



Business Intelligence

PUC
RIO

Rômulo de Araújo Matos

*Avaliação de acurácia de previsões de sucesso de
poços exploratórios*

Monografia de Final de Curso

28/01/2021

***Monografia apresentada ao Departamento de Engenharia Elétrica da
PUC/Rio como parte dos requisitos para a obtenção do título de
Especialização em Business Intelligence.***

Orientador:

Prof. Anderson Silva do Nascimento

Agradecimentos

O desenvolvimento deste trabalho de conclusão de curso contou com o apoio e incentivo de instituições e pessoas, dentre as quais agradeço:

À Petrobras por todo o suporte e fornecimento de parte dos dados utilizados e mais especificamente à Gerência Geral de Reservatórios e o seu gerente Luís Otávio Schmall que através do Programa E&P Competências possibilitou a minha participação no curso.

À toda equipe do Laboratório de Inteligência Computacional Aplicada da PUC-Rio, aos professores e instrutores do BI Master pelo rico aprendizado ao longo do curso essencial ao meu desenvolvimento profissional e em especial ao Professor Anderson Silva do Nascimento pela sua orientação.

Aos meus colegas de curso, pelo companheirismo e pela rica troca de experiências, principalmente enquanto foram possíveis as atividades presenciais.

Aos meus colegas e amigos do grupo D2 do Programa E&P Competências, pelo ambiente amistoso no qual convivemos, solidificamos nossos conhecimentos e enfrentamos os desafios propostos pelo Programa.

Aos gerentes imediatos com os quais trabalhei ao longo do curso e do período de desenvolvimento deste trabalho, Luiz Antonio Freitas Trindade e Vanessa Yuri Matsunaga, pela compreensão sem a qual não poderia administrar e atender a todas as demandas.

RESUMO

A avaliação de fatores de sucesso e de estimativas volumétricas são importante parte do trabalho de exploração da indústria de óleo e gás e insumos fundamentais às tomadas de decisões estratégicas. A análise sistemática da qualidade das estimativas realizadas frente aos resultados obtidos tem o potencial de melhorar as previsões ao possibilitar identificação de pontos críticos, padrões e/ou fatores chaves que determinam as diferenças de acurácia das estimativas para os diversos elementos avaliados. Aplicando uma metodologia originalmente desenvolvida para quantificar a acurácia de previsões meteorológicas, ao medir a distância entre o resultado observado e a previsão probabilística de eventos binários, foram obtidos os Índices de Brier e seus componentes relativos às previsões das probabilidades de descoberta de uma acumulação de hidrocarboneto e da ocorrência dos elementos do sistema petrolíferos associados a uma determinada oportunidade exploratória. Contando com os dados e avaliações de centenas de poços e oportunidades testadas foi possível verificar a aplicação dos índices na identificação de diferentes níveis de acurácia, na caracterização e constatação de eventuais padrões e vieses presentes nas avaliações. A identificação dos diferentes níveis de assertividade e os fatores a ela associados permitiria a melhor alocação de recursos e tem potencial de auxiliar as estimativas do valor da informação (VOI).

ABSTRACT

The assessment of success factors and volumetric estimates are an important part of the exploration work of the oil and gas industry and key inputs for strategic decision making. The systematic analysis of the quality of the estimates made in view of the results obtained has the potential to improve the predictions by enabling the identification of critical points, patterns and key factors that determine the differences in accuracy of the estimates for the various elements evaluated. Applying a methodology originally developed to quantify the accuracy of weather forecasts, when measuring the distance between the observed result and the probabilistic prediction of binary events, The Brier Indices and their components were obtained for the predictions of the probabilities of discovering a hydrocarbon accumulation and the occurrence of the elements of the petroleum system associated with a given exploratory opportunity. With data and evaluations of hundreds of wells and tested opportunities, it was possible to verify the application of the indices in the identification of different accuracy levels, in the characterization and verification of any patterns and biases present in the evaluations. The identification of the different levels of assertiveness and the factors associated with it would allow for a better allocation of resources and have the potential to help estimate the value of information (VOI).

Sumário

1 INTRODUÇÃO	5
2 CONTEXTUALIZAÇÃO DO PROBLEMA.....	7
3 METODOLOGIAS	10
4 ARQUITETURA DO SISTEMA PROPOSTO.....	13
5 RESULTADOS	17
6 CONCLUSÕES E TRABALHOS FUTUROS	23
Referências Bibliográficas	25

1 INTRODUÇÃO

O caminho até a descoberta de uma significativa reserva de hidrocarboneto é longo e envolve riscos. Como típica atividade extrativista, a exploração de óleo e gás (O&G) tende naturalmente a tornar-se mais complexa e custosa à medida que os recursos se tornam mais escassos ou menos acessíveis e, portanto, a melhora dos processos de avaliação do potencial petrolífero e dos riscos envolvidos é premente.

O risco faz parte da atividade exploratória e não pode ser eliminado. No entanto, o bom aproveitamento dos dados e o acúmulo de informações e de conhecimento tem o potencial de reduzir as incertezas e melhorar as estimativas de risco ou, vindo por outro lado, as estimativas das chances de sucesso da descoberta de uma nova acumulação de hidrocarboneto (um sucesso geológico).

Entre a definição da área ser prospectada e a perfuração, com a consequente constatação do resultado do poço pioneiro, há todo um período exploratório durante o qual são desenvolvidas atividades de prospecção geológica e geofísica que envolvem a aquisição de dados sísmicos, gravimétricos, magnetométricos, geoquímicos, dentre outros, visando aprimorar o conhecimento a respeito da área investigada e subsidiar a interpretação exploratória.

De forma geral, a fase de interpretação é onde todos os dados geológicos e geofísicos se integram. Os dados sísmicos constituem a base do processo de interpretação, seja uma interpretação sísmica de linhas 2D's regionais com objetivo de estudo do potencial de uma determinada bacia; a interpretação sísmica de um volume 3D para definição de um prospecto e a locação de um poço exploratório; ou uma interpretação sísmica 3D/4D em detalhe que permita a caracterização de um reservatório a ser desenvolvido.

Na condução do processo de interpretação exploratória os dados sísmicos são integrados aos dados não-sísmicos, aos dados de poços, paleontológicos, geoquímicos, entre outros, e analisados de forma crítica e multidisciplinar visando a construção de um modelo que descreva a sucessão e evolução dos eventos geológicos (tectônica, deposição, diagênese, geração e migração de hidrocarboneto) e reflita o cenário que se acredita existir em subsuperfície.

Por fim, o principal objetivo do processo de interpretação é determinar a locação de um poço exploratório, agregando o máximo de informação e conhecimento orientado à mitigação dos riscos exploratórios e, consequentemente, imprimindo ao projeto maior chance de sucesso. De forma subordinada, caso atinja o nível de detalhe suficiente, o modelo interpretado

permitiria estimar reservas e curvas de produção fundamentais aos estudos de desenvolvimento e produção e à avaliação econômica.

Assim, a precisão e qualidade da interpretação, imprescindível ao sucesso do processo exploratório, depende da quantidade e qualidade dos dados de qualquer natureza (sísmicos, multifísicos, de poços etc.), do acesso e da integração desses dados e das equipes multidisciplinares envolvidas nos estudos e análises.

A avaliação de fatores de sucesso e de estimativas volumétricas são importante parte do trabalho de exploração da indústria de óleo e gás. A determinação destes fatores é de fundamental importância para as tomadas de decisões estratégicas tais como posicionamento, desenvolvimento dos projetos exploratórios, definição de carteira e portfólio dentre tantas. Portanto, trata-se de um tema com significativos impactos econômicos para uma companhia de O&G.

Apesar de tal relevância e ainda que haja muitas iniciativas no sentido de melhorar tais estimativas, principalmente, a partir do maior uso e integração de dados, não são comuns as análises sistemáticas pretéritas sobre a qualidade das estimativas realizadas frente aos resultados obtidos. Com a avaliação de quais previsões foram sistematicamente bem sucedidas, ou não, é esperada a identificação de pontos críticos, padrões e/ou fatores chaves que determinam as diferenças de acurácia das estimativas para os diversos elementos avaliados.

A clara identificação dos diferentes níveis de assertividade e os fatores a ela associados permitiria a melhor alocação de recursos (aquisição de dados e informação, tempo, estudos, consultorias etc.), que poderiam ser direcionados para os segmentos de menor performance em determinado contexto, e tem potencial de auxiliar as estimativas do valor da informação (VOI).

Neste trabalho é apresentada uma proposta de avaliação das previsões das probabilidades de sucesso de poços exploratórios seguida pelo desenvolvimento de uma prova de conceito (POC) contando com uma base de dados constituída por centenas de poços exploratórios pioneiros e suas respectivas previsões e resultados.

2 CONTEXTUALIZAÇÃO DO PROBLEMA

2.1 Definição do Problema

A definição de uma locação, ou seja, da posição de um poço exploratório, é uma das últimas fases de um programa exploratório. O seu resultado muitas vezes determinará o sucesso ou não de uma campanha de estudos realizada ao longo de anos por equipes multidisciplinares e, normalmente, exigindo elevados investimentos. Significativa parcela desses investimentos corresponde ao próprio custo de execução do poço, assim a decisão de seguir ou não com a perfuração de um poço exploratório ou a escolha de determinados poços dentre as diversas oportunidades em uma carteira exploratória são decisões críticas na indústria de O&G (Figura 1).

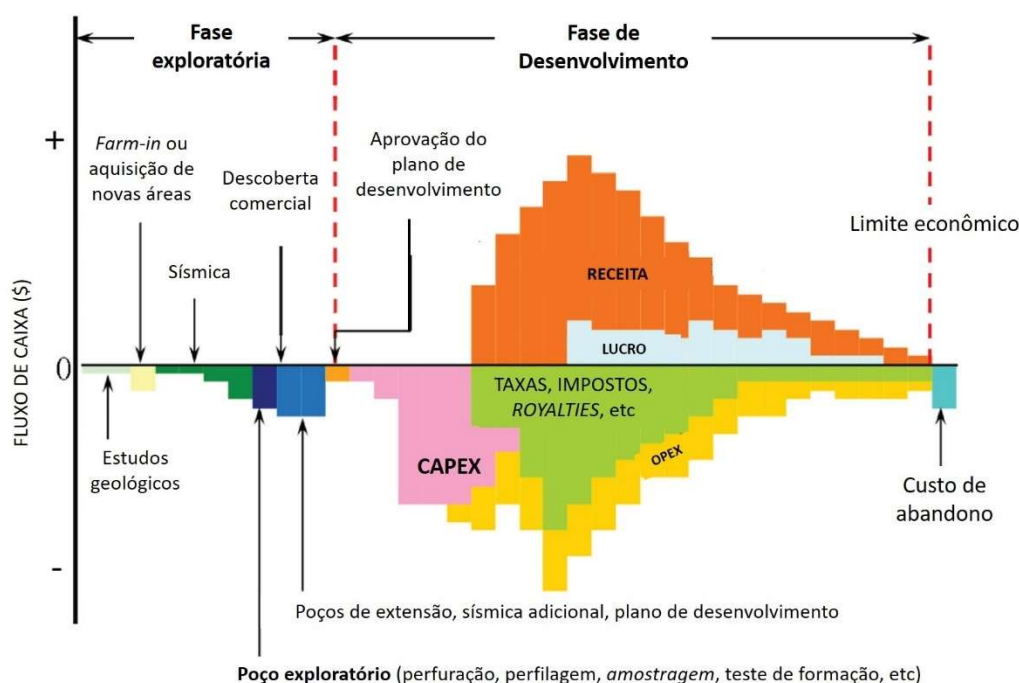


Figura 1 - Fluxo de caixa típico de um projeto de exploração e produção E&P sob regime fiscal do Brasil (adaptado de Suslick, 2009).

Para suportar tais decisões são fundamentais duas avaliações: *i)* das chances de sucesso geológico de um poço exploratório, ou seja, a probabilidade de que tal poço resulte na descoberta de uma acumulação de hidrocarboneto; e *ii)* das estimativas das dimensões ou volume dos recursos associados à eventual descoberta, avaliação fundamental para a determinação da comercialidade de um eventual descoberta.

A primeira avaliação, ainda que subsidiada por um grande volume de dados e informações de diferentes naturezas, reflete o grau de confiança da equipe de *exploracionistas* e guarda, portanto, certo grau de subjetividade (Newendorp, 2014). A chance ou fator de sucesso geológico de uma oportunidade descreve a probabilidade da descoberta de uma acumulação

de hidrocarboneto definindo um único valor entre 0 e 1 ou, em termos de porcentagem, entre 0 e 100%. Uma vez perfurado o poço exploratório, seu resultado será binário: 1, caso seja um sucesso resultando em uma descoberta, ou 0, caso não encontre uma acumulação. Assim, a verificação da avaliação de risco é imediata.

Quanto às estimativas de volumes das eventuais acumulações, estas são descritas em termos de uma distribuição de probabilidades cuja verificação não é tão simples e, mesmo com alto grau de incerteza, geralmente não ocorre antes da perfuração de ao menos mais um poço de extensão ou delimitação. À rigor, a definição mais precisa das reservas para a sua comparação com as estimativas da fase exploratória ocorreria apenas ao término da exploração do campo.

Tomadas tais considerações e com o objetivo de analisar as predições frente aos resultados obtidos pelos poços, serão considerados no presente trabalho apenas as probabilidades ou chances de sucesso esperados para os poços exploratórios e os seus respectivos resultados.

2.1 Definição das Chances ou Probabilidades de Sucesso

O objetivo da análise de risco geológico é determinar a probabilidade da perfuração de um prospecto – um alvo potencial - resultar em uma descoberta ou não, com base nos critérios geológicos disponíveis. O modelo mais comum adotado na indústria de O&G consiste na decomposição do problema em componentes ou fatores menores utilizando o conceito de *sistema petrolífero* onde são identificados os diversos elementos que controlam a existência de uma jazida de hidrocarboneto em uma bacia sedimentar. Um sistema petrolífero ativo compreende a existência e o funcionamento síncronos de quatro elementos - rochas geradoras maduras, rochas reservatório, rochas selantes e trapas, e dois fenômenos geológicos dependentes do tempo - migração e sincronismo (Milani et al., 2000).

Assim, à semelhança de uma solução de Fermi (Asorey, 2011), a complexa avaliação da probabilidade de uma descoberta pode ser obtida pela multiplicação das estimativas das probabilidades individuais para um conjunto de fatores mais restritos e, portanto, mais simples. Cada companhia tem sua própria metodologia e cultura acerca dos fatores geológicos considerados e como estes são definidos, mas tipicamente são utilizados entre quatro e seis fatores que normalmente incluem alguns dos elementos e fenômenos que definem um sistema petrolífero.

No presente estudo são considerados seis fatores, e suas respectivas probabilidades de ocorrência P , na composição da probabilidade de uma descoberta (Fator de chance ou Probabilidade de Sucesso), a saber: *Geração (gr)*, *Migração (mg)*, *Reservatório (rs)*,

Geometria (gm), *Retenção (rt)* e *Sincronismo (sn)*. Assim, para uma dada oportunidade *i* a Probabilidade de Sucesso $PS(i)$ é dada pela simples multiplicação:

$$PS(i) = Pgr(i) * Pmg(i) * Prs(i) * Pgm(i) * Prt(i) * Psn(i)$$

As avaliações dos termos à direita na equação, as probabilidades dos fatores geológicos condicionantes de uma acumulação, são mais específicas e podem ser mais objetivas a depender dos dados, das informações e conhecimento técnico sobre a área investigada.

Nestes termos, as probabilidades estimadas, sejam dos fatores geológicos ou de sucesso, vão variar entre 0 e 1 e representam um julgamento sobre o risco de ocorrência de um evento cujo resultado real, obtido ao final da perfuração do poço, será 0 ou 1.

2.2 Proposta da Prova de Conceito

A Prova de Conceito (POC) proposta consiste em partir de um conjunto de dados compostos por centenas de poços exploratórios perfurados nos últimos anos em diversas bacias, em sua maioria brasileiras, abrangendo diferentes *plays* e contextos geológicos e então avaliar as previsões relativas à ocorrência de todos os elementos do sistema petrolífero e a probabilidade de sucesso frente aos resultados constatados após a perfuração.

Uma primeira análise pretende estimar a acurácia das previsões caracterizando-as e identificando a presença de eventual viés. Em seguida, espera-se identificar padrões e fatores chaves ou comuns aos diferentes grupos de previsões. Tais padrões e fatores poderiam agregar alguma informação ao identificar elementos críticos das avaliações e auxiliar no reconhecimento dos aspectos das previsões que podem ser melhorados.

3 METODOLOGIAS

3.1 Avaliação de Previsões Probabilísticas

Tendo um conjunto de previsões e conhecendo os resultados podemos avaliar a acurácia dessas estimativas. O problema das previsões de resultados de poços exploratórios é binário, seu resultado é 0, não há acumulação, ou 1, houve uma descoberta, e se assemelha a um outro problema também binário bastante estudado em meteorologia, a previsão da ocorrência de chuvas em determinado tempo e local. Em ambos os casos há uma estimativa da chance de ocorrência de um evento, uma descoberta de hidrocarboneto ou a precipitação, variando entre 0 e 1 ou em termos de porcentagem de 0 a 100%. Ao final a *natureza* definirá se tal evento ocorre (1) ou não (0).

3.2 Calibração

Uma das primeiras e mais simples análises que podem ser feitas a partir de um conjunto de dados dessa natureza, constituído por pares de previsões de probabilidades P_j e resultados binários observados O_j , a Calibração descreve o alinhamento entre as previsões de probabilidade e as taxas dos resultados quando agrupados segundo algum critério (Figura 2). Tal análise permite a caracterização das previsões de probabilidades quanto a falta ou o excesso de confiança ou como otimistas ou pessimistas.

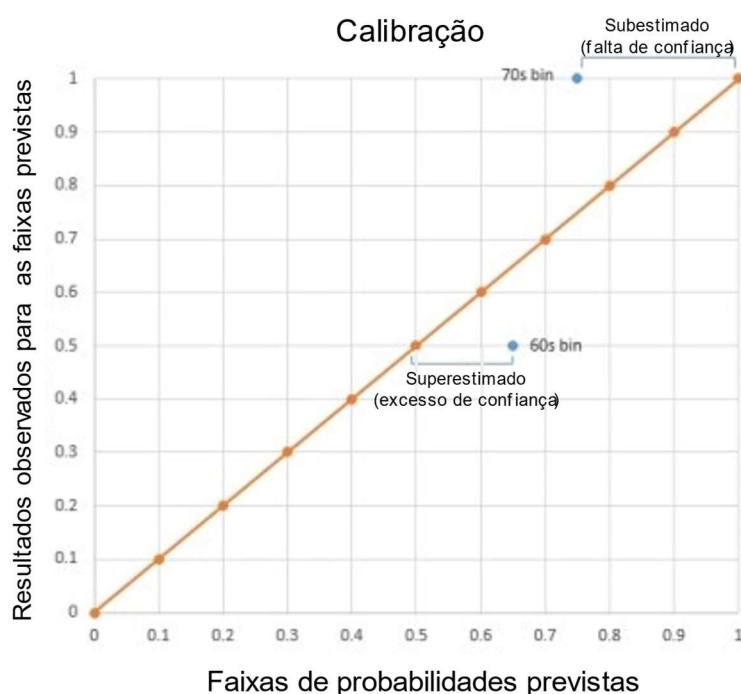


Figura 2 - Calibração

3.3 Índice de Brier

Proposto originalmente, ainda nos anos 50, para quantificar a acurácia de previsões meteorológicas, o **Índice de Brier** ou **Brier Score** (Brier, 1950) pode ser utilizado para verificar a acurácia de qualquer previsão probabilística de eventos binários ao medir a distância entre o resultado observado e a previsão.

Matematicamente, o Índice de Brier (BS) é a média quadrada da diferença entre as previsões probabilísticas P_j e o resultado binário observado O_j , dada uma amostra de N eventos j com os pares P_j e O_j , tal que

$$BS = \frac{1}{N} \sum_j^N (P_j - O_j)^2$$

Assim, o Índice de Brier é orientado negativamente, ou seja, quanto menor o seu valor melhor será a avaliação da previsão. A melhor avaliação possível é um BS igual a 0, e a pior possível é 1. Vale observar que um conjunto de previsões aleatórias tende a um valor de BS igual a 0,5 ao passo que um conjunto de hipotético de previsões “em cima do muro”, todas com 50% de chance, resultaria em um Índice de Brier de 0,25.

3.4 Decomposição do Índice de Brier

Como proposto por Murphy (1973), o Índice de Brier pode ainda ser decomposto na soma de três componentes: **Incerteza**, **Confiabilidade** e **Resolução**, caracterizando diferentes atributos das previsões e fornecendo informações úteis para o previsor sobre como as previsões podem ser melhoradas (Stephenson, 2008).

Assim,

$$BS = \text{Confiabilidade} - \text{Resolução} + \text{Incerteza}$$

Para definir os componentes é importante considerar a segmentação das previsões em m compartimentos ou subconjuntos k onde cada compartimento contém n_k termos com as mesmas probabilidades ou dentro de uma faixa comum de probabilidades, por exemplo, um mesmo decil.

A Confiabilidade é uma medida geral de calibração, sendo calculada como a média quadrada das diferenças entre as probabilidades previstas para um dado compartimento (P_k) e as taxas base dos resultados referentes ao mesmo compartimento (O_k),

$$\text{Confiabilidade} = \frac{1}{N} \sum_{k=1}^m (\overline{P}_k - \overline{O}_k)^2 n_k$$

O índice de confiabilidade também é orientado negativamente, sendo melhor quando a média quadrática é minimizada, portanto, quanto menor o seu valor maior o grau de confiabilidade do conjunto de previsões.

A Resolução se refere ao conjunto de previsões e mede o grau em que as previsões se enquadram em subconjuntos k com taxas de resultados \overline{O}_k diferentes da taxa base de resultados geral \bar{O} , calculada como a diferença média quadrática.

$$Resolução = \frac{1}{N} \sum_{k=1}^m (\overline{O}_k - \bar{O})^2 n_k$$

A Resolução mede a variação na média de resultados em diferentes subconjuntos indicando o grau em que a previsão pode separar diferentes situações. Assim, quanto maior a Resolução, melhor. Um previsor com maior resolução faz previsões que são mais distantes da taxa base geral do que um previsor com uma pontuação mais baixa e, nesse sentido, mais interessante ou ousado.

A Incerteza mede a variabilidade das observações e indica o grau de predição inerente aos resultados observados, sendo, portanto, função apenas das observações e não das previsões de probabilidades. O grau de Incerteza é calculado a partir da taxa base de ocorrência do resultado considerado \bar{O} ,

$$Incerteza = taxa\ base * (1 - taxa\ base)$$

ou

$$Incerteza = \bar{O} * (1 - \bar{O})$$

Assume o valor máximo, máxima incerteza igual 0,25, quando os resultados possíveis ocorrem 50% das vezes cada um e valor mínimo, igual a 0, quando apenas um dos resultados ocorre em todas as oportunidades. Quanto maior o grau de Incerteza de um domínio, maior será o Índice de Brier, portanto, não devem ser feitas comparações diretas dos BS entre distintos conjuntos de previsões a menos que se refiram aos mesmos eventos ou que tenham o mesmo nível de Incerteza.

4 ARQUITETURA DO SISTEMA PROPOSTO

Neste trabalho é proposto uma avaliação ou análise das previsões de poços exploratórios que permita o exame do conjunto de dados de forma interativa, com ferramentas de seleção e segmentação e utilizando recursos gráficos e analíticos buscando fornecer o máximo de informação ao analista, para tanto optou-se pela utilização de uma ferramenta ou sistema de *Business Intelligence*.

Business Intelligence (BI) é um processo, baseado em tecnologia, que compreende um conjunto de técnicas e ferramentas utilizadas na coleta, organização, transformação, integração e disseminação das informações úteis. Utilizados de forma integrada, tais recursos, que possibilitam o monitoramento de resultados e auxiliam a interpretação e análise de dados e informações, permitem a identificação de oportunidades e riscos e suportam a tomada de decisão.

A hierarquia DIKW (acrônimo do inglês **D**ata, **I**nformation, **K**nowledge e **W**isdom) ou Pirâmide do Conhecimento representa as características do desafio da transformação de dados em sabedoria para o suporte a tomada de decisões (Figura 3).



Figura 3 - Hierarquia DIKW ou DICS (Dados, Informação, Conhecimento e Sabedoria).

Apresentadas como soluções de gerenciamento de informações e visualização de dados que permitem a rápida análise, as plataformas de BI são intuitivas e oferecem diversos recursos customizáveis, com elementos interativos, como filtros e seleções, e possibilidade de atualização em tempo real apresentando uma visão geral da informação de forma rápida e dinâmica.

Ao permitirem o acesso, a análise conjunta, a exibição e o compartilhamento de dados das mais diversas fontes, sejam arquivos, bases de dados, *web* etc., as plataformas de BI são ferramentas fundamentais de suporte a tomada de decisão.

O *software* de BI utilizado para o desenvolvimento deste trabalho foi o *Power BI* da *Microsoft*.

- Base de dados

A principal base de dados utilizada neste trabalho é composta por uma amostra de poços exploratórios perfurados ao longo de 15 anos pela PETROBRAS em terra ou em mar em diversas bacias sedimentares do Brasil, predominantemente, e de outros países. Para cada oportunidade há, além de informações como a identificação do poço e do objetivo, as previsões ou expectativas das probabilidades de ocorrência de hidrocarboneto e de cada um dos elementos do sistema petrolífero bem como os respectivos resultados constatados.

O conjunto de dados em questão provêm do Banco de Dados Oficial de Oportunidades Exploratórias (BDOPEX), foi exportado em formato de planilha *Microsoft Excel* (*.xlsx) e conta com quase 400 poços pioneiros, por vezes com mais de um objetivo totalizando cerca de 580 oportunidades avaliadas.

Apesar de composta por cerca de 150 colunas, com dados mais que suficientes para atender a análise proposta, tal planilha não conta com alguns dados relevantes relativos aos poços tais como localização geográfica precisa e datas de início e conclusão da perfuração. Considerando que essas outras informações são públicas e que poderiam enriquecer o estudo recorreu-se a uma base complementar a ser integrada ao trabalho. Avaliando fatores como disponibilidade, estruturação e semelhanças entre as bases calhou a utilização da tabela de poços do Monitor de Perfuração disponibilizado *web* pela agência **epbr** em formato de texto delimitado (*.csv) (<https://monitor-perf.herokuapp.com/assets/data.csv>).

- Estrutura

Por se tratar de uma prova de conceito simples, com o principal objetivo de verificar a validade da metodologia proposta, alimentada basicamente por uma única fonte de dados, a planilha proveniente do BDOPEX, e por não haver perspectivas de atualização dos dados uma vez carregados optou-se pela elaboração de um projeto de ETL (*Extract, Transform and Load* ou Extrair, Transformar e Carregar) simplificado e realizado diretamente no *software* Microsoft Power BI.

A extração dos dados foi realizada de maneira que permitisse a montagem de uma estrutura dimensional no formato esquema estrela (Figura 4 - A estrutura dimensional desenhada no formato do esquema estrela (star schema).Figura 4) composta por uma tabela fato

(Oportunidades_ft), relativa às oportunidades exploratórias testadas, e as demais dimensões relativas às previsões, resultados e localização dos poços considerados. Esta última dimensão, de caráter complementar, mas não fundamental, conta com as coordenadas geográficas dos poços e foi obtida a partir do Monitor de Perfuração (epbr) e integrada à estrutura utilizando alguns recursos de transformação dos dados.

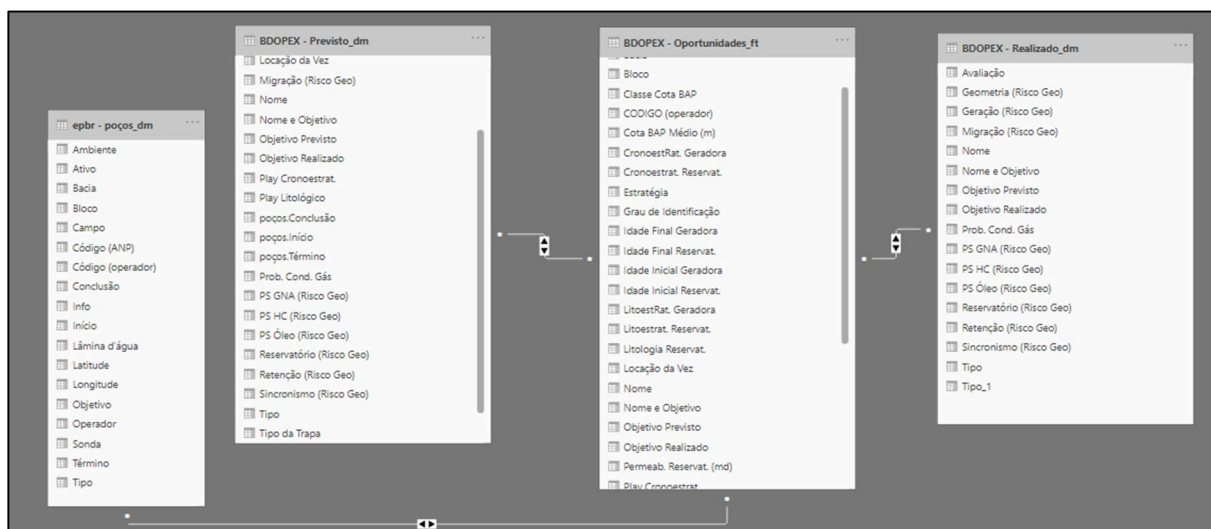


Figura 4 - A estrutura dimensional desenhada no formato do esquema estrela (star schema).

Uma vez carregados os dados e realizadas as transformações necessárias, foi criada uma tabela dinâmica para a alocação dos Índices de Brier e seus componentes, adicionando mais uma dimensão ao esquema estrela (Figura 5). Utilizando fórmulas DAX (*Data Analysis Expressions* ou Expressões de Análise de Dados), foram implementados os cálculos dos índices e componentes para as avaliações das previsões de sucesso de cada oportunidade e da ocorrência dos seus respectivos elementos do sistema petrolífero (Figura 6).

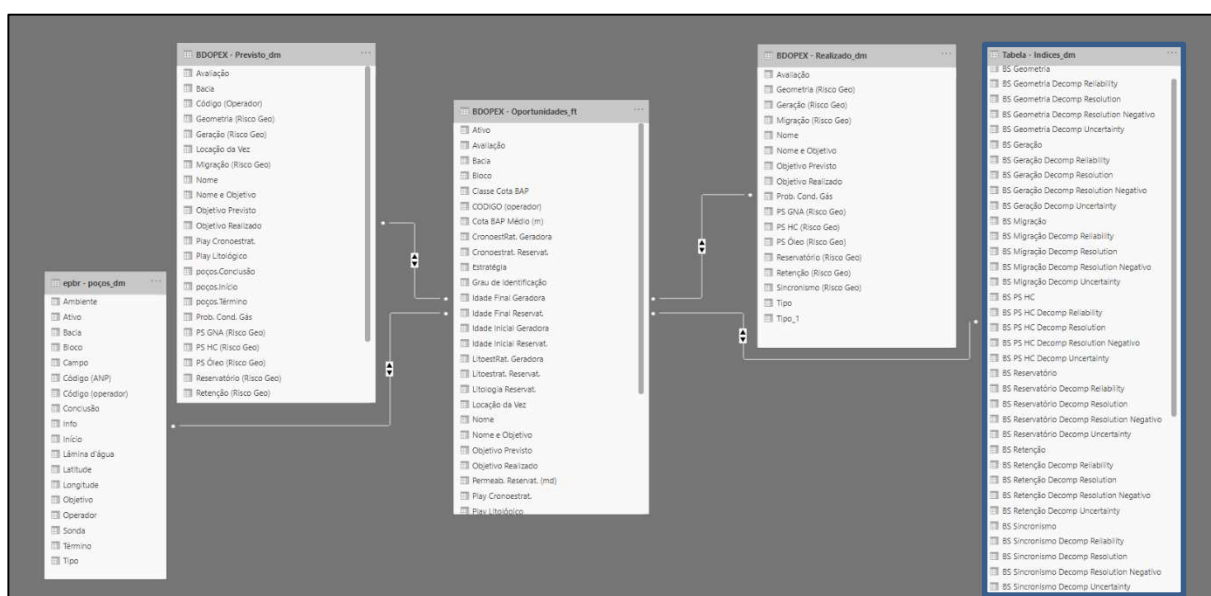


Figura 5 - A estrutura dimensional no formato do esquema estrela com a adição da tabela dinâmica com Índices de Brier e seus componentes (Tabela - Indices_dm), em destaque.

```

1 Tabela Índices = SUMMARIZE('DOPEX - Previsto', 'DOPEX - Previsto'[Nome e objetivo], 'DOPEX - Previsto'[Tipo de Play], 'DOPEX - Previsto'[Bacia], 'DOPEX - Previsto'[Nome], 'DOPEX - Previsto'[Objetivo Realizado], 'DOPEX - Previsto'[Objetivo Previsto],
2
3 "Gerado Result", AVERAGE(FILTER('DOPEX - Realizado', CONCATENATE(CONCATENATE('DOPEX - Realizado'[Nome], 'DOPEX - Realizado'[Objetivo Previsto]), 'DOPEX - Realizado'[Objetivo Realizado])=CONCATENATE(CONCATENATE('DOPEX - Previsto'[Nome], 'DOPEX - Previsto'[Objetivo Previsto]), 'DOPEX - Previsto'[Objetivo Realizado])), 'DOPEX - Realizado'[Gerado (Risco Geo)]),
4
5 "Migração Result", AVERAGE(FILTER('DOPEX - Realizado', CONCATENATE(CONCATENATE('DOPEX - Realizado'[Nome], 'DOPEX - Realizado'[Objetivo Previsto]), 'DOPEX - Realizado'[Objetivo Realizado])=CONCATENATE(CONCATENATE('DOPEX - Previsto'[Nome], 'DOPEX - Previsto'[Objetivo Previsto]), 'DOPEX - Previsto'[Objetivo Realizado])), 'DOPEX - Realizado'[Migração (Risco Geo)]),
6
7 "Desenvolvimento Result", AVERAGE(FILTER('DOPEX - Realizado', CONCATENATE(CONCATENATE('DOPEX - Realizado'[Nome], 'DOPEX - Realizado'[Objetivo Previsto]), 'DOPEX - Realizado'[Objetivo Realizado])=CONCATENATE(CONCATENATE('DOPEX - Previsto'[Nome], 'DOPEX - Previsto'[Objetivo Previsto]), 'DOPEX - Previsto'[Objetivo Realizado])), 'DOPEX - Realizado'[Desenvolvimento (Risco Geo)]),
8
9 "Geometria Result", AVERAGE(FILTER('DOPEX - Realizado', CONCATENATE(CONCATENATE('DOPEX - Realizado'[Nome], 'DOPEX - Realizado'[Objetivo Previsto]), 'DOPEX - Realizado'[Objetivo Realizado])=CONCATENATE(CONCATENATE('DOPEX - Previsto'[Nome], 'DOPEX - Previsto'[Objetivo Previsto]), 'DOPEX - Previsto'[Objetivo Realizado])), 'DOPEX - Realizado'[Geometria (Risco Geo)]),
10
11 "Retardo Result", AVERAGE(FILTER('DOPEX - Realizado', CONCATENATE(CONCATENATE('DOPEX - Realizado'[Nome], 'DOPEX - Realizado'[Objetivo Previsto]), 'DOPEX - Realizado'[Objetivo Realizado])=CONCATENATE(CONCATENATE('DOPEX - Previsto'[Nome], 'DOPEX - Previsto'[Objetivo Previsto]), 'DOPEX - Previsto'[Objetivo Realizado])), 'DOPEX - Realizado'[Retardo (Risco Geo)]),
12
13 "Sincronismo Result", AVERAGE(FILTER('DOPEX - Realizado', CONCATENATE(CONCATENATE('DOPEX - Realizado'[Nome], 'DOPEX - Realizado'[Objetivo Previsto]), 'DOPEX - Realizado'[Objetivo Realizado])=CONCATENATE(CONCATENATE('DOPEX - Previsto'[Nome], 'DOPEX - Previsto'[Objetivo Previsto]), 'DOPEX - Previsto'[Objetivo Realizado])), 'DOPEX - Realizado'[Sincronismo (Risco Geo)]),
14
15 "PS Óleo Result", AVERAGE(FILTER('DOPEX - Realizado', CONCATENATE(CONCATENATE('DOPEX - Realizado'[Nome], 'DOPEX - Realizado'[Objetivo Previsto]), 'DOPEX - Realizado'[Objetivo Realizado])=CONCATENATE(CONCATENATE('DOPEX - Previsto'[Nome], 'DOPEX - Previsto'[Objetivo Previsto]), 'DOPEX - Previsto'[Objetivo Realizado])), 'DOPEX - Realizado'[PS Óleo (Risco Geo)]),
16
17 "PS Gás Result", AVERAGE(FILTER('DOPEX - Realizado', CONCATENATE(CONCATENATE('DOPEX - Realizado'[Nome], 'DOPEX - Realizado'[Objetivo Previsto]), 'DOPEX - Realizado'[Objetivo Realizado])=CONCATENATE(CONCATENATE('DOPEX - Previsto'[Nome], 'DOPEX - Previsto'[Objetivo Previsto]), 'DOPEX - Previsto'[Objetivo Realizado])), 'DOPEX - Realizado'[PS Gás (Risco Geo)]),
18
19 "PS HC Result", AVERAGE(FILTER('DOPEX - Realizado', CONCATENATE(CONCATENATE('DOPEX - Realizado'[Nome], 'DOPEX - Realizado'[Objetivo Previsto]), 'DOPEX - Realizado'[Objetivo Realizado])=CONCATENATE(CONCATENATE('DOPEX - Previsto'[Nome], 'DOPEX - Previsto'[Objetivo Previsto]), 'DOPEX - Previsto'[Objetivo Realizado])), 'DOPEX - Realizado'[PS HC (Risco Geo)]),
20
21 "ES Geometria", AVERAGE('DOPEX - Previsto', 'DOPEX - Previsto'[Geometria (Risco Geo)]) - AVERAGE('DOPEX - Previsto', 'DOPEX - Previsto'[Geometria (Risco Geo)])/2,
22
23 "ES Geometria Decomposição", (
24 (CALCULATE(AVERAGE('DOPEX - Previsto', 'DOPEX - Previsto'[Geometria (Risco Geo)]), 'DOPEX - Previsto'[Geometria (Risco Geo)]>0.1)/CALCULATE(AVERAGE('DOPEX - Realizado', 'DOPEX - Realizado'[Geometria (Risco Geo)]), 'DOPEX - Realizado'[Geometria (Risco Geo)]>0.1)/2)+CALCULATE(COUNTROWS
('DOPEX - Previsto', 'DOPEX - Previsto'[Geometria (Risco Geo)]>0.1)
25 (CALCULATE(AVERAGE('DOPEX - Previsto', 'DOPEX - Previsto'[Geometria (Risco Geo)]), 'DOPEX - Previsto'[Geometria (Risco Geo)]>0.1)/CALCULATE(AVERAGE('DOPEX - Realizado', 'DOPEX - Realizado'[Geometria (Risco Geo)]), 'DOPEX - Realizado'[Geometria (Risco Geo)]>0.1)/2)+CALCULATE(AVERAGE('DOPEX - Previsto', 'DOPEX - Previsto'[Geometria (Risco Geo)]), 'DOPEX - Previsto'[Geometria (Risco Geo)]>0.1)
26 (CALCULATE(AVERAGE('DOPEX - Previsto', 'DOPEX - Previsto'[Geometria (Risco Geo)]), 'DOPEX - Previsto'[Geometria (Risco Geo)]>0.2)/CALCULATE(AVERAGE('DOPEX - Realizado', 'DOPEX - Realizado'[Geometria (Risco Geo)]), 'DOPEX - Realizado'[Geometria (Risco Geo)]>0.2)+CALCULATE(AVERAGE('DOPEX - Previsto', 'DOPEX - Previsto'[Geometria (Risco Geo)]), 'DOPEX - Previsto'[Geometria (Risco Geo)]>0.2)
27 (CALCULATE(AVERAGE('DOPEX - Previsto', 'DOPEX - Previsto'[Geometria (Risco Geo)]), 'DOPEX - Previsto'[Geometria (Risco Geo)]>0.3)/CALCULATE(AVERAGE('DOPEX - Realizado', 'DOPEX - Realizado'[Geometria (Risco Geo)]), 'DOPEX - Realizado'[Geometria (Risco Geo)]>0.3)+CALCULATE(AVERAGE('DOPEX - Previsto', 'DOPEX - Previsto'[Geometria (Risco Geo)]), 'DOPEX - Previsto'[Geometria (Risco Geo)]>0.3)
28 (CALCULATE(AVERAGE('DOPEX - Previsto', 'DOPEX - Previsto'[Geometria (Risco Geo)]), 'DOPEX - Previsto'[Geometria (Risco Geo)]>0.4)/CALCULATE(AVERAGE('DOPEX - Realizado', 'DOPEX - Realizado'[Geometria (Risco Geo)]), 'DOPEX - Realizado'[Geometria (Risco Geo)]>0.4)+CALCULATE(AVERAGE('DOPEX - Previsto', 'DOPEX - Previsto'[Geometria (Risco Geo)]), 'DOPEX - Previsto'[Geometria (Risco Geo)]>0.4)
29 (CALCULATE(AVERAGE('DOPEX - Previsto', 'DOPEX - Previsto'[Geometria (Risco Geo)]), 'DOPEX - Previsto'[Geometria (Risco Geo)]>0.5)/CALCULATE(AVERAGE('DOPEX - Realizado', 'DOPEX - Realizado'[Geometria (Risco Geo)]), 'DOPEX - Realizado'[Geometria (Risco Geo)]>0.5)+CALCULATE(AVERAGE('DOPEX - Previsto', 'DOPEX - Previsto'[Geometria (Risco Geo)]), 'DOPEX - Previsto'[Geometria (Risco Geo)]>0.5)
30

```

Figura 6 – Exemplo da edição de funções DAX utilizadas na implementação dos cálculos dos índices e componentes que avaliam, para cada oportunidade, as previsões de sucesso e seus respectivos elementos do sistema petrolífero

- Descaracterização e adição de ruído

Por se tratar de dados sensíveis e confidenciais é necessário a descaracterização de algumas informações. Uma vez que os dados serão analisados, em geral, de forma conjunta não há a necessidade de anonimização das oportunidades sendo suficiente a codificação das categorias utilizadas nas análises. Por exemplo, os nomes das bacias sedimentares e a identificação dos tipos de *plays*, que são, ambos, importantes critérios para agrupar e analisar os dados, foram substituídos pelos nomes das letras do alfabeto grego (alfa, beta, gama etc.) e pelas letras do alfabeto romano (tipo A, tipo B, tipo C etc.), respectivamente; os elementos do sistema petrolíferos, por sua vez, não foram especificados, mas apenas enumerados de um a seis.

Ainda que tais medidas não permitam a identificação, no trabalho publicado, do poço ou da bacia a qual as avaliações consideradas fazem referência, as grandezas e índices obtidos através das análises teriam por si só alguma informação potencialmente relevante. Dada a sensibilidade dos dados em questão e considerando que o presente trabalho se restringe à proposição de uma metodologia de avaliação das previsões e não trata da análise de tais avaliações foi adicionado uma série de ruídos não-aleatórios.

O ruído, elaborado com cinco níveis, pode aumentar, artificialmente, de 20% a até 100% o número de oportunidades consideradas ao realizar uma determinada análise. Os elementos adicionais, à título de ruído, apresentam valores de previsões coerentes com a distribuição dos dados reais, mas propositalmente geram distorções e criam vieses. Portanto, a adição do ruído não altera a natureza do dado, mas o afasta da sua representação original.

5 RESULTADOS

Os resultados são apresentados através de tabelas e, principalmente, de gráficos dinâmicos disponibilizados por uma série de painéis desenvolvidos utilizando o Microsoft Power BI. Os recursos gráficos permitem a rápida análise dos Índices de Brier e dos seus componentes frente a diferentes critérios de agrupamento tais como as bacias sedimentares ou os diferentes tipos de *plays*.

A seguir são apresentados alguns exemplos dos painéis e gráficos mais relevantes com os quais é possível explorar os dados e analisar as avaliações das previsões das chances de sucesso (PS HC) das oportunidades e da ocorrência dos elementos do sistema petrolíferos (ESP). Vale lembrar que as análises apresentadas não correspondem aos dados originais. Considerando a sensibilidade dos dados utilizados e das informações potencialmente geradas houve a introdução de um viés aos dados, contaminados intencionalmente por diferentes ruídos, além da descaracterização das principais categorias que agrupam as oportunidades e dos diferentes elementos do sistema petrolífero.

Nos primeiros painéis é possível explorar os dados identificando o número e a distribuição de todos os poços exploratórios perfurados no Brasil de acordo com diferentes critérios como bacias sedimentares, ambiente, período de perfuração, dentre outros, bem como a verificação em mapa das posições geográficas correspondentes (Figura 7). Aplicando um filtro simples é possível restringir tais análises ao conjunto de poços exploratórios correspondentes às oportunidades cujas previsões foram consideradas no presente trabalho.

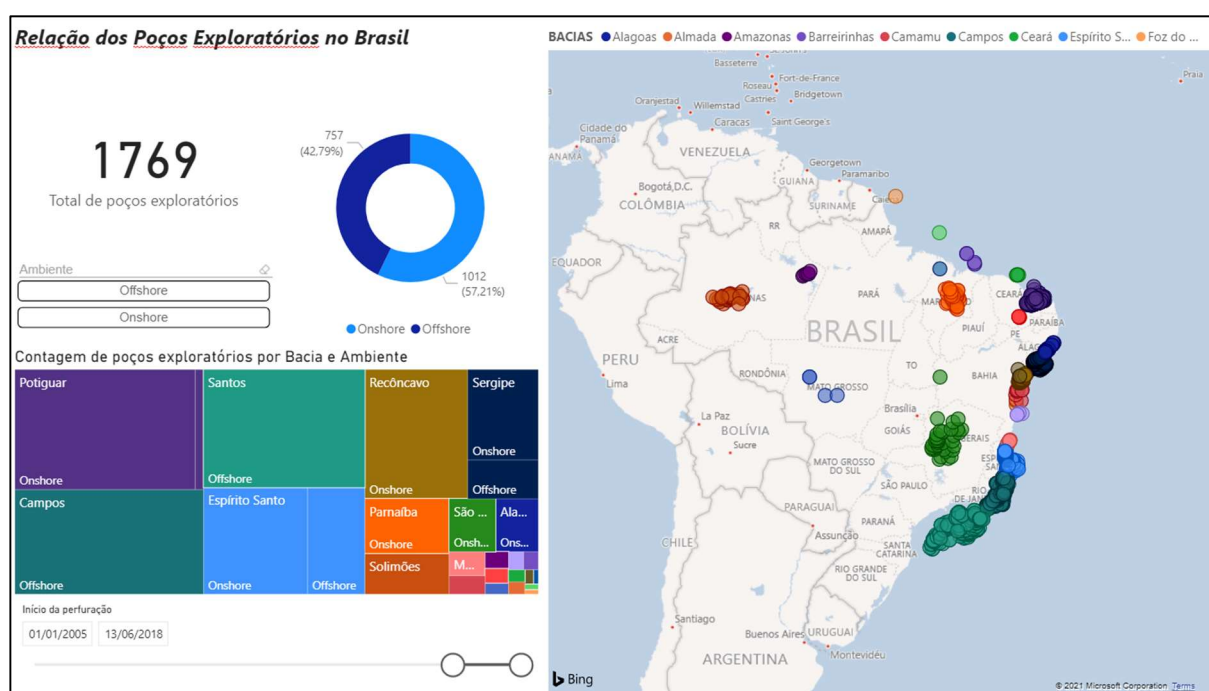


Figura 7 - Exploração dos dados referentes aos poços exploratórios, números, distribuição em função das bacias, ambiente onshore/offshore, data de perfuração e localização em mapa dos poços perfurados entre 2005 e 2018.

Considerando as avaliações das previsões das oportunidades, a primeira análise corresponde a calibração das estimativas de risco frente aos resultados obtidos pelos poços exploratórios (Figura 8). Essa simples avaliação permite a identificação do caráter otimista ou pessimista das estimativas, seja para todo o conjunto das oportunidades testadas ou individualizado por segmentos tais como bacias e tipos de plays. Além dos gráficos, ao consultar a tabela auxiliar é possível verificar a estimativas de chance de sucesso e resultados médios para cada uma das bacias, tipo de *plays* ou as previsões e resultados de cada uma das oportunidades.

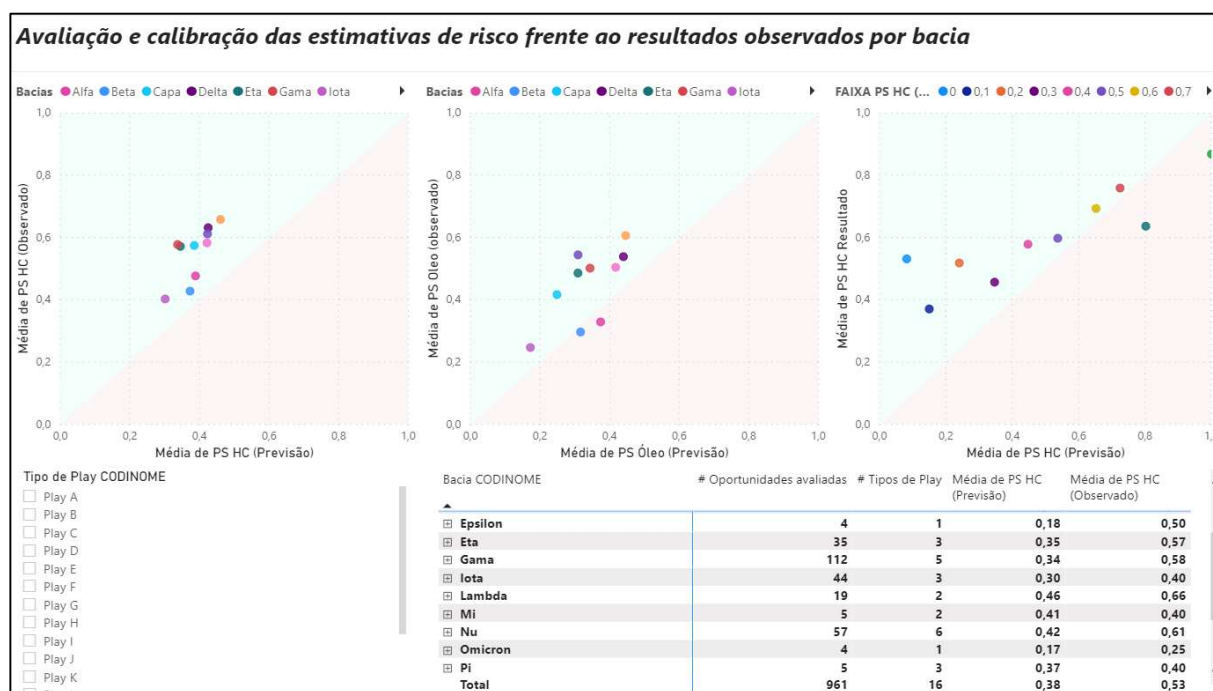


Figura 8 - Avaliação da calibração das estimativas ou previsões das chances de sucesso geológico (PS HC) frente aos resultados observados.

Na avaliação da calibração, idealmente, as previsões como o máximo de precisão e acurácia estariam distribuídas ao longo da diagonal. Previsões menos ou mais confiantes se afastariam da diagonal sendo a projeção horizontal de tal afastamento uma medida do excesso, à direita, ou da falta, à esquerda, de confiança.

Para exemplificar, a análise da calibração observada na Figura 8 sugere que o conjunto das estimativas médias tem um viés, predominantemente, pessimista ou pouco confiante uma vez que há o predomínio dos pontos na região acima da diagonal. Neste caso, as probabilidades de sucesso foram subestimadas ou, de outra forma, os riscos foram superestimados. Ainda seria possível identificar grupos com diferentes níveis ou graus de confiança.

Em outro painel são analisados os Índice de Brier e seus componentes, Confiabilidade Incerteza e Resolução, referentes às estimativas de probabilidade sucesso das oportunidades (PS HC) segmentados pelas diferentes bacias sedimentares e tipos de *plays* (Figura 9).



Figura 9 - Análise dos Índices de Brier e seus componentes, Confiabilidade, Incerteza e Resolução, relativos às previsões das chances de sucesso geológico PS HC para diferentes bacias sedimentares e tipos de plays.

Com o Índice de Brier é possível quantificar a acurácia das previsões probabilísticas dos eventos binários correspondentes à ocorrência ou não de hidrocarboneto. Quanto menor o seu valor, melhor será a avaliação da previsão. Já os componentes do índice caracterizam diferentes atributos das previsões e podem fornecer informações úteis sobre como as previsões podem ser melhoradas.

Em seguida são consideradas as avaliações de cada um dos elementos do sistema petrolífero (ESP). São as estimativas de ocorrência desses elementos que, conjuntamente, compõem a probabilidade de sucesso geológico de cada oportunidade. Podem ser analisadas as calibrações das estimativas de ocorrência de cada um dos elementos do sistema petrolífero frente ao constatado pelas análises dos poços exploratórios (Figura 10) bem como os respectivos Índices de Brier e componentes segmentados pelas bacias sedimentares (Figura 11) ou pelo tipo de *play* (Figura 12).

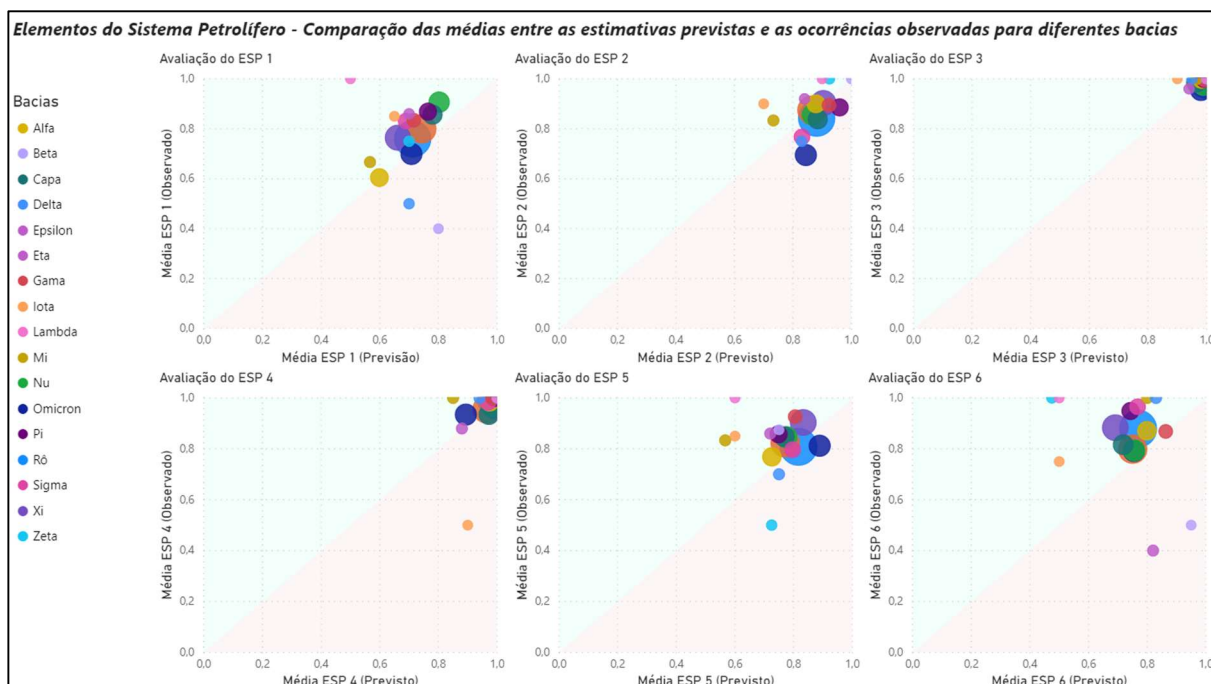


Figura 10 - Avaliação da calibração das estimativas de ocorrência de cada um dos elementos do sistema petrolífero (ESP).

Novamente a análise de calibração permite identificar, caso exista um, o viés das previsões de ocorrência de um determinado elemento do sistema petrolífero. Na Figura 10 há a sugestão de excesso de confiança na avaliação da ocorrência de alguns elementos para determinadas bacias sedimentares. Ao refinar a mesma análise utilizando outros critérios de agrupamento ou segmentação das oportunidades podem surgir elementos que expliquem ou identifiquem fatores que teriam, eventualmente, induzido uma avaliação superestimada.

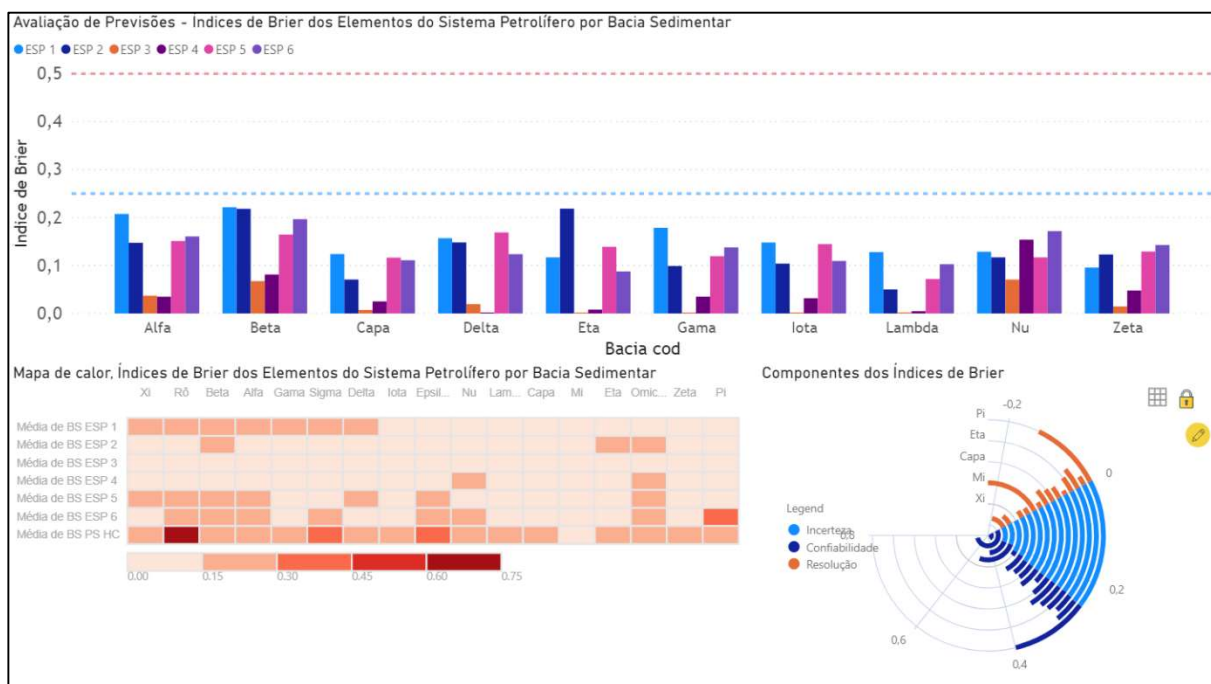


Figura 11 - Análise dos Índices de Brier e seus componentes, Confiabilidade, Incerteza e Resolução, relativos a cada um dos Elementos do Sistema Petrolífero (ESP) para diferentes bacias sedimentares. Matriz ou Mapa de Calor identifica os ESP críticos para cada bacia sedimentar.

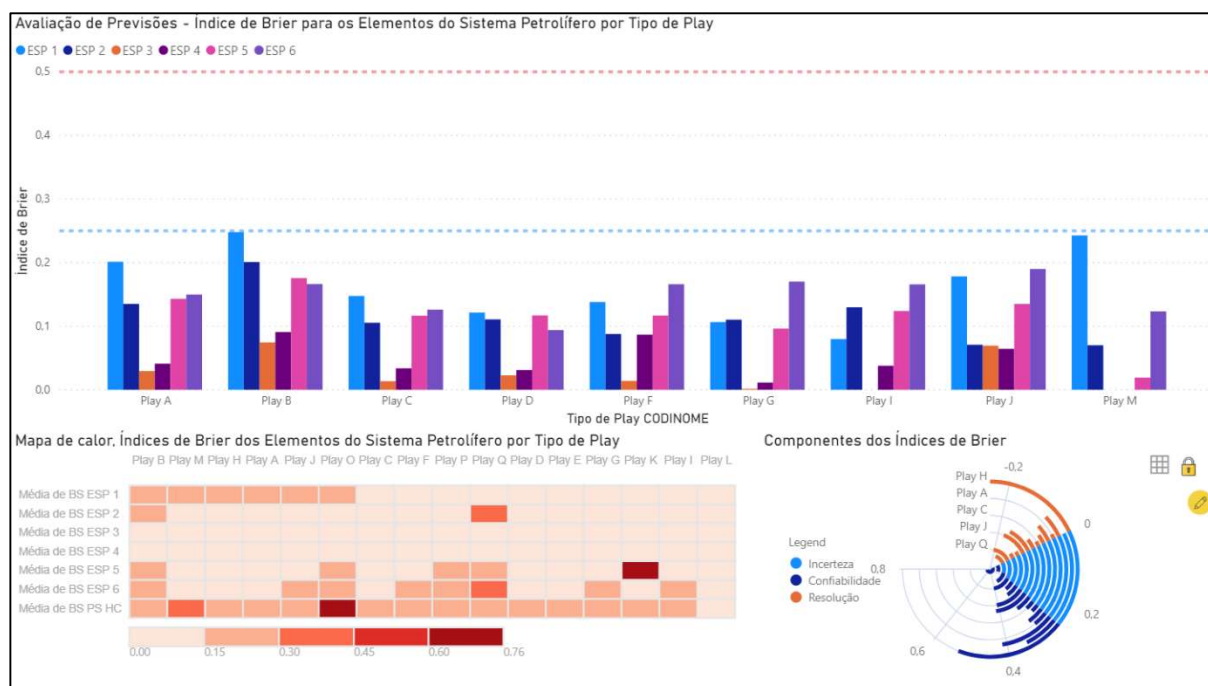


Figura 12 - Análise dos Índices de Brier e seus componentes, Confiabilidade, Incerteza e Resolução, relativos a cada um dos Elementos do Sistema Petrolífero (ESP) para diferentes tipos de plays. Matriz ou Mapa de Calor identifica os ESP críticos para cada tipo de play considerado.

Avaliações mais detalhadas dos Índices de Bier e componentes para as previsões dos elementos do sistema petrolífero, como as apresentadas nas Figura 11 Figura 12, permitem identificar de forma segmentada, seja para cada bacia sedimentar, para cada tipo de play, ou outra critério, quais elementos são críticos na determinação da probabilidade de sucesso atribuída a um oportunidade exploratória.

Aprofundando um pouco mais a análise dos elementos do sistema petrolífero é possível avaliar cada um através da calibração e dos Índices de Brier, acompanhando pelos seus componentes, com relação a diferentes critérios e fatores que fazem parte da base de dados utilizada ou que venham a ser integrados. No exemplo apresentado na Figura 13 as estimativas ou previsões das chances de ocorrência um determinado elemento do sistema petrolífero, identificado com ESP1, são avaliadas e comparadas para diferentes tipos de *plays*. Mas ainda no mesmo painel, utilizando o recurso *drill down* ou movendo através dos níveis de hierarquia, é possível avaliar o mesmo ESP 1 segundo diferentes critérios de segmentação alternando-os para bacias sedimentares ou para fatores mais específicos que possam, ocasionalmente, ser relevantes tais como *play* cronoestratigráfico, cronoestratigrafia e tipos de rochas tanto do reservatório quanto da geradora, tipo de trapas, entre outros (Figura 14).

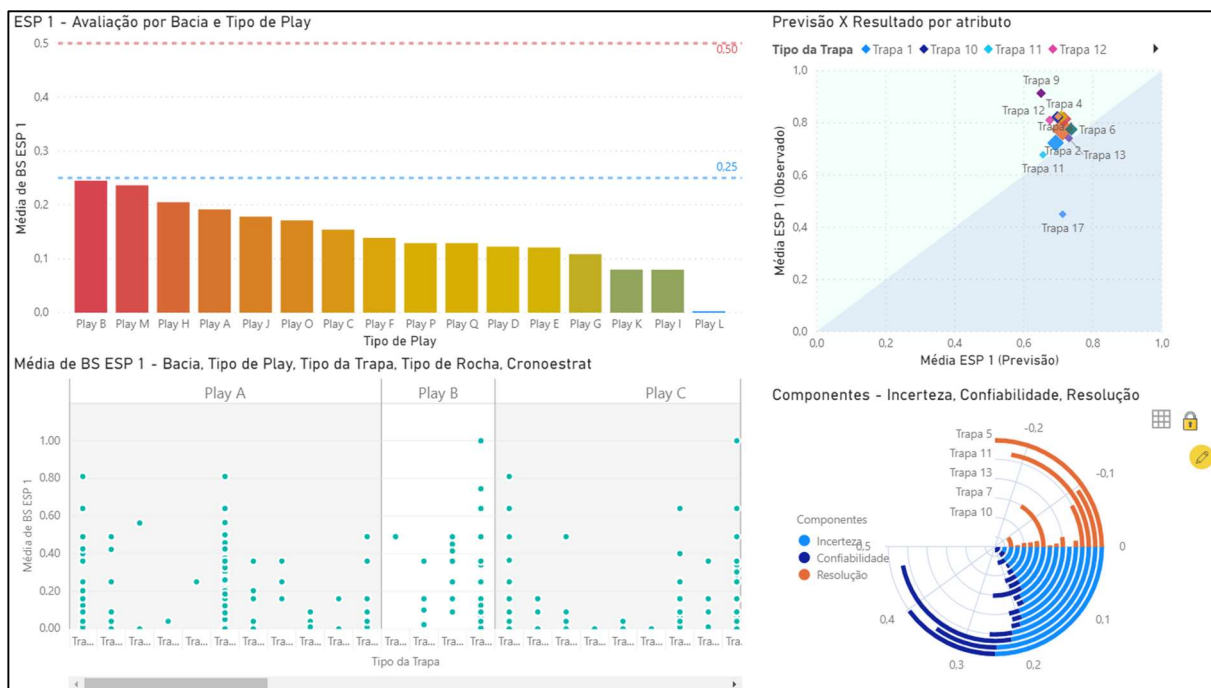


Figura 13 – Avaliação das estimativas de ocorrências de determinado elemento do sistema petrolífero ESP 1 a partir da calibração e dos Índices de Brier e seus componentes para diferentes tipos de plays. No gráfico inferior esquerdo, distribuição do Índices de Brier (BS) das oportunidades avaliadas em função dos tipos de plays e dos diferentes tipos de trapas.

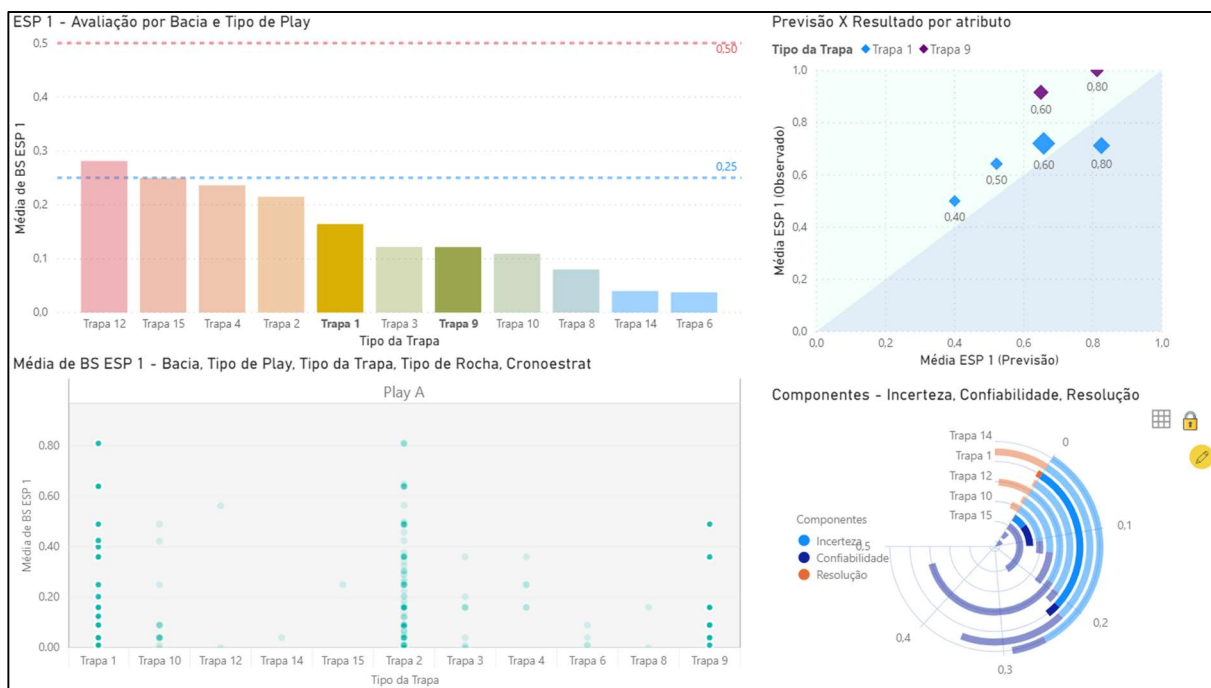


Figura 14 – Aplicando o recurso de drill down a avaliação mais detalhada das estimativas de ocorrências de ESP 1 e a variação dos correspondentes Índices de Brier (BS) para diferentes tipos de trapas. No exemplo, estão realçados os dados associados aos dois tipos de trapas seleccionados no gráfico superior esquerdo.

6 CONCLUSÕES E TRABALHOS FUTUROS

A prova de conceito desenvolvida atendeu às proposições fundamentais do presente trabalho ao apresentar uma avaliação relativamente simples das previsões de probabilidades de sucesso das oportunidades exploratórias. Através das análises, eminentemente gráficas dos Índices de Brier e seus componentes, foi possível verificar a acurácia das estimativas de ocorrência dos elementos do sistema petrolíferos e da acumulação de hidrocarboneto permitindo a identificação de eventuais padrões e vieses que poderiam caracterizar determinados conjuntos de avaliações. O modelo utilizado também possibilitou a comparação de tais aspectos frente a diferentes critérios de agrupamento das oportunidades, tais como bacias sedimentares, tipos de *plays*, tipos de rocha reservatório, dentre outros, demonstrando o potencial da utilização dos índices para a identificação de pontos críticos, vieses, padrões ou fatores-chaves que determinam as diferenças de acurácia das estimativas para os diversos elementos avaliados.

Por ora, não foi objetivo deste trabalho apresentar ou sugerir soluções para identificar eventuais elementos críticos para as avaliações, mas a partir da metodologia apresentada e dos resultados promissores seria interessante a integração dessas análises com outras bases de dados que agreguem mais informações sobre os diversos fatores que determinaram as previsões, permitindo a correlação e análise de cada um desses fatores com dos Índices. Espera-se, portanto, que esse tipo de avaliação permita a melhor alocação de recursos e auxilie a estimativa de valor da informação caso seja capaz de identificar os pesos de alguns fatores na determinação de acurácia das estimativas. Ao constatar vieses nas avaliações dos conjuntos de oportunidades testadas a aplicação também se apresenta como uma solução expedita para auxiliar na gestão de portfólio exploratório ao indicar ou sugerir que tipos de prospectos deveriam ser, preferencialmente, reavaliados.

Por fim, considerando que há um importante elemento subjetivo na elaboração das previsões de oportunidades e que não são comuns as análises sistemáticas da qualidade dos conjuntos de estimativas tais avaliações teriam importante papel na calibração dos previsores. Vale observar que em geral, os esforços de investigação são maiores por ocasião de um insucesso e se concentram objetivamente sobre os elementos que falharam. Em casos de sucesso, ainda menos atenção tende a ser dada às avaliações e, como observado com a introdução nos dados de um viés com falta de confiança, havendo um conjunto de previsões subestimadas (descalibradas) a aversão ao risco pode, eventualmente, afastar o exploracionista de boas oportunidades cujos riscos, neste caso, estariam superestimados.

Além da implementação de melhorias e integração de mais dados, uma linha a ser considerada em trabalhos futuros consiste na elaboração da avaliação de acurácia das estimativas de grandezas contínuas, sejam propriedades como porosidade e permeabilidade ou os volumes de reservas *in place*. Ainda que mais desafiadores,

Referências Bibliográficas

- Asorey, H., López Dávalos, A. (2011). Fermi Problem: Power developed at the eruption of the Puyehue-Cordón Caulle volcanic system in June 2011. arXiv:1109.1165v2 [physics.ed-ph] <https://arxiv.org/abs/1109.1165v2>
- Brier, G. W. (1950). Verification of forecasts expressed in terms of probabilities. *Monthly Weather Review*. 78, 1–3.
- Milani, E. J., Brandão, J. A. S. L., Zalán, P. V., & Gamboa, L. A. P.. (2000). Petróleo na margem continental brasileira: geologia, exploração, resultados e perspectivas. *Revista Brasileira de Geofísica*, 18(3), 352-396. <https://doi.org/10.1590/S0102-261X2000000300012>
- Murphy, A.H. 1973. A new vector partition of the probability score. *J. Appl. Meteorol.* 12: 595– 600, doi:10.1175/1520-0450(1973)012<0595:anvpot>2.0.CO;2
- Newendorp, P. & Schuyler, J.. (2015). *Decision Analysis for Petroleum Exploration*. (3th Edition). Leipzig, Germany: Planning Press.
- Stephenson, D. & Coelho, Sandra & Jolliffe, Ian. (2008). Two Extra Components in the Brier Score Decomposition. *Weather and Forecasting*. 23. 10.1175/2007WAF2006116.1.
- Suslick, S.B. & Schiozer, Denis & Rodriguez, M.R.. (2009). Uncertainty and risk analysis in petroleum exploration and production. *Terræ*. 6. 30-41.

