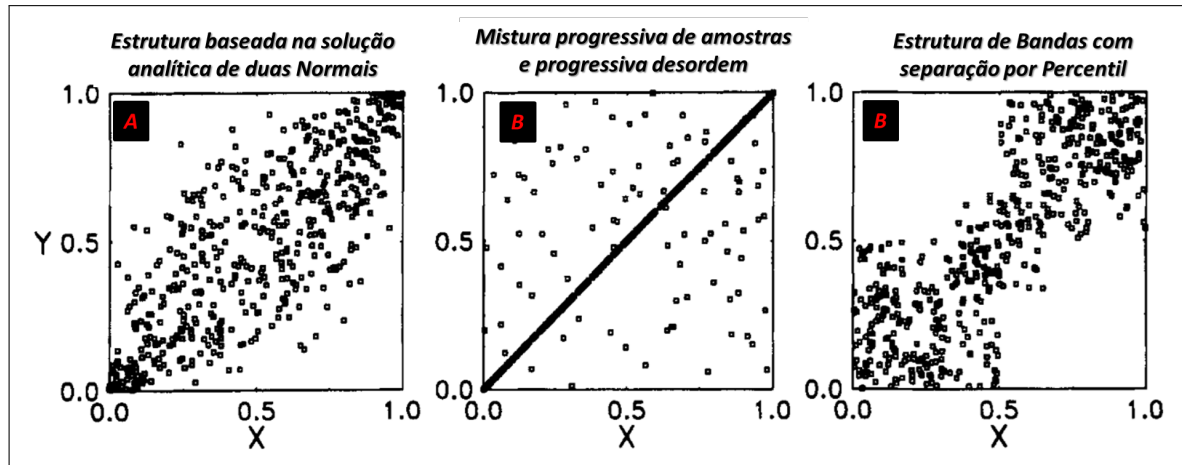


Figura 2.58: Séries Correlacionadas geradas pelos métodos propostos por Miller (1998) em seu artigo original (legendas foram traduzidas). Reamostragem da função normal analítica (A). O problema que abordaremos aqui é referente à figura na parte central (B). Amostragem estratificada (C). Todos os exemplos com as mesmas distribuições marginais e com correlação de Pearson aproximadamente 0.8.

O gráfico do centro na Figura 2.58 reflete os resultados obtidos pelo método de reordenamento numérico discutido nesta seção. Observa-se nesta figura uma forte concentração de pontos no eixo de correlação máxima e a presença de pontos dispersos (que geram a redução da correlação). O comportamento modelado se assemelha bastante com uma variável perfeitamente correlacionada, sob efeito de uma espécie de ruído aleatório.

Apesar de possível, esta dispersão não representa o comportamento de tendência em dispersão, que capta as complexidades verificadas em fenômenos da natureza discutidos neste trabalho e, portanto, não é a metodologia mais indicada na análise volumétrica.

É interessante notar que esta metodologia de simulação de variáveis correlacionadas não é propriamente documentada. Por isso, é recorrentemente publicada como "nova técnica" (e.g. (Vorechovsky, 2002; Hantschel and Kauerauf, 2009; Kenett et al., 2015)).

Vorechovsky (2002) cita a metodologia de reordenamento como um "problema de otimização de ordenamento e combinação para controle da estrutura de correlação". A metodologia é baseada no princípio de que a troca de posições de amostras degrada a correlação entre distribuições (positiva ou negativa) convergindo-a a valores próximos de zero. Ordena-se as distribuições em pares de variáveis em mesma direção (quando se quer obter correlações positivas) ou em ordem inversa (quando se quer obter correlações negativas). Pares de variáveis são sorteados aleatoriamente, uma das variáveis tem seus elementos trocados e aos poucos a correlação entre múltiplas variáveis é reduzida até o valor alvo (algoritmo traduzido na Figura 2.59).

Uma das vantagens deste método é poder lidar com o coef. de correlação de Pearson e não o coeficiente de correlação de Spearman ou Kendall (como nos métodos analíticos já abordados). Outra vantagem deste algoritmo também é poder realizar os experimentos com diversas geometrias de distribuições sem a necessidade de transformações.

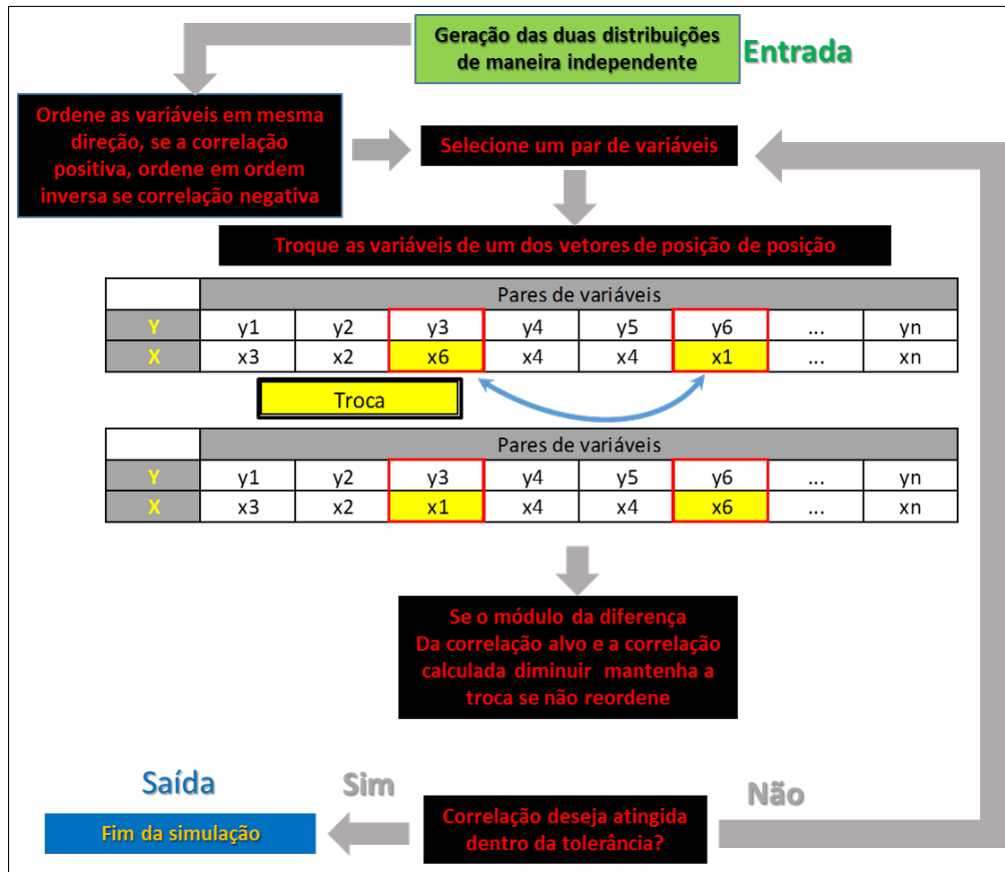


Figura 2.59: Fluxograma de mudança de correlação numericamente por troca de pares de variáveis.

Uma preocupação antes de iniciar o algoritmo é verificar se o coeficiente de correlação de Pearson desejado para as distribuições é atingível. Para isso é necessário utilizar o método GSC (já discutido anteriormente) ou alguma solução analítica para distribuições conhecidas.

É claro que esse procedimento de permutação certamente tem desvantagens se comparado com os métodos analíticos já discutidos (Iman Conover e Cópulas). Um deles é que a convergência é lenta em distribuições grandes se comparado aos métodos já citados. Outra limitação é que a aproximação precisa dos coeficientes de correlação pode levar a oscilações em mínimos locais gerando um *loop* infinito.

Os experimentos realizados no âmbito deste trabalho indicam que para simulações da ordem de 40000 iterações, a precisão de pelo menos duas casas decimais é atingida com rapidez e boa aproximação. Esta constatação está em linha com Vorechovsky (2010) que estabelece métricas empíricas que determinam o limite de precisão para o método de reordenamento simples proporcional a \sqrt{n} , onde n é o número de simulações.

Vorechovsky (2010) também traz expressões que relacionam a precisão com o número de distribuições simulações e de variáveis a serem correlacionadas. Entretanto, como tratamos do problema bivariado exclusivamente não precisamos de uma abordagem mais profunda sobre o assunto.

O Método GSSC- *Generate Sort Shuffle and Correlate*

Como uma extensão à metodologia GSC (Generate, Sort and Correlate) para avaliação da correlação máxima ou mínima proposta no trabalho de Demirtas and Hedeker (2011) e aos trabalhos de reordenamento numérico já mencionados.

Criamos um método de convergência mais rápida feita como produto desta dissertação designada de *Generate, Sort, Shuffle and Correlate* (GSSC), onde propomos uma otimização

relacionada ao uso de uma janela móvel de embaralhamento (*shuffle*) e/ou ordenação das variáveis dos dados seguida de sucessivos ajustes na janela apenas como um "ajuste fino" da correlação (Figura 2.60).

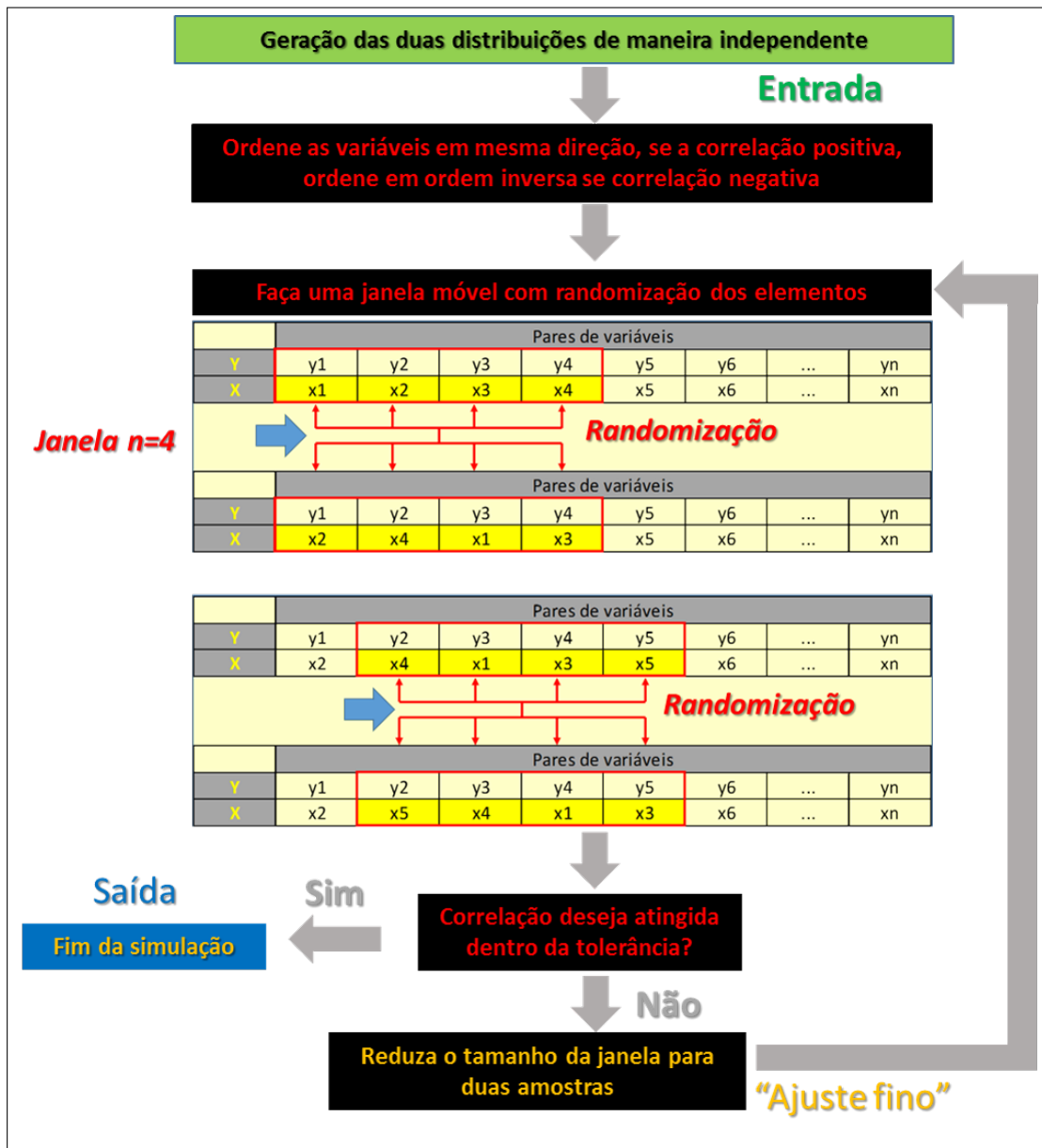


Figura 2.60: Fluxograma de mudança de correlação numericamente por troca de variáveis com o uso de janela móvel.

O objetivo aqui é acelerar a convergência do algoritmo e adicionalmente permitimos que ele tenha diferentes geometrias de dispersão ao longo do eixo de correlação máxima controladas pela janela móvel de reordenamento.

O tamanho da janela dependerá de dois parâmetros: (1) o tamanho da distribuição (quantidade de simulações) e; (2) O nível de correlação desejado. Através de testes empíricos, verificamos que a janela de embaralhamento pode ser dada por:

$$janela = 1 + (n - 1) * (1 - |\rho_{desejada}|^k) \quad (2.115)$$

Onde n é o número de amostras, $|\rho_{desejada}|$ é o módulo do coeficiente de correlação de Pearson desejado entre as duas distribuições e k é um expoente adimensional utilizado para ajusta a

convergência, para distribuições que variam de 30.000 a 40.000 amostras, k pode é uma constante definida como 2.5.

A Figura 2.61 demonstra a diferença dos resultados pelo reordenamento numérico "simples" apenas com a troca de variáveis em pares (conforme proposto em Dalenius and Reiss (1982); Miller (1998); Vorechovsky (2002)) e do reordenamento com a janela móvel proposto no método GSSC (proposto neste trabalho).

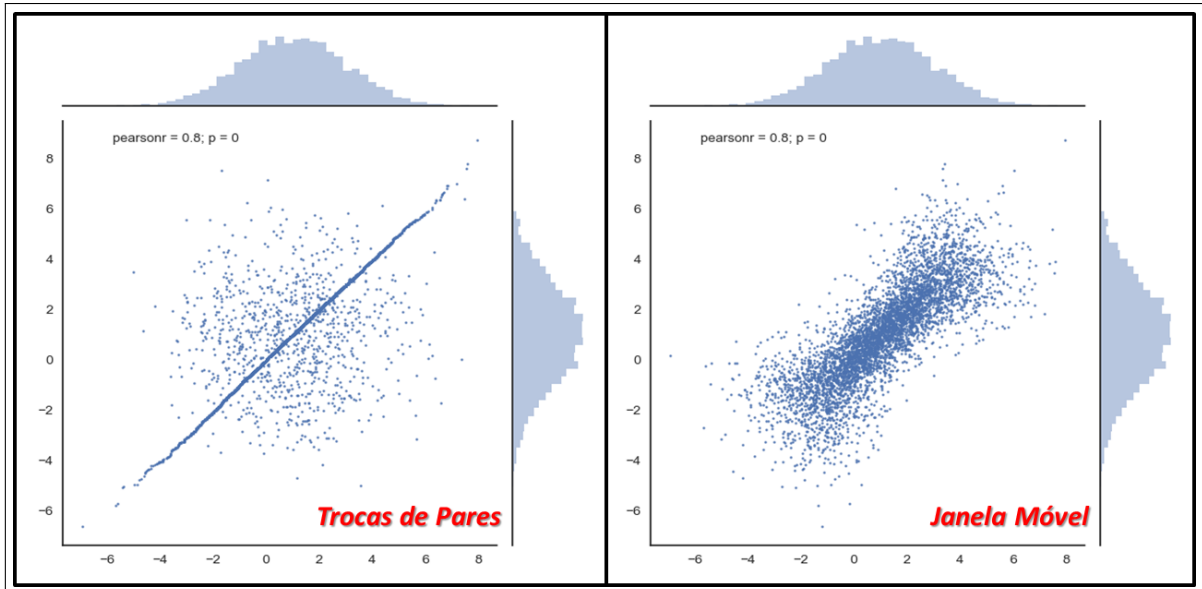


Figura 2.61: *Crossplot* de pontos com coeficiente para comparar as dispersões de pontos utilizando a a troca de pares e método GSSC. Ambas as figuras tem coeficiente de correlação de Pearson igual a 0,8 .

A dispersão de pontos utilizando o método GSSC com embaralhamento por janelas é bastante semelhante à distribuição obtida com a cópula de Frank (Figura 2.62).

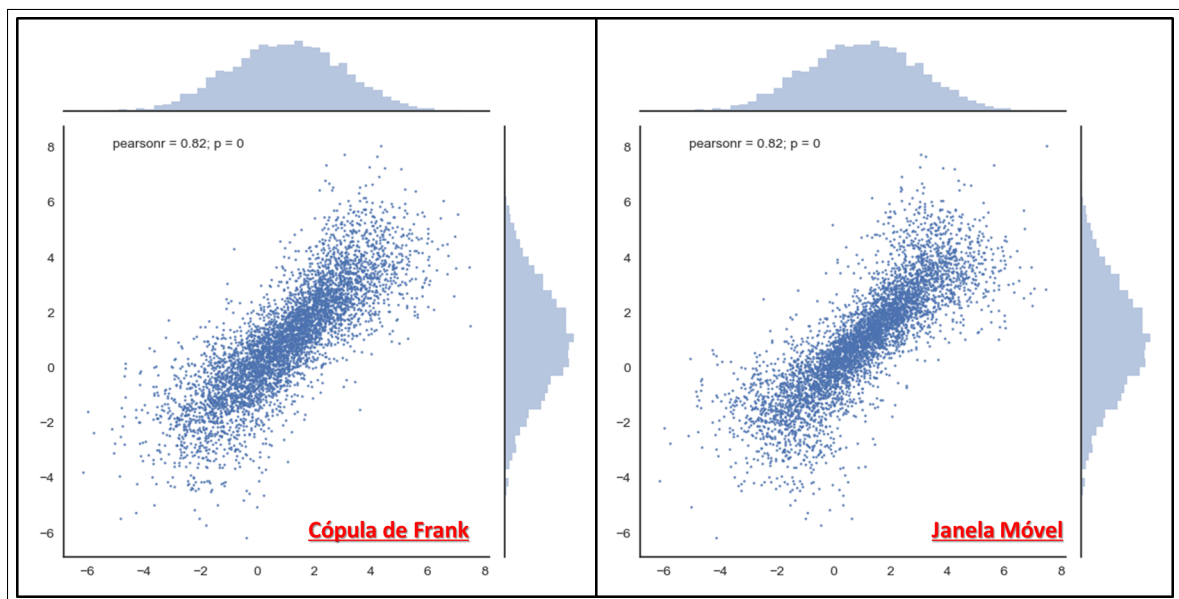


Figura 2.62: *Crossplot* de pontos com coeficiente para comparar as dispersões de pontos utilizando a cópula de Frank e GSSC. Ambas as figuras tem coeficiente de correlação de Pearson igual a 0,82 .

É possível gerar funções semelhantes a cópulas com reordenamento numérico. Trottini et al. (2011) propõe utilizar o estado inicial das amostras geradas pela cópula, obtendo assim uma

primeira aproximação baseada no coeficiente de correlação de Spearman. Desta forma, a cópula obtida entre as distribuições poderia ser recursivamente alterada, de maneira a preservar uma estrutura geral da cópula e ajustar apenas alguns pontos. Este procedimento pode ser estendido para qualquer cópula como estágio inicial.

É importante destacar que a preservação da estrutura de correlação original dependerá do coeficiente de correlação de Pearson desejado (quão menor o coeficiente de correlação menor a semelhança da estrutura original).

2.4.4 Limitações em Correlações Multivariadas

Apesar da correlação ser definida e medida em termos de combinações de pares de variáveis, é necessário considerar correlações multivariadas (quando mais de duas distribuições apresentam correlações entre si). Pode-se, por exemplo, ter um grupo de seis variáveis aleatórias onde todas são fortemente e positivamente correlacionadas entre si.

Existem limitações matemáticas à maneira pelas quais variáveis aleatórias podem ser correlacionadas num ambiente multivariado. Por exemplo, a seguinte correlação de três variáveis é estatisticamente impossível:

- Correlação (A,B) = -0,9;
- Correlação (A,C) = -0,9;
- Correlação (B,C) = -0,9;

As duas primeiras expressões indicam que grandes valores de A tendem a ser combinados com valores pequenos de ambos B e C. Enquanto a última declaração implica que os pequenos valores de B tendem a ser combinados com grandes valores de C. Esta condição não pode ser realizada por nenhum esquema de simulação real.

O risco de combinações impossíveis aumenta com a quantidade de variáveis quando há entre elas correlações negativas, o que acarretaria em *loops* infinitos ao utilizar algoritmos de ajuste iterativo sem critérios específicos de parada relacionados ao número de iterações Nelsen (2006).

É importante notar que situações como as das Distribuições A,B,C citadas acima, serão representadas por uma matriz que não será positiva definida. Daí a importância de fazer a verificação desta condição mesmo em algoritmos de reordenamento que não usem esta computação como pré-requisito. No caso geológico normalmente trabalhamos com correlações positivas, neste cenário múltiplas correlações tendem a ser realizáveis, podemos por exemplo ter correlações positivas de mais de 2 variáveis utilizadas na volumetria (*GRV*, *S_{hc}*, ϕ , *NTG,TF* e *RGO, FR*) entre si.

2.4.5 A questão da escala em Correlações Multivariadas

É importante destacar a importância da escala dos parâmetros avaliados e o valor de correlação utilizado. No caso da avaliação de prospectos os valores de correlação normalmente são obtidos de uma escala regional e aplicados no elemento local (o prospecto). Entretanto, nestes casos normalmente estamos trabalhando com escalas diferentes.

Vamos imaginar uma correlação de duas distribuições normais (A,B) com estrutura de correlação que segue a cópula de Clayton. Agora vamos extrair subamostras desta população e verificar como é a estrutura de correlação destas subamostras em relação à população (Figura 2.63)

A partir deste exemplo podemos aprender duas lições: (1) A escala pode mudar os valores nominais de correlação entre duas variáveis de maneira sutil (retângulo vermelho) ou severa (retângulo azul) e ;(2) A escala de observação pode mudar também a estrutura de correlação entre as variáveis ainda que os seus valores nominais não seja severamente alterados.

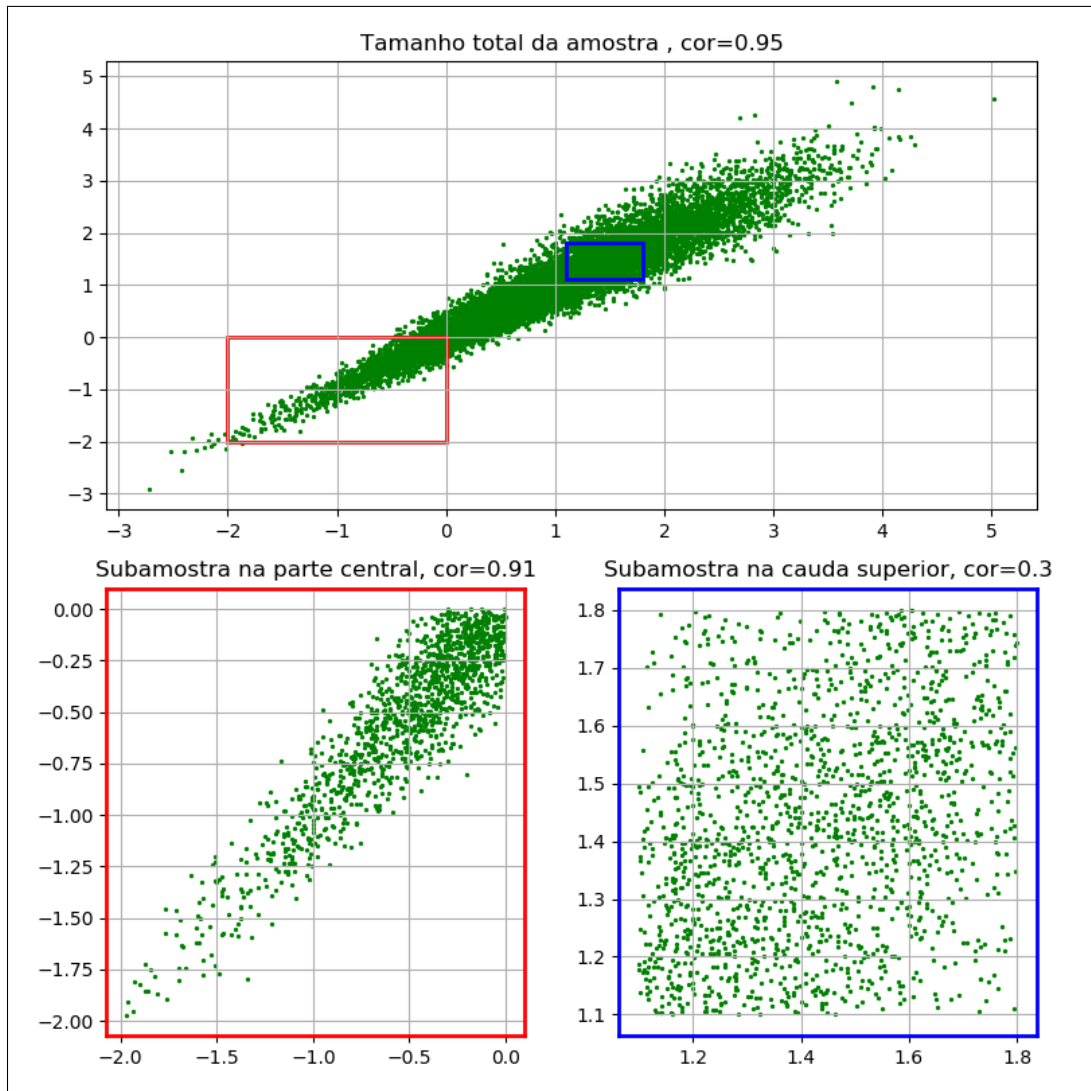


Figura 2.63: *Crossplot* de pontos de duas distribuições normais. Acima temos a distribuição completa com coeficiente de correlação de Pearson de 0.95, seguindo uma estrutura de correlação que segue a cópula de Clayton ($\theta=1$). Os retângulos nas figuras são as sub amostras selecionadas e apresentadas na parte inferior. Note na parte inferior esquerda que a subamostra preserva a alta correlação mas apresenta uma estrutura semelhante à cópula de Gumbel enquanto a figura da parte central (quadrado inferior direito) apresenta um baixíssimo coeficiente de correlação.

Esta observação nos leva a uma conclusão importante. Não podemos simplesmente aplicar correlações regionais em prospectos, se a escala de incerteza determinada pela razão p_{10}/p_{90} das variáveis análogas são diferentes. Destaco que, mesmo que as escalas sejam iguais, contextos geológicos diferentes possivelmente vão levar a estruturas e valores de correlação entre variáveis diferentes. Portanto, neste trabalho evitaremos determinar valores sugeridos de correlação entre variáveis, ao invés disso indicaremos faixas apropriadas e apenas apontaremos exemplos de como determinar estas correlações bem como o impacto delas a análise volumétrica.

Valores muito baixos (inferiores a 0.3) podem ser pouco efetivos e indicarem na verdade uma ausência de correlação confiável Cherubini et al. (2004). Dessa forma, recomendamos o uso da faixa entre 0.3 e 0.8. Uma vez que valores superiores a 0.8 podem representar um grau de certeza infactível sobre os dois parâmetros objeto da correlação.

Capítulo 3

Metodologia

Neste capítulo abordamos as metodologias e premissas aplicadas na correlação de parâmetros de estimativa volumétrica. Em seguida mencionaremos a correlação entre parâmetros de volume e variáveis de risco geológico.

Num segundo momento abordaremos a construção das ferramentas e implementação dos códigos produzidos. Apesar de comentarmos os algoritmos utilizados, não entraremos em detalhe sobre os códigos. Os programas em python mais relevantes poderão ser encontrados nos anexos do trabalho.

3.1 Correlação de Parâmetros

A correlação entre parâmetros de avaliação volumétrica é bastante intuitiva entre os geocientistas, principalmente entre aqueles que trabalham na avaliação de reservatórios. No âmbito da exploração, alguns autores comentam sobre o assunto mas, não vão a fundo em metodologias ou estimativas de correlação (Harbaugh et al., 1995; Rose, 2001; Favoreto et al., 2016).

Conforme comentamos no Capítulo 2, o coeficiente de correlação de Pearson nos traz duas informações: (1) a direção de relação entre as variáveis e; (2) o grau de dispersão em relação ao eixo linear de maior correlação (para um mesmo coeficiente de correlação é possível ter várias estruturas de correlação).

Uma das metodologias de estimar os valores de correlação das variáveis envolvidas na volumetria seria: obter valores médios dos parâmetros de campos próximos e/ou dados de poços e fazer uma regressão linear das variáveis. Este processo demandaria o licenciamento de informações normalmente confidenciais. Por esta razão, estimamos a ocorrência de correlação, com base em simulações numéricas, em equações empíricas amplamente utilizadas na indústria e dados públicos.

Nesta seção evidenciaremos principalmente a direção de correlação entre as variáveis com modelos. Com isso tentamos propor metodologias de identificar e estimar correlações entre parâmetros de análise volumétrica.

É importante destacar que o componente de dispersão dos pontos não é geológico (no fim das contas existe apenas um valor de média real para o que ocorre em subsuperfície em um campo de petróleo), mas sim intimamente relacionado ao nosso desconhecimento dentro da variedade de cenários geológicos possíveis e adequados aos nossos dados. Por outro lado, cenários geológicos mais complexos aumentam a nossa incerteza sobre a área, gerando a necessidade de valores mais baixos de correlação bem como valores maiores de razão p10/p90.

A incerteza nos leva à demanda por mais dados que possam mitigar as dúvidas que levarão a novas perguntas...(uma estrutura de causalidade mútua). Uma consequência disso é que dados em abundância e de boa qualidade vão levar a maior confiabilidade e relações mais precisas (e não necessariamente altas) das variáveis e suas correlações e questões mais particulares.

Apesar da correlação poder ocorrer em múltiplos parâmetros simultaneamente, neste trabalho iremos nos concentrar em pares de variáveis. Ao construir múltiplos pares de variáveis correlacionadas poderemos calcular a matriz de covariância/correlação entre as variáveis conjuntamente. Isto irá demonstrar naturalmente as correlações que eventualmente foram deixadas de lado.

Inicialmente trataremos das correlações relacionadas à rocha (*Area* vs *Espessura*) do ponto de vista geométrico. Em seguida lidaremos com a relação de (*Area* vs *Fácies*).

As fácies discutidas neste trabalho serão discutidas no âmbito geológico. Entretanto, o que temos no contexto exploratório, normalmente, são as fácies sísmicas que poderemos delimitar com base em mapas interpretados. Dessa forma, o experimento envolvendo as fácies, quando aplicado a dados sísmicos, podem ser estendidos para os parâmetros porosidade, *net pay* ou saturação (todos tem relação direta com a amplitude).

Apresentamos também a relação entre Porosidade (ϕ) vs *Fator de Recuperação*. Em seguida avaliamos as correlações relacionadas a fluido (e.g. B_o vs *RGO*). É importante destacar que outras correlações (e.g. B_o vs B_g) podem ser exploradas. Entretanto, neste trabalho tentaremos limitar o número de correlações apresentadas.

Um ponto importante a ser ressaltado é que não estamos advogando o uso irrestrito das correlações na etapa exploratória. O que propomos neste trabalho é uma metodologia que possibilite à equipe de interpretação, utilizar e explorar a relação de variáveis ao avaliar uma oportunidade, quando há segurança e informações *a priori* que permitam este ajuste.

Por fim, tentaremos evidenciar a correlação entre risco de retenção e os parâmetros espessura e chance de ocorrência de gás em trapas estruturais contra a falha.

3.1.1 Área e *Gross pay*

A relação entre área e *gross pay*¹ pode ser vista de duas formas: (1) Devido ao aumento/diminuição natural de espessura causada por fatores como erosão/não-deposição e; (2) incertezas geométricas da superfícies de topo e base do reservatório geradas por imprecisão dos dados sísmicos sísmica.

As variações geométricas são normalmente relacionadas ao modelo de velocidade devido à imprecisão na sua estimativa. As variações de espessura em função da variação lateral de fácies reservatório são encontradas em trapas estruturais e/ou estratigráficas penalizadas pelo NTG². Consideraremos os casos de incerteza relacionadas ao modelo de velocidade. Apesar de não comentar sobre as conversões especificamente, avaliaremos modelos sintéticos que são decorrentes da incerteza deste elemento. A metodologia aqui empregada consiste nos seguintes passos:

1. Definir o topo do reservatório (e base quando aplicável);
2. Fazer cortes simulando diferentes contatos;
3. Computar a área cortada pelo contato;
4. Fazer a computação da média do atributo desejado do topo do reservatório até a base ou contato (o que for maior) por toda a área definida;

Trataremos agora de um exemplo clássico que consiste de um prospecto estrutural quaquaversal, totalmente preenchido de reservatório como num depocentro invertido (Figura 3.1). Nesta pseudo-oportunidade verificaremos numericamente que ocorre o aumento de espessura à medida que aumentamos a área. Este caso é comum em prospectos do tipo casco de tartaruga³ no Golfo do México, ele também poderia ser equivalente a acumulações já verificadas no Brasil, nos cascos de tartarugas de idade Albiana da Bacia de campos.

¹Espessura de Rocha Prospectiva na trapa

²Net-to-Gross: Razão de espessura de reservatório (*net*) vs espessura total prospectiva *gross-pay*.

³Estruturas de Rollover gerados pelo deslocamento do sal e aumento do espaço lateral de acomodação

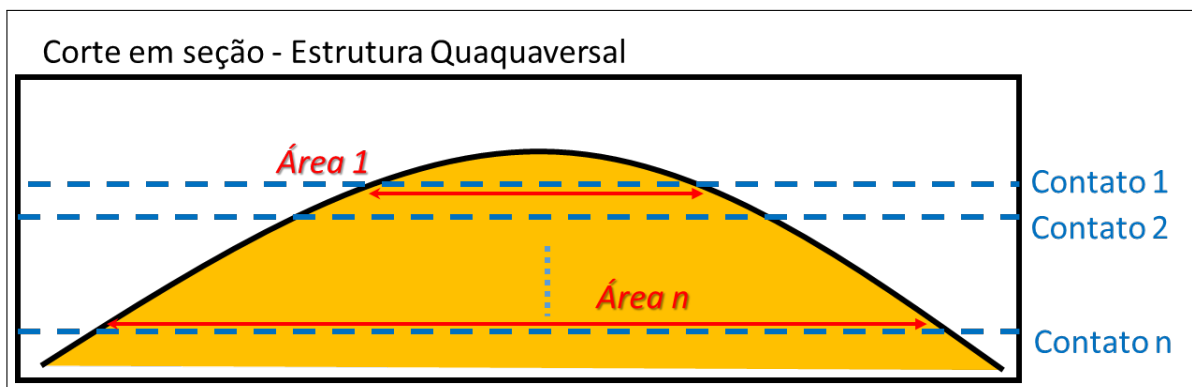


Figura 3.1: Modelo esquemático em seção de prospecto quaquaversal com espessura variável com aumento da área.

A seguir criamos um modelo sintético (Modelo ES1, Figura 3.2) da superfície no topo da trapa para exemplificar a metodologia. Basicamente calculamos o volume entre o topo do reservatório e o contato através de uma integração numérica (utilizando o software Petrel), calculamos também a área numericamente da estrutura acima do contato hidrocarbonetos/água (Figura 3.1). Baseado na razão de volume e área, calculamos a espessura média para cada cenário de corte de contato.

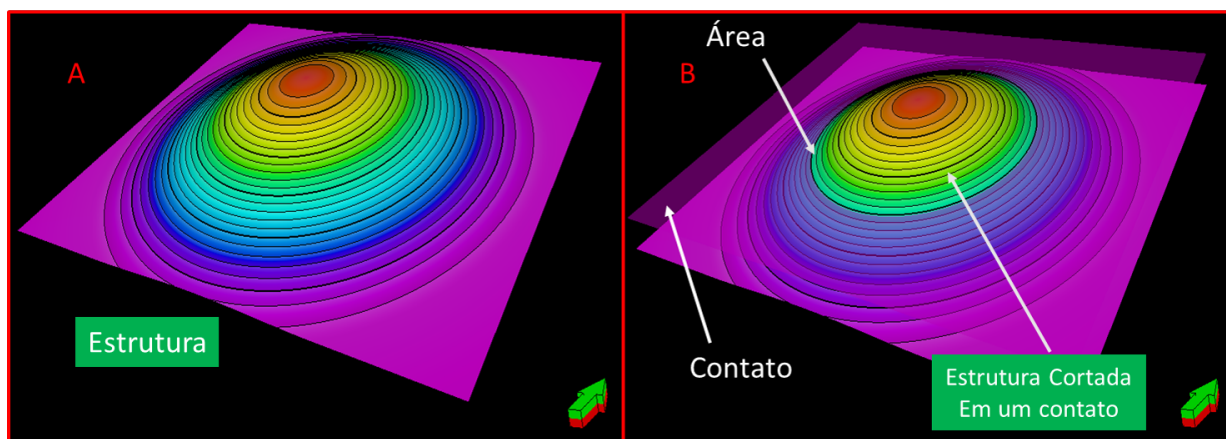


Figura 3.2: Topo do reservatório do Modelo esquemático (ES1) que constitui um prospecto quaquaversal radialmente simétrico (A), superfície com valor de z constante simulando um contato cortando a estrutura (B).

A partir deste modelo podemos plotar o gráfico de dispersão de pontos para a área vs espessura média (e volume de rocha em cores) para a estrutura representada na Figura 3.2.

Neste cenário verifica-se um comportamento praticamente linear entre área e espessura média, onde os pontos estão bastante alinhados e portanto, a correlação é bem próxima de 100% (99,1%). É possível verificar uma leve curvatura nos pontos, que variará com a esfericidade da geometria da estrutura.

Em um cenário de máxima simetria e nenhuma esfericidade (um paralelepípedo com base retangular), obteríamos uma distribuição de pontos totalmente linearizada (coeficiente de correlação 100%) de espessura média e área.

Considere agora um cenário onde o topo do reservatório possui deformações importantes (Modelo ES2), conferindo assimetria forte ao mapa estrutural (Figura 3.4).

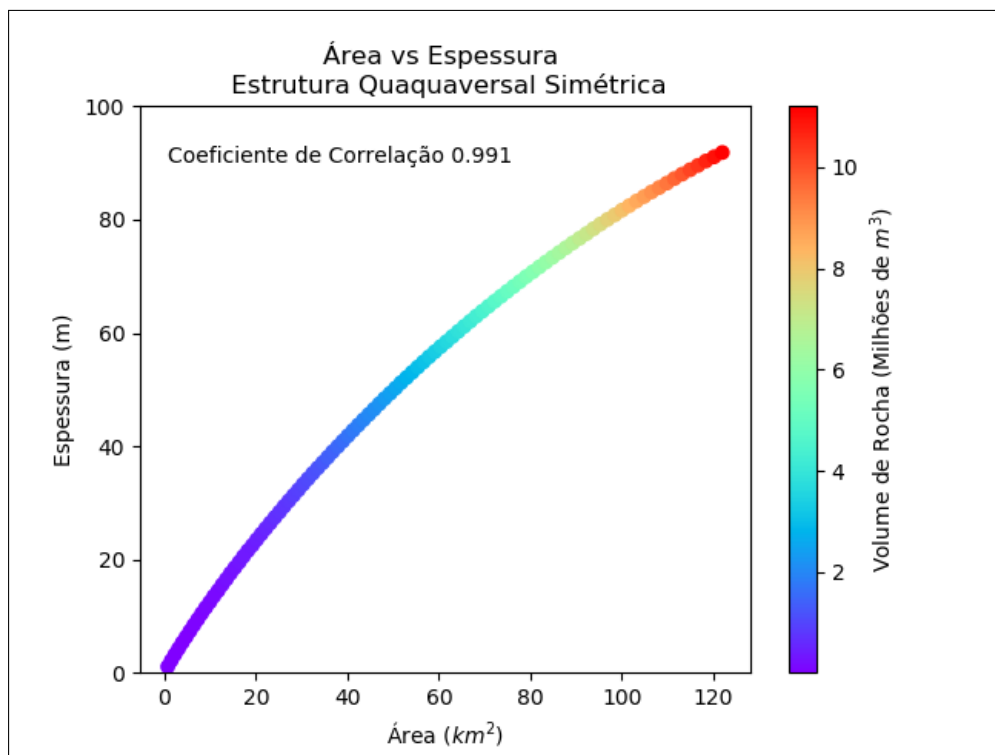


Figura 3.3: *Crossplot* de área vs espessura baseado em cortes do contato da estrutura do Modelo ES1 (representado na Figura 3.2), volume de rochas em cores.

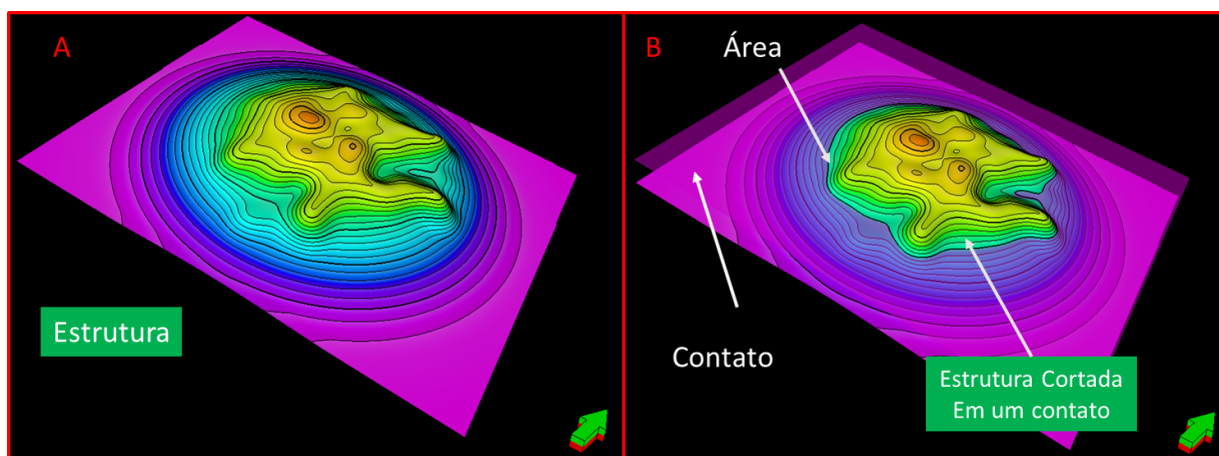


Figura 3.4: Topo do reservatório do Modelo esquemático (ES2) de prospecto quaquaversal **deformado** (A), superfície simulando um contato cortando a estrutura (B).

O Modelo ES2 poderia ser visto também como uma ilustração de variações de topo do reservatório do Modelo ES1, num cenário em que o modelo de velocidade sob a estrutura não é de alta confiança. A partir da metodologia de cortes de contatos podemos avaliar novamente a área vs espessura (e volume de rocha em cores).

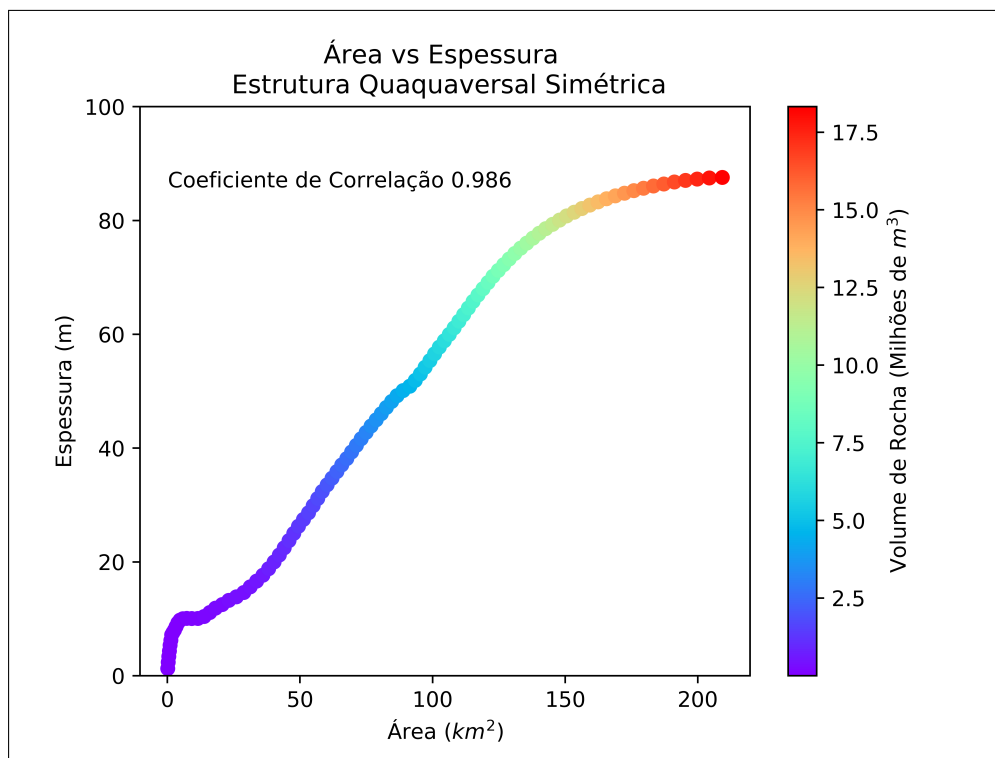


Figura 3.5: *Crossplot* de área vs espessura baseado em cortes de contato do Modelo ES2 conforme representado na Figura 3.4, volume de rochas em cores. Note que a diferença da geometria da estrutura impacta a geometria de correlação mas ela continua com forte comportamento linear e preserva um alto coeficiente de correlação (98.6%).

É importante destacar que, com exceção de condições específicas como na ocorrência de corpos de sal sotopostos ao prospecto, mudanças de uma geometria claramente simétrica e suave normalmente são causadas por variações suaves no modelo de velocidade. Isto mudaria muito mais a curvatura dos flancos da trapa como um todo (Figura 3.6) do que pequenas variações em seu ápice.

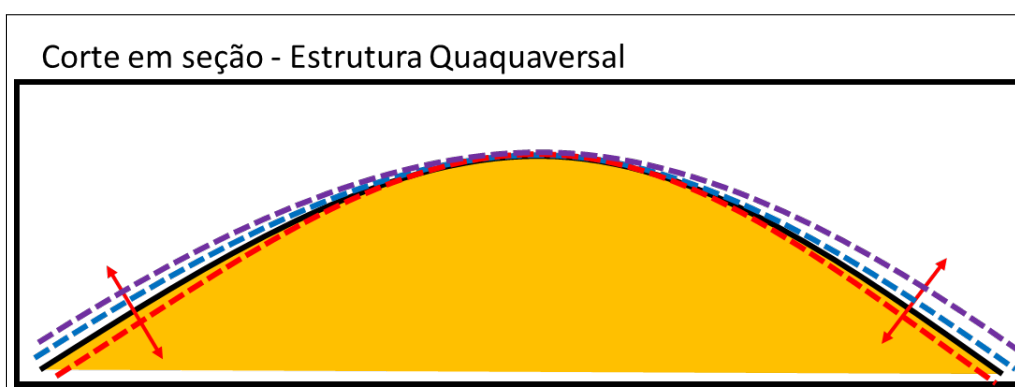


Figura 3.6: Modelo esquemático em seção de prospecto quaquaversal simétrico com variação nos flancos da estrutura causada por uma possível variação de velocidade (20% a -20%).

Para lidar com o cenário de variações nos flancos da estrutura, aplicamos os cortes em contatos num cenário semelhante ao aventado na Figura 3.6. Mudamos tanto os contatos quanto a curvatura e chegamos ao *crossplot* da Figura 3.7.

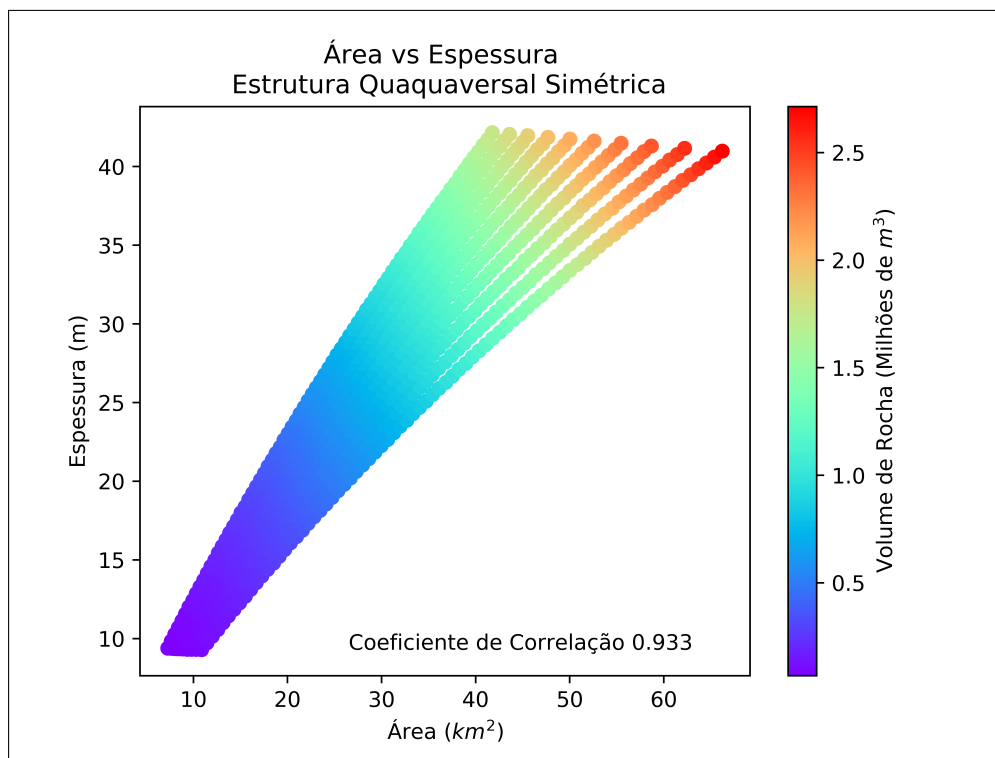


Figura 3.7: *Crossplot* de área vs espessura baseado em cortes de contato da estrutura simétrica esquematizado na Figura 3.8, volume de rochas em cores. Note que há uma replicação do gráfico da Figura 3.3 onde a espessura cresce até atingir assintoticamente a diferença entre topo e base do reservatório (nesse caso definido como 10m).

Verifique na Figura 3.7 que para pequenas áreas temos uma concentração de pequenas espessuras enquanto para espessuras maiores temos uma maior dispersão semelhante ao que obtemos com a estrutura de correlação de Clayton.

Apesar deste modelo ser comum ele não é universal. Um cenário igualmente comum consiste na ocorrência de intervalos reservatórios com espessuras aproximadamente constante posteriormente deformados. Este cenário é ilustrado na Figura 3.8.

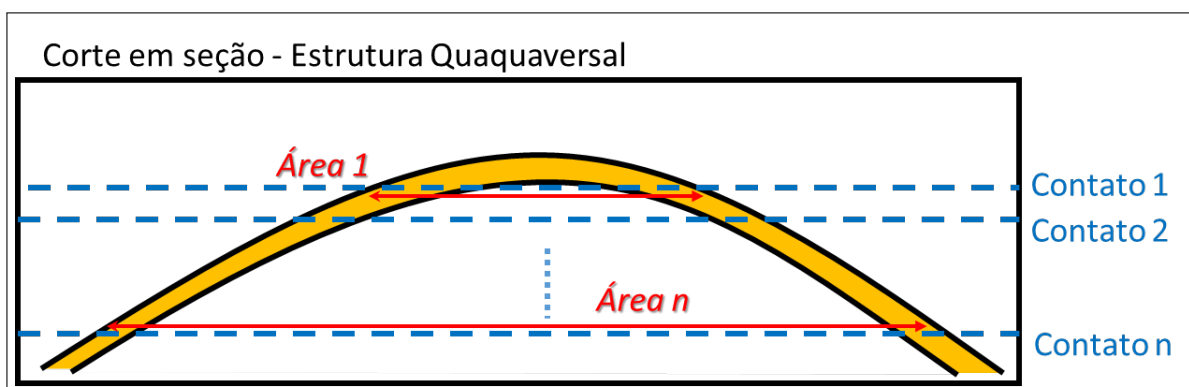


Figura 3.8: Modelo esquemático em seção de prospecto quaquaversal com espessura de reservatório aproximadamente constante.

Foi feito então um modelo de topo semelhante ao do Modelo ES1 e a base através do deslocamento vertical do topo para baixo em 10m, gerando o Modelo ES3. Fizemos então o *crossplot* a partir da metodologia de cortes de contatos e plotamos novamente a área vs espessura (e volume de rocha em cores) ilustrada na Figura 3.9 .

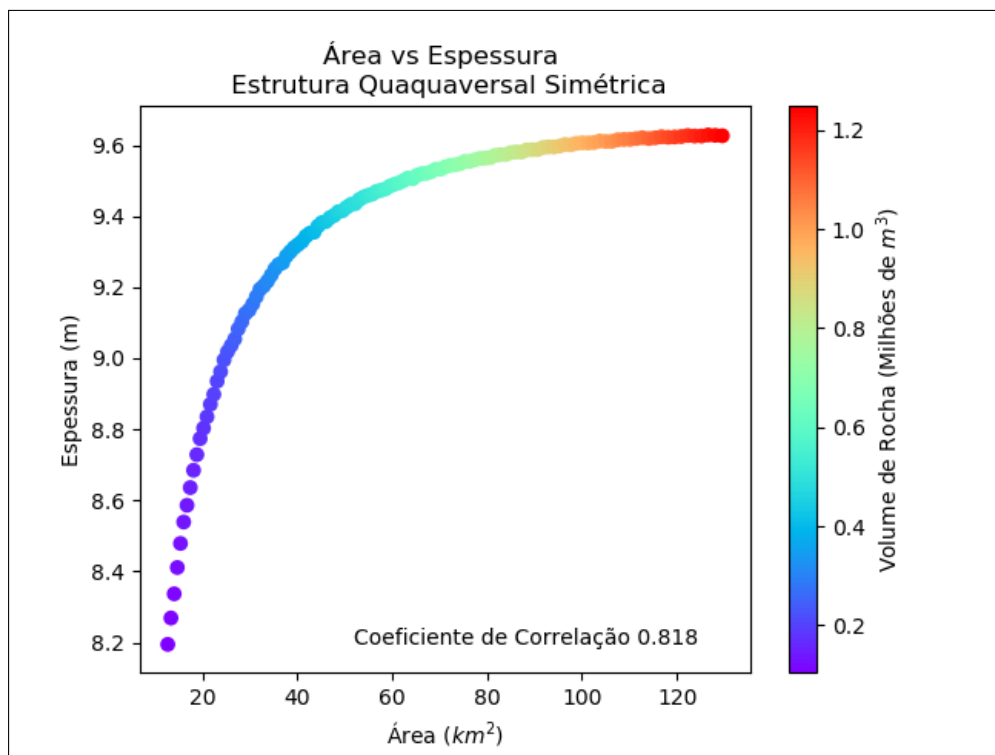


Figura 3.9: *Crossplot* de área versus espessura média baseado em cortes de contato do Modelo ES3, volume de rochas em cores. Note que a espessura cresce até atingir assintoticamente o valor de espessura da camada (nesse caso definido como 10m).

Podemos verificar um progressivo aumento da espessura média para os valores de pequenas áreas até atingir o valor assintótico de 10m (que é a espessura da camada). Este comportamento ocorre devido à geometria de calota semi-esférica. Quando a coluna de óleo é menor do que a espessura da trapa que se torna menos importante na média, à medida que a área aumenta.

De maneira análoga ao modelo de depocentro invertido fizemos agora o experimento numérico de variar a geometria dos flancos da trapa, mas mantendo a espessura aproximadamente constante ainda em 10m (Figura 3.10).

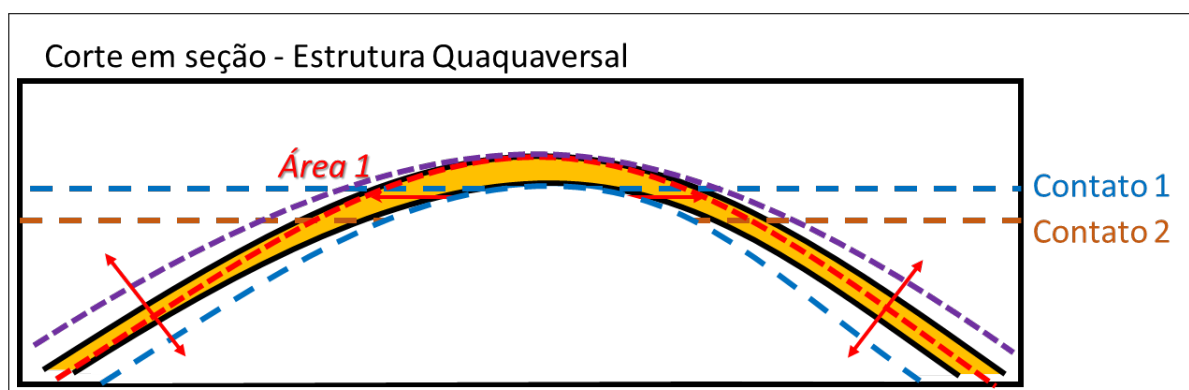


Figura 3.10: Modelo esquemático em seção de prospecto quaquaversal com espessura de reservatório aproximadamente constante definida em 10m (Modelo ES3), com variação nos flancos da estrutura de +20% à -20% .

O *crossplot* para este experimento está representado na Figura 3.11.

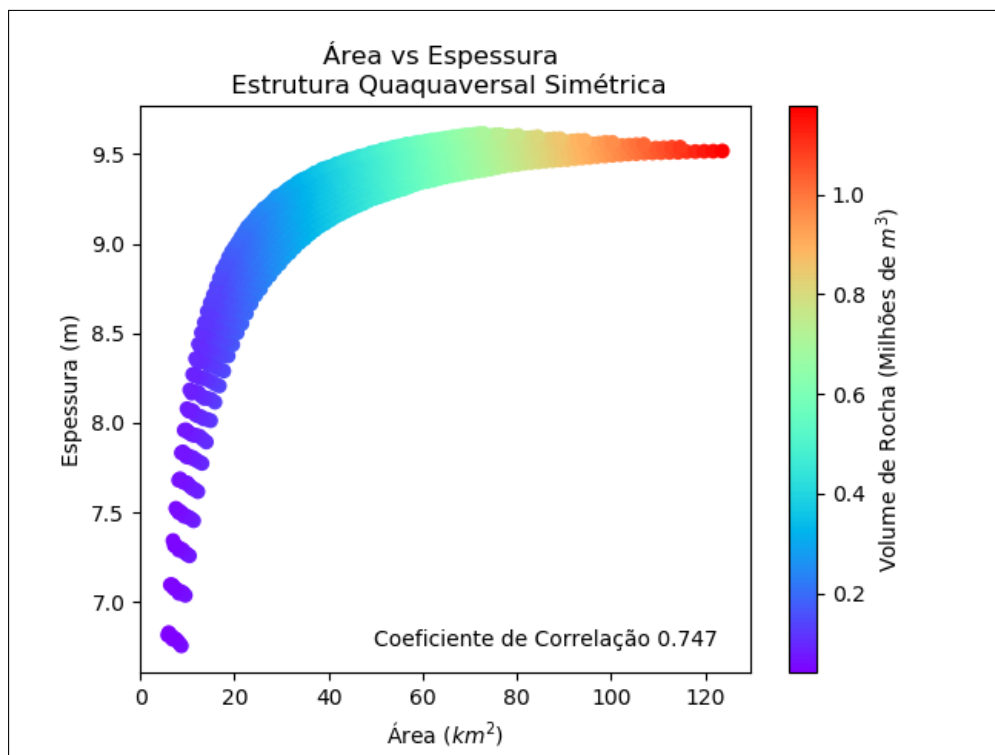


Figura 3.11: *Crossplot* de área vs espessura média baseado em cortes de contato da estrutura simétrica esquematizado na Figura 3.10, volume de rochas em cores. Note que espessura cresce até atingir assintoticamente a entre topo e base do reservatório (nesse caso definido como 10m).

Por fim variamos a espessura (constante ao longo da estrutura em cada iteração) e a curvatura do topo do reservatório conforme o modelo esquemático da Figura 3.12.

Neste tipo de experimento simulamos a variação de velocidade na porção superior à seção reservatório (mudando a curvatura dos flancos da trapa) e as velocidades dentro da própria camada (variando a espessura da própria camada). O *crossplot* para este ultimo teste está representado na Figura 3.13.

Note que há um preenchimento da parte central do gráfico em relação à Figura 3.11, isso ocorre devido ao aumento de ocorrência da calota esférica, que não sofre influência da base (contemplada nos cenários de maior espessura) e desta forma o modelo de espessura fixa começa a se aproximar do modelo de depocentro invertido ganhando maior linearidade.

Neste experimento, variamos a espessura do reservatório de 5 a 55m e a curvatura de +20% a -20%. Destaco um comportamento interessante aqui verificado. O leitor atento perceberá que o coeficiente de correlação das Figuras 3.3 e 3.7 em relação aos respectivos modelos com um único cenário de curvatura tendem a diminuir entre as variáveis uma vez que adicionamos mais cenários e mais dispersão. Entretanto, este comportamento é diferente no caso do cenário representado na Figura 3.11. Isto ocorre porque os cenários adicionados aproximam a geometria a um comportamento linear e não logarítmico verificado na Figura 3.11 aumentando, portanto, o coeficiente de correlação de Pearson.

Para facilitar a visualização dos resultados em planta e em 3D, foi utilizado o próprio *software* de interpretação (Petrel 2016) para simular as variações de velocidade. Estas foram feitas no pacote básico Petrel Workflow. Abaixo exibimos os passos do *workflow* para o ultimo caso (o mais geral) que pode ser adaptado para gerar os resultados anteriores:

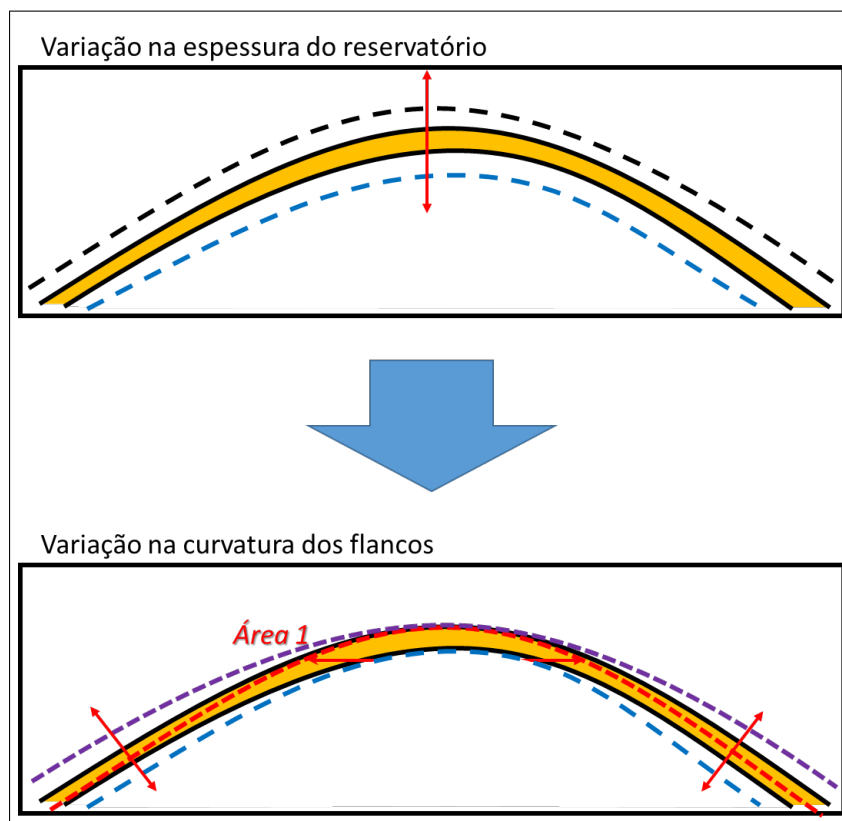


Figura 3.12: Modelo esquemático em seção de prospecto quaquaversal com variação da espessura de reservatório em cada iteração seguida da variação da curvatura.

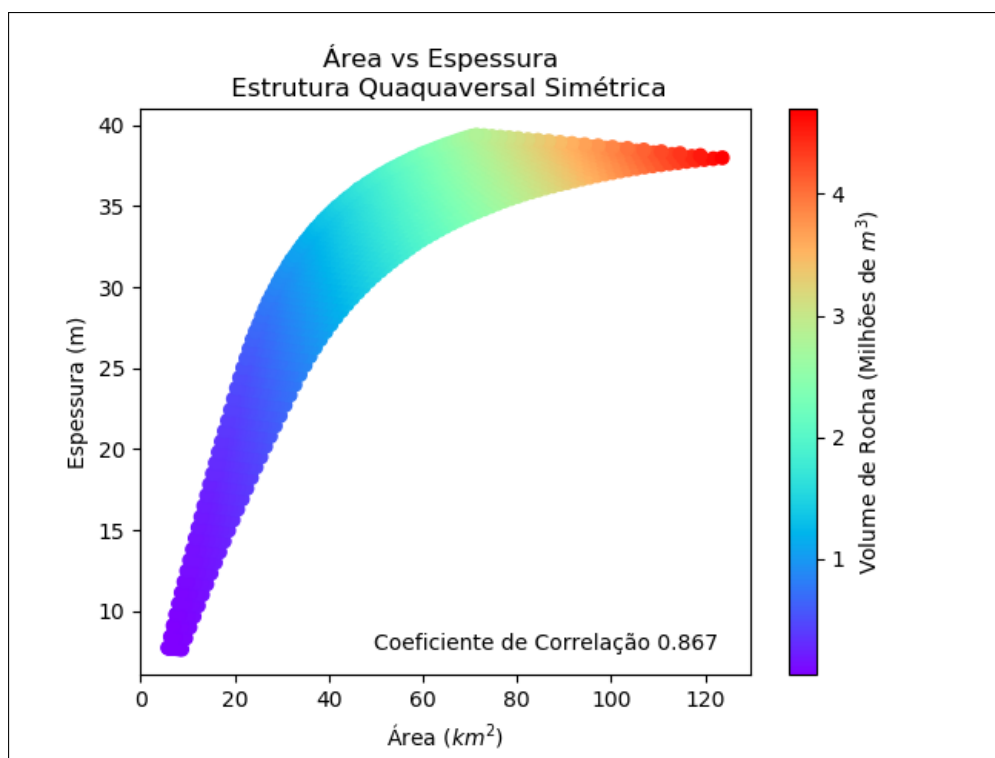


Figura 3.13: *Crossplot* de Área versus espessura média, baseado em cortes de contato da estrutura simétrica esquematizado pelos passos da Figura 3.12, variando espessura e curvatura. Volumes de rochas em cores.

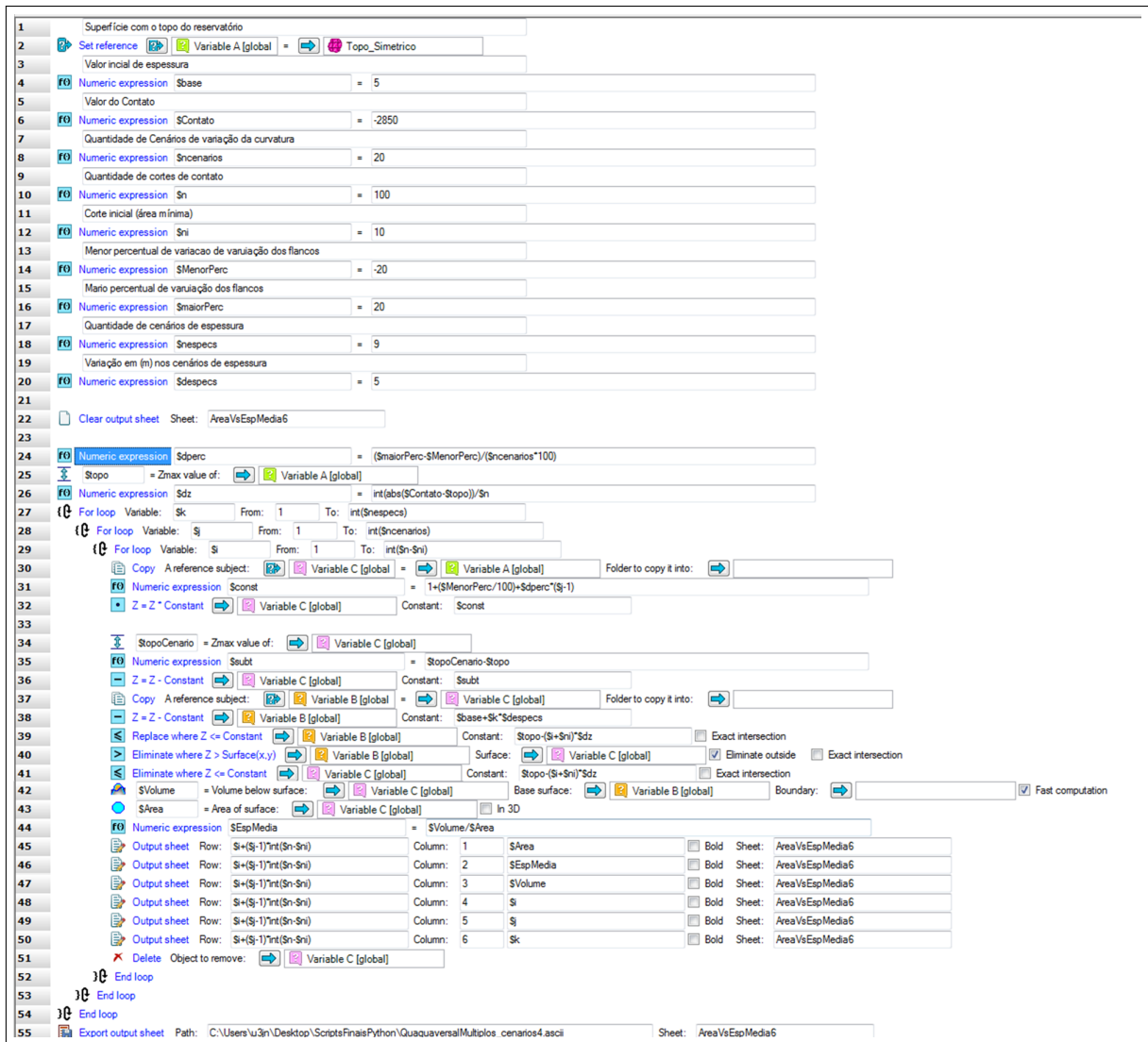


Figura 3.14: Workflow Petrel 2016 utilizado para representar variações no topo do reservatório de um prospecto estrutural com espessuras que variam uniformemente em cada iteração. Juntamente implementamos a variação lateral nos flancos de -20% a +20% . O workflow está comentado de maneira a possibilitar a reprodução.

Área vs Fácies

Até aqui consideramos que a trapa estava completamente preenchida de reservatório de maneira uniforme ao longo da estrutura e em diversos cenários verificamos uma correlação positiva entre área e espessura. Entretanto, muitas vezes (na verdade na maioria delas) os prospectos apresentam variações laterais de fácies inerentes às dos sistemas deposicionais ao longo da trapa (3.15).

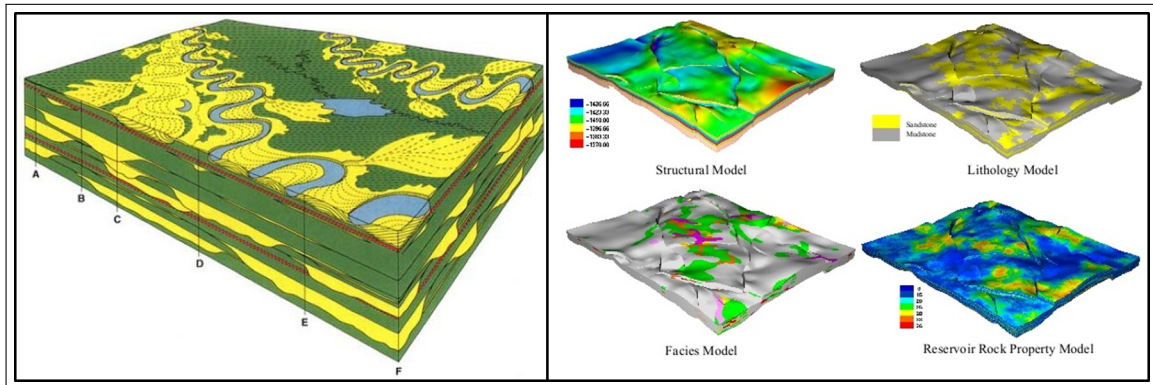


Figura 3.15: Exemplo de modelo de variação lateral de fácies (esquerda) e exemplo de incorporação de variações no modelo de reservatório (direita).

Estas variações podem resultar na ausência de intervalo considerado *net pay* e resultando em variação lateral desta propriedade. Pode-se ainda representar as variações de porosidade que degradam a qualidade do reservatório mas ainda não o desqualificam como *net pay*. Estas variações podem representar melhoras que eventualmente podem ocorrer em algumas propriedades. Abordaremos aqui primeiro a variação de *net pay* e depois de porosidade. Para ilustrar este caso, por simplicidade, consideraremos uma estrutura quaquaversal simétrica igual àquela estudada no exemplo da área vs espessura (Figura 3.16).

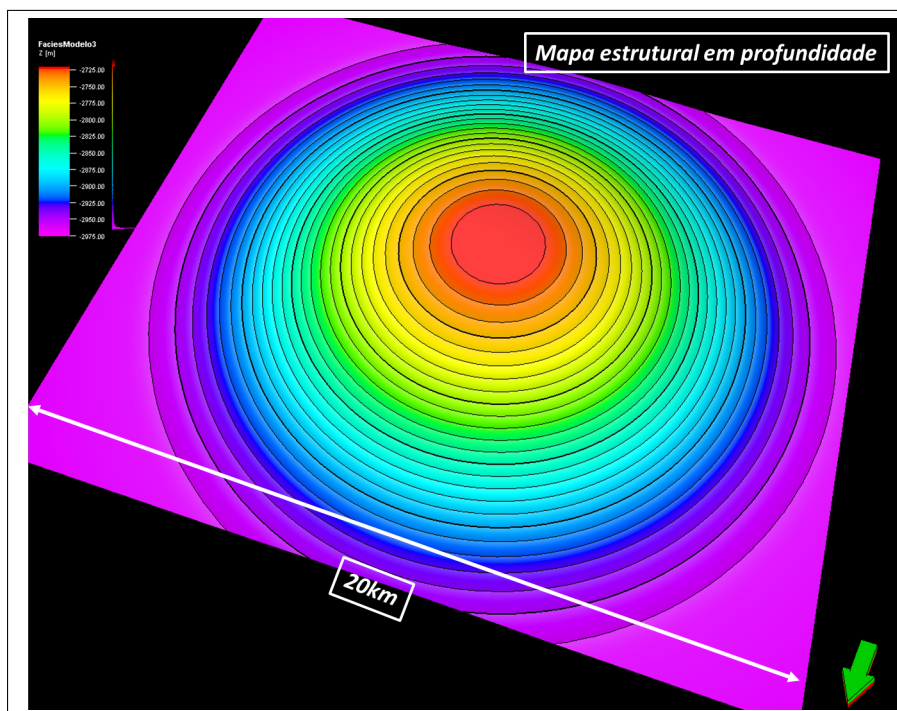


Figura 3.16: Mapa estrutural do topo do reservatório do pseudo-prospecto quaquaversal, o mesmo do modelo ES1 .

Preencheremos a estrutura com 3 feições estratigráficas de um sistema de mesma idade, duas delas compostas por canais nos flancos e por um lobo no ápice da estrutura (Figura 3.17), gerando o modelo FR1.

Em termos de geologia e preenchimento sedimentar, neste exemplo considera-se que a deformação é pós-deposicional e por isso não tem nenhum impacto na distribuição de fácies e/ou espessuras. Consideramos que a qualidade sísmica é boa o suficiente a ponto de conseguirmos detectar estas variações com segurança. Assumimos a premissa de que todas as feições tem espessuras visualizadas pela escala de cores.

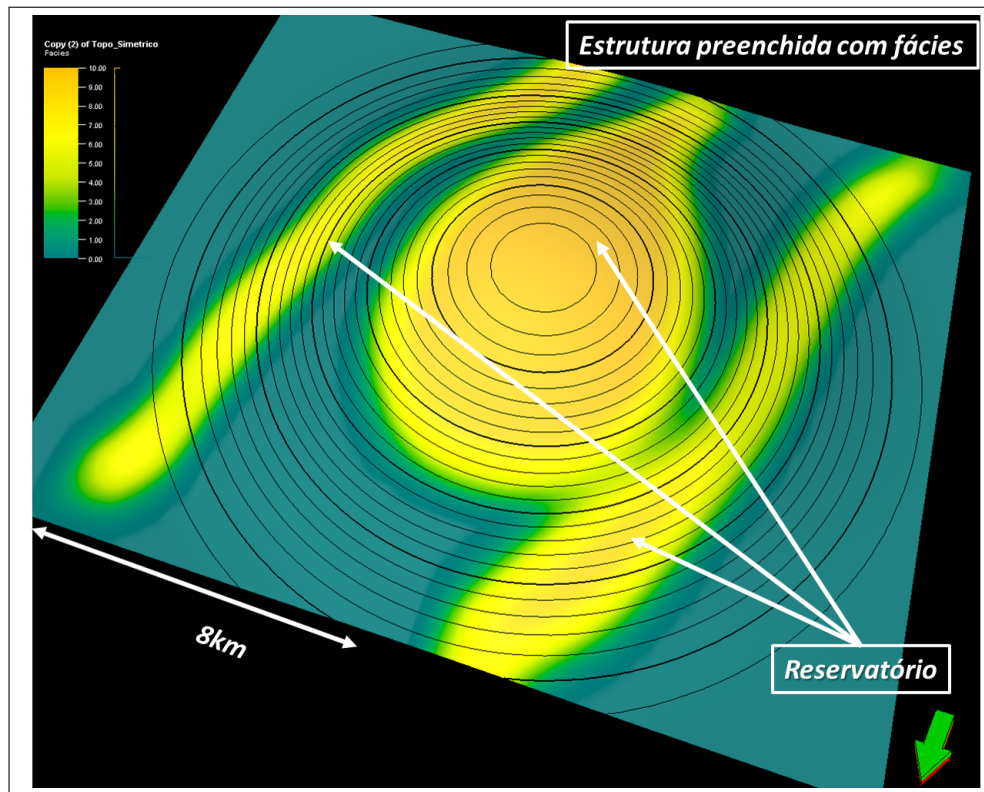


Figura 3.17: Ilustração do Modelo FR1, relevo do mapa estrutural do topo do reservatório do pseudo-prospecto quaquaversal superposto com feições estratigráficas propostas neste trabalho. As cores representam os diferentes valores de *net pay*.

Utilizamos a metodologia de fazer sucessivos cortes em profundidades constantes simulando diferentes contatos (Figura 3.18).

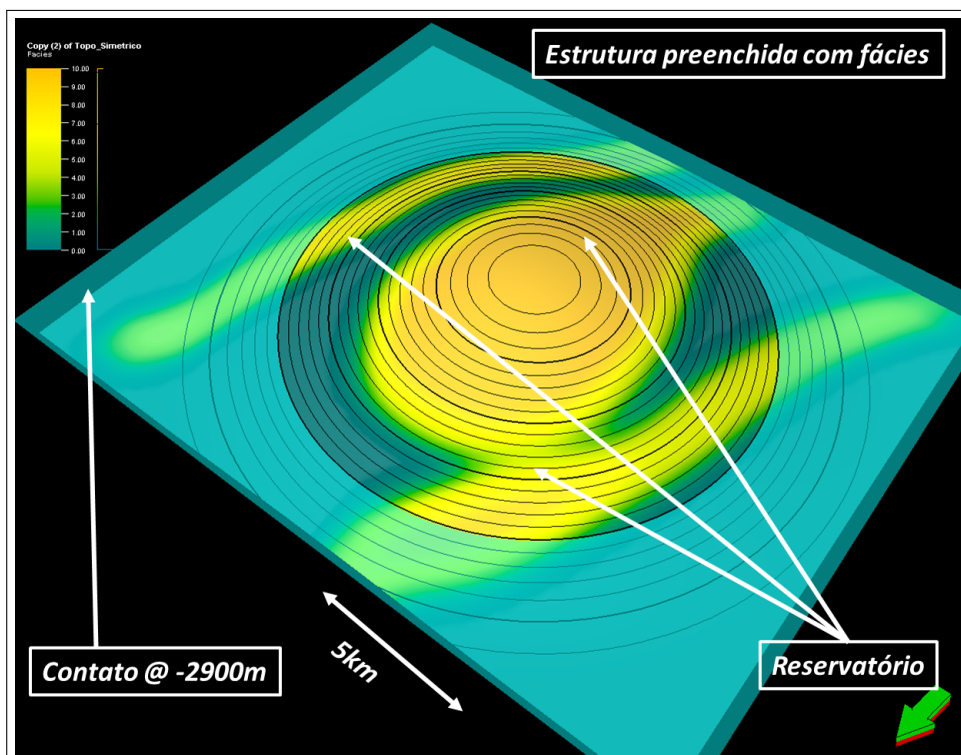


Figura 3.18: Modelo FR1 e ilustração do corte no contato máximo @ 2900m de profundidade. Superposto ao modelo estão as respectivas fácies reservatórios. O contato está representado pelo plano constante.

Com base nos valores obtidos de área e net pay médio obtidos com os cortes em contatos podemos fazer o crossplot da área vs o *net pay* (Figura 3.19). Note que aqui a escala de cores representa a profundidade do contato utilizado.

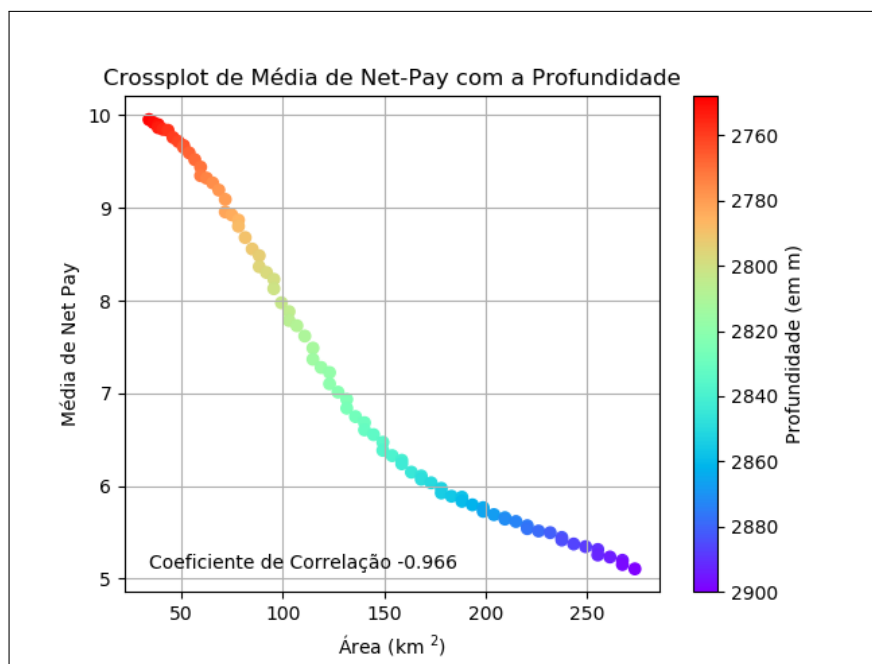


Figura 3.19: *Crossplot* de Área versus *net pay* médio baseado em cortes de contato considerando as fácies representadas no Modelo FR1, Figura 3.18. Profundidade do contato em cores.

Como tínhamos um lobo no ápice da estrutura saímos de um valor máximo de *net pay* médio

para valores menores (que incorporaram fácies não reservatório). Esta variação representa uma correlação negativa desta propriedade com a área, assim como pode ser visto na Figura 3.19.

Vamos agora experimentar o cenário em que o lobo na porção central tem uma espessura bastante reduzida (5m) em relação aos canais nos flancos (30m), gerando o Modelo FR2. O modelo está ilustrado na Figura 3.20.

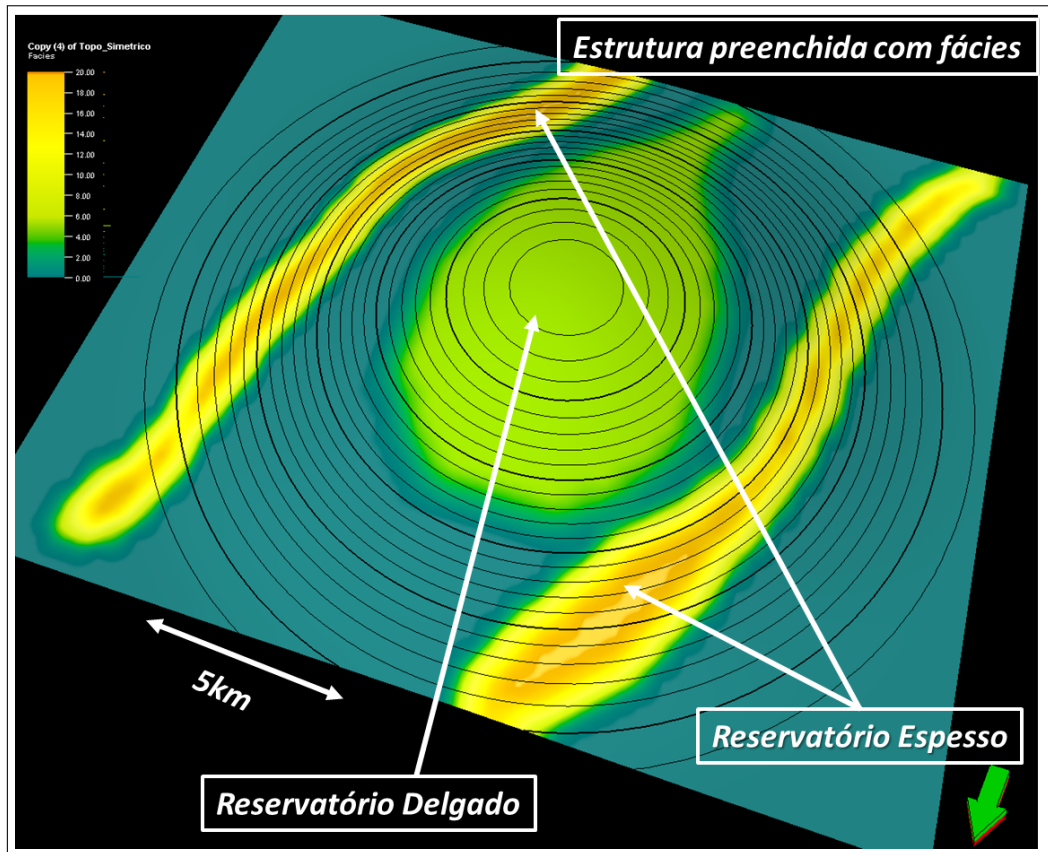


Figura 3.20: Modelo FR2, Cenário em que temos fácies mais empobrecida no ápice da estrutura e canais mais ricos em areias nos flancos.

Utilizamos a metodologia de fazer sucessivos cortes em profundidades constantes simulando diferentes contatos.

Em seguida foi feito o *crossplot* da área vs o *net pay* para o Modelo FR2 (Figura 3.21).

Note que temos uma inversão do comportamento de *net-pay* versus área. Esta inversão em formato de ("U") gera um coeficiente de correlação linear para os dois parâmetros em toda a estrutura próximo a zero.

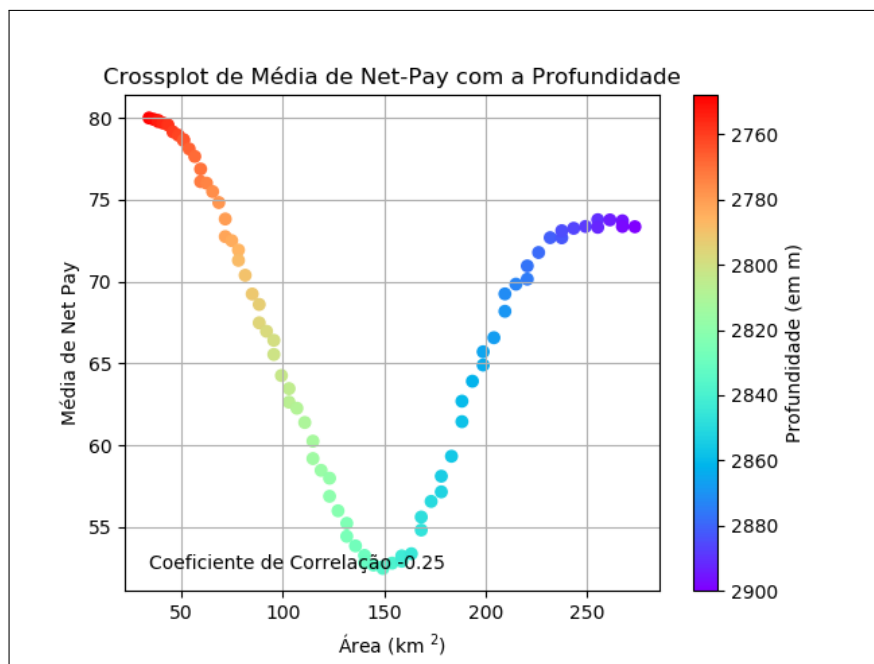


Figura 3.21: *Crossplot* de área versus *netpay* do Modelo FR2.

Considere agora o Modelo FR3 que compreende o cenário em que os reservatórios são apenas os canais nos flancos (30m), ilustrado na Figura 3.22. Este cenário poderia representar uma eventual estrutura sin-deposicional e portanto as fácies reservatórias estariam mais concentradas nos flancos, enquanto o ápice da estrutura teria sido preenchido por fácies mais pobres oriundas de extravasamento do eixo principal de deformação.

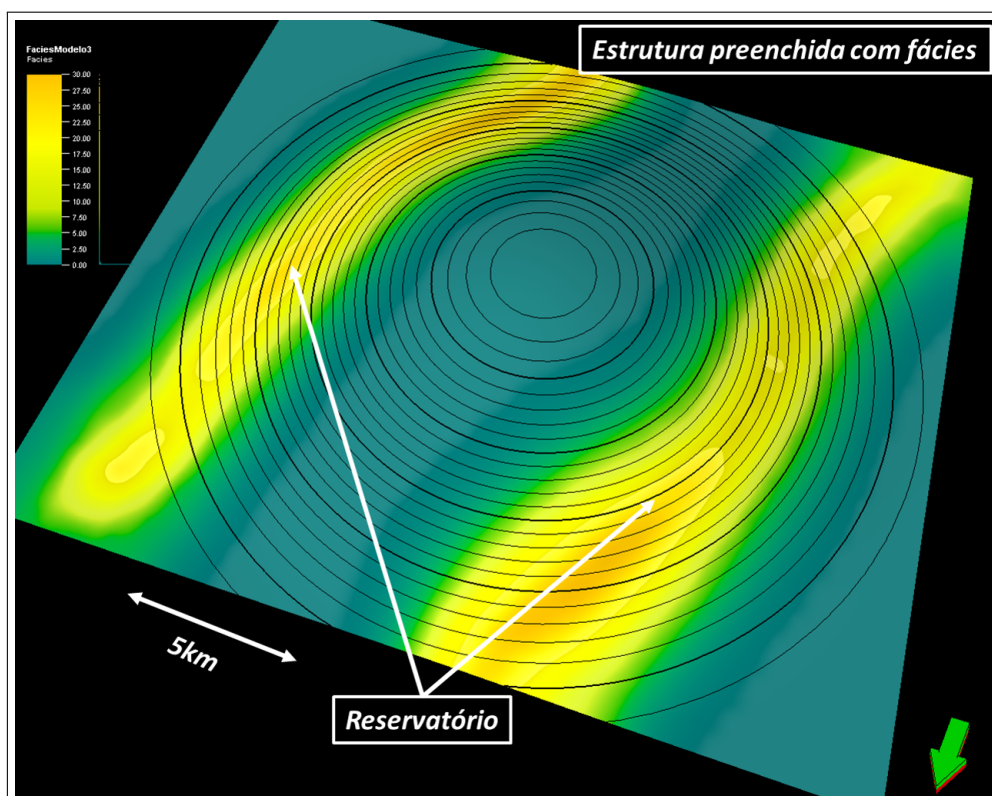


Figura 3.22: Modelo FR3, cenário em que temos uma fácies reservatórios concentradas nos flancos da estrutura.

Novamente aplicamos a metodologia de fazer sucessivos cortes em profundidades constantes simulando diferentes contatos assim como fizemos para os outros dois modelos.

Com base nesta metodologia podemos fazer o crossplot da área vs o *net pay* (Figura 3.23).

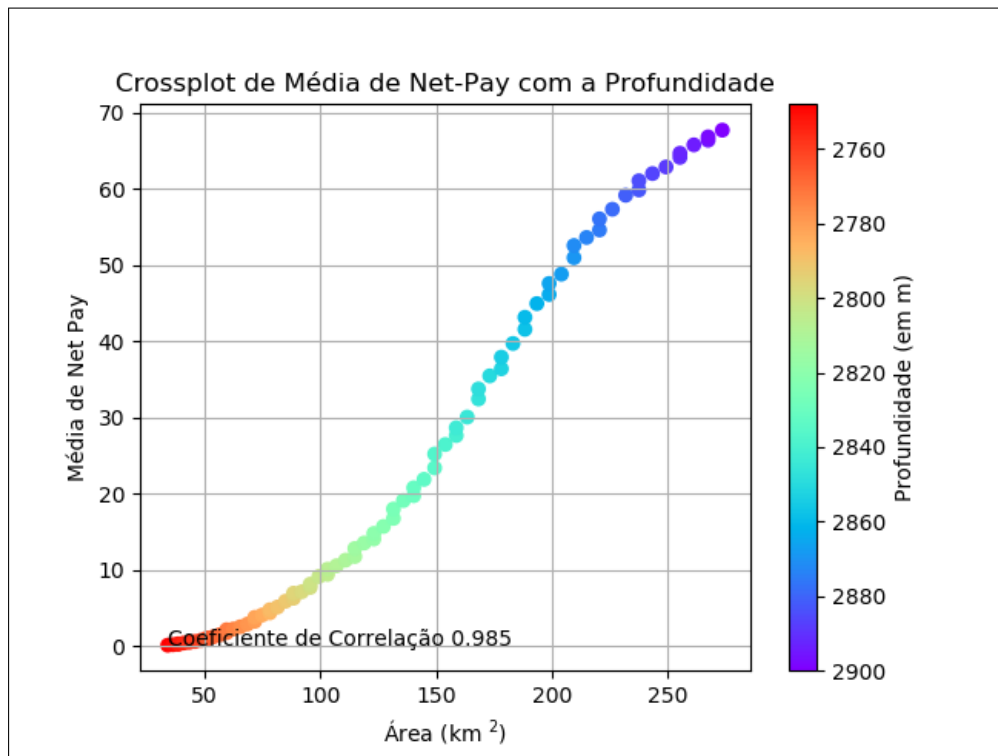


Figura 3.23: *Crossplot* de área versus *net pay*, para os cenários em que os reservatórios estão nos flancos da estrutura (para o modelo FR2). Profundidade do contato em cores.

Note que agora temos um comportamento positivo de correlação onde um aumento de área gera um aumento médio de espessura. Desta forma exemplificamos os casos em que poderíamos ter coeficiente de correlação linear de Pearson altamente positivo, praticamente zero e negativo devido apenas à variação de fácies.

As correlações utilizando as fácies para representar o *net pay* também podem ser aplicadas a parâmetros como porosidade ou saturação (se há um modelo geológico que suporte esta relação lateral). O maior problema aqui é que as estimativas de porosidade/*net pay*/saturação de hidrocarboneto, na fase exploratória, são feitas principalmente com base em dados sísmicos (migrados, em faixas de ângulo etc). Por isso podemos ter múltiplas soluções destes 3 parâmetros com respostas de amplitude semelhantes. Algumas técnicas de inversão podem ser utilizadas no sentido de separar o efeito da porosidade do efeito da saturação entretanto, há uma outra limitação, a sísmica detecta propriedades de interface e não de camada. Isto adiciona mais duas variáveis desconhecidas no sistema (as camadas selante sotoposta e sobreposta ao reservatório).

É importante destacar que ao utilizar parâmetros de correlação para as fácies sísmicas deve-se evitar (a não ser que esteja calibrado por muitos dados) altos valores de correlação para os três parâmetros (porosidade/*net pay*/saturação) simultaneamente. Esta abordagem pode gerar valores máximos, excessivamente otimista de volumes.

3.1.2 Correlação B_o e RGO .

O fator de formação de óleo (B_o) corresponde ao encolhimento do óleo causado pela perda da fração gasosa e pela variação de pressão e temperatura das condições de reservatório para as condições de superfície (Standing, 1977).

Conforme exposto em Ahmed (2016) e Kamari et al. (2016), como definição, o RGO é a quantidade de gás dissolvido em óleo com respeito a uma determinada pressão e temperatura. Deve-se notar que óleos leves têm mais frações de gás dissolvido do que óleos pesados (Standing, 1977). Com um aumento na pressão, a RGO aumenta de forma aproximadamente linear até atingir o ponto de bolha, condição em que a saturação de gás será controlada pela solubilidade do óleo 100% saturado. Por isso, intuitivamente, podemos tomar a fração de gás representada no RGO como um fator direto na expressão do B_o .

Os dados de B_o e RGO normalmente são obtidos com base em dados de PVT⁴ e/ou calibração de laboratório com base em simulações numéricas de fluido. Esta relação é amplamente discutida na literatura. De fato, esta relação é verificada nas diferentes expressões matemáticas empíricas ajustadas por dados observados (Standing (1977), Elsharkawy and Alikhan (1996), Kamari et al. (2016)). Nesta dissertação trataremos o cenário nos casos em que o óleo está 100% saturado e acima do ponto de bolha, portanto o RGO será equivalente à solubilidade do gás no óleo (R_s) dada as composições do óleo e do gás, bem como as condições ambientais.

Com base nas relações de Kamari et al. (2016) construímos os valores de RGO (dado pela solubilidade de gás) e com base nas expressões empíricas de Standing (1977) de Arp e de Hanafy et al. (1997) calculamos o B_o .

Expressão empírica de Kamari et al. (2016) é dada por:

$$R_s = A + B - 15.849 \quad (3.1)$$

Onde:

$$B = 0.0064332 * (p + API) * (API * \gamma_g - 14.811) \quad (3.2)$$

$$A = 0.14624 * (p - API) + 802.44/p + ((2.727 * p - API * T)/(API - 995.53))^2 \quad (3.3)$$

Onde API é o grau api do óleo, p é a pressão (em psi), T a temperatura (em graus célcus) e γ_g amassa específica do gás em relação ao ar. A Expressão empírica de Standing (1977) é dada por:

$$B_o = 0.9759 + 0.00012 * (R_s * \sqrt{(\gamma_g/\gamma_o)} + 1.25 * T)^{1.2} \quad (3.4)$$

Onde γ_o é a massa específica do óleo em relação à água. A expressão empírica de Arp é dada por:

$$B_o = 1.05 + 0.0005 * R_s \quad (3.5)$$

Expressão empírica de Hanafy et al. (1997):

$$B_o = 1.079 + 0.0006 * R_s \quad (3.6)$$

Dando origem à Figura 3.24:

Nas equações de Kamari et al. (2016) e Standing (1977) precisamos de outras variáveis que não apenas o B_o e o RGO. Foram utilizadas distribuições normais com 10.000 amostras de valores tipicamente obtidas na janela de óleo traduzidas na tabela 3.1. Deve-se ressaltar que os parâmetros ambientais (temperatura e pressão) devem ser coerentes com os *trends*⁵ regionais e os parâmetros relacionados ao óleo devem ser condizentes com estudos de modelagem de sistema petrolífero e amostras de campos próximos ao prospecto (conforme comentado na seção sistema petrolífero do Capítulo 2).

⁴Ensaio em amostras de fluido, onde são controlados Pressão, Volume e Temperatura

⁵Direção de isocomportamento de uma propriedade ao longo de um eixo, que pode ser profundidade ou direção de deposição, no caso relacionado ao soterramento.

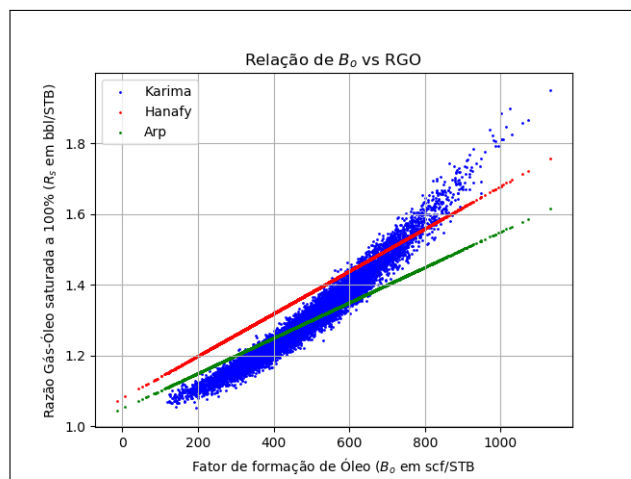


Figura 3.24: Relação de RGO com B_o dependente baseado nas leis de Standing e equação empírica de Kamari et al. (2016).

Propriedade	unidade	Média	Desvio padrão
API	-	25	2
Temperatura	$^{\circ}F$	200	20
Pressão	psi	2800	200
Gravidade específica do gás (γ_g)	-	0.8	0.3

Tabela 3.1: Parâmetros de distribuições normais utilizadas na construção da Figura 3.24. Os valores de γ_o foram obtidos com base nos dados de API.

Diferentemente da relação de fácies e área, a correlação de RGO e B_o dado por $c(RGO, B_o)$ será estritamente maior ou igual a zero [$c(RGO, B_o) \geq 0$]. Isso ocorre porque as variáveis são intrinsecamente dependentes. É impossível aumentarmos o conteúdo de gás na solução de óleo e diminuirmos o fator de formação uma vez que há uma relação de causalidade mútua entre elas.

A variação dos valores de correlação dependerá principalmente da escala de valores de p_{10}/p_{90} e da incerteza relacionada aos parâmetros das equações utilizadas para modelar. Dessa forma propomos que a metodologia utilizada para avaliação deste parâmetro seja feita ou através da correlação de dados regionais ou pela modelagem com simulação de Monte Carlo dos parâmetros assim como apresentados neste trabalho (e.g. Tabela 3.1).

3.1.3 Correlação Porosidade (ϕ) vs Saturação de hidrocarbonetos (S_{hc}).

O preenchimento de uma trapa por hidrocarbonetos (o processo de imbibição) se dá progressivamente pela remobilização da água pelo óleo no reservatório até o valor de saturação de água irreduzível (S_{wir}), por vezes também chamada de água conata, ou seja $S_{hc} = 1 - S_{wir}$. Dentre os fatores que determinam o valor de saturação de óleo/água irreduzível. Dois fatores são preponderantes: (1) a porosidade e (2) a molhabilidade do reservatório.

As saturações de água irreduzíveis podem variar de aproximadamente de 5% à quase 100%. Se a molhabilidade da rocha é conhecida (obtida a partir do conjunto de dados de poços e campos mais próximos do prospecto), alguns limites podem ser definidos na saturação. Conforme descrito em Baker et al. (2015), a saturação de água irreduzível é normalmente menor que 15% em rochas molháveis a óleo e maior que 20% em rocha molhável a água. Os arenitos produtivos molháveis a água geralmente têm saturações de água irreduzíveis entre 20% e 40%. Em geral, a saturação de água irreduzível aumenta à medida que a permeabilidade diminui.

Em alguns campos, há uma relação hiperbólica entre a saturação de água irreduzível e a

porosidade seguindo a expressão empírica 3.7 proposta por Buckles (1965) ⁶.

$$S_{wir} = \kappa/\phi \quad (3.7)$$

A Figura 3.25 representa alguns valores típicos da constante κ utilizada para diferentes reservatórios.

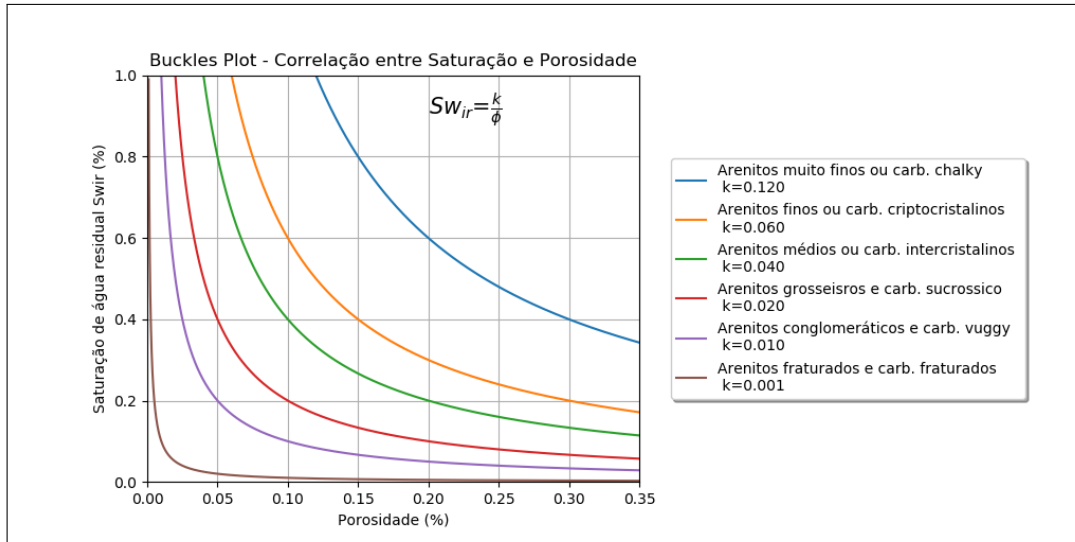


Figura 3.25: Buckles plot com a relação de S_{wir} e ϕ com decrescimento hiperbólico definido pelo método de Buckles para diferentes constantes k.

É importante destacar que κ está relacionado com a permeabilidade mas não é ela propriamente dita. Conforme Baker et al. (2015), na ausência de dados básicos, podemos estimar o coeficiente κ , a partir de uma regressão baseada em dados de campos análogos ao prospecto de interesse, um exemplo pode ser visto na Figura 3.26.

Note que os valores de S_{wir} e ϕ foram dados na escala de 0 a 100, por isso $\kappa=641$. Esta relação evidencia que quanto maior a porosidade menor a saturação irreduzível de água e, portanto, maior a saturação de hidrocarbonetos que a rocha pode armazenar. Isto nos leva a uma correlação positiva entre saturação de hidrocarbonetos e porosidade Figura 3.27.

⁶em seu artigo original ele apresentou uma relação mais genérica do tipo $S_{wir} = \kappa/(\phi + b)$.

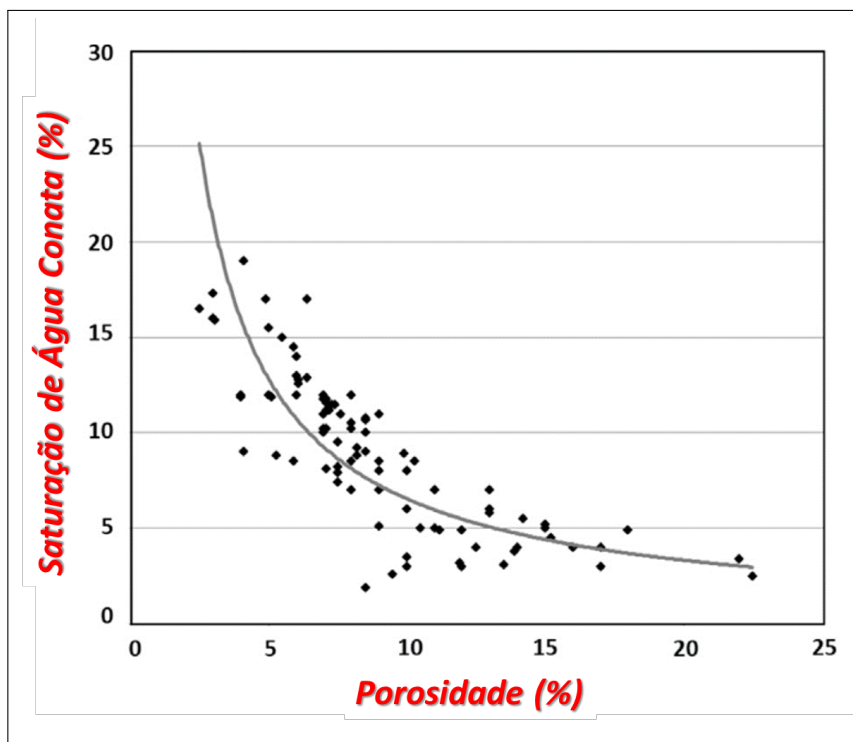


Figura 3.26: Buckles plot com a relação de S_{wir} e ϕ com decrescimento hiperbólico conforme extraído de Baker et al. (2015) p 272. Os dados foram extraídos de testemunhos de recifes carbonáticos com óleo obtidos de da região de Bonnie Glen D-3A Pool, Alberta, Canada.

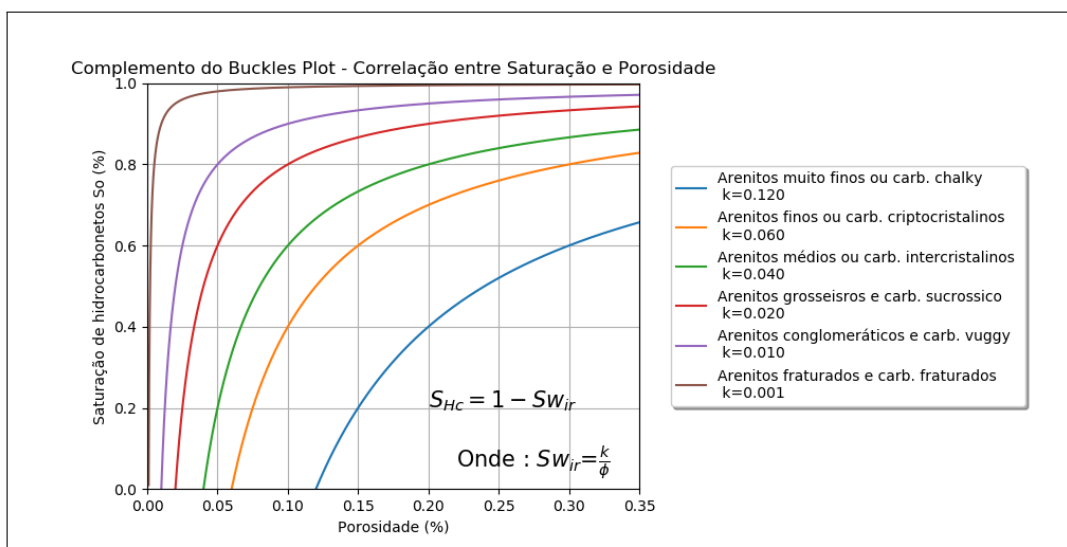


Figura 3.27: Complemento do Buckles plot com a relação de S_{Hc} e ϕ com crescimento hiperbólico definido pelo método de Buckles para diferentes constantes κ .

3.1.4 Correlações envolvendo o *Fator de Recuperação*

Conforme comentado no Capítulo 2, o fator de recuperação (F_r) é determinado por fatores geológicos, fatores relacionados ao conceito de desenvolvimento do campo e fatores relacionados à operação. Os fatores relacionados à engenharia são guiados pelo problema geológico, pelas tecnologias disponíveis e as premissas econômicas que, em última análise, definirão a viabilidade do projeto. Alguns dos pontos chaves dos fatores supracitados são ilustrados na Tabela 3.2:

Verificamos no item anterior que quanto maior a porosidade maior a saturação de hidrocar-

Fatores Geológicos	Desenvolvimento	Operação
Porosidade	N de Poços	Eficiência de Injeção
Permeabilidade	Conceito de desenvolvimento	Otimização da planta
Conectividade	Razão de Injetores/Produtores	Intervenção nos poços
Propriedade dos fluidos	Tamanho da facilidade de produção	Paradas de Manutenção
Saturação de fluidos	Tempo de exploração	Gerenciamento do campo

Tabela 3.2: Alguns fatores que impactam no fator de recuperação de uma acumulação.

bonetos que o reservatório pode armazenar. Verificamos também nos fundamentos que quanto maior a porosidade mais provável que os poros estejam interconectados e portanto maior a permeabilidade. Com base nas premissas que acabamos de citar podemos inferir que quanto maior a porosidade haverá uma tendência a termos um maior fator de recuperação. Contudo, existem tantos elementos que influenciam o fator de recuperação que eles acabam reduzindo a correlação de cada elemento independente. Compilei alguns dados de campos de águas profundas em que o principal tipo de porosidade é a intergranular resultando na Figura 3.28.

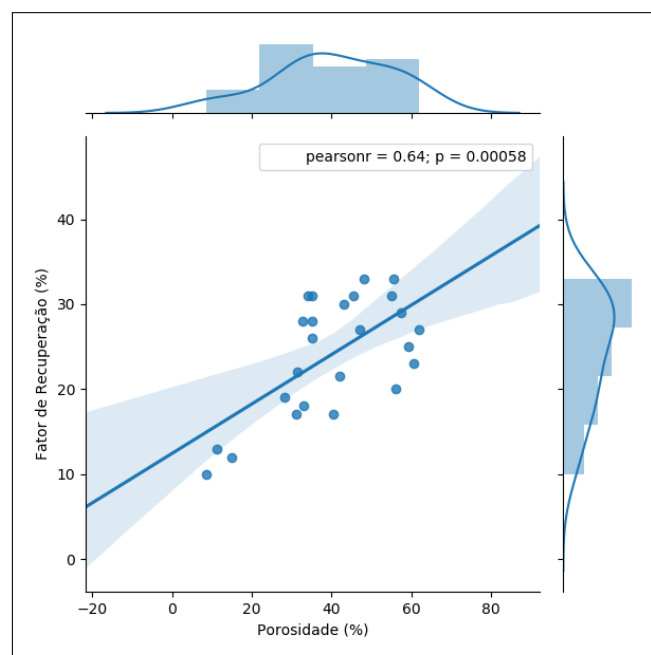


Figura 3.28: Gráfico de Fator de Recuperação vs Porosidade para reservatórios de águas profundas portadores de óleo. As acumulações estão localizadas no Golfo do México, compostos de canais e lobos no mundo Amado (2013).

É importante destacar que esta relação não é biunívoca em todas as litologias. Folhelhos podem ter porosidades excessivamente grandes ($> 25\%$) e permeabilidades reduzidas ($< 1\text{md}$). Carbonatos fechados fraturados podem ter porosidade extremamente baixas ($< 6\%$) e altas permeabilidades ($> 1000\text{md}$).

Ao adicionar mais dados a este *crossplot* referentes a outros ambientes deposicionais e locais (e.g. *onshore*) em que a estratégia de produção admite mais poços, a correlação persiste mas fica bastante piorada (Figura 3.29).

Dessa forma, caso seja utilizada a correlação entre porosidade e fator de recuperação recomendaríamos o uso de valores de correlação na faixa de 0.3-0.75, a não ser que existam evidências que suportem valores maiores⁷.

⁷Valores inferiores a 0.3 podem indicar uma falta de correlação.

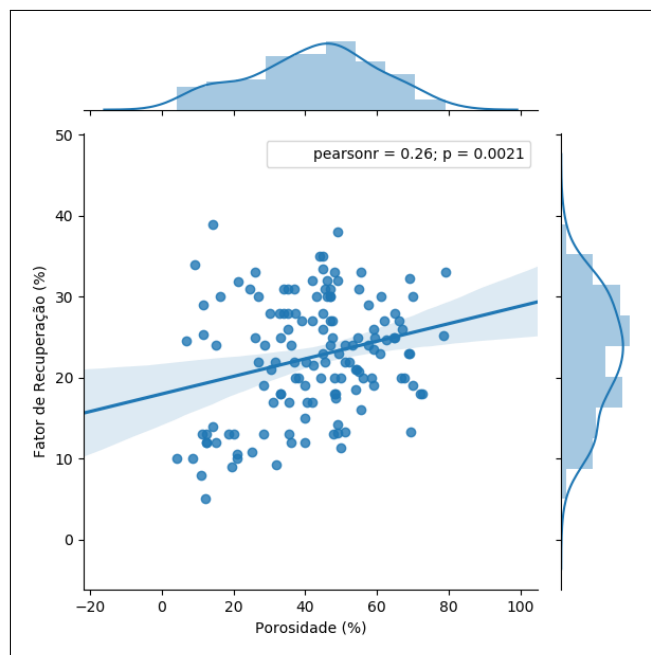


Figura 3.29: Gráfico de Fator de Recuperação vs Porosidade para diversos tipos de reservatório portadores de óleo (incluindo os de águas profundas da Figura 3.28) no mundo.

Vamos agora adicionar os dados de alguns campos portadores de gás seco ao gráfico da Figura 3.29.

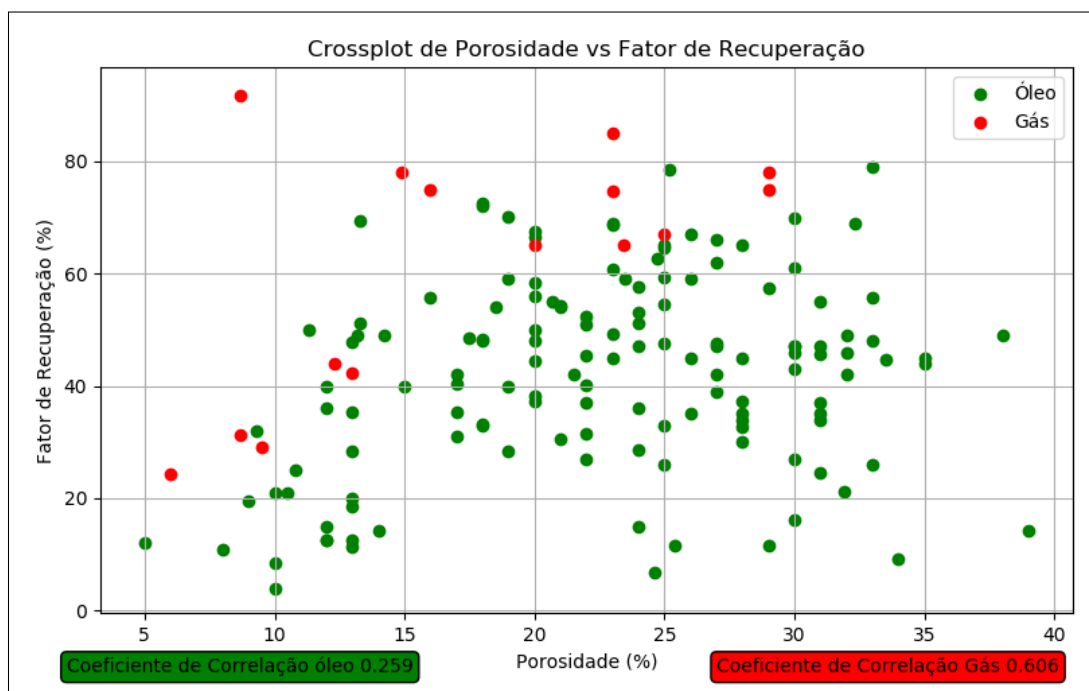


Figura 3.30: Gráfico de Fator de Recuperação vs Porosidade para diversos campos portadores de óleo e gás abrangendo diversos tipos de reservatório incluindo os de águas profundas da Figura.

Note na Figura 3.30 que para uma mesma porosidade temos fatores de recuperação maiores para gás (o que já era esperado). O que nos leva a crer que em prospectos onde temos a dúvida sobre o preenchimento de gás ou óleo. Devemos considerar Fatores de Recuperação maiores quando na ocorrência de gás seco.

A metodologia proposta aqui é que além de aplicar valores de correlação devamos utilizar um valor limítrofe de que ao simular acumulações de gás os valores de correlação assumam um fator de recuperação superior segundo o seguinte fluxograma:

1. Simular os cenários;
2. Se for sorteado um cenário com óleo - aplicar o fator de recuperação do óleo;
3. Se for sorteado um cenário com gás - aplicar o fator de recuperação do gás;

A chance de ocorrência de gás é um fator importante na percepção de valor do prospecto, principalmente porque os dois produtos tem valores de venda e necessidades logísticas de escoamento bastante diferentes. Faremos agora uma discussão sobre um possível impacto na percepção da chance de ocorrência de gás e coluna relacionada ao risco de retenção.

3.1.5 Correlações Verificadas

Nesta seção apresentamos metodologias para estimar os valores de correlação entre os parâmetros volumétricos. Com base no já exposto podemos chegar a uma tabela de correlações que pode ser utilizada como guia para a análise de prospectos. O objetivo aqui não é esgotar o assunto mas consolidar os parâmetros que foram expostos nesta seção.

	Area	Net Pay	So	TF	Φ	Bo	Bg	RGO	Chance Gás	FR
Area	1									
Net Pay		1								
So			1							
TF				1						
Φ					1					
Bo						1				
Bg							1			
RGO								1		
Chance Gás									1	
FR										1

	Pode ter correlação que varia de positiva a negativa.
	Correlação fracamente positiva.
	Correlação positiva forte

Tabela 3.3: Tabela com correlações entre elementos volumétricos.

Apresentamos também a relação entre risco de retenção, ocorrência de gás, e a altura da coluna de hidrocarbonetos modelada em um caso bastante particular. Na hipótese que estes parâmetros sejam relevantes é necessário avaliar a estratégia de avaliação do prospecto (se serão utilizados riscos ponderados para cada cenário ou se serão utilizados os mesmos riscos e parâmetros volumétricos diferentes).

3.2 Metodologia de avaliação do impacto da correlação.

Discutimos até agora metodologias para avaliar a existência de correlação entre duas variáveis com distribuições normais (apresentadas por uma questão de simplicidade). Entretanto o problema que abordamos nesta dissertação consiste num produto de distribuições lognormais (cenários) que originará volumes calculados. Com o objetivo de simplificar o entendimento apresentaremos a metodologia de avaliação em apenas duas distribuições onde naturalmente será aplicada à multiplicação destes no restante dos parâmetros.

Em seguida apresentaremos os gráficos que subsidiam a metodologia de análise de construção dos gráficos de controle de qualidade apresentados neste trabalho.

3.3 Análise de Sensibilidade

Para realizar a análise de sensibilidade, além da avaliação matemática, podemos nos apoiar em diversos gráficos como por exemplo *crossplots* e gráficos de barra.

Um dos gráficos de barra é o Tornado, que é uma ferramenta comum usada para representar a sensibilidade de um resultado a alterações nas variáveis selecionadas sendo, portanto, uma ferramenta analítica de análise de incerteza (Figura 3.31). Ele mostra o efeito simulado de variar uma variável de entrada de cada vez, mantendo todas as outras variáveis de entrada em valores considerados pelo caso base ou mais provável (consideramos aqui a média de cada variável).

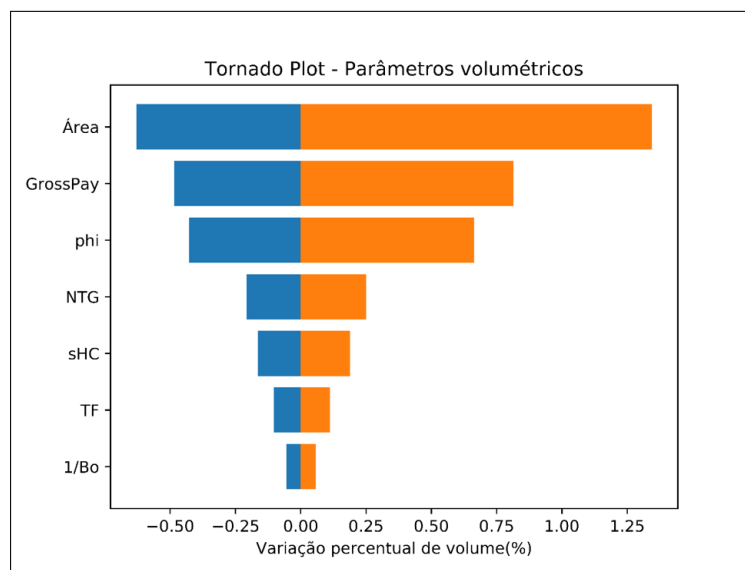


Figura 3.31: Exemplo de gráfico tornado aplicado na análise de sensibilidade de um prospecto.

Normalmente, escolhemos um valor otimista e um conservador para cada entrada. O resultado é exibido como um tipo especial de gráfico de barras horizontais centradas em zero (caso base). As barras para cada variável de entrada exibem a variação do caso base em função dos diferentes cenários calculados. É comum ver esta análise de sensibilidade em função dos valores máximos e mínimos de cada variável. Entretanto, como utilizamos a distribuição lognormal, os casos otimistas podem assumir valores excessivamente altos. Por isso nesta dissertação modelamos os casos otimista e pessimista em função do P90 da distribuição modelada⁸.

O gráfico Tornado pode ter sua variação em função dos valores nominais (barris, R\$ etc) ou variações percentuais em relação ao caso base. Matematicamente podemos estimar os valores de cada variável nominal:

⁸Devido aos truncamentos utilizados, o P90 da distribuição pode ser ligeiramente diferente daquele utilizado como dado de entrada.

$$T_{min}[x] = f(x_{P90}, y) - f(x_{media}, y) \quad (3.8)$$

$$T_{max}[x] = f(x_{P10}, y) - f(x_{media}, y) \quad (3.9)$$

Onde y constitui o caso base de todas as outras variáveis do problema, x é a variável de sensibilidade, x_{P90} é o caso conservador da variável, x_{P10} é o caso otimista da variável de análise. f é a função de análise, T_{min} e T_{max} é a variação nominal da subtração da função utilizando a variável de sensibilidade em cenário pessimista e otimista em relação à média.

Podemos também abordar o problema de forma normalizadas em função da média da função considerando o caso base ($f(x_{media}, y)$) como elemento normalizador.

$$Tn_{min}[x] = \frac{f(x_{P90}, y) - f(x_{media}, y)}{f(x_{media}, y)} \quad (3.10)$$

$$Tn_{max}[x] = \frac{f(x_{P10}, y) - f(x_{media}, y)}{f(x_{media}, y)} \quad (3.11)$$

É prática padrão plotar as barras horizontalmente, classificadas de forma que a barra mais larga seja colocada na parte superior. Quando desenhado dessa maneira, o diagrama assume a aparência de um tornado, daí seu nome.

Note que o gráfico Tornado tende a refletir a razão p10/p90 das variáveis. Entretanto algumas variações podem ocorrer no problema proposto nesta dissertação. O primeiro deles é que as variáveis de análise podem afetar apenas frações dos hidrocarbonetos calculados como é o caso da razão gás-óleo (RGO), razão líquido-gás (RLG), Fator de Recuperação do Gás Associado (FRGA) Fator de Recuperação do óleo (FrO) e Fator de recuperação de condensado (Figura 3.32).

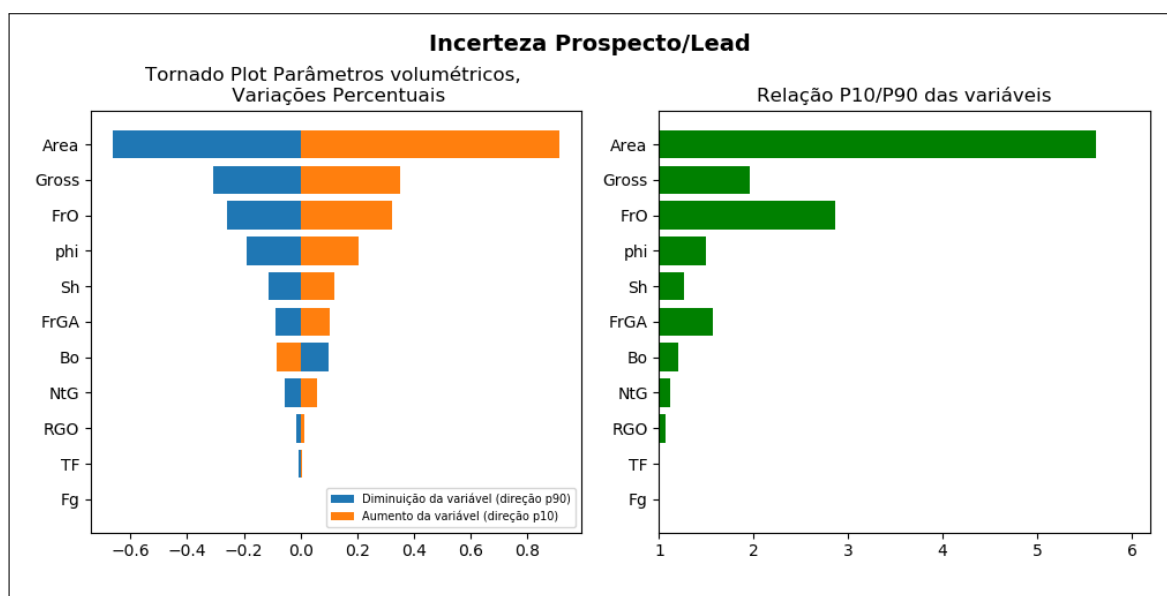


Figura 3.32: Exemplo de gráfico tornado aplicado na análise de sensibilidade de um prospecto.

Nestes casos devemos calcular a sensibilidade apenas para o fluido em análise e considerar os outros volumes de hidrocarbonetos constantes. Outra forma de visualizar a importância de uma variável na avaliação de incerteza de volumes calculados é através de *crossplots*. Em variáveis cuja a importância é significativa temos uma estratificação significativa de volumes ao longo do eixo referente à variável (Figura 3.33).

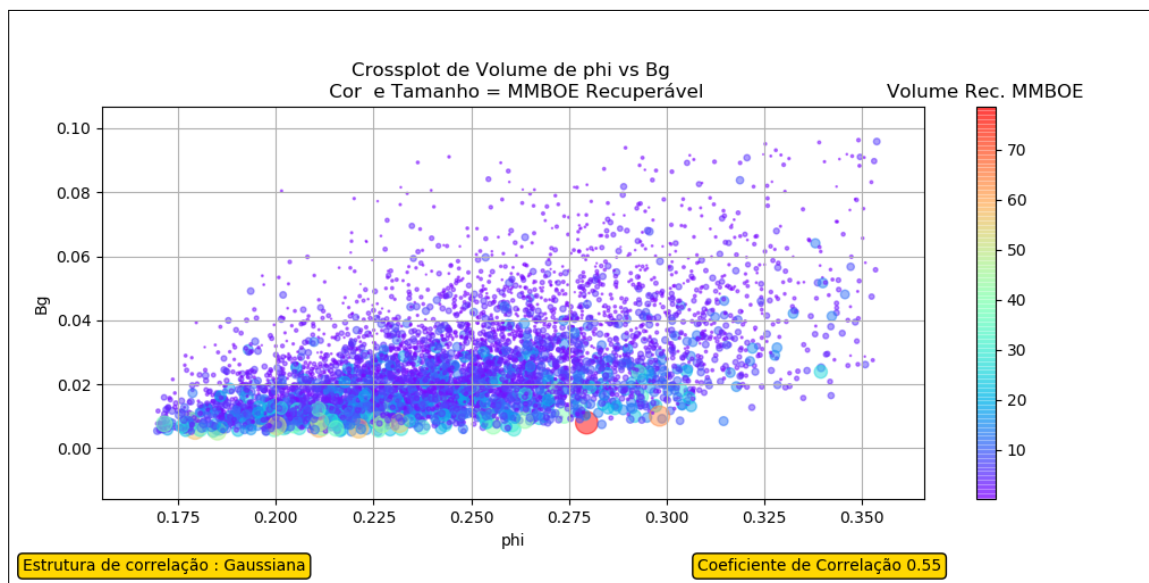


Figura 3.33: Exemplo de crossplot de Bg vs ϕ . Note a estratificação vertical mostrando que a incerteza de Bg tem grande impacto no cálculo de volumes.

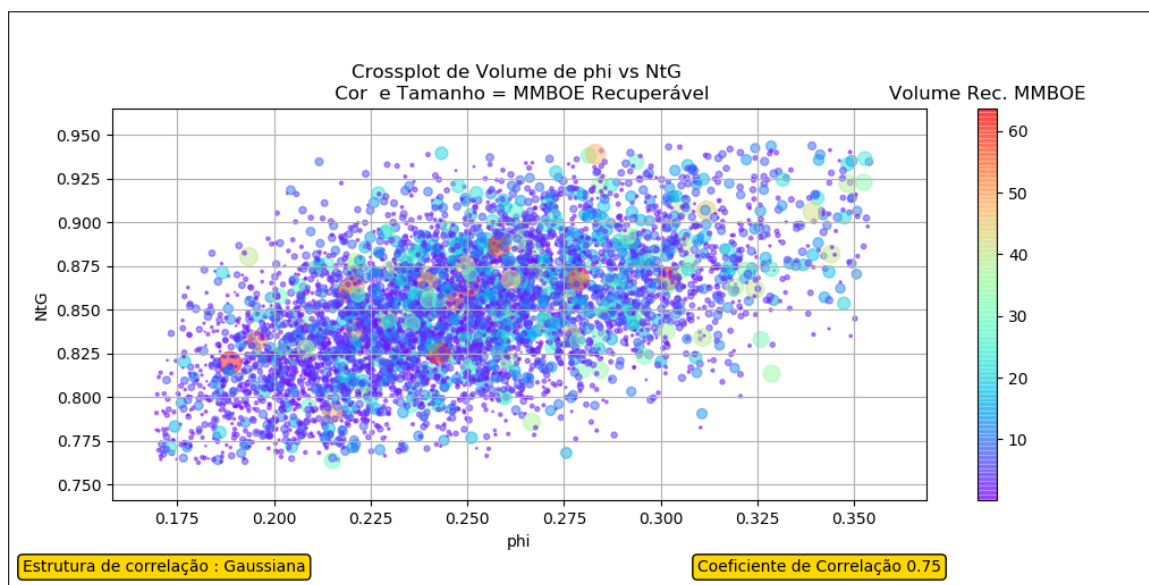


Figura 3.34: Exemplo de crossplot de NTG vs ϕ . Note que não há estratificação vertical em nenhum dos quadrantes mostrando que a incerteza de Bg e NTG tem pouco impacto no cálculo de volumes para esta oportunidade.

No caso de variáveis menos significativas no cálculo de volumes é possível verificar uma distribuição uniforme ao longo do crossplot (Figura 3.34).

Caso ambas as variáveis tenham forte impacto verificamos a concentração dos pontos de maior volume em um dos quadrantes do crossplot (Figura 3.35).

Note que esta avaliação aponta o impacto das variáveis feitas de maneira independente. Precisamos então avaliar o impacto das correlações de maneira separada.

A estratégia utilizada neste trabalho consiste de calcular o volume com os pares de variáveis independentes, seguido do cálculo de variáveis com dependência e dessa forma quantificar os volumes:

$$\Delta Vol = Vol_{dep} - Vol_{ind} \quad (3.12)$$

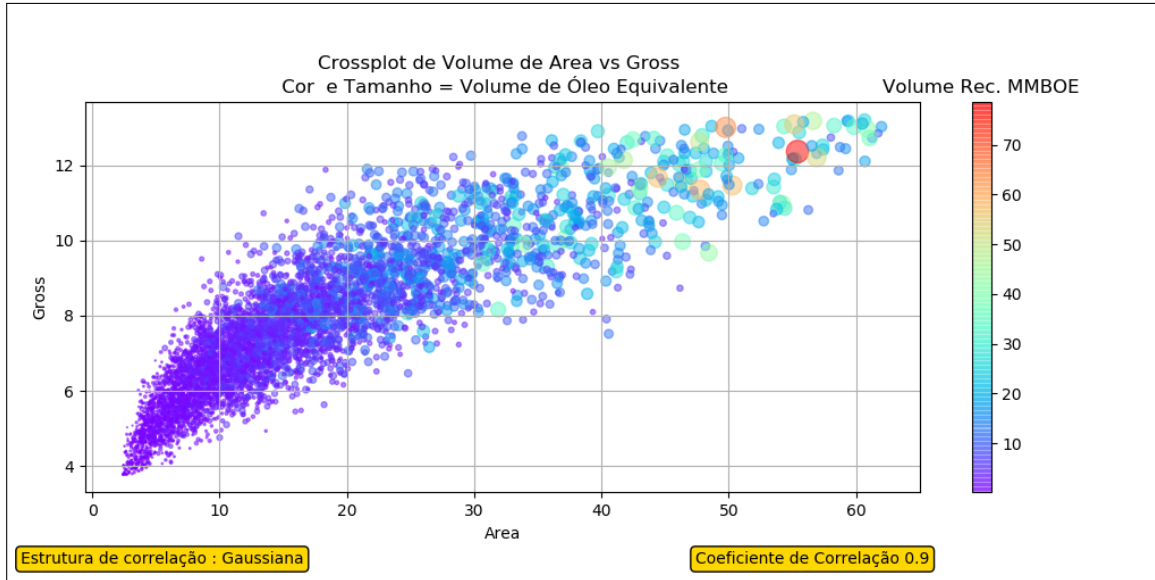


Figura 3.35: Exemplo de crossplot de área versus Gross pay. O tamanho e cor dos pontos refletem o volume de hidrocarbonetos recuperáveis. Note que a concentração de pontos de maior volume no 1.o quadrante indica que as duas variáveis são muito relevantes para o cálculo da oportunidade o que indica que a correlação entre elas terá bastante impacto nos volumes.

Onde ΔVol é a variação de volume, Vol_{dep} é o volume calculado com dependência (correlação) das variáveis e Vol_{ind} é o volume calculado considerando as variáveis totalmente independentes. E percentualmente pode ser calculado da seguinte forma:

$$\Delta Vol = \frac{Vol_{dep} - Vol_{ind}}{Vol_{ind}} \quad (3.13)$$

Chegamos então a um exemplo como o da Figura 3.36.

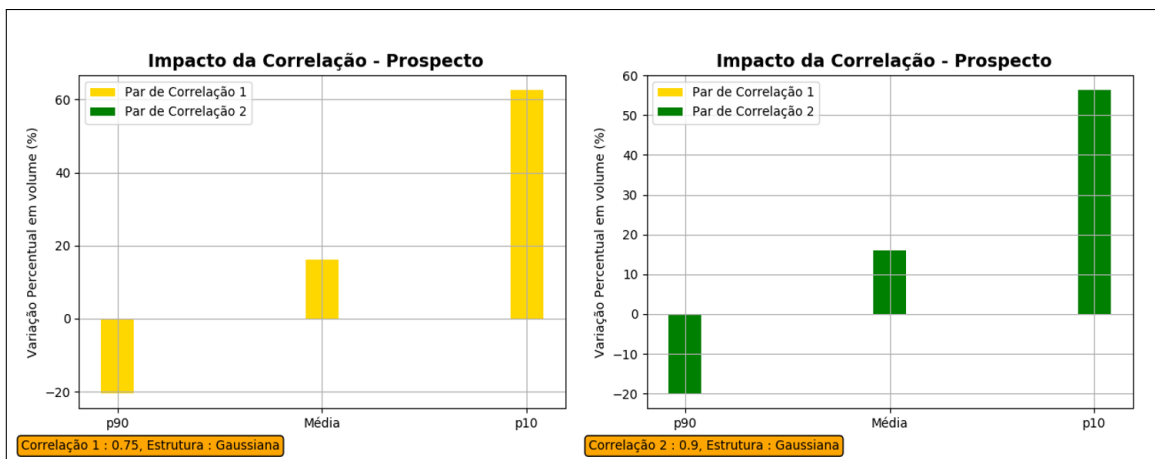


Figura 3.36: Exemplo de gráfico de barras para avaliar o impacto da correlação das duas distribuições.

Note que as equações 3.12 e (3.13) são aplicáveis para pares de distribuições isolados, contudo, ao utilizarmos mais de uma par de variáveis os efeitos não se combinam linearmente. Para avaliar o impacto de cada dependência das duas distribuições, podemos aplicar uma regra distributiva:

$$\Delta Vol_a = \frac{\alpha}{\alpha + \beta} \cdot \gamma \quad (3.14)$$

$$\Delta Vol_b = \frac{\beta}{\alpha + \beta} \cdot \gamma \quad (3.15)$$

Considere os pares de variáveis (a,b) e (c,d). Neste caso, α é o ΔVol da equação (3.13) aplicando o par (a,b) independentemente. β é o ΔVol da equação (3.13) aplicando a correlação nas variáveis (c,d) independentemente e γ diz respeito à equação 3.13 considerando os pares de correlação (a,b) e (c,d) conjuntamente. Dessa forma chegamos à Figura 3.37 :

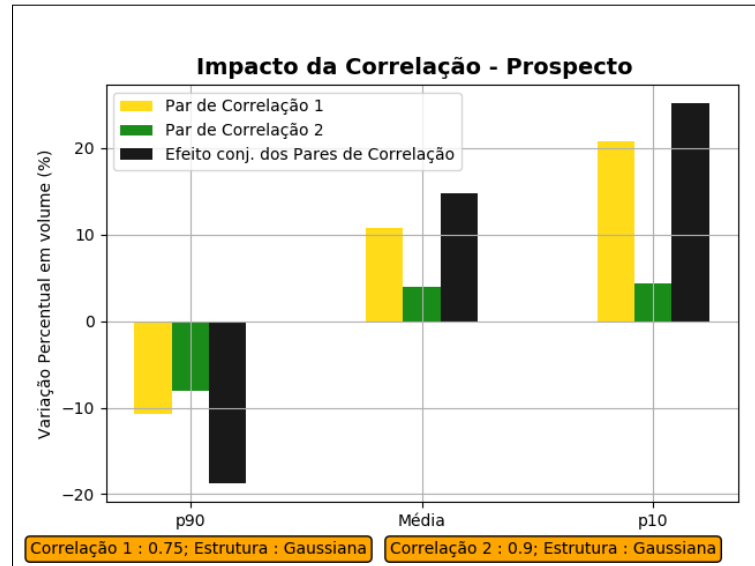


Figura 3.37: Exemplo de gráfico de barras para avaliar o impacto da correlação das duas distribuições.

Onde as barras amarelas dizem respeito ao impacto da correlação (a,b) e as barras verdes dizem respeito à variação de volume nos pares de variáveis (c,d). As barras pretas dizem respeito ao impacto da correlação das variáveis (a,b) e (c,d) conjuntamente. Como utilizamos a regra distributiva podemos dizer que as barras pretas constituem a soma da barra amarela e verde em cada cenário (P90,média e P10).

3.4 Programa e Interface - ReVBR (Risco e Volume)

Além das bibliotecas de cálculo disponibilizadas nos anexos, como sugestão do professor e orientador Dr. Luiz Alberto Santos, foram desenvolvidas interfaces gráficas de maneira a permitir que usuários pouco familiarizados com programação pudessem utilizar a ferramenta livremente. Batizamos o programa de ReV (Risco e Volume), em função do seu objetivo principal. O termo BR explicita o país de origem do trabalho (Brasil).

Todo o trabalho foi implementado utilizando a linguagem de programação Python 3 - versão 3.6. Esta escolha foi feita por ser uma linguagem gratuita, independente de sistema operacional, ser interpretada e possuir grande variedade de bibliotecas gratuitas disponíveis. Isto permitiu melhor velocidade de implementação e que os resultados aqui produzidos pudessem ser replicados em ambiente acadêmico livremente.

A linguagem Python foi criada no início dos anos 90 por Guido van Rossum na *Stichting Mathematisch Centrum (CWI)*, na Holanda como um sucessor de uma linguagem chamada ABC. Todas as versões do Python são de código aberto. Historicamente, a maioria das versões do Python são *GPL-compatible*. Em linhas gerais, a única diferença da licença Python para a licença GPL pura é que ela permite que se faça alterações no software e que o código produzido seja redistribuído sem que as alterações sejam obrigatoriamente feitas de forma aberta. Isto permite que produtos comerciais sejam produzidos com maior facilidade.

Toda a construção da interface fez uso da biblioteca Tkinter (licença de código aberto BSD -*Berkeley Software Distribution*), e portanto, tem reprodução gratuita. Aconselhamos aos usuários que queiram utilizar os programas construídos neste trabalho a recompilarem os códigos utilizando a instalação Anaconda em versão 3.6 - 64 bits para máxima compatibilidade com o código originalmente escrito pelo autor.

Espera-se que as aplicações sejam especialmente úteis no meio acadêmico tanto na produção de projetos de pesquisa em que os produtos alvo desta dissertação sejam abordados, quanto em aulas de disciplinas que envolvam os conteúdos aqui apresentados como por exemplo em geologia econômica ou geofísica de prospecção. Por isso nesta seção iremos explicar a metodologia de uso da interface criada no âmbito desta dissertação. Esta seção funciona como um pequeno manual para os usuários desta ferramenta.

A interface com o Usuário.

A interface criada foi pensada de modo a ser simples e funcional (Figura 3.38). Evitamos o uso de múltiplas janelas para simplificar o programa e a implementação. Tivemos especial preocupação em facilitar a produção de gráficos univariado (histograma ou frequência acumulada) ou multivariado (crossplot) de maneira a aumentar a capacidade analítica da equipe de interpretação sobre todos os parâmetros de entrada. Foram utilizados menus do tipo *dropdown* nos elementos de seleção, reduzindo a possibilidade de erros de execução devido a digitação.

Outra preocupação foi deixar os botões facilmente reconhecíveis. Por isso, fizemos uso de diversas cores para chamar atenção aos botões rotineiramente utilizados. Uma vantagem da estrutura de interface utilizada, é que erros de execução raramente interrompem o programa.

A janela principal do software é dividida em cinco grandes grupos Figura (3.39):



Figura 3.38: Exemplo de interface Gráfica do Software ReV BR.

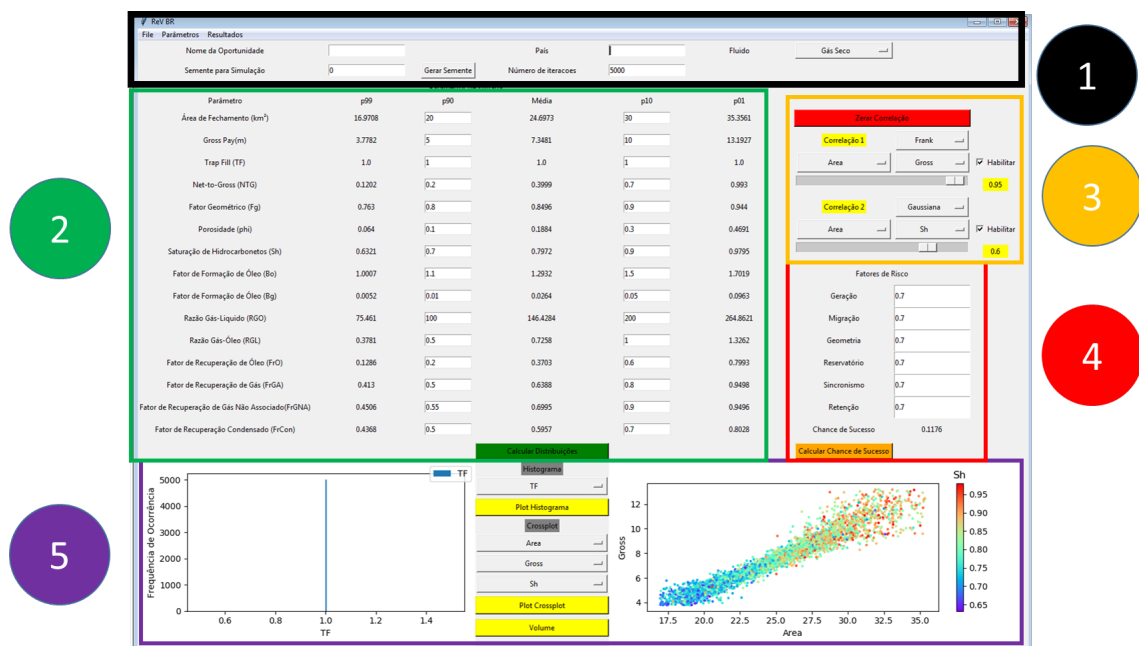


Figura 3.39: Exemplo de interface Gráfica do Software ReV BR.

Estas áreas podem ser definidas como:

1. Área para características gerais da oportunidade;
2. Área para preenchimento de parâmetros volumétricos;
3. Área de correlação entre parâmetros volumétricos;
4. Área de definição de parâmetros de risco geológico;
5. Área de plotagem de gráficos;

Na parte superior da aplicação foram implementadas algumas funções úteis no menu *Arquivo*, como a opção de começar um novo trabalho **Novo**, bem como as opções **Salvar** e **Importar** o trabalho realizado. As opções de salvar/importar são especialmente úteis para dar continuidade e realizar comparações de diferentes prospectos. Os arquivos são salvos em formato json. O nome do arquivo não é utilizado na estrutura de carregamento dos dados, mas recomendamos veementemente que não sejam alterados os nomes dos mesmos após serem salvos uma vez que eles guardam informações bastante descritivas sobre o prospecto.

No menu Resultados é possível obter diferentes representações dos resultados, em forma de tabela (Figura 3.40) onde podem ser visualizados os valores de média e percentis principais além da média incondicional (média multiplicada pelo risco), ou na forma gráfica como uma c.d.f. visualmente truncada em p99 e p01 (Figura 3.41). A representação gráfica pode ser obtida também a partir do botão Volumes na porção mais inferior da interface.

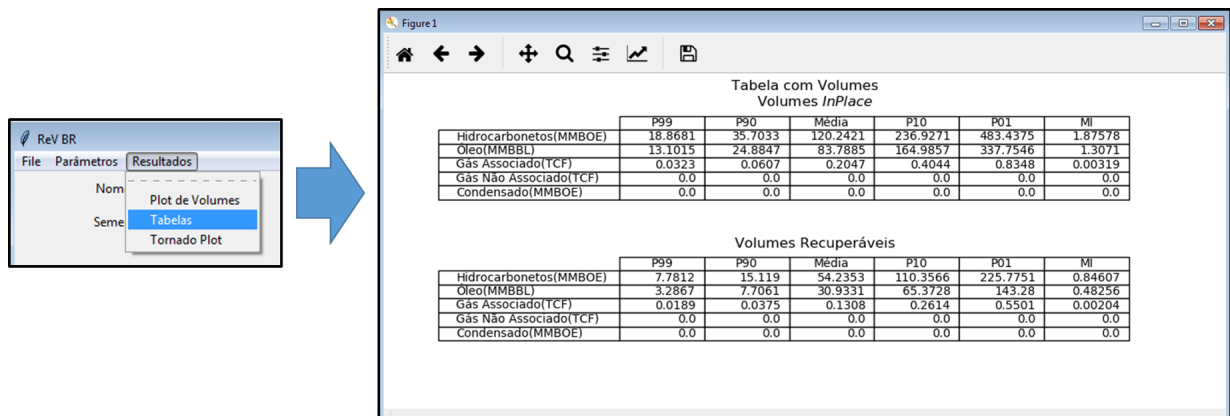


Figura 3.40: Exemplo de resultados de volumes calculados extraídos na forma de tabela.

Já no painel principal do software é possível definir o nome da oportunidade, o país onde ela se encontra e o tipo de fluido esperado, a semente para a simulação de Monte Carlo (que deve ser um número inteiro), o botão gerar semente permite que uma semente aleatória seja gerada, e o número de simulações. É aconselhável que em notebooks pessoais ou computadores de baixa performance sejam utilizadas 5000 simulações enquanto que em estações científicas de trabalho 40.000 possa ser um bom número para distribuições mais bem amostradas. O número de iterações máxima é definida em 50.000 (cinquenta mil), de maneira a prevenir que erros de digitação de zeros adicionais tornem o programa instável. Este número se mostrou razoável na modelagem de prospectos em diversos testes realizados.

Foi definido um menu *dropdown*, que permite determinar o tipo de fluido da acumulação. Foram definidos os principais : Gás Seco, Gás Úmido, Condensado e Óleo. Note que a depender do parâmetro de tipo de fluido a unidade do eixo horizontal do gráfico de saída pode variar conforme os itens abaixo. A utilização do eixo das abcissas em MMBOE (milhões de barris de óleo equivalente), permitem que gás e óleo sejam computados de forma conjunta.

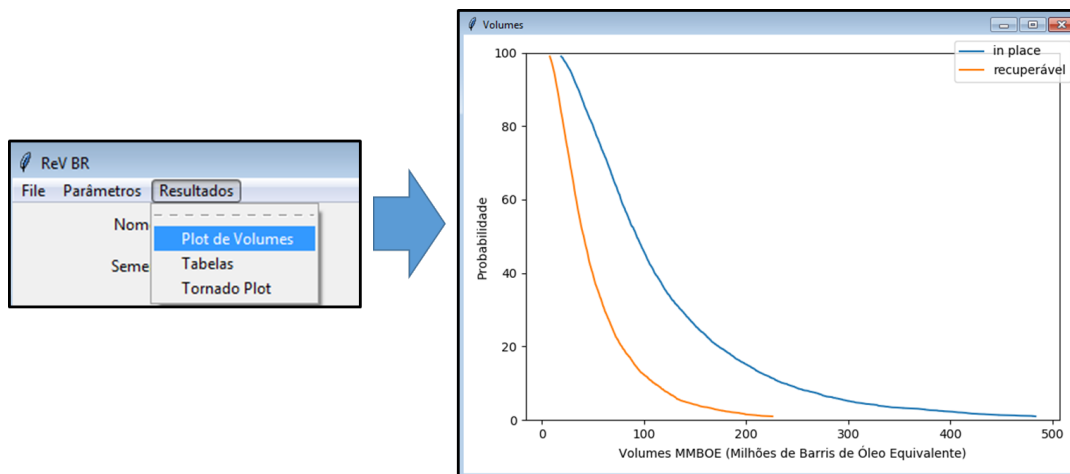


Figura 3.41: Exemplo de resultados de volumes calculados extraídos na forma de função de curva de probabilidade cumulativa.

- Gás Seco - eixo horizontal TcF (*Trillion Cubic Feet*).
- Gás Úmido - eixo horizontal MMBOE (*Million Barrels of Oil Equivalent*);
- Condensado - eixo horizontal MMBOE;
- Óleo - eixo horizontal MMBOE.

É comum, ao avaliar um prospecto exploratório, que haja alguma divergência entre as diversas equipes de interpretação que avaliam um mesmo projeto. Entretanto, nem sempre é possível verificar o impacto real daquela variável no resultado final de um prospecto. Por este motivo implementamos também o Tornado Plot (Figura 3.42), que mede a sensibilidade do volume a cada parâmetro de entrada. Esta visualização é útil para chamar atenção às variáveis com maior impacto na percepção de volume de um prospecto. Ressaltamos que este gráfico não deve ser encarado como um indicativo de que podemos concentrar menos esforços nas variáveis de menor incerteza, porque a nossa percepção de incerteza sobre os parâmetros pode estar errada.

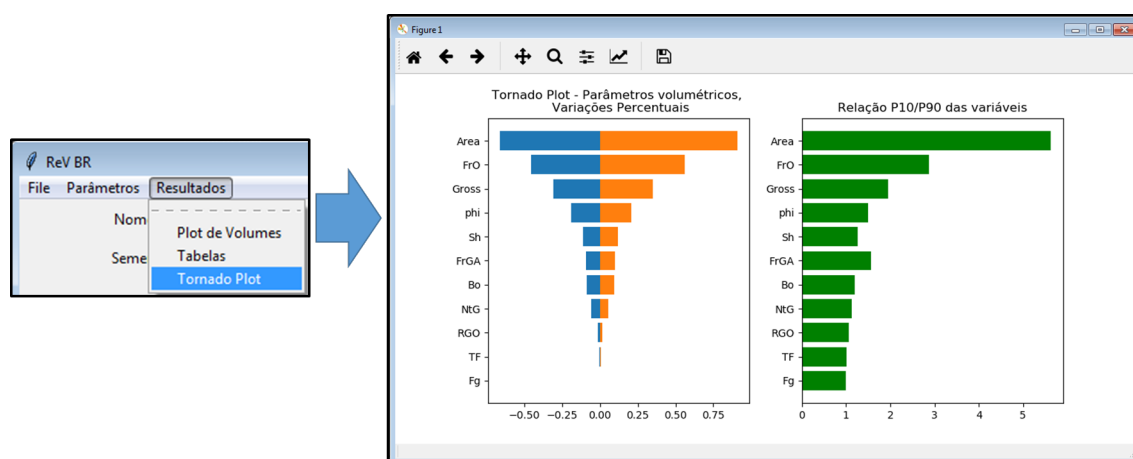


Figura 3.42: Exemplo de Tornado plot do Software ReV BR.

Note que o *tornado plot* avalia apenas as variáveis de forma independentes. Por isso implementamos também um gráfico que mede o impacto das correlações aplicadas em termos de variação de volume de hidrocarbonetos recuperáveis (Figura 3.43). A escolha do uso do volume

de hidrocarbonetos recuperável se deu porque ele é sensível a todos os parâmetros envolvidos na avaliação (incluindo obviamente o fator de recuperação)⁹.

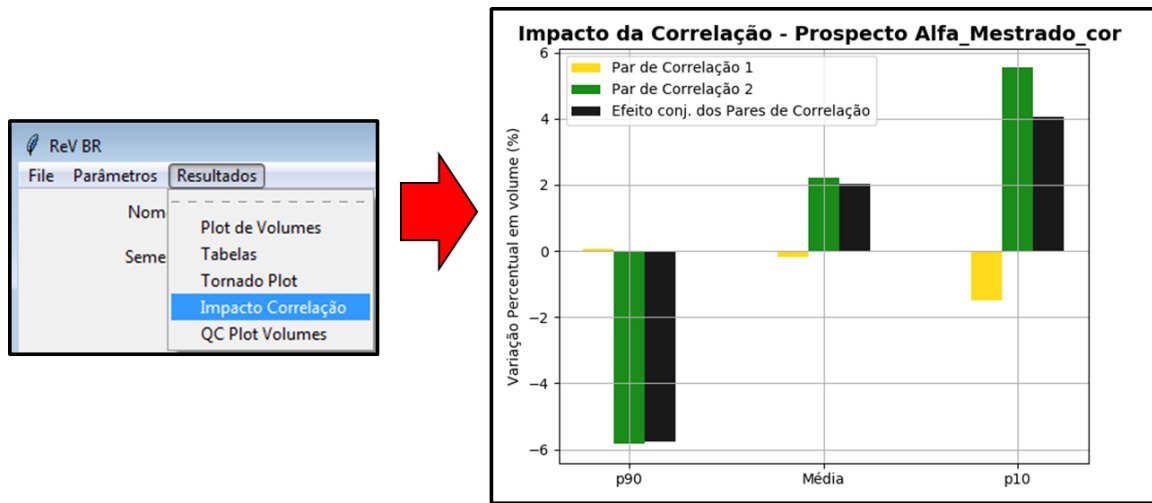


Figura 3.43: Exemplo de gráfico de avaliação de controle de impacto da correlação.

Neste gráfico avaliamos o impacto conjunto das correlações (barras pretas) nos valores de p10, média e p90, do primeiro par de correlação (barras amarelas) e do segundo par de correlações (barras verdes).

Por fim implementamos também a opção de avaliar o par de correlação com o volume recuperável através de um crossplot. Com isso esperamos avaliar o impacto da correlação com o volume calculado (Figura 3.44).

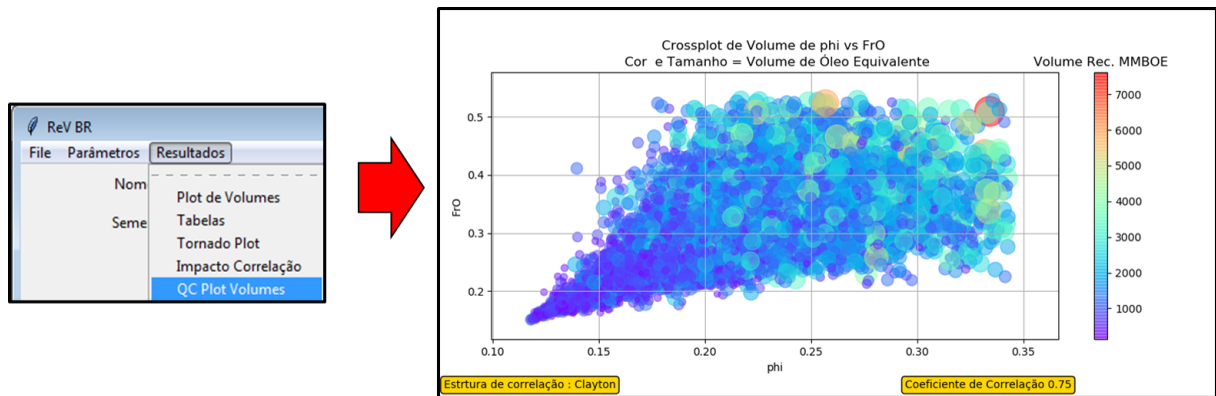


Figura 3.44: Exemplo de crossplot do par de correlação avaliado em relação com respeito ao volume de hidrocarbonetos calculado (hidrocarbonetos são a cor e o tamanho dos pontos).

Construção da distribuição lognormal truncada

Conforme discutido no Capítulo 2, assumimos que incerteza representada pelas distribuições utilizadas na avaliação individual de cada parâmetro pode ser aproximada por distribuições lognormais truncadas em p01 e p99. A questão que se põe é: Como truncar distribuições que não foram geradas?

A estratégia aqui utilizada se baseia em calcular os valores teóricos de p99 e p01 com base nos parâmetros de P10 e P90. Conforme o passo-a-passo a seguir:

⁹Se utilizássemos os volumes *in place* não conseguiríamos fazer a avaliação dos parâmetros de fator de recuperação.

1. Com base nos parâmetros P10 e P90 calcular os valores teóricos de parâmetros p99 e p01 pela transformada z-normal;
2. Gerar as simulações uma a uma descartando os valores que excedem os limites de p99 e p01 ou valores máximos e mínimos possíveis para cada parâmetro até que o número de simulações válidas seja alcançado;

Note que existem outras estratégias de fazer esta estimativa sem a necessidade dos valores teóricos (no caso dos parâmetros de entrada seguirem diferentes distribuições), as quais poderíamos citar:

- Gerar amostras por *bootstrapping* ou simulação sem truncamento com 102% da quantidade de simulações desejadas e clipar o primeiro e ultimo percentil;
- Gerar amostras da normal padrão truncada nos valores de p01 e p99; transforma-la em lognormal padrão e apenas reescalar as distribuições de acordo com os parâmetros de p10 e p90;

Utilizamos a primeira estratégia citada pois com isso teríamos melhor controle dos valores de p99 e p01, calculados analiticamente e, portanto, estaríamos menos vulneráveis a dispersões numéricas principalmente quando as simulações tem poucas iterações (<1000).

O painel inicial do programa apresenta os campos p10 e p90 utilizados para criar as distribuições. Foram utilizados alguns valores tipicamente encontrados na indústria de petróleo, para popular os parâmetros iniciais. Os valores de p99, Média e p01 são calculados a partir dos valores de p90 e p10, são inicialmente definidos como *nan* (*not a number*), como uma forma de demonstrar para o usuário que nenhuma distribuição foi calculada até aquele momento.

Parâmetro	Parâmetros de Entrada				
	p99	p90	Média	p10	p01
Área de Fechamento (km ²)	16.9708	<input type="text" value="20"/>	24.6973	<input type="text" value="30"/>	35.3561
Gross Pay(m)	3.7782	<input type="text" value="5"/>	7.3481	<input type="text" value="10"/>	13.1927
Trap Fill (TF)	1.0	<input type="text" value="1"/>	1.0	<input type="text" value="1"/>	1.0
Net-to-Gross (NTG)	0.1202	<input type="text" value="0.2"/>	0.3999	<input type="text" value="0.7"/>	0.993
Fator Geométrico (Fg)	0.763	<input type="text" value="0.8"/>	0.8496	<input type="text" value="0.9"/>	0.944
Porosidade (phi)	0.064	<input type="text" value="0.1"/>	0.1884	<input type="text" value="0.3"/>	0.4691
Saturação de Hidrocarbonetos (Sh)	0.6321	<input type="text" value="0.7"/>	0.7972	<input type="text" value="0.9"/>	0.9795
Fator de Formação de Óleo (Bo)	1.0007	<input type="text" value="1.1"/>	1.2932	<input type="text" value="1.5"/>	1.7019
Fator de Formação de Óleo (Bg)	0.0052	<input type="text" value="0.01"/>	0.0264	<input type="text" value="0.05"/>	0.0963
Razão Gás-Líquido (RGO)	75.461	<input type="text" value="100"/>	146.4284	<input type="text" value="200"/>	264.8621
Razão Gás-Óleo (RGL)	0.3781	<input type="text" value="0.5"/>	0.7258	<input type="text" value="1"/>	1.3262
Fator de Recuperação de Óleo (FrO)	0.1286	<input type="text" value="0.2"/>	0.3703	<input type="text" value="0.6"/>	0.7993
Fator de Recuperação de Gás (FrGA)	0.413	<input type="text" value="0.5"/>	0.6388	<input type="text" value="0.8"/>	0.9498
Fator de Recuperação de Gás Não Associado(FrGNA)	0.4506	<input type="text" value="0.55"/>	0.6995	<input type="text" value="0.9"/>	0.9496
Fator de Recuperação Condensado (FrCon)	0.4368	<input type="text" value="0.5"/>	0.5957	<input type="text" value="0.7"/>	0.8028

1

Figura 3.45: Painel de parâmetros de entrada utilizados no cálculo das distribuições de entrada. O botão verde é utilizado para calcular as distribuições.

De forma transparente ao usuário, foram definidos os valores máximos e mínimos para cada parâmetro de forma a limitar os valores obtidos em cada distribuição Tabela 3.4. Os valores

mínimos foram definidos como 0.001 por limitação da distribuição logarítmica que não assume valores iguais a zero.

Parâmetros	Valor Mínimo	Valor Máximo
Área de Fechamento(km^2)	0.001	9999999
Gross Pay(m)	0.001	9999999
Trap Fill (TF)	0.001	1
Net-to-Gross (NTG)	0.001	1
Fator Geométrico (Fg)	0.001	1
Porosidade (ϕ)	0.001	0.5
Saturação de Hidrocarbonetos (Sh)	0.001	0.98
Fator de Formação de Óleo (B_o)	1	350
Fator de Formação de Óleo (B_g)	0.001	1
Razão Gás-Líquido (RGO)	0.001	3000
Razão Líquido-Gás (RLG)	0.001	1
Fator de Recuperação de Óleo (FrO)	0.001	0.8
Fator de Recuperação de Gás (FrGA)	0.001	0.95
Fator de Recuperação de Gás Não Associado(FrGNA)	0.001	0.95
Fator de Recuperação Condensado (FrCon)	0.001	0.95

Tabela 3.4: Alguns fatores que impactam no fator de recuperação de uma acumulação.

É possível visualizar a distribuição de qualquer parâmetro através da janela gráfica inferior esquerda (Figura 3.46).

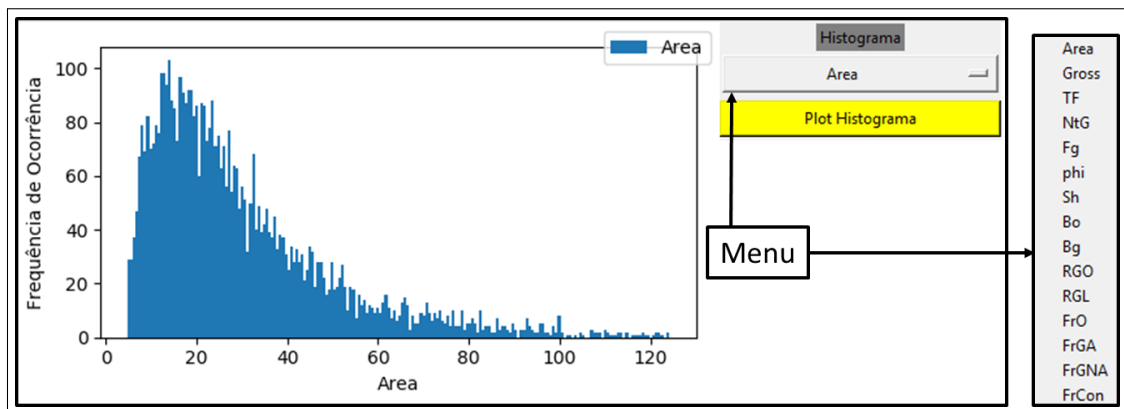


Figura 3.46: Exemplo de plotagem de histograma com distribuição.

Como dados de entrada, são utilizados os valores de p10 (otimista) e p90 (pessimista) para cada distribuição. O caso especial em que os parâmetros de entrada p10 e p90 são iguais, a distribuição assume a forma constante. Não foram implementados muitos códigos de tratamentos de exceção mas tivemos o cuidado de impedir que os valores de p10 sejam menores do que os de p90 sinalizando o par de variável em vermelho (Figura 3.47).

Painel para aplicação das correlações

No software é possível gerar até dois pares de variáveis correlacionadas. Esta limitação evita que sejam construídas matrizes de correlação que não sejam positivas definidas. O painel referente à Figura 3.48 se encontra na porção direita superior da tela:

Os itens destacados na Figura são:

Parâmetro	Parâmetros de Entrada			
	p99	p90	Média	p10
Área de Fechamento (km ²)	2.4197	20	14.9288	10
Gross Pay(m)	3.7782	5	7.3481	10
Trap Fill (TF)	1.0	1		
Net-to-Gross (NTG)	0.1202	0.2		
Fator Geométrico (Fg)	0.763	0.8		
Porosidade (phi)	0.064	0.1		
Saturação de Hidrocarbonetos (Sh)	0.6321	0.7		

Dados de Entrada

Favor rever parâmetros de entrada em vermelho.

OK

Figura 3.47: Exemplo de parâmetros p10 e p90 mal definidos (onde $p90 > p10$).

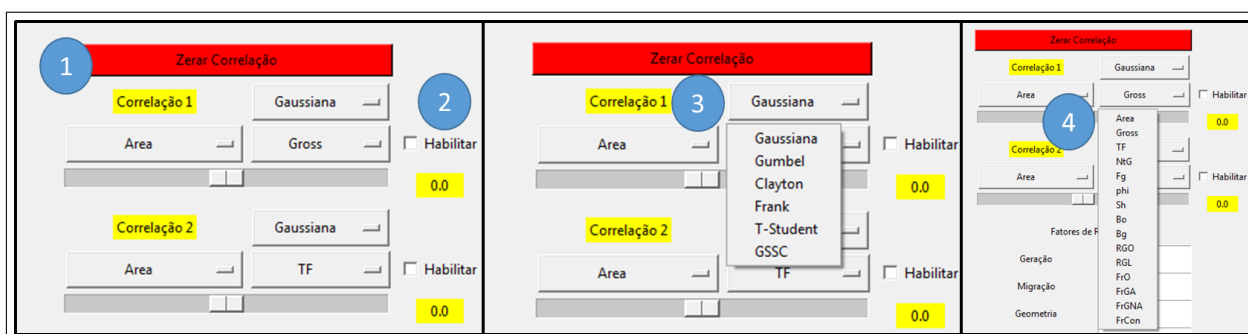


Figura 3.48: Painel referente a correlação entre pares de variáveis.

1. Botão Zerar correlação: Permite que o usuário retorno as configurações de correlação à posição inicial;
2. *CheckBox*: Habilitar os valores de correlação selecionados;
3. *Drop-Down menu* : Selecionar estrutura de correlação;
4. *Drop-Down menu* : Selecionar variável para correlação;

Abaixo de cada variável é possível selecionar os valores de correlação que variarão de -1 a 1 de acordo com um seletor. No caso do algoritmo GSSC, antes de tentar gerar as distribuições é feito um teste de limite de correlação com o método GSC para verificar se o valor de correlação é possível entre as duas distribuições. Ao utilizar as estruturas de correlação que seguem alguma função cópula os valores de θ são normalizados de forma a representar aproximadamente o valor de coeficiente de correlação desejado.. Na hipótese de duas correlações iguais serem habilitadas, a ultima será utilizada e a primeira descartada.

É possível visualizar a distribuição de pontos geradas pela correlação no gráfico de dispersão de pontos na porção inferior direita da Figura 3.49.

Os itens apontados na figura são:

1. *Drop-Down menu* : Selecionar parâmetro do eixo X;
2. *Drop-Down menu* : Selecionar parâmetro do eixo Y;
3. *Drop-Down menu* : Selecionar parâmetro referente às cores;
4. Botão Plot Crossplot : Gerar o gráfico de dispersão.

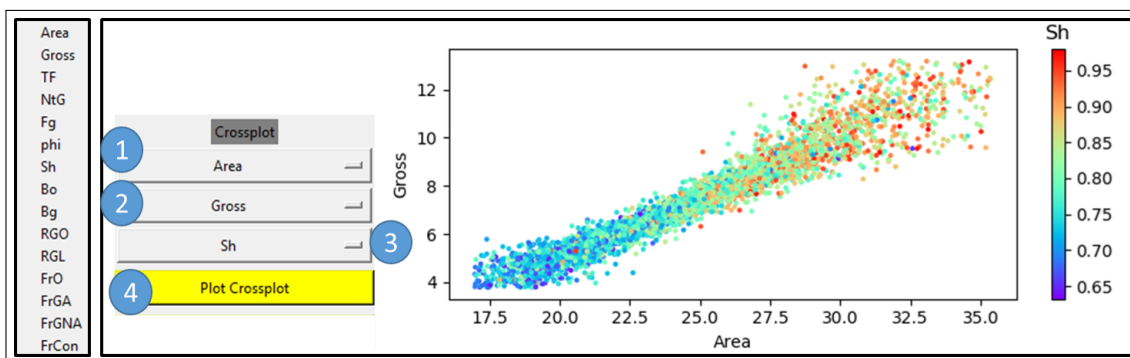


Figura 3.49: Exemplo de plotagem de *crossplot* com distribuições de Área (eixo x), *Gross Pay* (eixo y) e Saturação (cores).

A escala de cores selecionada foi a *rainbow*. A escolha foi feita para tentar determinar a máxima variabilidade de cores.

Cálculo de Chance de Sucesso

Além dos parâmetros volumétricos é possível também associar o risco exploratório geológico. O cálculo de chance de sucesso consiste basicamente no produtório dos elementos considerados Figura 3.50. Caso o usuário acredite que os elementos de chance são muito extensos, é possível definir alguns elementos como 1, que é o elemento neutro do produtório.

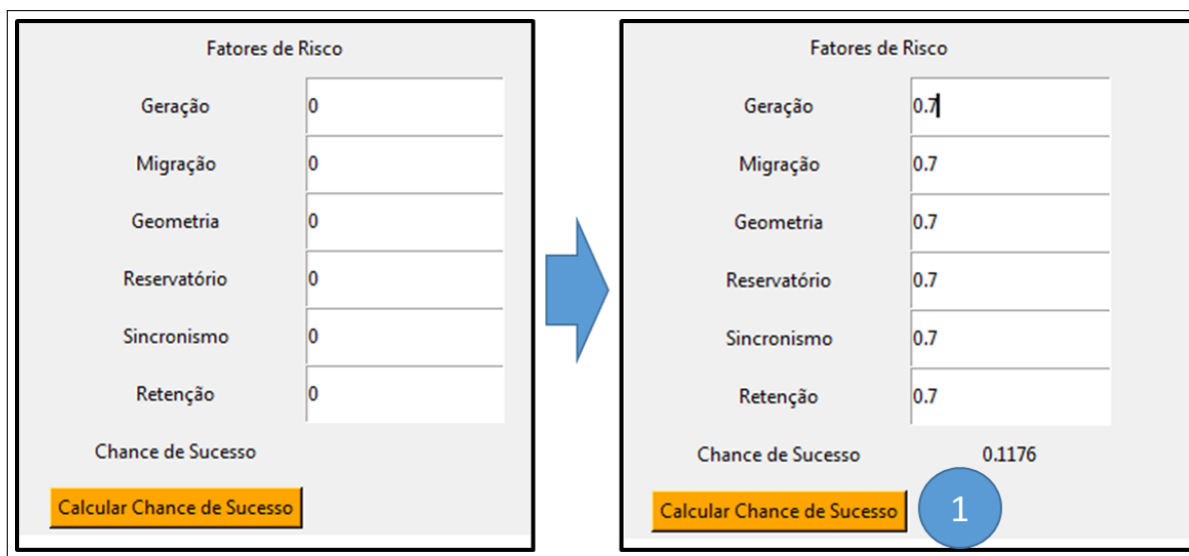


Figura 3.50: Painel referente a determinação dos fatores de chance de sucesso geológico.

A chance de sucesso deveria seguir pelo menos uma tendência semelhante àquela proposta por Otis and Schneidermann (1997) apresentada no Capítulo 2. Os campos preenchidos inicialmente como 1 (100%). Note que os elementos do sistema petrolífero devem ser definidos entre 0 e 1. A chance de sucesso é calculada quando o botão laranja (botão (1) da Figura 3.50), é pressionado.

Capítulo 4

Aplicações

Este capítulo é dedicado a aplicações práticas do uso de correlações em prospectos. Do ponto de vista geológico exploraremos a relação entre área e *net pay* explorando a interpretação de como as fácies se distribuem ao longo de um prospecto. Avaliaremos também o impacto da correlação nos parâmetros de fluido conforme abordado na seção de metodologia.

Num primeiro momento faremos um breve exercício no *prospecto* Libra, hoje campo de Mero, com base em dados públicos disponíveis na literatura e sob licença que permitam a distribuição e publicação. Estes exemplos serão meramente ilustrativos e não tem o compromisso com uma avaliação comercial da oportunidade. Num segundo momento faremos a avaliação de um lead identificado em dados públicos da região do Golfo do México.

4.1 Relação entre Área *Net Pay* em Prospectos do Play Pré-Sal - Santos Brasil

Sem dúvida o pré-sal é um dos *plays* de maior sucesso exploratório da última década, seja por seus expressivos volumes ou por sua alta taxa de sucesso. O objetivo desta seção não é abordar em detalhe os assuntos relacionados ao play pré-sal, mas apenas demonstrar como a correlação de variáveis poderia ser utilizada na avaliação de abordagens probabilística deste tipo de oportunidade.

O play pré-sal carbonático se estende ao longo do atlântico sul dos litorais do Brasil e Angola abaixo da extensa coluna de sal Aptiano (Figura 4.1) Moreira et al. (2007).

Este *play* tem sido estudado desde 2007 devido às grandes descobertas na margem brasileira em estruturas com mais de $500km^2$ de área que excedem 1Bbbl de óleo recuperável.

No continente Sul-Americano ele se estende da Bacia de Santos (Campos de Lula, Mero etc), a sul, até a Bacia do Espírito Santo (Jubarte e parque das baleias) a norte, em uma área de cerca de $150,000 Km^2$, denominada de picanha azul, a profundidades geralmente superiores a 1500m de Lâmina d'água. Na margem africana este *play* está localizado desde a bacia de Benguela a sul até a Bacia de Kwanza (campos de Lontra e Cameia). A norte a sua extensão em área não é bem definida devido a pequena quantidade de poços perfurados na bacia até o momento e a influência do delta do Baixo Congo (Ceraldi and Green (2016)).

O sistema petrolífero compreende a geradora lacustre syn-rifte de idade equivalente ao andar bioestratigráfico Jiquiá (ostracodes série Recôncavo) Formação Lagoa Feia, e reservatórios divididos em dois intervalos (Figura 4.2): os microbialitos/estromatólitos do andar Alagoas, depositados em um ambiente equivalente a um grande lago hipersalino, em uma fase de quiescência tectônica geometricamente denominada de seção *sag*. Há também uma seção adicional de reservatórios carbonáticos composta de coquinas (bancos de conchas depositados sob condições lacustres) no topo da seção syn-rifte.

Estes dois alvos são separados por uma seção de sedimentos carbonáticos fechados preenchidos por material fino denominado de seção talco-estivensita (Ceraldi and Green, 2016).

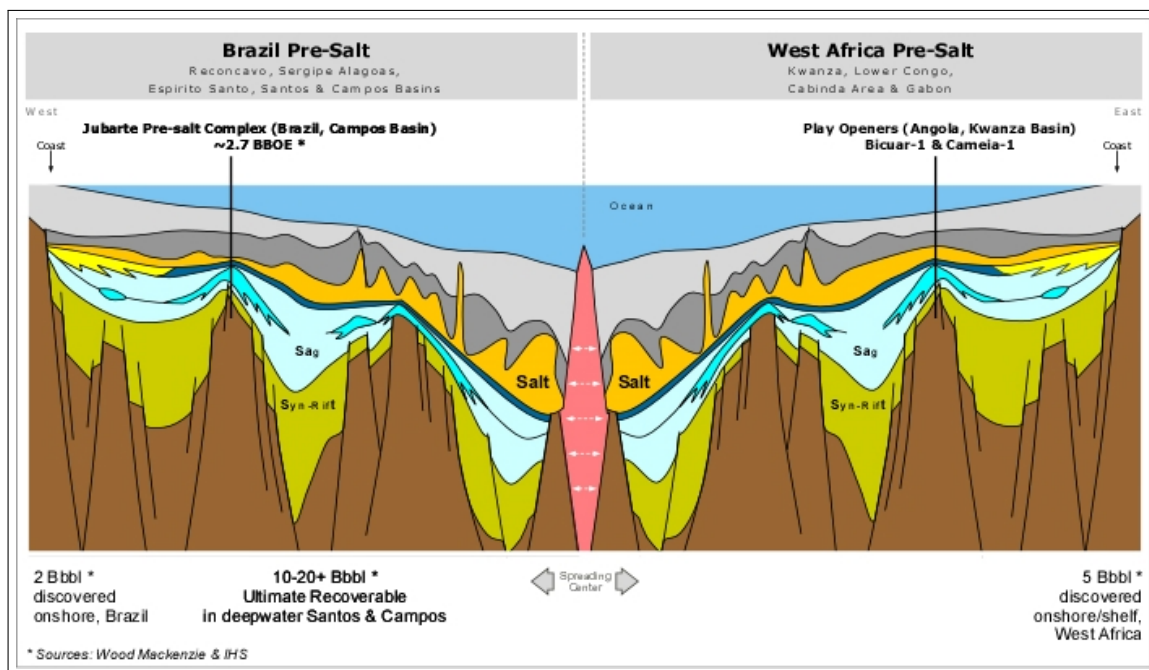


Figura 4.1: Seção geológica esquemática com correlação do play pré-sal na margem brasileira e angolana (fonte: <http://www.gasandoil.com/news/2013/08/west-africa-deep-water-discovery-disappoints>).

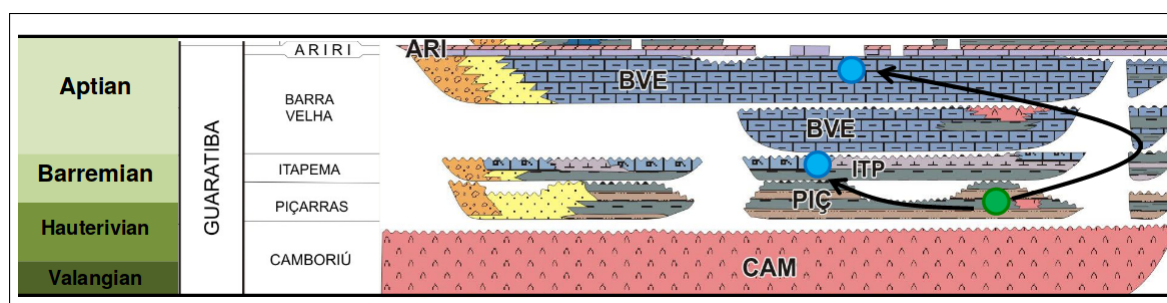


Figura 4.2: Parte da carta estratigráfica da Bacia de Santos, apenas com a seção pré-sal, o círculo verde indica a geradora Jiquiá, enquanto os reservatórios são os estromatólitos da seção sag (Formação barra Velha) e as coquinas da Formação Itapema. Fonte Moreira et al. (2007) apud. Petersohn and Abelha (2013).

As Estruturas normalmente são relacionadas a fechamentos na base do sal relacionados a altos estruturais, que se replicariam nos intervalos estratigráficos inferiores (Figura 4.3). A ocorrência de boas fácies reservatórios normalmente são controladas também pela ocorrência de altos externos em conjunção com variações no espaço de acomodação. Dessa forma ambientes mais profundos, seja por estarem estruturalmente mais baixo seja pela subida do nível do lago, tendem a gerar fácies reservatórios piores. Em contexto proximal a ocorrência de boas fácies reservatórios é também controlada pela ausência do influxo de terrígenos.

Considerando que os altos estruturais atualmente verificados em subsuperfície correspondem aos paleo-altos estruturais, podemos considerar que a medida que nos afastamos do ápice da estrutura, maior a chance de obtermos fácies piores traduzidas em menores porosidades efetivas e proporção inferior de *net pay*.

Note que na avaliação de prospectos desta natureza normalmente associamos os valores de área P90 como correspondentes às cotas mais altas e os valores de P10 englobaria as cotas mais baixas, por isso inferimos que os valores de área tem correlação negativa com as fácies. Contudo, a subsidência nos flancos da estrutura pode ter taxas de subsidências maiores ou menores do que

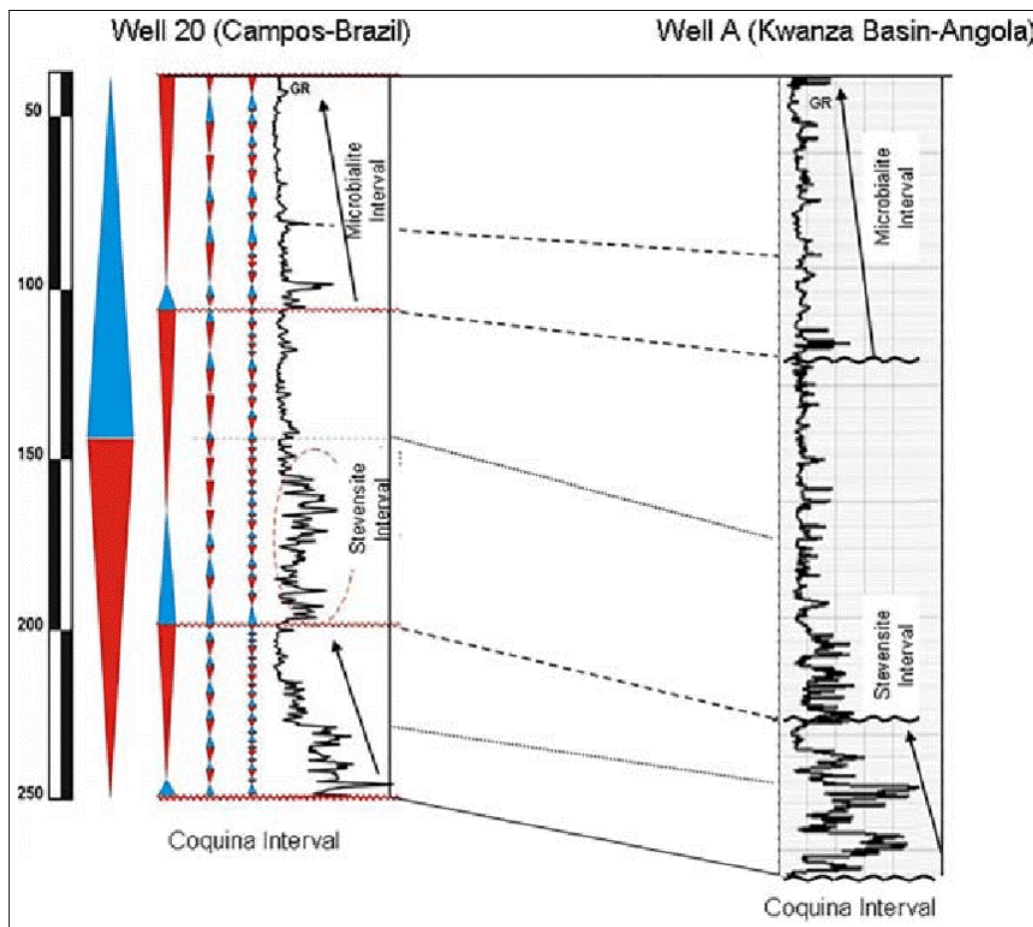


Figura 4.3: Correlação de poços e respectivos ciclos de Karagodin da seção pré-sal carbonática entre poço das Bacias de Campos, Brasil e Kwanza, Angola (Ceraldi and Green (2016)).

o ápice, o que afetaria diretamente a nossa percepção de variação de fácies e portanto o grau de correlação entre estas duas variáveis.

É importante destacar que nesta conjectura estaríamos considerando que há chance de que a estrutura esteja parcialmente preenchida. Esta variação poderia ser feita com o parâmetro área ou com o parâmetro *Trap-Fill*, que considera justamente o preenchimento parcial da trapa.

Faremos agora o exercício de calcular os volumes estimados do prospecto Libra (que engloba o atual campo de Mero), com e sem as características de correlação propostas neste trabalho. Como sabemos que o prospecto já se transformou numa acumulação de hidrocarbonetos em produção, não faz sentido fazer a avaliação do risco geológico ou comercial (este último não coberto neste trabalho).

O campo de Mero foi descoberto em 2010. É uma das maiores acumulações de petróleo *offshore* do mundo, com reservas que foram estimadas de 8 a 15 bilhões de boe Gaffney (2010), segundo dados da ANP. Encontra-se no bloco BM-S-11 que cobre uma área de 1.547,76km², na Bacia de Santos, a aproximadamente 230 km do litoral do Rio de Janeiro, próximo ao campo de Búzios. A acumulação se encontra em lâmina d'água que vai de 2000m à 3000m, e soterrada por aproximadamente 5.000m de sedimentos, rocha e sal. Libra é uma estrutura de tendência alongada de nordeste a sudoeste situada a leste do campo de Franco.

O poço pioneiro 2-ANP-2A-RJS foi perfurado pela Petrobras, em nome do governo federal brasileiro, para testar o intervalo do pré-sal. Parâmetros análogos de reservatório podem ser vistos entre Franco e Libra devido a suas semelhanças em ambientes deposicionais e proximidade um do outro. O poço encontrou o reservatório superior dentro da sequência sag altamente prolífica a uma profundidade de aproximadamente 5.370 m. A coluna de óleo é caracterizada por

um forte aumento no perfil de resistividade localizado dentro das coquinas syn-rift superiores. Além disso, CGGVeritas (2013) apontam um contato água-óleo que poderia ser observado a partir dos dados sísmicos, destacados por um aumento plano na impedância.

Ainda segundo CGGVeritas (2013), a descoberta difere do gigantesco campo Lula, no sul de Santos, devido à natureza e extensão dos hidrocarbonetos encontrados. No campo de Lula, os reservatórios preenchidos de hidrocarbonetos são compartimentados e presos dentro de selos intra-formacionais, enquanto que em Libra existe uma coluna de óleo razoavelmente contínua, abrangendo pelo menos duas mega-sequências. Essas observações destacam as diferenças entre a estratigrafia e as propriedades do reservatório na bacia.

Os dados utilizados aqui são advindos única e exclusivamente de material público, adquirido da literatura e sempre referenciado. Atualmente há uma extensiva quantidade de dados disponibilizados sobre o prospecto. Podemos citar como exemplo a apresentação de Petersohn and Abelha (2013) na rodada de partilha 1 e o relatório oferecido por Gaffney (2010) encomendado pela ANP, uma fonte muito rica de informações publicadas para avaliação de 10 prospectos/acumulações do pré-sal. Ainda de acordo com Gaffney (2010), o campo deve contar com reservas que podem variar de 3.7Bbbl à 15 Bbbl com valor mais provável de 7.9 Bbbl. Utilizaremos esta estimativa como base, nesta seção.

O mapa do topo do reservatório sag (coincidente com a base do sal) em profundidade, pode ser verificada na Figura 4.4. O poço pioneiro 2-ANP-2A-RJS se encontra na culminância oeste da estrutura.

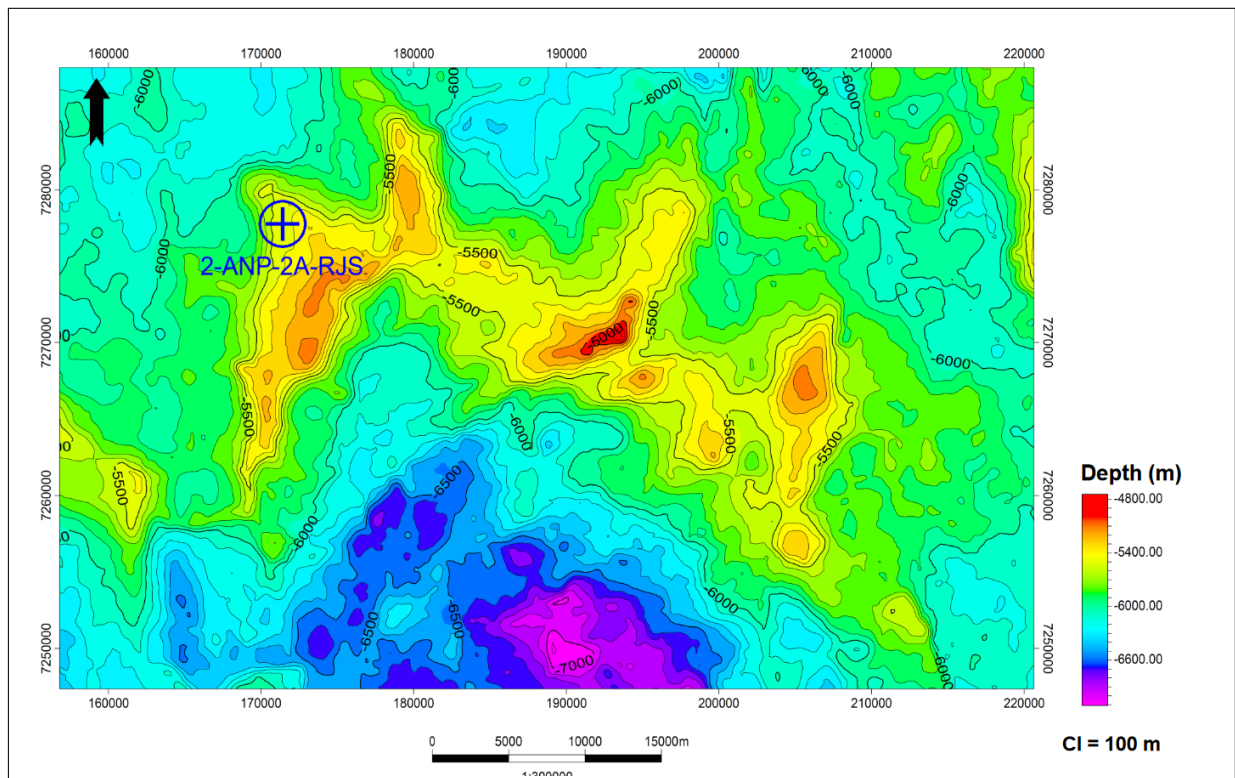


Figura 4.4: Estrutura da base sal, topo do reservatório, para o prospecto Libra (atual campo de Mero). Fonte: (Petersohn and Abelha, 2013).

As seções sísmicas que passam pela estrutura (e.g. Figura 4.5) indicam uma configuração altamente heterogênea, com possíveis altos do embasamento passando para níveis acima do contato e com variações importantes de espessura das unidades estratigráficas investigadas. Note que, conforme já mencionado, a sísmica indica também a presença de um possível *flat spot* o que reduziria significativamente as chances da estrutura estar parcialmente preenchida. Entretanto, como não temos informações que suportem esta interpretação não a usaremos.

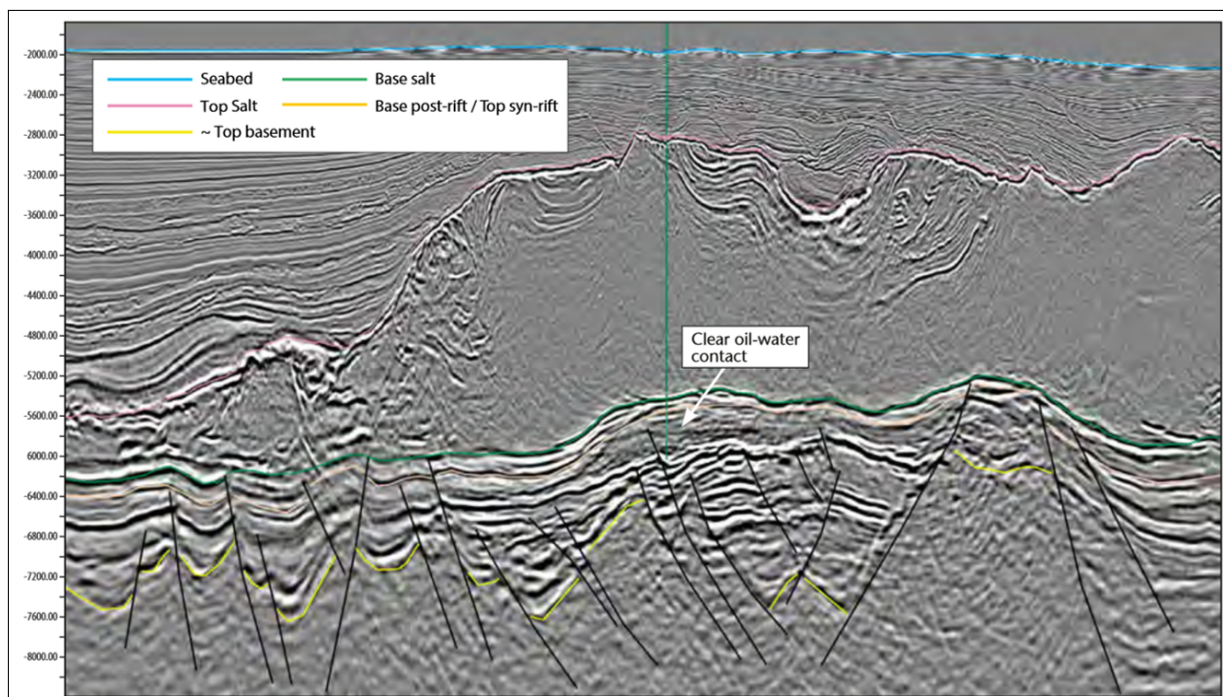


Figura 4.5: Seção sísmica passando pela estrutura de Libra (CGGVeritas, 2013). Note que a figura faz referência a um possível *flat spot* coerente com o que já foi publicado sobre o campo.

Como temos o mapa estrutural, o poço 2-ANP-2A-RJS determinando o contato óleo água, podemos inferir que os valores de área divulgados de aproximadamente 500km^2 é coerente com o *spill point* que se encontra aproximadamente em -5700m TVD . O topo do reservatório se encontra a -4750m e portanto o *gross pay* máximo é de aproximadamente 950m conforme divulgado por Petersohn and Abelha (2013) variarão em função de mudanças do modelo de velocidades. Dessa forma, consideramos uma pequena dispersão no parâmetro área.

Os valores de porosidade extraídos dos poços para este intervalo variam de 0.05 à 0.2. Utilizaremos esta distribuição para representar as possíveis médias do campo não por sua representatividade mas pelos valores coerentes a uma estrutura com carbonatos que podem ser altamente heterogêneo.

O fator de recuperação médio estimado está entre 0.08 e 0.35. Isto reflete as baixas porosidades médias, que terão como consequência altos valores de saturação irreduzível a complexidade dos reservatórios carbonáticos. Os reservatórios podem ser altamente conectados, por fraturas por exemplo, ou pobremente conectados, com porosidade *vuggy* isoladas disseminadas por todo o reservatório.

A partir de Gaffney (2010), que fez uma análise de 10 campos e oportunidades do pré-sal brasileiro, foi estimado um valor deB_o que varia de 1.25 à 1.45 o que representaria um óleo leve de 27 graus API com RGO que varia de $410\text{-}450\text{ m}^3/\text{m}^3$. Infelizmente a estrutura conta com um alto teor de CO_2 (44%), Gaffney (2010).

As variações de *net pay* e *trap fill*, devem refletir os riscos de encontrar fácies de baixa qualidade, principalmente nas bordas da estrutura e compartimentos subpreenchidos nas posições distantes do poço perfurado naquela altura. Dessa forma foram estimados valores de *net-to-gross* que variam de 10% à 40% e de *trap fill* de 80% à 90%.

Faremos agora a estimativa de volumes inicialmente sem os parâmetros de correlação e depois com as correlações de maneira a avaliar o impacto desta métrica nos volumes.

A Tabela 4.1 contém os valores utilizados na modelagem do prospecto que traduz os parâmetro já discutidos.

Com base nestes parâmetros podemos calcular os volumes estimados para o prospecto Libra

Parâmetro	P90	P10	unidade
Área	400	480	(km^2)
<i>Gross Pay</i>	100	200	(m)
<i>Trap Fill</i>	0.8	0.9	0-1 (%)
<i>Net Pay</i>	0.5	0.8	(m)
Porosidade (ϕ)	0.07	0.14	0-1 (%)
Saturação de Hidrocarbonetos (S_{hc})	0.7	0.9	0-1 (%)
Fator de Formação de óleo (B_o)	1.25	1.45	0-1 (%)
Razão Gás Óleo (RGO)	410	450	m^3/m^3
Fator de Recuperação de Óleo (FrO)	0.15	0.35	0-1 (%)
Fator de Recuperação de GA ($FrGA$)	0.3	0.6	0-1 (%)

Tabela 4.1: Parâmetros de entrada para modelar prospecto Libra.

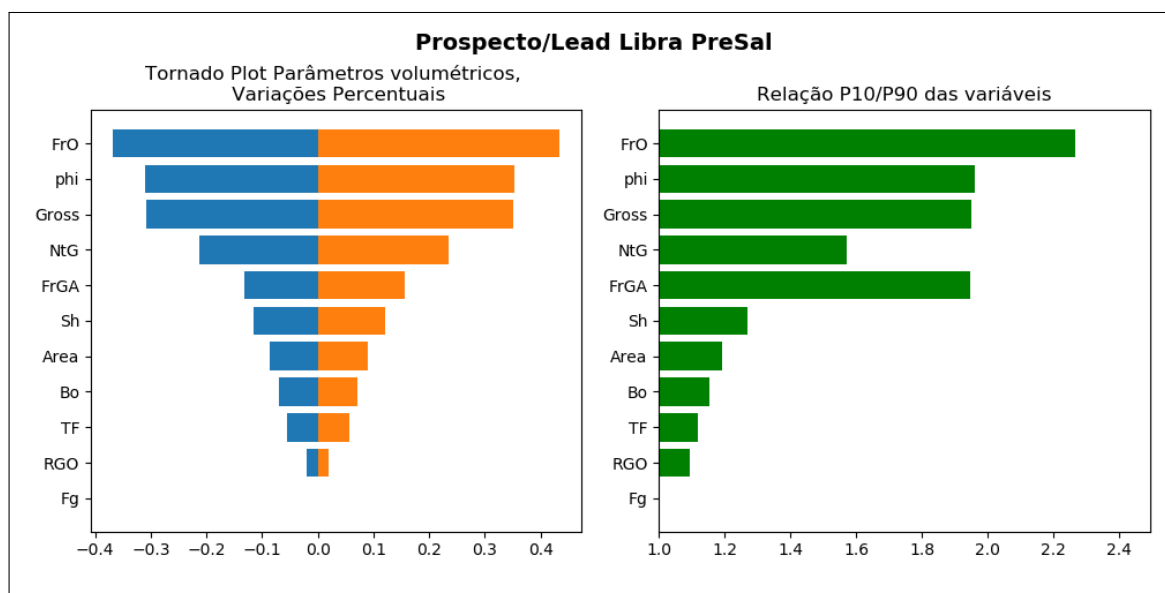


Figura 4.6: Tornado plot com os parâmetros do prospecto Libra.

conforme ilustrado na forma da *cdf* (*cumulative density function*) heurística na Figura ??.

Note que os valores são expressivos, e estão alinhados com as diversas avaliações já publicadas na literatura sobre esta acumulação. Vamos avaliar agora quais parâmetros tem maior impacto na avaliação econômica do prospecto (Figura 4.6).

Com base nos parâmetros utilizados, ao contrário do que acontece recorrentemente na avaliação de oportunidades, o parâmetro área não é o maior determinante na percepção de incerteza quanto ao volume de hidrocarbonetos. Os principais fatores aqui são o fator de recuperação de óleo e a porosidade média. Estes parâmetros refletem a incerteza relacionada à complexidade do reservatório e a extensão das propriedades verificadas no poço se replicarem ao longo de toda a área da acumulação.

Vamos agora explorar duas correlações, a relação positiva entre área e *gross pay* médio (conforme discutido no Capítulo 3) e a correlação negativa de *gross pay* médio e *net-to-gross* (NTG) médio. Com estas relações queremos determinar que, quanto maior a área maior a espessura e quanto maior a espessura total mais fácies ruins são adicionadas ao volume cubado (reduzindo o NTG). Utilizaremos nesta primeira abordagem a correlação com estrutura gaussiana e valores absolutos de correlação de 0.8, indicando que existe correlação entre os parâmetros Figura 4.7.

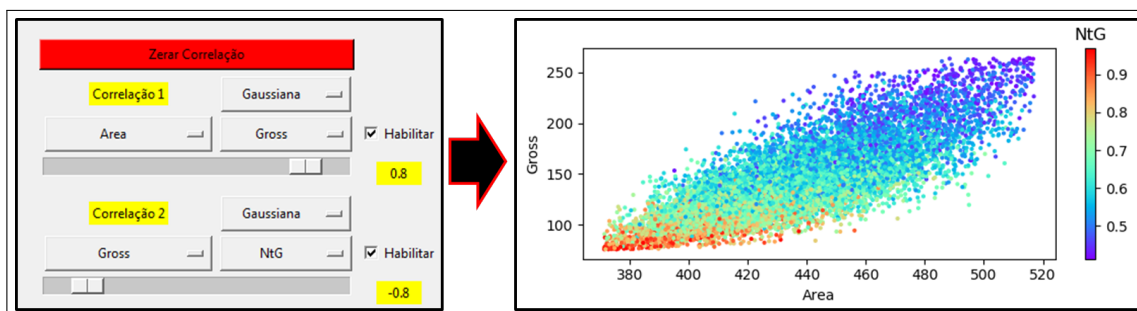


Figura 4.7: Parâmetros de correlação aplicados ao prospecto Libra, positiva entre área e *gross pay* e negativa entre *gross pay* médio e NTG médio. Note que em um mesmo *crossplot* é possível visualizar os dois pares de correlação uma vez que o *gross pay* está presente em ambos os pares de variáveis.

É interessante notar também que como o *gross pay* tem um vínculo tanto com NTG quanto com área, criamos um vínculo indireto de Área com o NTG¹ (Figura 4.8).

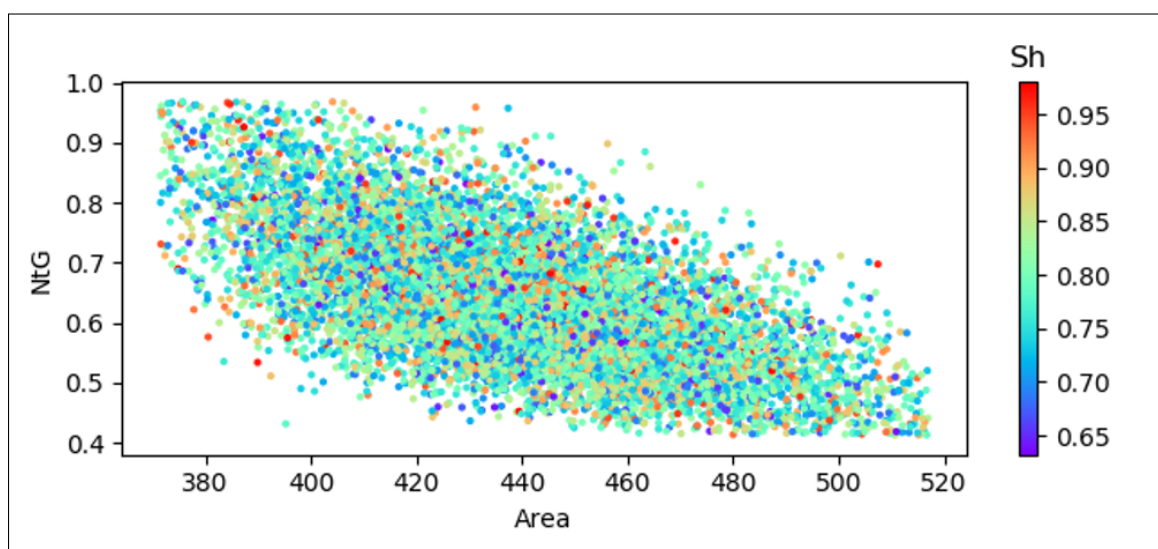


Figura 4.8: Correlação gerada indiretamente entre Área e NTG para o prospecto Libra a partir da correlação positiva entre área e *gross pay* e negativa entre *gross pay* médio e NTG médio. As cores representam o volume de argila (V. Shale) com comportamento independente.

Note que temos uma correlação geral negativa (aumento do P90 e diminuição da média e P10), isso significa que a correlação negativa entre área e NTG atuou de maneira mais efetiva do que a correlação entre área e *gross pay*. Este comportamento era esperado uma vez que área apresentou menor dispersão do que o NTG, (veja a Figura 4.6). A tradução desta relação pode ser visualizada no gráfico de impacto da correlação no volume (Figura 4.9)

Observe na Figura 4.9 ao fato de que, se a anticorrelação entre *gross pay* e NTG não tivesse sido aplicada, teríamos um acréscimo de até 4% ao caso base e 20% no volume p10 estimado. Vamos agora visualizar os pares de variáveis correlacionadas geradas neste exercício em função dos volumes recuperáveis calculados (Figura 4.10).

Note que apesar da correlação entre NTG e GP dominar a correlação entre área e GP em termos de distribuição de volumes, os maiores valores de área e GP ainda concentram os maiores volumes calculados.

¹Este exemplo é semelhante ao que discutimos sobre a relação entre tubarão e venda de sorvetes.

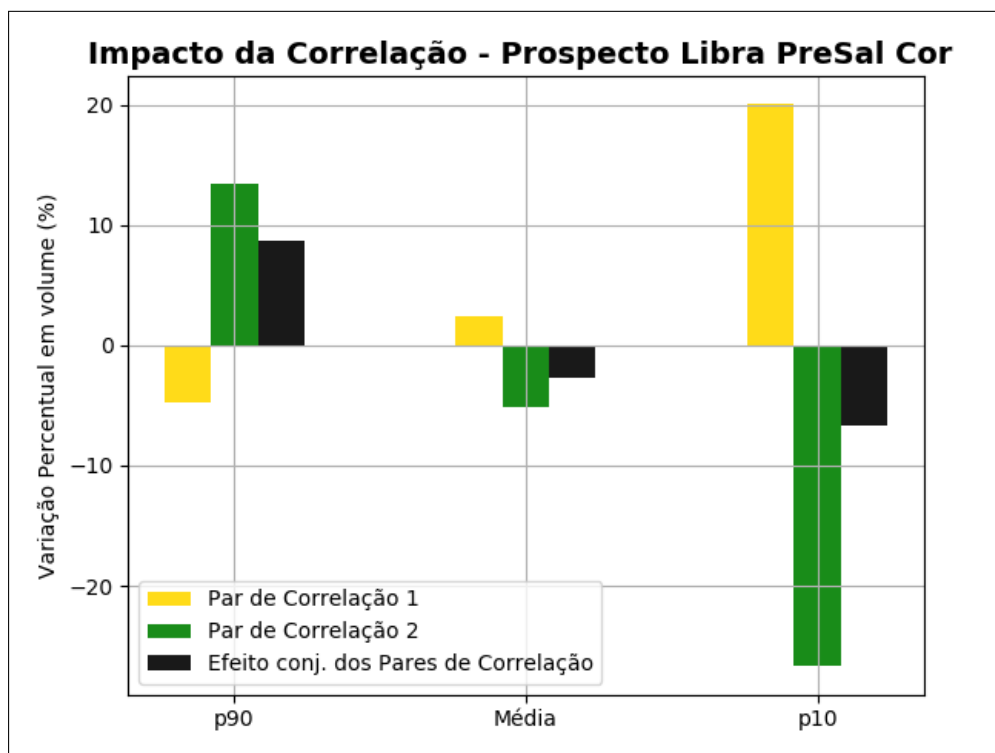


Figura 4.9: Gráfico com o impacto das correlações entre área e *gross pay* médio (gaussiana positiva : +0.8) e NTG vs *gross pay* médio, (gaussiana negativa : -0.8).

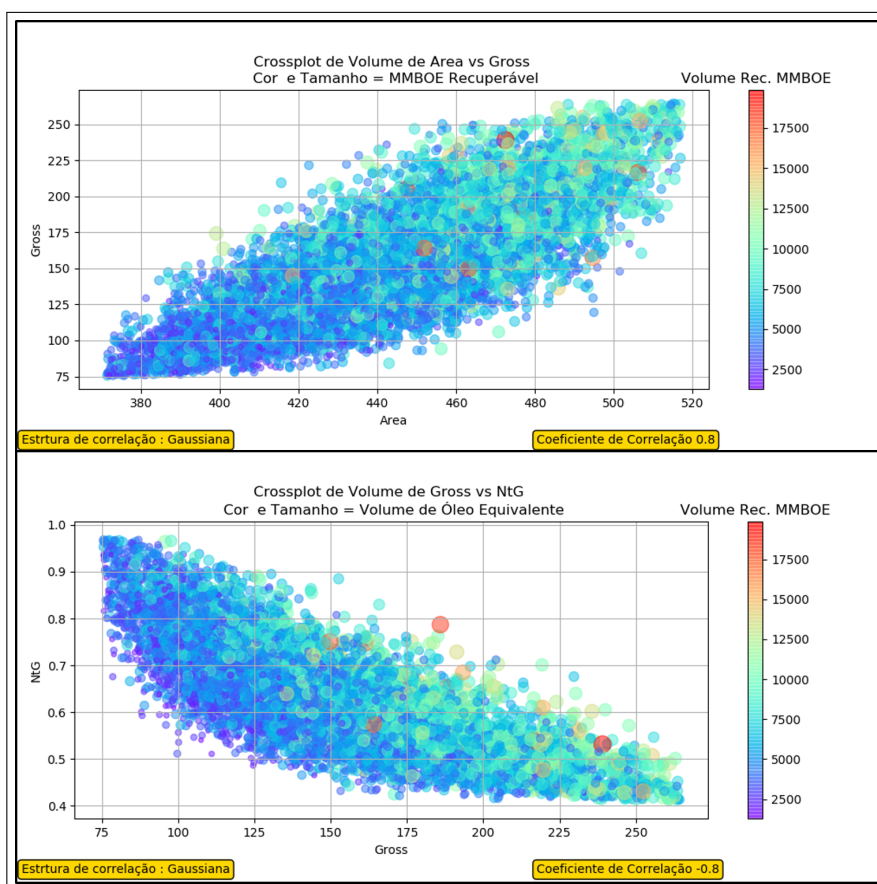


Figura 4.10: *Crossplots* das correlações entre área e *gross pay* (gaussiana positiva : +0.8) e NTG vs *gross pay*, (gaussiana negativa : -0.8). Veja em ambos os gráficos os maiores valores de volume se concentram nos maiores valores de área e GP

Ainda utilizando o mesmos valores de correlação (aplicando apenas o par área vs GP) e mantendo os parâmetros volumétricos, testamos as 6 estruturas de correlação apresentadas neste trabalho (Gaussiana, T-student, Gumbel, Copula, Frank e GSSC). Conforme esperado, todas as estruturas apresentam um aumento dos volumes com o aumento de área e espessura.

Para estas estruturas de correlação temos diferentes volumes (Figura 4.11).

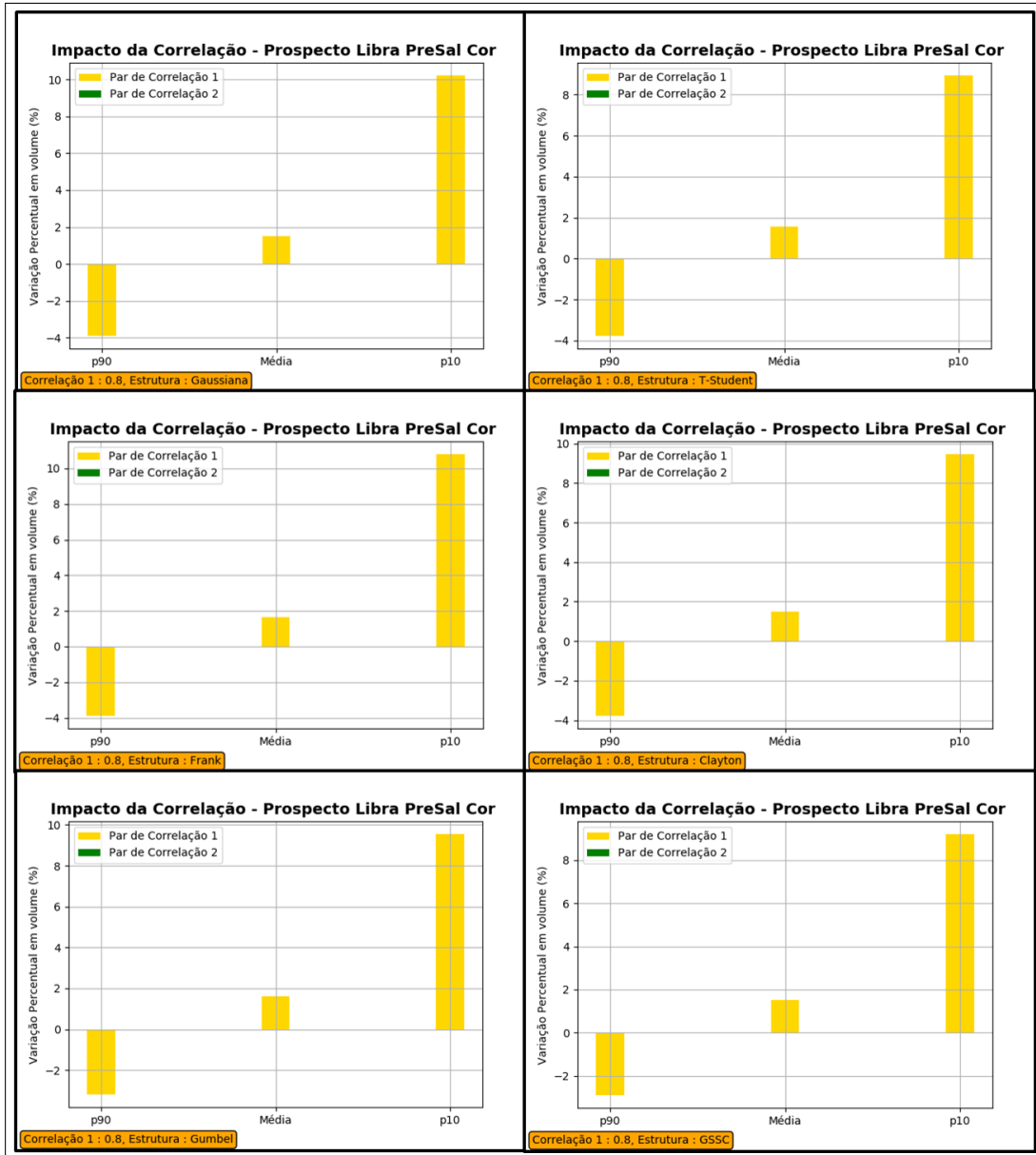


Figura 4.11: Verifique os impactos das diferentes estruturas de correlação para um valor de correlação Pearson de 0.8. São exibidas aqui a estrutura de correlação gaussiana (superior esquerda), T-student (superior direita), Frank (centro esquerda), Clayton (centro Direita), Gumbel (inferior direita), GSSC (inferior esquerda)

Note que em todas as estruturas de correlação as variações são significativas (alteram a média em mais de 1% e os valores de P10 em mais de 6%). Este efeito poderia ser ainda ampliado, contudo devido à pequena razão P10/P90 de área o limita as variações em função dos valores médios. A Tabela 4.2 resume os resultados obtidos.

Estrutura	P90	Média	P10
Gaussiana	9699	19229	31118
Gumbel	9741	19244	31379
Clayton	9684	19219	31237
Frank	9697	19250	31375
T-Student	9723	19234	31197
GSSC	9764	19249	31282

Tabela 4.2: Volumes de óleo equivalente *in place* (em MMBoe) calculados para os diferentes cenários.

Verifique que dentre as cópulas utilizadas, conforme esperado pelos experimentos no Capítulo 3, a estrutura de correlação de Gumbel apresenta valores mais otimistas para o cenário P10, P90 e média. Para o cenário P90 a estrutura de Clayton juntamente à estrutura de Frank. Note que o método numérico GSSC, apresentou P90 inferior à cópula de Clayton (a mais conservadora).

Na Tabela 4.2. Perceba que as variações entre as simulações com correlação e independentes são expressivas mas entre as cópulas é bastante suave. Isto provavelmente ocorre por que o parâmetro área tem razão P10/P90 relativamente baixa e portanto sua variação tem impacto menor sobre os volumes calculados.

4.2 Avaliação de Prospecto no Dataset 3D no Golfo do México

Faremos agora o exercício de identificar uma oportunidade e fazer a avaliação, assim como fazemos em projeto típico de interpretação durante a avaliação de oportunidade em um *dataroom*² para uma área pouco conhecida. Aplicamos esta metodologia a um prospecto chamado de prospecto Alfa. A oportunidade está localizada no Golfo do México Profundo - Setor Americano Figura 4.12.

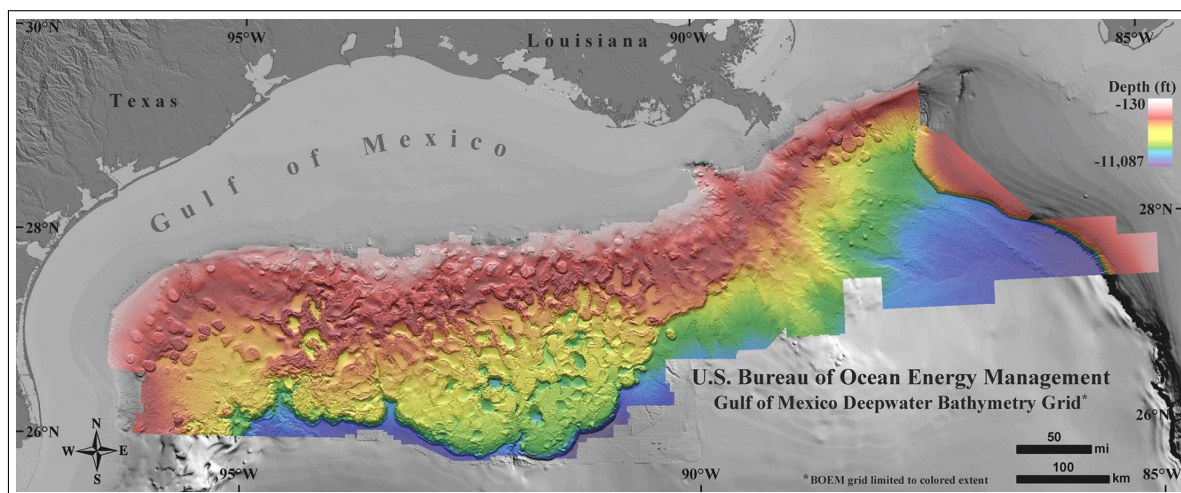


Figura 4.12: Batimetria em cores evidenciando a localização da Bacia do Golfo do México profundo. Fonte : BOEM (2017).

Geologia Regional

A origem do Golfo do México é relacionada a separação do supercontinente Pangea e a evolução do Atlântico e região do proto-Caribe. A evolução do Golfo do México pode ser dividida em 2 estágios primários: 1) Estágio *Syn rift*, estiramento NW-SE da crosta continental conduzindo a uma deposição inicial durante o Triássico ao Caloviano; e 2) Estágio *drift*, nova crosta oceânica formada, seguida por uma migração rotacional da placa Yucatan para sua posição atual (Pindell, 2010).

Os estágios Mesozoicos de evolução tectônica do Golfo do México incluem o típico *Syn rift*, *drift*, e desenvolvimento de margem passiva de um cenário de margem continental passiva. Durante o Cenozóico, a subsidência da bacia foi resultado do sobrecarregamento da crosta com espessas cunhas sedimentares progradantes. Elas influenciaram a mobilização do sal Jurássico da Fm. Louann.

A quebra do Pangea iniciou no Triássico Tardio continuando até o Jurássico Médio. A sedimentação relacionada ao rift e atividade ígnea foram associadas com este evento de extensão crustal intracontinental. Bacias rift com geometria de Graben e semi-Graben poderiam ter sido preenchidas com sedimentos lacustres e fluviais. A transição do estágio *syn-rift* para *drift* marcou o começo da origem do Golfo do México.

Expansão do fundo do mar e formação da crosta oceânica ocorreu enquanto o Pangea continuou a se separar ao longo do rift da zona Oeste Central. O sal Regional (Fm. Louann) foi depositado do Caloviano ao Oxfordiano Recente. Esse sal espesso, quando remobilizado mais tarde, desempenhou um papel crítico na maturação, migração, e traqueamento de hidrocarbonos na Bacia DWGoM (*Deep Water Gulf of Mexico*).

A idade em que ocorre o final da expansão da crosta oceânica varia entre autores na literatura. (Winker and Buffler, 1988; Marton and Buffler, 1994; Bird et al., 2005) defendem que a expansão

²Processo de avaliação de oportunidade exploratória oferecida por uma empresa/governo à outra.

teria terminado durante o Berriasiano e o Oxfordiano Médio (Salvador (1991)). Entretanto segundo J.L.Pindell and Kennan (2010) o final da expansão chegaria até o Valangeniano.

A bacia começou a sofrer subsidência térmica devido à redução da atividade mantélica e carga sedimentar desde o Cretáceo Recente. Uma sucessão de folhelhos ricos em querogênio e siltitos seriam depositados sobre o Sal Louann.

Plataformas carbonáticas espessas foram desenvolvidas durante o Cretáceo ao longo das margens do Golfo, se estendendo do México até a Flórida. A maior queda do nível do mar ocorreu durante o Cretáceo Médio (Middle Cenomanian) e é registrado por uma descontinuidade regional ao longo das margens da Bacia.

A área experimentou um grande período de agradação durante o Cenozóico. Regionalmente, grandes depocentros siliciclásticos foram estabelecidos no noroeste do Golfo, refletindo suprimento e drenagem dos sedimentos provenientes da Orogenia Laramidiana Figura 4.13. Na parte central e leste da bacia, um influxo de sedimentos terrígenos estava relacionado ao delta do Rio Mississippi.

A grande sobrecarga dos sedimentos do Cenozóico causou deformação do sal de sua forma autóctone original para forma alóctone em diversos níveis formando canopies e linguas de sal.

Sedimentos batiais de mesma idade consistindo de sistemas turbidíticos foram depositados em uma série de *minibasins* sobre o topo do sal autóctone e entre os corpos de sal alóctones. O Sal formou complexas paredes, domos, e nappes coalescentes que por vezes impactam a arquitetura e distribuição das areias submarinas depositadas no período Terciário, atuaram também como selo lateral e vertical em diversas acumulações (e.g. Campo de Mars, Tubular Bells etc).

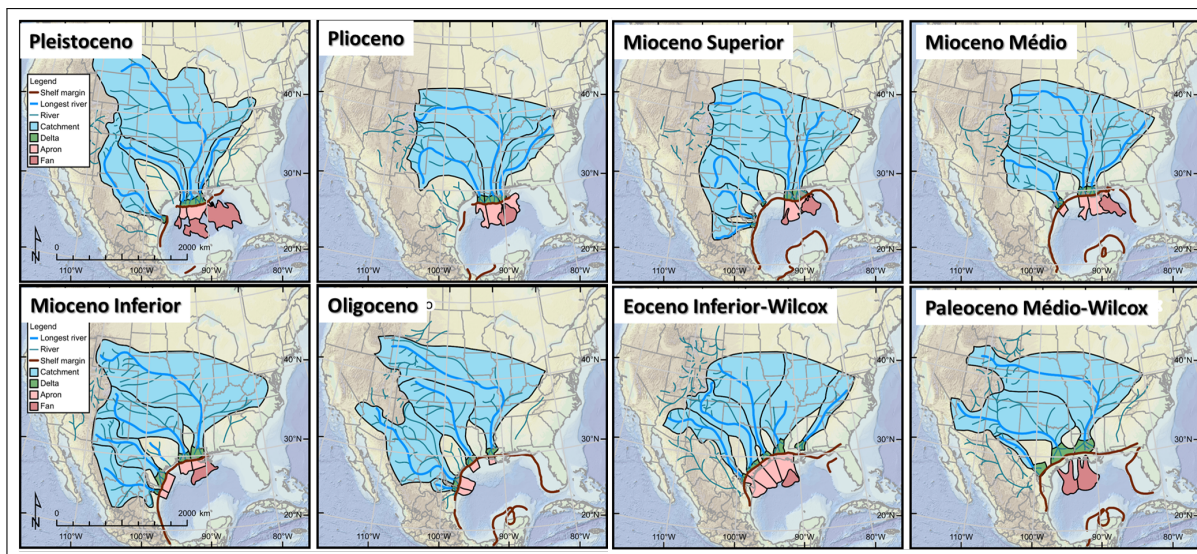


Figura 4.13: Mapas de drenagem expressando a grande área de drenagem que foi depositada no GoM (*Gulf of Mexico*) ao longo do Terciário-Neógeno. Fonte: Galloway (2009).

Dada a maturidade exploratória do GoM (*Gulf of Mexico*), foram feitas diversas descobertas em múltiplos intervalos reservatório, incluindo arenitos eólicos do Jurássico Superior, turbiditos do Terciário e Pleistoceno. Um total de 17 *plays* foram estabelecidos na Bacia. Entre esses *plays* o Terciário Inferior, o Mioceno Superior, Mioceno Inferior, Plioceno, Mioceno Médio são os *plays* mais prolíficos até o momento. Cada um contém reservas entre 1 e 7 Bboe. Os *plays* estruturais contêm a maior parte das reservas, e *plays* estritamente estratigráficos e normalmente mais arriscados contribuem pouco para as reservas provadas no GoM. Em termos de idade de reservatórios o Mioceno contém a maior quantidade de campos e reservas (Figura 4.14).

Os reservatórios mais significativos são de areias do Mioceno, seguido por areias do Plioceno e Terciário Inferior. Reservatórios turbidíticos em leques submarinos são amplamente distribuídos e possuem boa qualidade e espessura (*net pay* >20m). A arquitetura das areias dentro dos turbiditos varia de *levees/overbank* a corpos tabulares ou canais amalgamados. Corpos tabulares

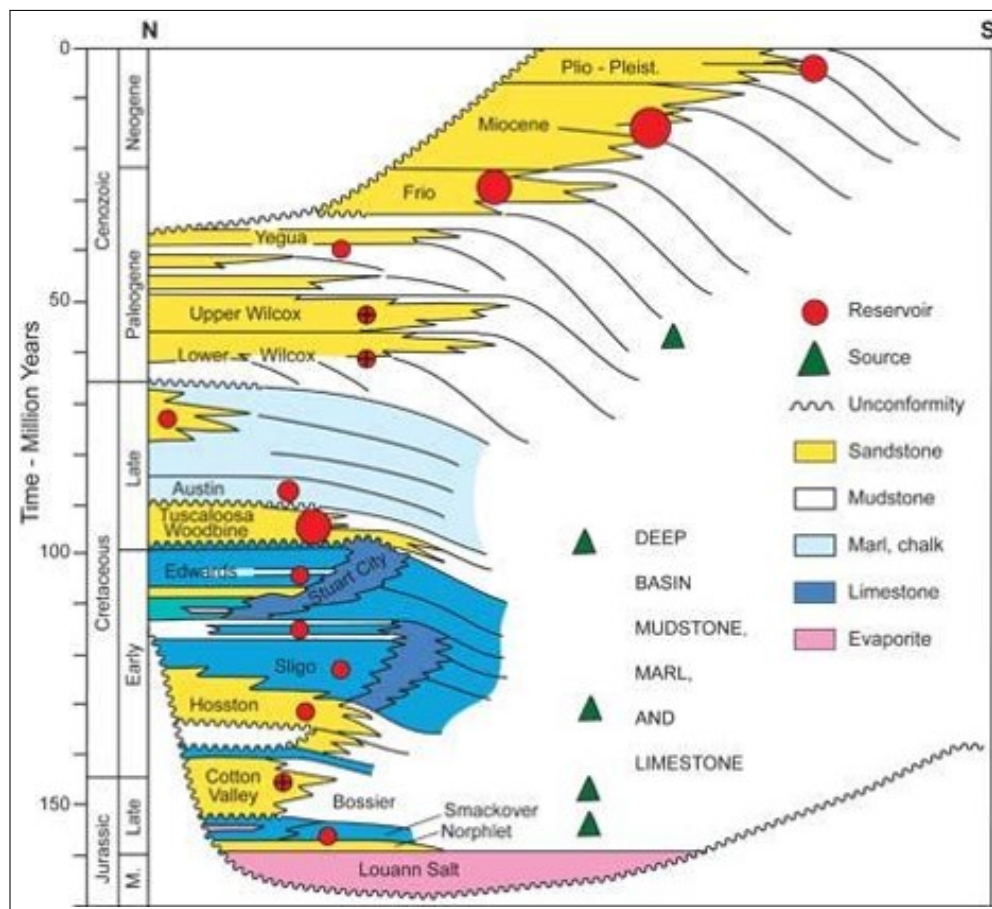


Figura 4.14: Carta Estratigráfica do Golfo do México com os principais reservatórios e geradoras testados. Fonte: Galloway (2009).

e amalgamados fazem o melhor reservatório em termos de produtividade. Descobertas recentes nos arenitos do Terciário Inferior, correlacionadas ao Grupo Wilcox, confirmaram a existência de um *trend* Terciário Inferior em águas ultraprofundas que foi estimado em 2,8 BBo de recuperáveis. Os arenitos eólicos do Jurássico Superior estão apenas distribuídos na parte leste da Bacia.

As principais rochas geradoras do GOM profundo foram depositadas durante o período Jurássico Tardio. A ampla ocorrência de acumulações ao longo da bacia correlacionadas a esta geradora, sustentam a hipótese de que o risco de geração na bacia é baixíssimo do ponto de vista regional, enquanto avaliações locais devem ser calibradas por estudos de modelagem de sistema petrolífero e avaliação de fácies sísmicas que comprovem de fato a presença da geradora.

O GOM é uma bacia dominada por sedimentos finos, além do sal. Por isso podemos afirmar que selos são abundantemente intercalados e truncados com fácies potencial reservatório. As complexas arquiteturas deposicionais e estruturais se combinam para criar um conjunto diversificado de tipos de armadilhas. Algumas das trapas incluem as estratigráficas simples, fechamentos quaquaversais, fechamentos contra domos de sal e anticlinais relacionados a estruturas de casco de tartarugas, falhas de crescimento e anticlinais de *rollover*. A interação entre a deformação sin-deposicional e a deposição aumentou o empilhamento vertical dos reservatórios nas trapas estruturais.

4.2.1 Interpretação do Prospecto Alfa

A interpretação sísmica foi feita em dados 3D PSTM no domínio do tempo adquiridos na década de 1990 e tornados públicos após 25 anos de aquisição. Os dados foram baixados do site <https://walrus.wr.usgs.gov/namss/>.

Não foram utilizadas as localizações reais dos dados de maneira a não agregar nenhum conteúdo

comercial ao problema proposto aqui. A região de estudo é caracterizada por uma grande sucessão sedimentar de idade predominantemente Miocênica. Este grande influxo sedimentar relacionado com a deposição do cone do Mississippi, deformou os corpos de sal verificados a sudoeste do levantamento.

Para esta rápida interpretação, foram mapeados 6 horizontes, que são:

- Fundo do Mar - Azul;
- Pleistoceno - Verde Escuro;
- Plioceno - Azul Turquesa;
- Mioceno Médio 100 - Amarelo;
- Mioceno Médio 200 - Vermelho;
- Oligoceno - Verde Musgo.

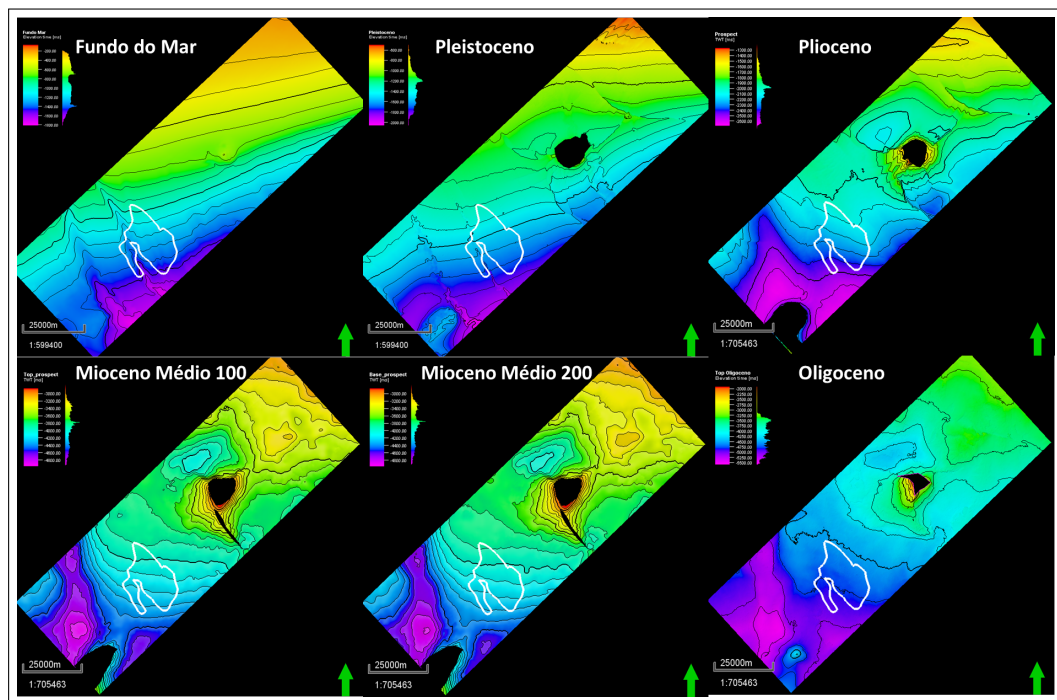


Figura 4.15: Mapas estruturais, em branco contorno p01 do prospecto Alfa.

O fundo do mar indica a ocorrência de muitos paleo-cânions e a deformação gerada pela ascensão do sal. O próximo horizonte mapeado foi o pleistoceno, que é caracterizado como base da seção refletiva. Ela é rica em fácies relacionadas a possíveis entradas de canais mais novos. Abaixo do horizonte do pleistoceno, temos uma seção pouco refletiva seguida de uma sequência de complexos deslizamentos de massas.

O plioceno é marcado como o topo da zona prospectiva. A partir deste nível estratigráfico, é possível verificar a ocorrência de diversos campos no Golfo do México. A seção do Plioceno e Mioceno de fato são as que concentram maior quantidade de acumulações, principalmente na porção central da bacia.

Na seção miocênica é possível verificar sucessões de fácies sísmicas pouco refletivas e fácies de canais. Entretanto, é possível verificar a presença de anomalias alongadas que motivaram a geração das superfícies estratigráficas do topo e base do reservatório nos horizontes definidos aqui como Mioceno Médio 100 e Mioceno Médio 200.

Por fim é possível verificar as fácies refletivas caracterizadas pelo horizonte verde na Figura 4.18, reconhecida como o topo da seção oligocênica da bacia. Sedimentos desta idade são normalmente caracterizados por margas e material orgânico pobres em fácies reservatórios.

O prospecto Alfa constitui um complexo turbidítico em trapa estratigráfica identificado com base em mapa de amplitudes RMS.

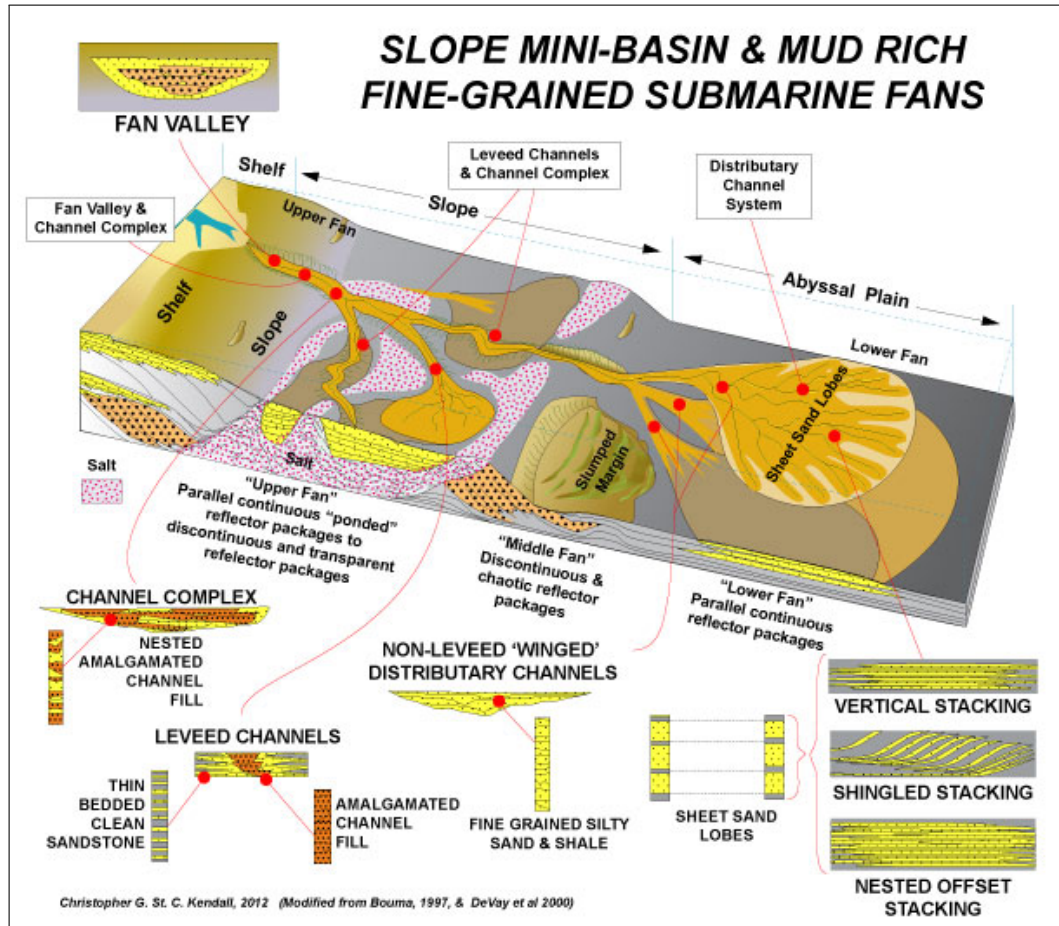


Figura 4.16: Modelo estratigráfico de deposição dos lobos e canais em ambiente de água profunda em bacias ricas em sal (com o GoM). Fonte : Buckles (2000).

Como o objetivo do trabalho é só avaliar o efeito da correlação nos parâmetros de entrada do cálculo volumétrico faremos uma abordagem *fast track*. Consideraremos o tempo como pseudo-profundidade e não serão feitas análises típicas do processo exploratório como análise estrutural-estratigráfica, calibrações com dados de poço.

Do ponto de vista do risco exploratórios consideraremos uma chance de geração de 70% devido à ocorrência de acumulações próximas à anomalia mas ausência de modelagens para a subbacia onde se encontra a oportunidade.

A migração foi definida em 70% de PS, uma vez que o prospecto se encontra em um nariz estrutural, mas não no alto propriamente dito. Prospectos estratigráficos normalmente tem menor poder de focalização quando não associados a feições estruturais próximas.

A geometria foi definida em 90%, pela qualidade do dado sísmico, mas ausência de modelagens de física de rochas.

Definiu-se o fator de chance do Reservatório em 90% pois temos indicativos de que o reservatório está presente, mas temos uma pequena dúvida sobre sua qualidade.

Os fatores sincronismo e retenção foram definidos em 70% devido ao fato do prospecto estratigráfico estar soterrado com seção de baixa refletividade, indicando possível presença de selo, mas não termos dados e modelagens que calibrem esta percepção.

Apresentaremos agora uma seção *strike* à bacia com e sem interpretação passando pelo prospecto.

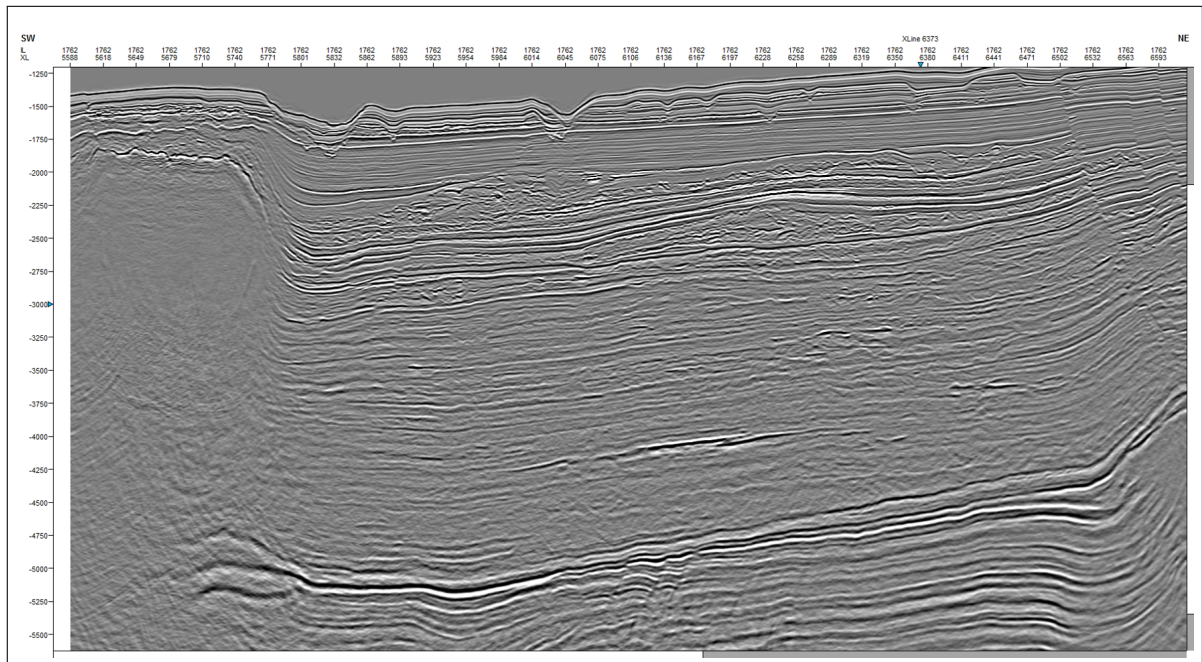


Figura 4.17: Seção sísmica *strike* sem interpretação ao longo do "Prospecto Alfa".

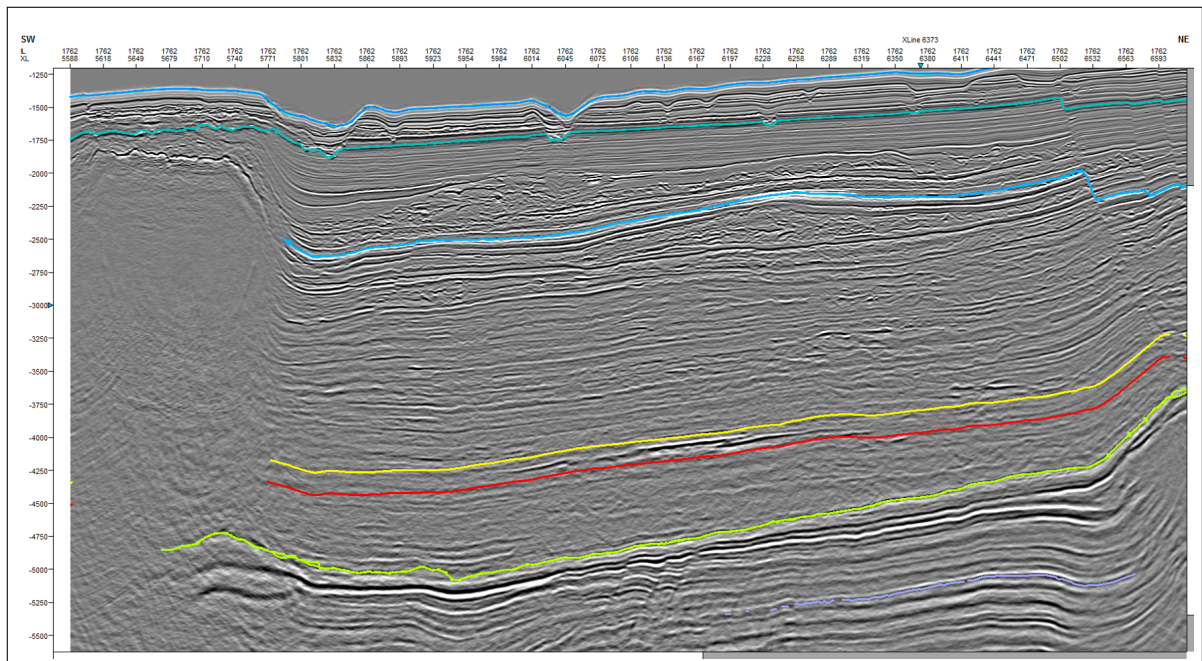


Figura 4.18: Seção sísmica *strike* interpretada, o horizonte em vermelho representa a base da seção alvo e em amarelo o topo da seção alvo do "Prospecto Alfa".

Foram extraídos mapas de amplitude RMS entre os dois horizontes de topo e base do reservatório. É possível ver claramente as feições de lobos e canais de direção norte-sul, conforme esperado para a área. As linhas pretas nas Figuras 4.19 e 4.20 indicam as curvas de contorno de profundidade.

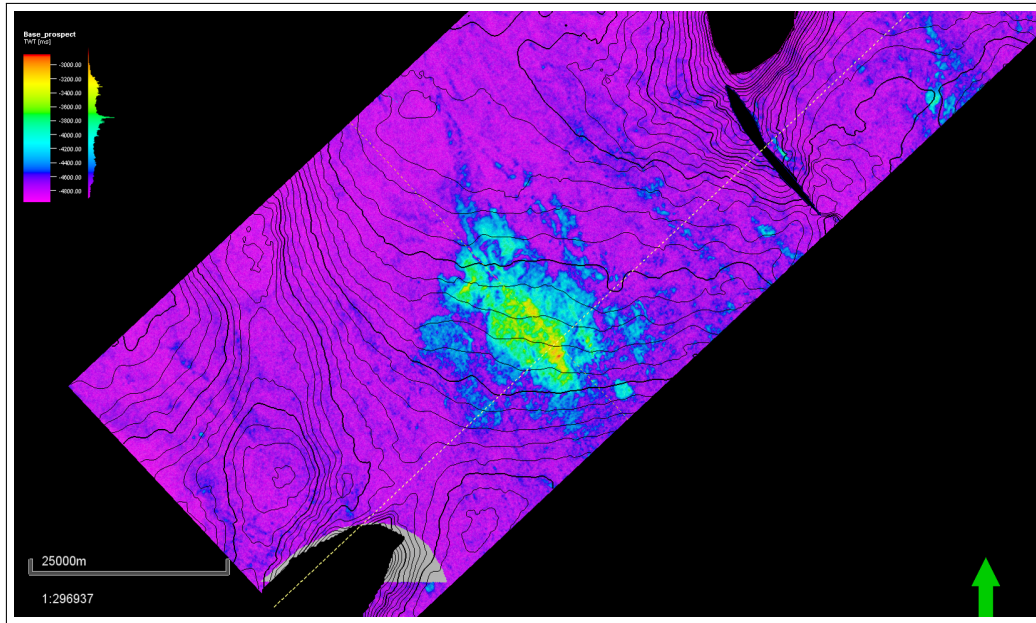


Figura 4.19: Mapa de amplitudes RMS extraído entre os horizontes do topo e base da seção alvo (aproximadamente 200ms).

Foram definidos os contornos de p90 (local com maior probabilidade de ocorrência de uma acumulação de hidrocarbonetos) com base nos valores mais altos de amplitude RMS. O intervalo p01 (máximo possível) foi definido no limite mais otimista de ocorrência de reservatório. Devido a ausência de dados de empilhamentos parciais não foram feitas análises mais detalhadas de amplitude como Análises de AVO (*Amplitude versus Offset*), que não as extrações de amplitude empilhadas já apresentadas.

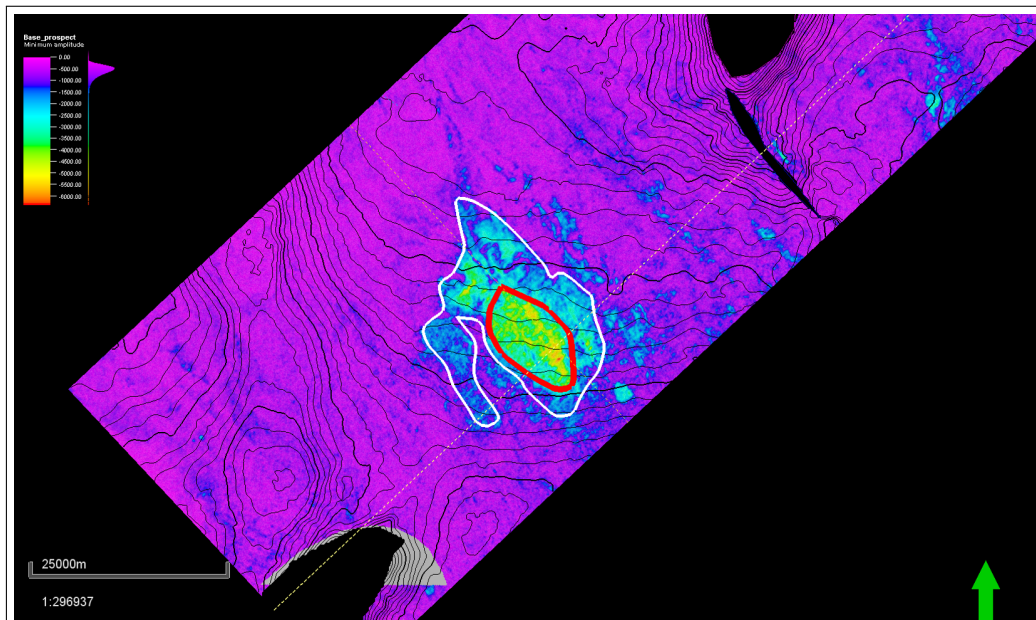


Figura 4.20: Mapa de amplitudes RMS extraído entre os horizontes do topo e base da seção alvo (aproximadamente 200ms), o contorno em branco representa a área de amplitudes p01 ($360km^2$) e em vermelho a área de amplitudes P90 ($86km^2$).

Para confirmar que o prospecto é estratigráfico exibimos o contorno das áreas definidas como p01 (polígono branco) e p90 (polígono vermelho) sobre o mapa estrutural do topo do reservatório

(Figura 4.21). Fizemos um rápido mapeamento apenas dos corpos turbidíticos e verificamos que, aparentemente, temos um sistema siliciclástico retrogradacional onde se nota uma série de pequenos canais seguida de leques amalgamados Figura 4.22.

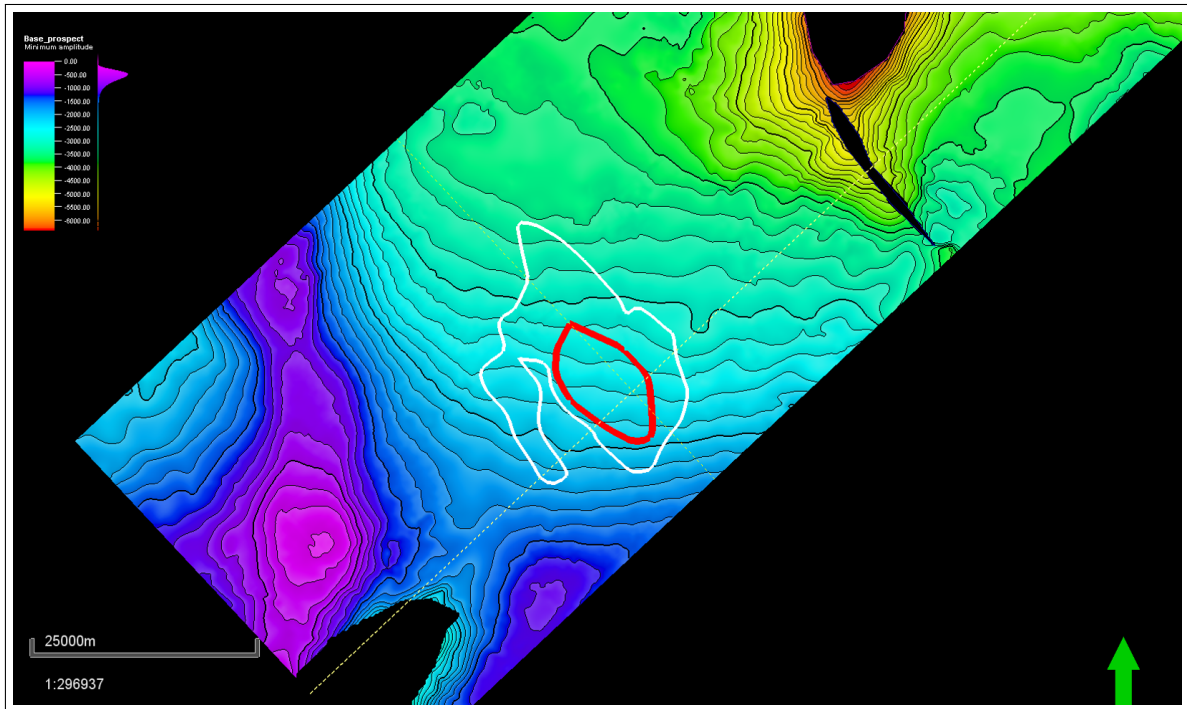


Figura 4.21: Mapa estrutural do topo da seção objetivo, o contorno em branco representa a área máxima de ocorrência de areias p01 (360km^2) e em vermelho a área de amplitudes P90 (86km^2).

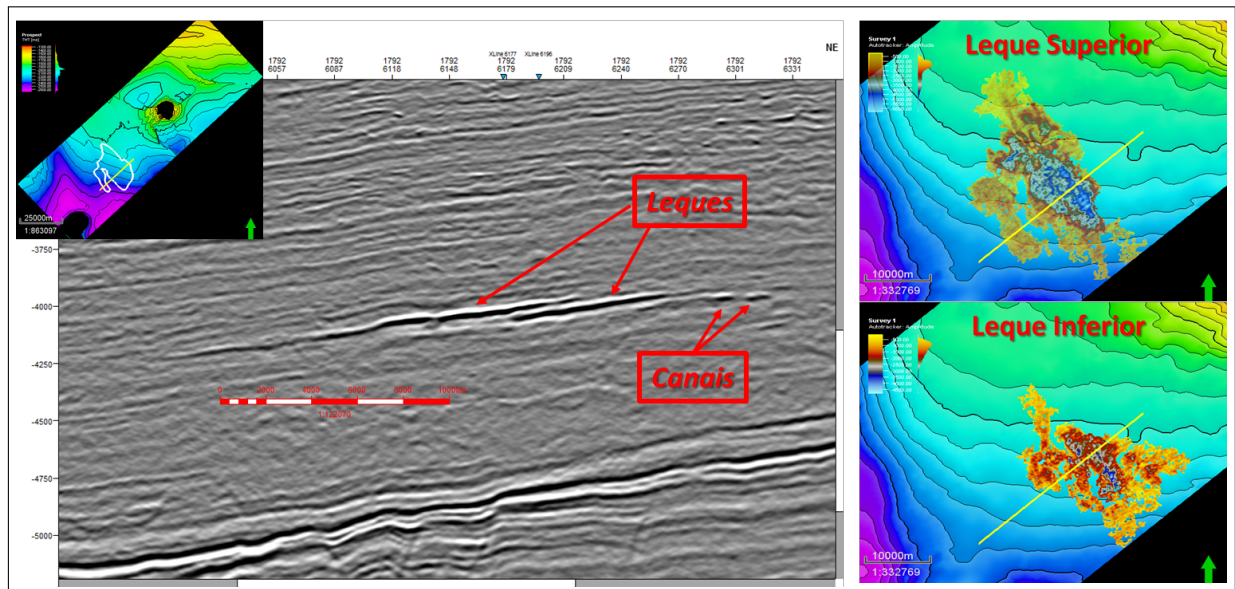


Figura 4.22: Seção sísmica com maior detalhe sobre o prospecto Alfa, note as feições de pequeno comprimento de onda típicas de canal e sobre elas feições alongadas típicas de leques.

Para os parâmetros de entrada do cálculo volumétrico as áreas e espessuras foram extraídas diretamente da sísmica e para os outros parâmetros utilizaremos valores típicos para este *play*. Muitos parâmetros foram extraídos dos *trends* regionais de Amado (2013) (apresentadas no capítulo 2) chegando à Tabela 4.3.

O primeiro passo foi calcular o valor de P10 de área com base no p01 e P90. Para isso podemos utilizar a calculadora de percentis ou montarmos um sistema de equações a partir do valor de z-normal de cada percentil:

$$P_{01} = e^{\mu+2.32635.\sigma} \quad (4.1)$$

$$P_{90} = e^{\mu+1.282\sigma} \quad (4.2)$$

Aplicando a propriedade dos logaritmos e resolvendo o sistema de equações acima obtemos os valores de média e desvio padrão da distribuição não truncada ($\mu = 4.96$ e $\sigma = 0.39$), utilizando novamente a equação 4.2 obtemos o valor de P10=237,8 .

Parâmetro	P90	P10	unidade
Área	86	237.8	(km^2)
Gross Pay	20	40	(m)
Trap Fill	0.8	0.9	0-1 (%)
Net to Gross	0.7	0.9	(m)
Porosidade (ϕ)	0.15	0.27	0-1 (%)
Saturação de Hidrocarbonetos (S_{hc})	0.7	0.85	0-1 (%)
Fator de Formação de óleo (B_o)	1.165	1.24	0-1 (%)
Razão Gás Óleo (RGO)	142	356	m^3/m^3
Fator de Recuperação (Fr)	0.20	0.40	0-1 (%)
Fator de Recuperação do GA ($FrGA$)	0.5	0.8	0-1 (%)

Tabela 4.3: Parâmetros de entrada do prospecto Alfa

Com os dados de entrada podemos gerar as distribuições truncadas em P99 e P01 a partir de 10 mil simulações (Figura 4.23).

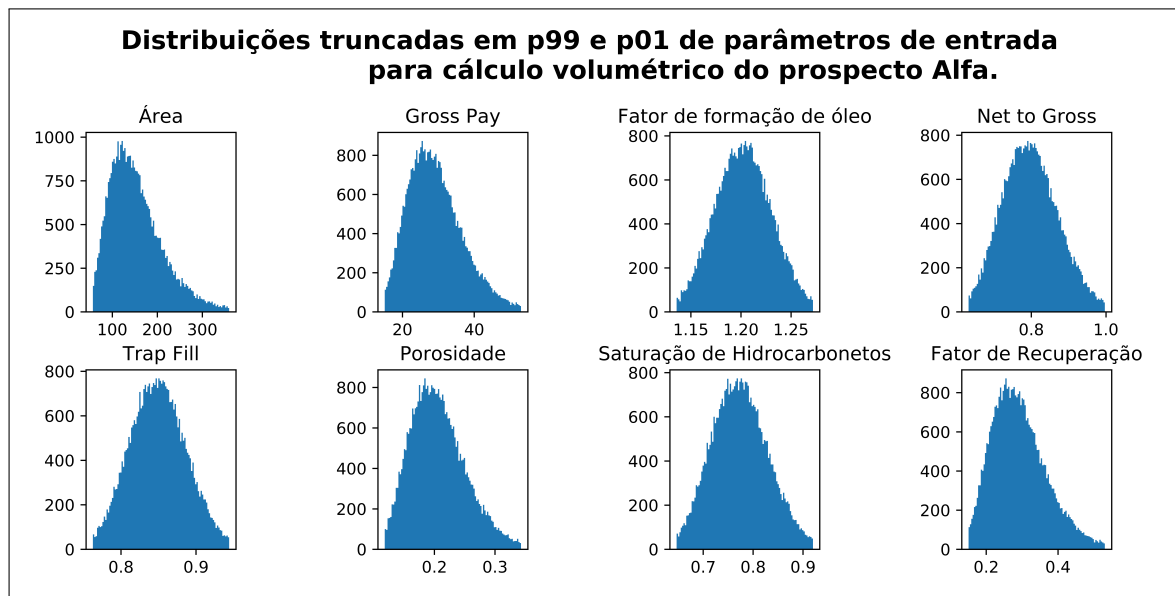


Figura 4.23: Distribuições de entrada utilizadas para o cálculo de distribuições de volumes geradas para o prospecto Alfa.

Na Figura 4.23 é possível verificar em algumas das distribuições como, *trap fill* (TF), fator de formação de óleo (B_o) e Saturação de hidrocarbonetos (S_{hc}), que a proximidade dos valores de P10 e P90 reduz sobremaneira a curtose da distribuição lognormal sendo a distribuição resultante

próxima da normal (conforme comentado no capítulo de Fundamentos Teóricos). Podemos assim preencher a tabela 1 com mais parâmetros das distribuições (p99, média e p01) conforme podemos ver na Tabela 4.4.

Parâmetro	p99	P90	media	P10	p01	unidade
Área	56.86	86	153.4	237.8	360	(km^2)
<i>Gross Pay</i>	15.08	20	29.24	40	53.04	(m)
<i>Net to Gross</i>	0.63	0.7	0.8	0.9	0.99	(m)
<i>Trap Fill</i>	0.76	0.8	0.85	0.9	0.94	0-1 (%)
Porosidade (ϕ)	0.157	0.15	0.21	0.27	0.285	0-1 (%)
Saturação de Hidrocarbonetos (S_{hc})	0.65	0.7	0.77	0.85	0.92	0-1 (%)
Fator de Formação de óleo (B_o)	1.136	1.165	1.202	1.24	1.272	0-1 (%)
Razão Gás Óleo (RGO)	238.16	97.7	238.16	356	517.5	m^3/m^3
Fator de Recup. de Óleo (Fr)	0.15	0.20	0.29	0.40	0.53	0-1 (%)
Fator de Recuperação do GA ($FrGA$)	0.413	0.5	0.6393	0.8	0.9498	0-1 (%)

Tabela 4.4: Parâmetros de entrada do prospecto Alfa. Tabela Estendida com base nas distribuições geradas.

A partir destes parâmetros obtivemos a distribuição de volumes *in place* e recuperáveis (Figura 4.24).

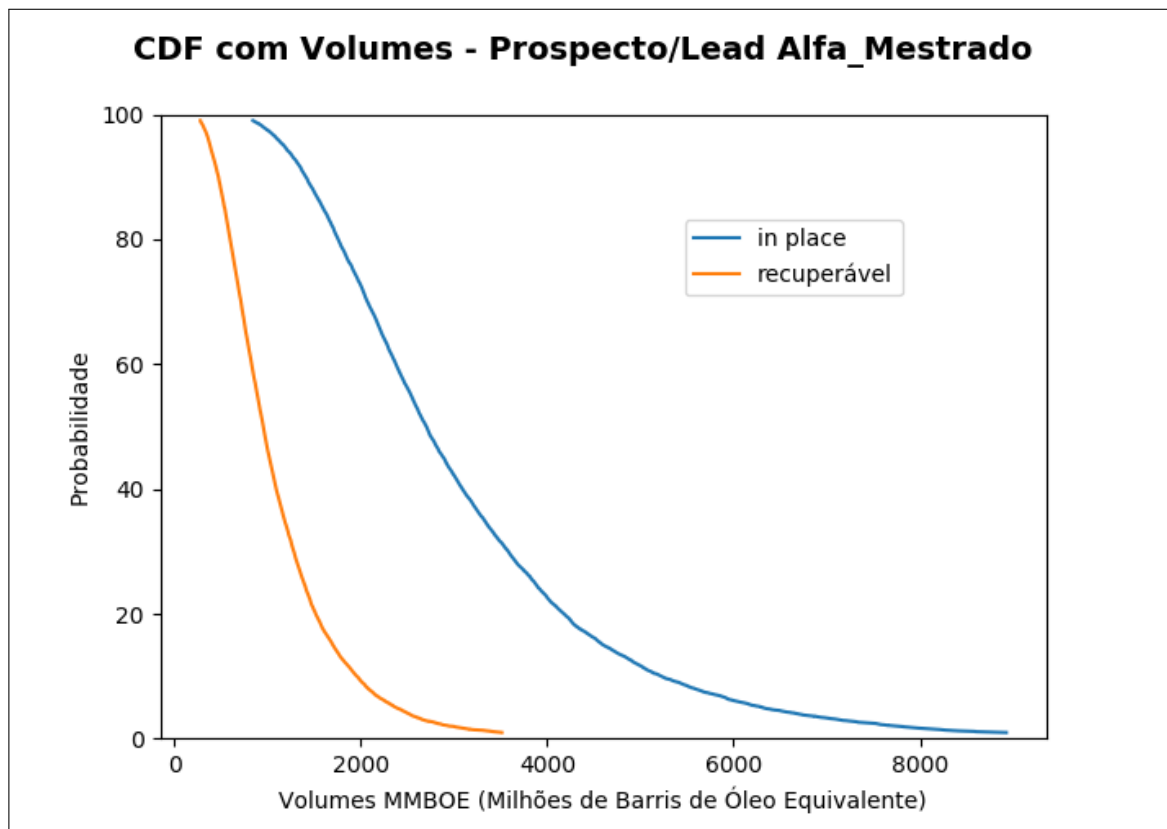


Figura 4.24: Volumes gerados para o prospecto Alfa. Em azul, curvas da *cdf* heurística de distribuições de volumes *in place* e em laranja a *cdf* heurística de volumes recuperáveis.

Estes volumes estão traduzidos na tabela da Figura 4.25. Em seguida faremos a avaliação de incerteza considerando os elementos independentemente organizados de maneira decrescente dos que mais influenciaram na variação de volume no Tornado plot (Figura 4.26). Neste gráfico

adicionamos também a razão P10/P90 que é uma medida de incerteza de cada variável.

Tabela com Volumes - Prospecto/Lead Alfa_Mestrado						
Volumes InPlace						
	P99	P90	Média	P10	P01	MI
Hidrocarbonetos(MMBOE)	842.916	1410.2537	3098.5674	5225.9586	8925.5753	602.67136
Óleo(MMBBL)	685.9235	1148.1434	2504.3464	4225.6957	7047.7002	487.09537
Gás Associado(ICF)	0.7036	1.2626	3.3363	6.0268	11.3692	0.64891
Gás Não Associado(ICF)	0.0	0.0	0.0	0.0	0.0	0.0
Condensado(MMBOE)	0.0	0.0	0.0	0.0	0.0	0.0

Volumes Recuperáveis						
	P99	P90	Média	P10	P01	MI
Hidrocarbonetos(MMBOE)	278.0593	472.4286	1114.3563	1958.4373	3512.7527	216.7423
Óleo(MMBBL)	171.4489	300.3338	734.2906	1296.1059	2391.3072	142.81952
Gás Associado(ICF)	0.4174	0.7738	2.1339	3.9369	7.5527	0.41504
Gás Não Associado(ICF)	0.0	0.0	0.0	0.0	0.0	0.0
Condensado(MMBOE)	0.0	0.0	0.0	0.0	0.0	0.0

Figura 4.25: Tabela com os volumes *in place* e recuperáveis do prospecto Alfa sem o uso de correlação. Esta tabela traduz os dados da Figura 4.24.

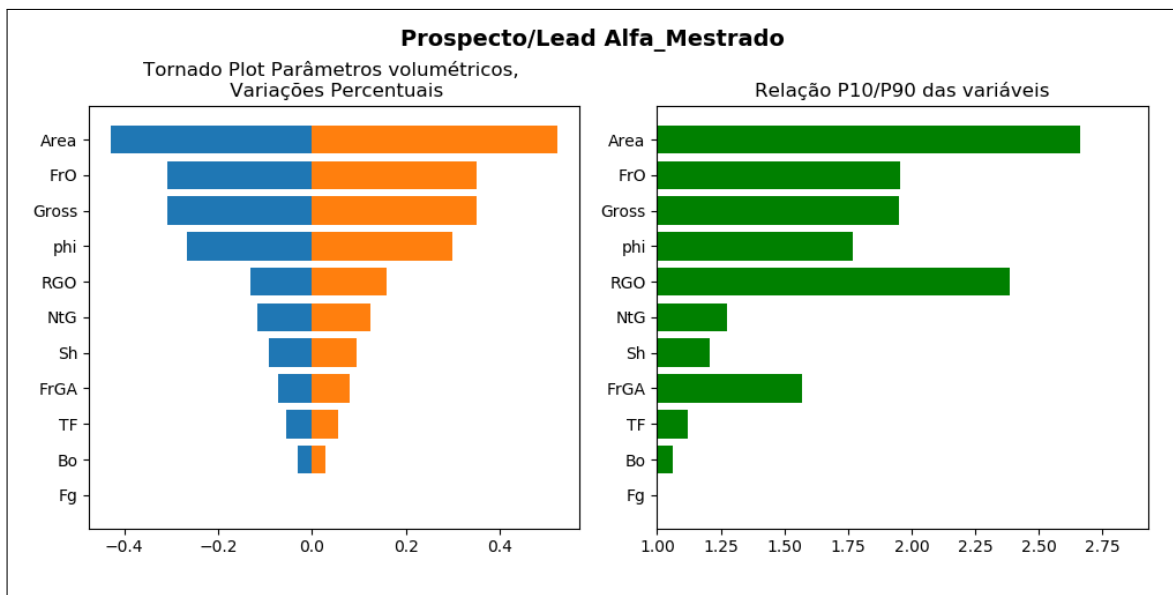


Figura 4.26: Na esquerda o Tornado plot com o impacto de cada variável no volume final (em MMBOE) e na direita as razões p10/p90 de cada variável.

Note que a faixa de variação do tornado plot não é perfeitamente fiel às razões de p10/p90. Isso ocorre porque os valores de RGO e fator de recuperação do gás associado (FrGA) afetam apenas a fase gasosa da mistura de óleo considerada (não estamos considerando correlação entre Bo e RGO).

Foi simulada a correlação entre o fator de formação de óleo (Bo) e o RGO. Conforme discutido no Capítulo 2, o Bo corresponde à redução do volume de óleo pela variação de pressão e temperatura e perda de gás e, portanto, é diretamente ligado ao RGO.

Adicionaremos a correlação entre a porosidade e o fator de recuperação. Com isso simulamos uma melhor capacidade de escoamento do reservatório, seja por melhoria na capacidade de injeção ou por melhoria da permeabilidade do reservatório (relacionada indiretamente à porosidade). No primeiro exercício utilizaremos uma estrutura de correlação Gaussiana que não apresenta estrutura de correlação diferenciada para os seus extremos (Figura 4.27).

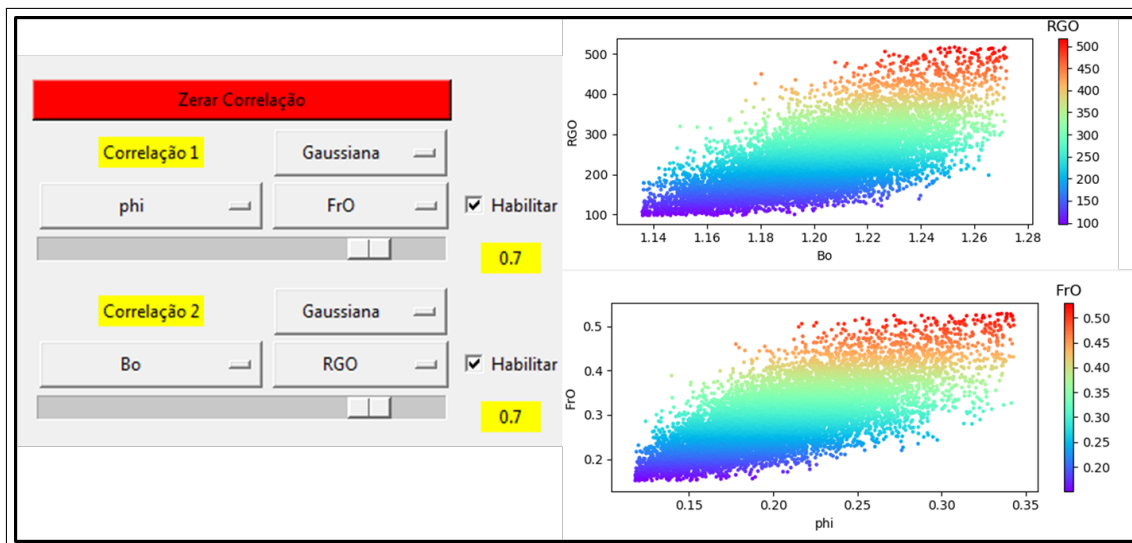


Figura 4.27: Parâmetros utilizado para aplicar a correlação entre Bo e o RGO e ϕ e FrO. Na direita o crossplot para controle de qualidade dos parâmetros utilizando. Note que mesmo com o alto valor de correlação (0.75) ainda temos uma dispersão considerável.

Tabela com Volumes - Prospecto/Lead Alfa_Mestrado_cor

Volumes *InPlace*

	P99	P90	Média	P10	P01	MI
Hidrocarbonetos(MMBOE)	828.7249	1404.6898	3098.4393	5267.637	8950.5681	602.64644
Óleo(MMBBL)	678.1135	1135.9581	2507.4592	4231.0356	7229.6569	487.70081
Gás Associado(TCF)	0.7128	1.2829	3.3181	6.0183	10.9339	0.64537
Gás Não Associado(TCF)	0.0	0.0	0.0	0.0	0.0	0.0
Condensado(MMBOE)	0.0	0.0	0.0	0.0	0.0	0.0

Volumes Recuperáveis

	P99	P90	Média	P10	P01	MI
Hidrocarbonetos(MMBOE)	239.0121	444.3445	1141.0315	2033.3217	3833.5748	221.93063
Óleo(MMBBL)	143.3814	273.6794	762.763	1403.8355	2680.8431	148.3574
Gás Associado(TCF)	0.4255	0.7863	2.1238	3.9248	7.5765	0.41308
Gás Não Associado(TCF)	0.0	0.0	0.0	0.0	0.0	0.0
Condensado(MMBOE)	0.0	0.0	0.0	0.0	0.0	0.0

Figura 4.28: Impacto nos volumes recuperáveis causados pela escolha dos parâmetros de correlação Bo e o RGO (0.75) e ϕ e FrO (0.75) com estrutura gaussiana.

Vamos verificar agora o impacto de cada variável, nos volumes recuperáveis de correlação (Figura 4.29).

Podemos verificar (com as barras pretas) que há um aumento do volume em função da correlação entre porosidade e fator de recuperação de óleo (par de correlações 1 Figura 4.30).

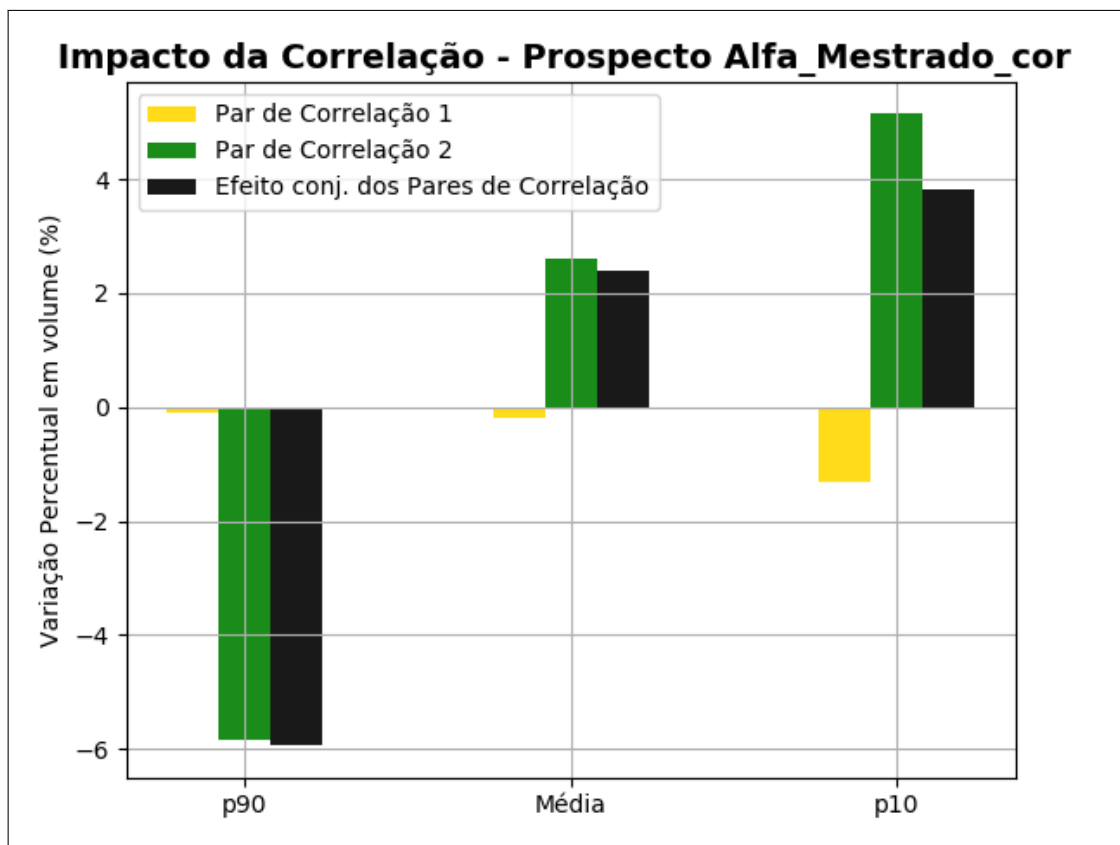


Figura 4.29: Impacto nos volumes recuperáveis causados pela escolha dos parâmetros de correlação Bo e o RGO (0.75) e ϕ e FrO (0.75) com estrutura gaussiana.

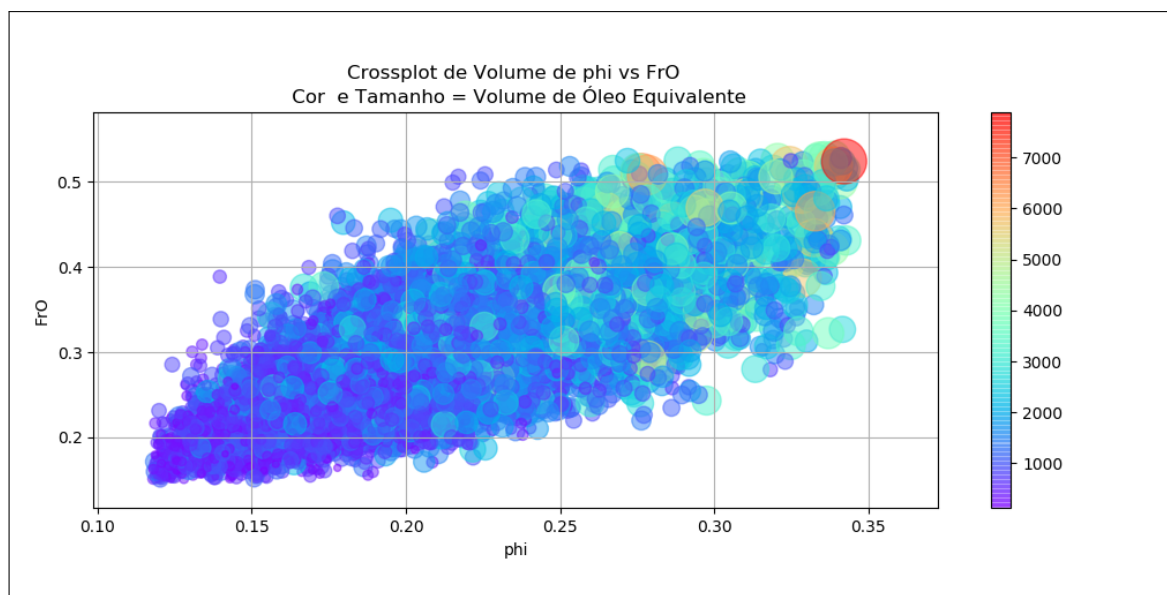


Figura 4.30: Gráfico de controle de qualidade (QC plot) do par de correlação (phi vs FrO) com estrutura gaussiana e correlação com intensidade de 0.75. Note o progressivo aumento do volume com o valor dos pares de correlação.

Um comportamento anormal pode ser verificado entre a correlação do Bo e RGO. Em todas os percentis verificados (p90, média e p10) há uma redução dos volumes de hidrocarbonetos. Este comportamento anômalo ocorre porque na correlação fazemos com que o aumento de RGO (que aumenta a fração gasosa) gere um aumento do Bo (que diminui a fração líquida de óleo). Dessa

forma esta correlação funciona como um mecanismo de compensação e não há um comportamento de correlação claro dos volumes com o aumento ou diminuição do par de variáveis (Figura 4.31).

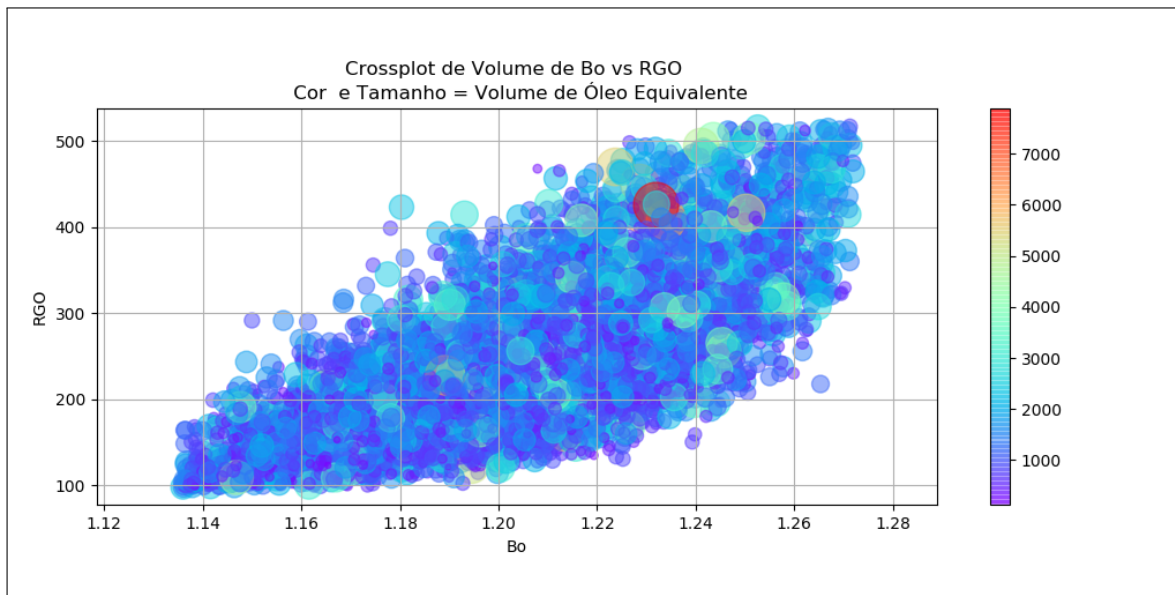


Figura 4.31: Gráfico de controle de qualidade (QC plot) do par de correlação (RGO vs Bo) com estrutura gaussiana e correlação com intensidade de 0.75. Note que não há um padrão claro de aumento ou diminuição de volumes com o valor dos pares de correlação.

Do ponto de vista da interpretação verificamos que a variação é reduzida no p90 justamente por ser o cenário em que a fração de gás (menores valores de RGO) teriam menor volume e como ele é apenas uma fração menor do volume de hidrocarbonetos tem menor impacto nas diferenças de volumes.

Vamos verificar agora o impacto das diferentes estruturas de correlação disponíveis no software (Figura 4.32).

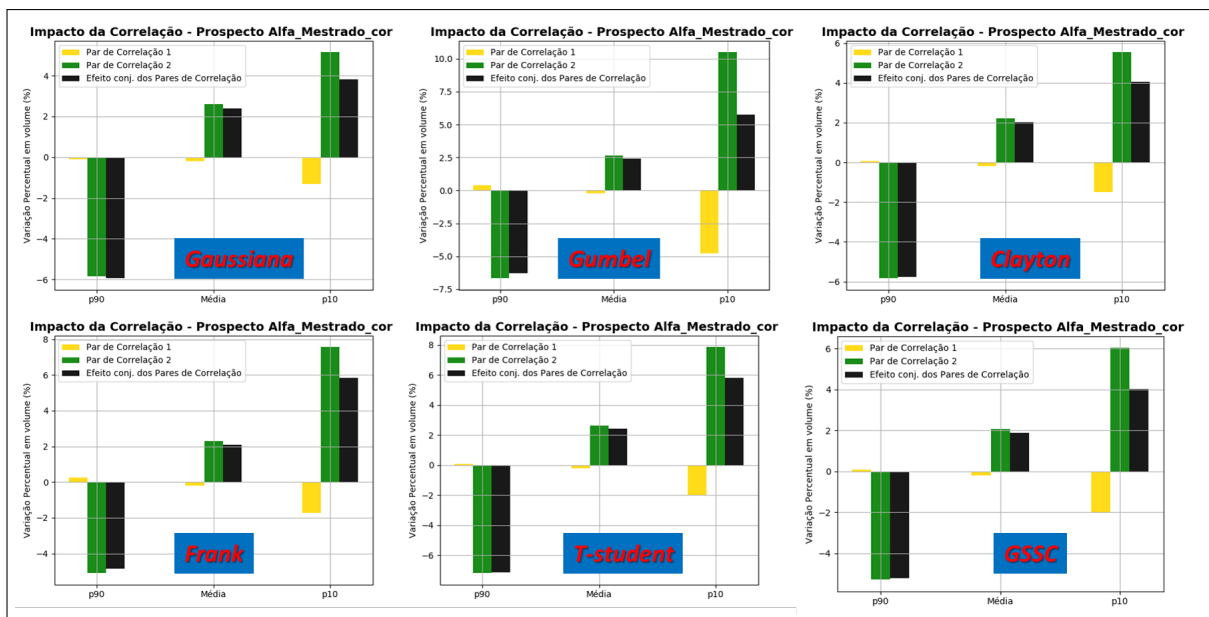


Figura 4.32: Impacto da correlação para diferentes estruturas de correlação disponível para os pares de correlação do prospecto Alfa (RGO vs Bo) e (ϕ vs FrO).

Note que, conforme esperado, a estrutura de correlação de Gumbel e T-Student (aproxima-

damente 5% dos volumes recuperáveis), que privilegiam o alinhamento dos pares no primeiro quadrante da correlação apresenta maior variação de volume.

As estruturas gaussianas e de Clayton, que não apresentam dependência no extremo maior da correlação são as que apresentam menor variação com a correlação (aproximadamente 4% dos volumes recuperáveis). Note que esta variação positiva é devido ao par phi vs FrO.

Testamos também diferentes correlações para verificar seu impacto final no volume das acumulações. Neste exercício utilizamos os valores de P90, média e P10 (valores que normalmente recebem maior peso na análise econômica de oportunidades). Veja o resultado na Figura (4.33).

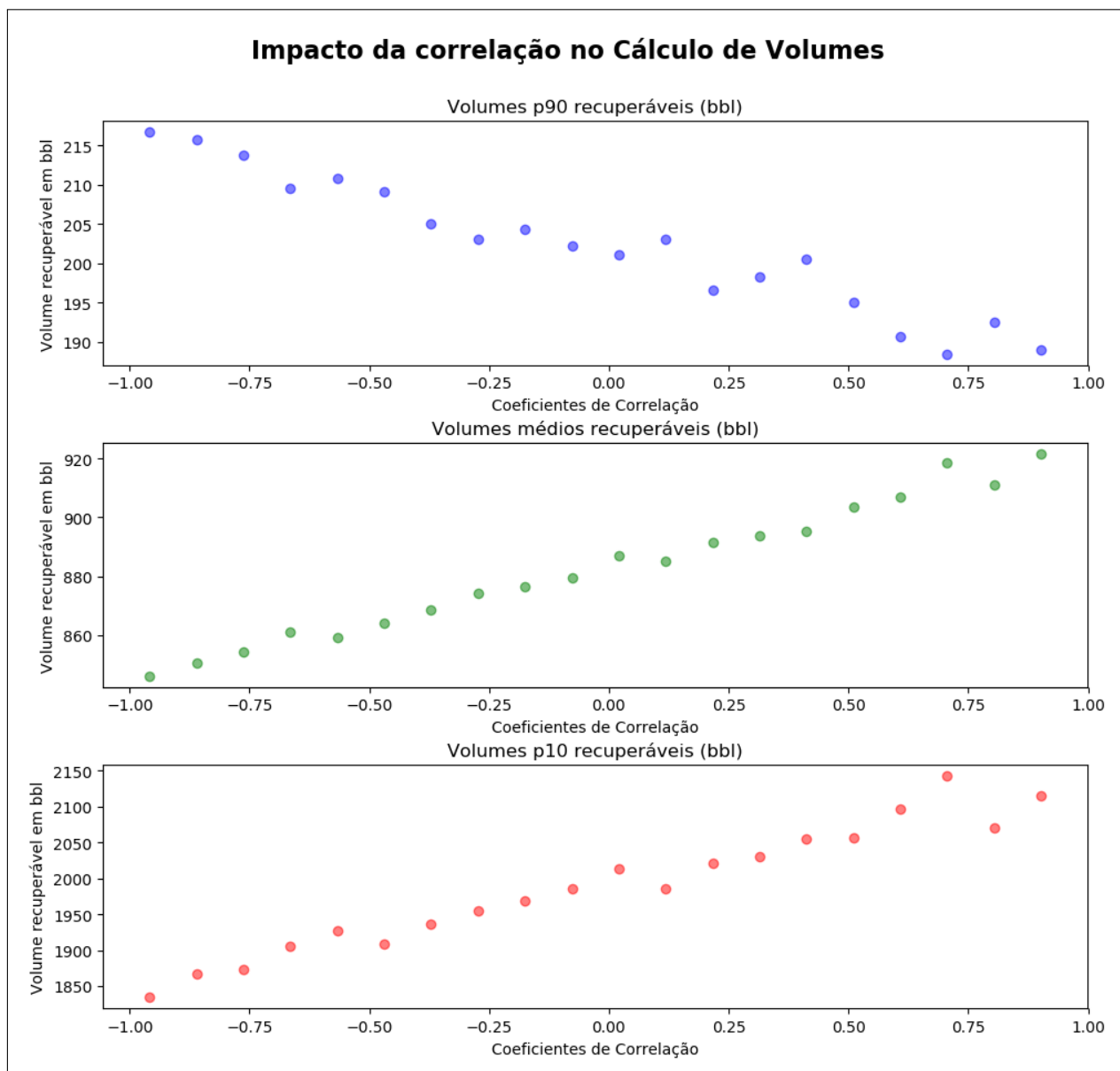


Figura 4.33: Distribuições de volumes P90 (acima), média (centro) e P10 (abaixo) para multiplicação de distribuições com diferentes correlações (eixo horizontal dos 3 gráficos).

Conforme esperado os valores de média aumentam com o aumento da correlação, os valores de P90 diminuem e os valores de P10 aumentam.

Conforme comentado no Capítulo 3, ao tomarmos correlações positivas o que ocorre é que os valores mínimos interagem com outros valores mínimos e valores máximos são multiplicados por outros valores máximos o que faz com que a distribuição mais alongada e, portanto, a dispersão medida pela razão P10/P90 aumente.

No caso de correlações negativas o comportamento é diametralmente oposto, ou seja, os valores máximos de uma distribuição são multiplicados pelos valores mínimos da outra e desta forma, os valores máximos e mínimos são achatados reduzindo a incerteza medida pelo P10/P90.

Capítulo 5

Resultados e Conclusões

Verificamos que a correlação de parâmetros deve ser utilizada em áreas e intervalos estratigráficos cujo grau de preditividade geológica suporta o uso de correlações;

Verificamos também, a variação do produto de duas distribuições, como função do ordenamento/desordenamento das amostras das mesmas, é resultado da dispersão de pontos por percentil;

Verificamos que correlações positivas tem o efeito de aumento do volume médio e aumento na dispersão de volumes modelados (refletidos nos valores de P10 e P90). Por outro lado, correlações negativas tem redução do volume médio esperado e redução na dispersão dos valores;

Com base nas modelagens sintéticas, concluímos que a obtenção de parâmetros de correlação com base em análogos regionais devem respeitar a escala do prospecto avaliado;

Com base nas aplicações práticas apresentadas, verificamos que utilizando o mesmo nível de correlação a estrutura de correlação de Gumbel se mostrou a mais otimista. A estrutura de correlação de Clayton se mostrou a mais pessimista.

O primeiro produto deste trabalho constitui as bibliotecas para análise volumétrica de prospectos exploratórios. As bibliotecas estão disponibilizadas no anexo, gratuitas e alinhadas com padrões internacionais de avaliação de oportunidades exploratórias. Os resultados estão calibrados com softwares como *@RISK* e *Cristal Ball* tipicamente utilizados por consultorias de E&P para análise volumétrica. Este produto pode ser utilizado para fins pedagógicos em ambientes universitários.

O segundo produto deste trabalho constitui o método GSSC. Este método traz como benefício o aumento da velocidade de convergência em relação aos métodos de mesma natureza anteriormente publicados na literatura. Verificou-se também que o método GSSC apresentou maior dispersão em relação à geração de amostras aleatórias correlatas do que a correlação normal como aquela obtida no método de Iman-Conover (que no caso de cálculo de volumes pode ser interpretada como maior abrangência de cenários geológicos).

O terceiro produto deste trabalho constitui a interface gráfica que além de facilitar o uso, agiliza a possível avaliação de oportunidades com a ferramenta desenvolvida nesta dissertação.

Além da revisão bibliográfica e breve explicação sobre a metodologia de análise em língua portuguesa sobre o tema, este trabalho deixa os códigos em python que podem ser utilizadas livremente para replicação dos resultados e/ou aplicação em outros espaços. O uso e melhora do software é encorajado pelo autor principalmente em ambiente acadêmico.

Bibliografia

- Abdul-Majeed, G. H. (2016). Estimation of solution gas oil ratio. *Petroleum and Coal*, page 1–12.
- Ahmed, D. S. S. (2016). *Sedimentary environments - Civil and Environmental Engineering Department*. disponível em : <https://faculty.mu.edu.sa/smohamed/Sedimentary%20environments>.
- Ahmed, T. H. (2010). *Reservoir engineering handbook*. ELSEVIER Inc.
- Al-Harthya, M., Beggb, S., and Bratvold, R. B. (2007). Copulas: A new technique to model dependence in petroleum decision making. *Journal of Petroleum Science and Engineering*, 55:195–208.
- Amado, L. (2013). *Reservoir Exploration and Appraisal*. Elsevier Inc. - Gulf Professional Publishing.
- Archie, G. E. (1942). The electrical resistivity log as an aid in determining some reservoir characteristics. *Trans. AIME*, 146:54–67.
- Armacosta, R. L. and Pet-Edwards, J. (1999). Integrative risk and uncertainty analysis for complex public sector operational systems. *Socio-Economic Planning Sciences*, 33:105–130.
- Armstrong, M., Galli, A., Bailey, W., and B.Couët (2004). Incorporating technical uncertainty in real option valuation of oil projects. *Journal of Petroleum Science and Engineering*, 31:67–82.
- Avseth, P., Mukerji, T., and Mavko, G. (2010). *An Introduction to: Velocity Model Building*, volume 2. Cambridge University Press.
- Azevedo, L. and Soares, A. (2017). *Geostatistical Methods for Reservoir Geophysics*, volume 1. Springer, Advances in Oil and Gas Exploration & Production.
- Baker, R. O., Yarranton, H. W., and Jensen, J. L. (2015). *Chp 8 - Special Core Analysis—Rock–Fluid Interactions extraído de Practical Reservoir Engineering and Characterization*, volume 2. SPE.
- Barnard, J., McCulloch, R., and X.Meng (2000). Modeling covariance matrices in terms of standard deviations and correlations, with applications to shrinkage. *Statistica Sinica*, 10:281–312.
- Beaumont, R. (2012). *An Introduction to statistics Correlation*. disponível em : <http://www.floppybunny.org/robin/web/virtualclassroom/stats/basics/part9.pdf>.
- Bird, D. E., Burke, K., Hali, S. A., and Casey, J. F. (2005). *Gulf of Mexico tectonic history: Hotspot tracks, crustal boundaries, and early salt distribution*, volume 89 (3). AAPG Bulletin.
- Bishop, R. (2012). Is the trap half full or half empty. *AAPG Explorer*.

- BOEM (2017). Us boem releases highest-resolution bathymetry map of the gulf of mexico... evah! <https://wattsupwiththat.com/2017/05/30/u-s-boem-releases-highest-resolution-bathymetry-map-of-the-gulf-of-mexico-evah/>. Accessed: 2019-01-30.
- Bradley, P. G. and Kaufman, G. M. (1971). Reward and uncertainty in exploration programs. *AAPG Notes*, pages 1–8.
- Brown, A. R. (1999). *Interpretation of Three-Dimensional Seismic Data 5th Edition*. Aapg Memoir 42.
- Buckles, R. S. (1965). Correlating and averaging connate water saturation data. *Journal Can. Petroleum Technology*, 4:42–52.
- Buckles, R. S. (2000). Correlating and averaging connate water saturation data. *Journal Can. Petroleum Technology*, 4:42–52.
- Capen, E. C. (1992). *Business of Petroleum Exploration*. AAPG, Dordrecht, Heidelberg, London, New York.
- Ceraldi, T. S. and Green, D. (2016). Evolution of the south atlantic lacustrine deposits in response to early cretaceous rifting, subsidence and lake hydrology. *Geological Society London Special Publications*, Special Publications,438.
- CGGVeritas (2013). The brazil santos basin: Unlocking the pre-salt potential. broadband technology: New benefits for pre-salt exploration. https://www.cgg.com/technicalDocuments/cggv_0000017532.pdf. Accessed: 2019-01-30.
- Cherubini, U., Luciano, E., and Vecchiato, W. (2004). *Copula Methods in Finance*, volume 1. Wiley, West Sussex, Inglaterra.
- Clayton, D. (1978). A model for association in bivariate life tables and its application in epidemiological studies of family tendency in chronic disease incidence. *Biometrika*, 65:141–151.
- Corridor-Resources-Inc (2016). Old harry, newfoundland/quebec. <https://www.corridor.ca/core-operating-areas/old-harry/>. Acessado: 2019-02-21.
- Costa, L. A. R., Zalán, P. V., and de Mendonça Nobre, L. P. (2013). Estimativa da chance de sucesso exploratório: uma abordagem em três passos consistente com a classificação de recursos petrolíferos. *B. Geoci. Petrobras, Rio de Janeiro*, 21:313–324.
- Crain, E. R. (2019). Crain’s petrophysical handbook - water saturation basics. <https://www.spec2000.net/03-13swbasics.htm>. Acessado: 2019-05-12.
- Dalenius, T. and Reiss, S. P. (1982). Data-swapping: A technique for disclosure control. *Journal of Statistical Planning and Inference*, 6:73–85.
- Demirtas, H. and Hedeker, D. (2011). A practical way for computing approximate lower and upper correlation bounds. *The American Statistician*, 65(1):05–17.
- Doveton, J. H. (1986). *Log Analysis of Subsurface Geology*. Wiley Interscience, New York.
- Dow, W. G. (1978). Application of oil-correlation and source-rock data to exploration in williston basin. *AAPG*, 58:1253–1262.

- Durante, F. and Sempi, C. (2009). *Copula Theory: An Introduction* extraído de *C1 em Copula Theory and Its Applications de Piotr Jaworski, Fabrizio Durante, Wolfgang Karl Härdle e Tomasz Rychlik*, volume 1. Springer, Warsaw.
- Dvorkin, J. and Nur, A. (2001). *Critical-Porosity Models em Pressure Regimes in Sedimentary Basins and Their Prediction - Memoir 76*. American Association of Petroleum Geologists.
- Ellis, D. V. and Singer, J. M. (2007). *Basic Well Log Analysis, Second Edition, AAPG Methods in Exploration Series, No.16*. Springer.
- Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *Quarterly Journal of Economics*, 75:1–39.
- Elsharkawy, A. M. and Alikhan, A. A. (1996). Correlations for predicting solution gas/oil ratio, oil formation volume factor, and undersaturated oil compressibility. *Journal of Petroleum Science and Engineering*, 17:291–302.
- Embrechts, P., Lindskog, F., and McNeil, A. (2010). Modelling dependence with copulas and applications to risk management. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.23.5130&rep=rep1&type=pdf>. Accessed: 2018-11-30.
- Favoreto, J., Rohn, R., Lykawa, R., and Okubo, J. (2016). CaracterizaÇÃo sedimentolÓgica dos carbonatos albianos do reservatÓrio quissamÃ na porÇÃo meridional da bacia de campos. *Revista de Geosciências da UNESP*, 35(1):1–15.
- Feller, W. (1971). An introduction to probability theory and its applications. *John Wiley & Sons, New York*, I:280–283.
- Ferdinando, D. and Longley, I. (2016). Permian play mapping in the northern perth basin. *Petroleum Division Publication*, 1(1):17–26.
- Fisher, R. (1926). The management of field experiments. *Journal Ministry of Agriculture*, 33:503.
- Frank, M. (1979). On the simultaneous associativity of $f(x, y)$ and $x+y-f(x, y)$. *Aequationes Mathematicae*, 19:194–226.
- FRED, U. (2019). Federal research economical data - crude oil prices: West texas intermediate (wti)- cushing, oklahoma. Acessado: 2019-04-10.
- Frick, T. (1962). *Petroleum Production Handbook, Volume II*, volume 14. SPE.
- Fréchet, M. (1956). Sur les tableaux de corrélation dont les marges sont données. *C. R. Acad. Science Paris*, 242:2426–2428.
- Féron, R. (1956). Sur les tableaux de corrélation dont les marges sont données. cas de l'espace a trois dimensions. *Publ. Inst. Stat. Univ. Paris*, 5:3–12.
- Gabriel, K. R. (1962). Ante-dependence analysis of an ordered set of variables. *Ann. Math. Statist*, 33:201–212.
- Gaffney, C. . A. G. (2010). Review and evaluation of ten discoveries and prospects in the pre-salt play of the deepwater santos basin, brazil. www.anp.gov.br/?dw=39137. Accessed: 2018-01-30.
- Galloway, W. E. (2009). Gulf of mexico. <https://www.geoexplor.com/articles/2009/03/gulf-of-mexico>. Accessed: 2019-01-30.

- Galton, F. (1879). The geometric mean, in vital and social statistics. *Proceedings of the Royal Society*, 29:365–367.
- Genest, C. and MacKay, J. (1986). The joy of copulas: Bivariate distributions with uniform marginals. *The American Statistician*, 40:280–283.
- Genest, C., Quessy, J., and Rémillard, B. (2006). Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scandinavian J Statistics*, 33:337–366.
- Glaso, O. (1980). Generalized pressure-volume-temperature correlations. *Journal of Petroleum Technology*, page 785–795.
- Glover, P. (2016). Petrophysics msc course notes - porosity chapter 2. http://homepages.see.leeds.ac.uk/~earpwjg/PG_EN/CD%20Contents/GGL-66565%20Petrophysics%20English/Chapter%202.PDF. Acessado: 2019-02-20.
- Goda, K. and Tesfamariam, S. (2015). Multi-variate seismic demand modelling using copulas: Application to non-ductile reinforced concrete frame in victoria, canada. *Structural Safety*, 56:39–51.
- Grimaldi, S., Kao, S., Castellarin, A., Papalexiou, S., Viglione, A., Laio, F., Aksoy, H., and Gedikli, A. (2011). *Statistical Hydrology, Vol.2 Cap 2 extraido de Treatise on Water Science*, volume 2. Elsevier Inc. All.
- Gumbel, E. (1958). Distributions à plusieurs variables dont les marges sont données. *C. R. Acad. Sci. Paris*, 246:2717–2719.
- Gumbel, E. (1960). Distributions des valeurs extrêmes en plusieurs dimensions. *Publ. Inst. Stat. Univ. Paris 9*, page 171–173.
- Gunver, M. G., Senocak, M. S., and Vehid, S. (2018). To determine skewness, mean and deviation with a new approach on continuous data. *International Journal of Sciences and Research*, 1:64–79.
- Hall, K. R. and Yarborough, L. (1973). A new equation of state of z-factor calculation. *The Oil and Gas Journal*, 146:140–149.
- Hanafy, H., Macary, S., ElNady, Y., Bayomi, A., and Batanony, M. E. (1997). A new approach for predicting the crude oil properties. *Soc. of Petroleum Engineering (SPE)*, 17:291–302.
- Hantschel, T. and Kauerauf, A. I. (2009). *Fundamentals of Basin and Petroleum Systems Modeling*. Springer, Dordrecht and Heidelberg and London and New York.
- Harbaugh, J. W., Davis, J., and J.Wendebourg (1995). *Computing Risk for Oil Prospects: Principles and Programs*, volume 14. Pergamon, 1 edition.
- Hastings, W. (1970). Monte carlo sampling methods using markov chains and their application. *Biometrika*, 57:97–109.
- hua Han, D. (2001). 'fizz water' and low gas saturated reservoirs. *SEG Int. Exposition and Annual Meeting San Antonio, Texas*, page 3–4.
- Iman, R. and Conover, W. (1982). A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics*, B11(3):311–334.
- J.C.W., B. M. and Rayner (1987). The comparison of sample covariance matrices using likelihood ratio tests. *Biometrika*, 74:841–847.

- J.L.Pindell and Kennan, L. (2010). *Tectonic evolution of the Gulf of Mexico, Caribbean and northern South America in the mantle reference frame: An update*. In: James, K.H. (ed), *Geological Society Special Publications 328: The Origin and Evolution of the Caribbean Plate*. Gulf Coast Association of Geological Societies, New Orleans, LA, United States.
- Jones, T. A. (1969). Skewness and kurtosis as criteria of normality in observed frequency distributions. *AAPG - Notes on Exploration*, pages 1–6.
- Jonhson, M. E. and Ramberg, J. S. (1977). Transformatinos of the multivariate normal distribution with aplications to simulation. *Simulation Tech Report* , *Los Alamos Scientific Laboratory, New Mexico*, LA-UR-77-2595.
- Kamari, A., Zendehboudi, S., Sheng, J. J., Mohammadi, A. H., , and Ramjugernath, D. (2016). Rigorous modeling of solution gas–oil ratios for a wide ranges of reservoir fluid properties. *Journal of Petroleum & Environmental Biotechnology*, 7 (6):1–7.
- Kareem, L. A., Iwalewa, T. M., and Al-Marhoun, M. (2014). New explicit correlation for the compressibility factor of natural gas: linearized z-factor isotherms. *Journal of Petroleum Exploration Production Technology*, 6:481–492.
- Kenett, D. Y., Huang, X., Vodenska, I., Havlin, S., and Stanley, H. E. (2015). Partial correlation analysis: applications for financial markets. *Quantitative Finance*, 15 (4):569–578.
- Knight, F. H. (1921). Risk, uncertainty and profit. *University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship.*, 1.
- Law, C. A. (1999). *Evaluating source rocks - Treatise in Petroleum Geology Exploring for Oil and Gas Traps* Edward A. Beaumont. AAPG.
- Lehoczky, J. P. (2015). *Distributions, Statistical: Special and Continuous, extraído de International Encyclopedia of the Social & Behavioral Sciences (Second Edition)*, volume 1. Princeton University Press, Princeton, New Jersey.
- Liang, K. and Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73:13–22.
- Limpert, E., Stahel, W., and Abbt, M. (2001). Log-normal distribution across the science: Keys and clues. *BioScience*, 51(51):17–26.
- Liner, D. (2016). The limitations of lognormal distributions: Using subsurface data to make more accurate resource estimations. *AAPG Annual Convention and Exhibition, Calgary, Alberta, Canada*, 1:1–2.
- Lucia, F. J. (2007). *Carbonate Reservoir Characterization : An integrated approach, 2nd Edition*, volume 2. Springer.
- Magoon, L. B. (1988). *Petroleum Systems of the United States*. U.S. Geological Survey, Washington, D.C., USA.
- Magoon, L. B. and Dow, W. G. (1994). *The Petroleum System- from source to Trap : AAPG Memoir 60*. AAPG.
- Mahfoud, M. (2012). Bivariate archimedean copulas: an application to two stock market indices. *Vrije Universiteit Amsterdam*, pages 141–151.
- Marton, G. and Buffler, R. (1994). Jurassic reconstruction of the gulf of mexico basin. in: International geology review. *Winston & Son, Silver Spring, MD, United States*, 36:545–586.

- Matsumoto, M. and Nishimura, T. (1988). Mersenne twister: A 693-dimensionally equidistributed uniform pseudorandom number generator. <http://www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/ARTICLES/mt.pdf>. Acessado: 2019-01-29.
- McKay, M., Beckman, R., and Conover, W. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics - American Statistical Association.*, 21 (2):239–245.
- McNeil, A. J., Frey, R., and Embrechts, P. (2005). *Quantitative Risk Management, Concepts, Techniques and Tools*, volume 1. Princeton University Press, Princeton, New Jersey.
- Mendoza-Torres, F., Díaz-Viera, M., and Erdely, A. (2017). Bernstein copula modeling for 2d discrete fracture network simulations. *Journal of Petroleum Science and Engineering*, 156:710–720.
- Metgasco (2018). Apresentação para investidores - annual general meeting.
- Metropolis, N. (1987). The beginning of the monte carlo method. *Los Alamos Science*, 1:125–130.
- Miller, J. O. (1998). Bivar: A program for generating correlated random numbers. behavior research methods. *Instruments & Computers*, 30 (3):720–723.
- Moor, R. (1996). Controlled data swapping techniques for masking public use micro-data sets. *U.S. Census Bureau*, I:280–283.
- Moreira, J. L., Madeira, C., Gil, J., and Machado, M. A. P. (2007). *Bacia de Santos.*, volume 15 (2). Boletim de Geociências da Petrobras.
- Muniz, M. C. and Bosence, D. (2017). Reservoir layering controlled by the taphonomy and taphofacies of coquinas from the early aptian campos basin, brazil. *AAPG ICE 2017*, 1(1):1–9.
- Nelsen, R. (2006). *An Introduction to Copulas*. Springer, New York, USA.
- Nery, G. G. (2013). *Perfilagem Geofísica em Poço Aberto: Fundamentos básicos com ênfase em petróleo*. Sociedade Brasileira de Geofísica (SBGF).
- Newendorp, P. (1976). A method for treating dependencies between variables in simulation risk-analysis. *Journal of Petroleum Technology*, 28:1145–1150.
- Nikolaidis, E., Mourelatos, P. Z., and Pandey, V. (2011). Decisions under uncertainty with limited information design. *CRC Press*, page 196.
- Nur, A., Mavko, G., Dvorkin, J., and Galmudi, D. (1998). Critical porosity: a key to relating physical properties to porosity in rocks. *The Leading Edge*, 17 (10):357–362.
- Osyrov, K., Ivanova, N., Yang, Y., Fournier, A., Nichols, D., Bachrach, R., Yarman, C. E., and Nikolenko, D. (2013). Uncertainty as a rosetta stone uniting geoscience knowledge and e&p business value. *International Petroleum Technology Conference*, pages 1–12.
- Otis, R. M. and Schneidermann, N. (1997). A process for evaluating exploration prospects. *AAPG Bulletin*, 81:1–7.
- Owen, D. (1956). Tables for computing bivariate normal probabilities. *Annals of Mathematical Statistics*, 27 (10):75–1090.
- Peacock, D. (2007). Is your trap filled to spill? *International Petroleum Technology Conference*, pages 1–6.

- Pearson, K. (1920). Notes on the history of correlation. *Biometrika*, 13:25–45.
- Petersohn, E. and Abelha, M. (2013). Avaliação geológica libra, brasil pré-sal. *1 Rodada de Licitação de Partilha*, 1:1–90.
- Pindell, J. (2010). *History of tectonic modeling and implications for depositional architecture in the Gulf of Mexico: Where we should go from here*. In: Hammes, U. (ed), *Transactions of the 60th Annual GCAGS Convention and the 57th Annual GCSSEPM Meeting, San Antonio, Texas*. Geological Society of London, London, United Kingdom.
- Piotr Jaworski, Fabrizio Durante, W. K. H. e. T. R. (2009). *Copula Theory and Its Applications*, volume 1. Springer, Warsaw.
- Pitman, E. (1937). Significance tests which may be applied to samples from any population, parts i and ii. *J.R. Stat. Soc. Suppl.*, 4:119–232.
- Pourahmadi, M. (2007). *Simultaneous modeling of covariance matrices: GLM, Bayesian and nonparametric perspective*. In *Correlated Data Modelling 2004 (D. Gregori et al., eds.)*. FrancoAngeli, Milão, Italia.
- Rachev, S. T. (2003). *Book 1, Handbook of Heavy Tailed Distribution in Finance University of California*, volume 1. Elsevier, AMSTERDAM & BOSTON & LONDON & NEW YORK & OXFORD & PARIS SAN DIEGO & SAN FRANCISCO & SINGAPORE & SYDNEY & TOKYO.
- Railsback, L. B. (2011). Petroleum geoscience and subsurface geology. prepared for geol 4320/6320 petroleum geology course. <http://www.gly.uga.edu/railsback/PGSG/PGSGmain.html>. Acessado: 2019-02-29.
- Rodgers, J. L. and Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician*, 42(1):59–66.
- Rose, P. R. (2001). Risk analysis and management of petroleum exploration ventures. *AAPG special series - CD/DVD*.
- Rose, P. R. (2007). Measuring what we think we have found: Advantages of probabilistic over deterministic methods for estimating oil and gas reserves and resources in exploration and production. *GEOHORIZONS, AAPG Bulletin*, 91 (1):21–29.
- Rose, P. R. (2017). Evolution of e&p risk analysis (1960-2017). *AAPG 100th Annual Convention and Exhibition, Houston, Texas*, 1:1–45.
- Rowe, W. D. (1977). An anatomy of risk. *John Wiley and Sons. Wiley Series on Systems Engineering and Analysis.*, 01.
- Rukhin, A., Soto, J., Nechvatal, J., Smid, M., Barker, E., Leigh, S., Levenson, M., Vangel, M., Banks, D., Heckert, A., Dray, J., and Vo, S. (2001). A statistical test suite for random and pseudo-random number generators for cryptographic applications - nist special publication 800-22. <https://www.random.org/statistics/SP800-22b.pdf>. Acessado: 2019-01-26.
- Salvador, A. (1991). *The Geology of North America: The Gulf of Mexico Basin*, volume J. Geological Society of America (GSA), Boulder, CO, United States.
- Schlumberger (2019). Schlumberger oilfield glossary. <https://www.glossary.oilfield.slb.com/>. Accessed: 2019-01-30.
- Schweizer, B. (1991). *Thirty years of copulas*. In: *Dall’Aglío G, Kotz S, Salinetti - Advances in Probability Distributions with Given Marginals*, volume 1. Kluwer.

- Shaw, W. and Lee, K. (2008). Bivariate student t distributions with variable marginal degrees of freedom and independence. *Journal of Multivariate Analysis*, 99:1276 – 1287.
- Shell, R. D. (2014). Play based exploration a guide for aapg’s imperial barrel award participants. *AAPG*, 1:1–52.
- Shepherd, M. (2009). *Factors influencing recovery from oil and gas fields*, in M. Shepherd, *Oil field production geology: AAPG Memoir 91*. AAPG.
- Sheriff, R. E. and Geldart, L. P. (1995). *Exploration Seismology*. Cambridge University Press.
- Sklar, A. (1959). *R.B. Nelsen*. PhD thesis, Publ. Inst. Statist. Univ.Paris 8.
- Smalley, P. C., Begg, S. H., Naylor, M., Johnsen, S., and Godi, A. (2008). Handling risk and uncertainty in petroleum exploration and asset management: An overview. *AAPG Bulletin*, 92 (10).
- Smith, M. D. and Jones, D. R. (1992). *Trend Analysis: Chapter 18: Part IV. Managing the Business capitulo de The Business of Petroleum Exploration*, Ed Capen. AAPG, Dordrecht, Heidelberg, London, New York.
- Smith, N., Ghous, A., and Bharatiya, S. (2018). Managing complex, laterally and vertically compartmentalized reservoirs in a subsea tie-back via smart well completions – case studies. *Society of Petroleum Engineers*, pages 1–14.
- SPE (2007). Guidelines for the evaluation of reserves and resources,. *Petroleum Resources Management System, Society of Petroleum Engineer*, 01(1):01.
- Standing, M. B. (1977). Volumetric and phase behavior of oil field hydrocarbon systems. *Society of Petroleum Engineers*, 1977:125–126.
- Standing, M. B. and Katz, D. L. (1942). Density of natural gases. *Trans. AIME*, 146:140–149.
- Strand, R. and Oughton, D. (2009). *Risk and Uncertainty - as a Research Ethics Challenge*. National Committees for Research Ethics in Norway, Tulsa, Oklahoma.
- Styan, G. (1968). *ference in multivariate normal populations with structure. Part I: inference when correlations are known. Technical report*. Department of Statistics, University of Minnesota, USA, Minnesota, USA.
- Sun, C., Liu, H., Yan, K., Chen, Q. M. B. L. G., Xiao, X., Wanga, H., Zheng, X., and Liu, S. (2012). Experiments and modeling of volumetric properties and phase behavior for condensate gas under ultra-high-pressure conditions. *Industrial and Engineering Chemistry Research*, 51 (19):6916–6925.
- Tiab and Donaldson, E. C. (2012). *Petrophysics, Theory and Practice of Measuring Reservoir Rock and Fluid Transport Properties, 3rd Edition*. Gulf Professional Publishing is an imprint of Elsevier.
- Toma, S.-V., Chitita, M., and Sarpea, D. (2012). Risk and uncertainty. *Emerging Markets Queries in Finance and Business*, 3:975 – 980.
- Trottini, M., Muralidhar, K., and Sarathy, R. (2011). Maintaining tail dependence in data shuffling using t copula. *Statistics and Probability Letter*, 81:420–428.
- Ungerer, P., Faissat, B., Leibovici, C., Zhou, H., Behar, E., Moracchini, G., and Courcy, J. (1995). High pressure-high temperature reservoir fluids: investigation of synthetic condensate gases containing a solid hydrocarbon. *Fluid Phase Equilibria*, 111 (2):287–311.

- Vorechovsky, M. (2002). New improvements to simulation technique latin hypercube sampling and possibilities of its utilization. *Problémy modelování (Problems of Modeling) conference, Faculty of Civil Engineering, Ostrava, Czech Republic, Brno University of Technology*, pages 83–90.
- Vorechovsky, M. (2010). On correlation control in monte carlo type sampling. *4th International Workshop on Reliable Engineering Computing*, pages 340–352.
- Walker, R. (1984). *Facies Models*, volume 2. Geoscience Canada.
- Wang, W. and Wells, M. (2000). Model selection and semi parametric inference for bivariate failure-time data. *J Amer Statistical Assoc*, 95:62–76.
- White, D. (1987). Methods of oil and gas prospect and play assessment course manual. *Oil & Gas Consultants International*.
- Whitford, A. B. (2005). *Encyclopedia of Social Measurement- Correlation*. Elsevier Inc. All, Dordrecht, Heidelberg, London, New York.
- Winker, C. and Buffler, T. (1988). *Palaeogeographic evolution of early deep-water Gulf of Mexico and margins, Jurassic to Middle Cretaceous (Commanchean)*, volume 72 (3). AAPG Bulletin, Tulsa, OK, United States.
- Winsauer, W., Shearin, H. J., Masson, P., and Williams, M. (1952). Resistivity of brine-saturated sands in relation to pore geometry. *American Association of Petroleum Geologists (AAPG) Bulletin*, 36 (2):253–277.
- Wolfe, J. A. (1963). Exploration in perspective. *AAPG Bulletin*, 47(5):1–8.
- Yule, G. (1907). On the theory of correlation for any number of variables, treated by a new system of notation. *Roy. Soc. Proc.*, 79:85–96.
- Yule, G. (1927). On a model of investigating periodicities in disturbed series with special reference to wolfer’s sunspot numbers. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 226:267–298.
- Zimmermann, H. (2000). An application-oriented view of modeling uncertainty. *European Journal of Operational Research*, 122:190–198.
- Zio, E. and Pedroni, N. (2012). Les cahiers de la sÉcuritÉ industrielle, risk analysis, uncertainty characterization in risk analysis for decision-making practice. *Foundation for an Industrial Safety Culture*, 1:190–198.