

## **Técnicas de aprendizaje bayesiano aplicadas a proyectos no convencionales de exploración y desarrollo de petróleo**

Alejandro Murgia  
Knowledge Lead, Caesar Systems, 14811 St. Mary's Lane, Suite 225  
Houston, Texas 77079  
amurgia@caesarsystems.com

**Resumen.** Uno de los desafíos de los proyectos no convencionales de petróleo y gas es la correcta estimación de lo que producirán los campos, ya que suelen requerir una inversión cuantiosa y los márgenes de ganancia son pequeños. Si se realiza un proyecto piloto antes de embarcarse en el desarrollo completo, perforando una serie de pozos, se plantean preguntas de difícil respuesta, como: ¿la información que proveerá el piloto justifica sus costos? Y ¿cuántos pozos deben perforarse? Este trabajo explica cómo el equipo de una empresa que debía prestar apoyo a la toma de decisiones en la planificación de proyectos no convencionales abordó este problema empleando conceptos de aprendizaje bayesiano, valor de la información, y comparación de escenarios probabilísticos.

### **1 Introducción**

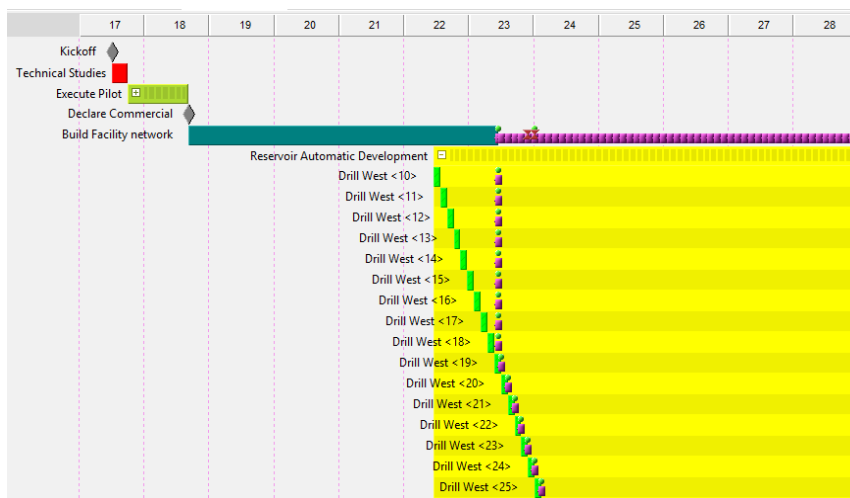
La irrupción de proyectos no convencionales de extracción de petróleo y gas ha puesto en entredicho reglas de oro sólidamente establecidas en la industria a la hora de tomar decisiones sobre cómo explorar, evaluar y desarrollar campos [1]. En particular, las fronteras entre actividades que anteriormente se veían como radicalmente diferentes y separadas en el tiempo (exploración, evaluación, desarrollo) han quedado en los últimos años desdibujadas. Otro de los desafíos de estos campos es que suelen requerir una inversión cuantiosa y los márgenes de ganancia son sustancialmente menores que los que la industria manejaba tradicionalmente; una correcta estimación de lo que producirán los campos se vuelve entonces crucial, y en ese contexto es frecuente que se considere la posibilidad de realizar un proyecto piloto antes de embarcarse en el desarrollo completo. Ese piloto suele consistir en la perforación de una serie de pozos que provean una información mayor sobre las posibilidades productivas del campo, y la instalación de un sistema de producción temprana (Early Production System) [2]. Pero el piloto mismo plantea preguntas de difícil respuesta, comenzando por dos muy básicas: ¿vale la pena realizarlo? Y ¿cuántos pozos debe incluir? Los costos asociados a cada perforación son significativamente altos, y por lo tanto el número óptimo de pozos es un concepto que no puede tomarse a la ligera.

En este trabajo se describe cómo el equipo de una empresa que debía realizar un análisis inicial del desarrollo de un campo de las características mencionadas abordó este problema empleando conceptos de aprendizaje bayesiano, Valor de la Información, y comparación de escenarios probabilísticos.

## 2 Modelando el proyecto

Para modelar el proyecto el equipo empleó el software PetroVR, un sistema de simulación del ciclo completo del negocio del Upstream que provee herramientas de apoyo para la toma de decisiones, análisis probabilístico y cálculo económico. El sistema permite comparar distintos escenarios que modelan estrategias de evaluación y desarrollo, lo que en este proyecto en particular resultaba esencial.

El objetivo fue realizar un modelo lo más sintético posible que permitiera identificar los conceptos más relevantes para la toma de decisiones y evitara detalles superfluos que oscurecieran la comprensión de los desafíos planteados. En cuanto al cronograma de tareas a realizar, se incluyeron Estudios Técnicos preliminares, la ejecución del piloto, la Decisión de declarar al proyecto comercialmente viable, la construcción de la red definitiva de instalaciones y la campaña de perforación de los pozos de desarrollo (**Figura 1**).



**Fig. 1.** Gantt Chart de las tareas a realizar. La sucesión de trabajos en cascada es la campaña de perforación de los pozos de desarrollo, realizados sucesivamente con un equipo de perforación.

Para una información detallada de todos los inputs de este modelo, consúltese el Anexo I.

### 3 Determinando el número de experimentos

El trabajo denominado “Execute Pilot” contiene un número de sub-tareas igual a la cantidad de pozos de evaluación que se desean realizar. Pero ¿cuál es ese número? La determinación del número de pozos del piloto es una de las funcionalidades que se le pide al modelo. Para ello se apeló a técnicas de aprendizaje bayesiano.

El aprendizaje bayesiano es una técnica que permite establecer el peso de nueva información combinándola con información previa según el teorema de Bayes:

$$P(H_0|E) = \frac{P(E|H_0)P(H_0)}{P(E)}$$

Donde  $H_0$  representa una hipótesis que ha sido inferida antes de que la nueva evidencia  $E$  estuviera disponible.  $P(H_0)$  es la probabilidad a priori de  $H_0$ .  $P(H_0|E)$  es la probabilidad condicional de que se cumpla la evidencia  $E$  si la hipótesis  $H_0$  es verdadera.  $P(E)$  es la probabilidad de observar la nueva evidencia  $E$  bajo todas las hipótesis mutuamente excluyentes.  $P(H_0|E)$  es referida también como la probabilidad a posteriori de  $H_0$  dado  $E$ .

Las técnicas derivadas del teorema de Bayes han captado la atención de la industria del petróleo [3] y han motivado algunas propuestas metodológicas, que no puede decirse que hayan arraigado suficientemente en las empresas aún. El enfoque que se propone aquí es original y requiere del apoyo de un sistema que –además de incorporar conceptos de aprendizaje bayesiano– proporcione modelado y simulación del ciclo completo del negocio, escenarios, y Monte Carlo.

#### 3.1 Conocimiento previo

Antes de comenzar el piloto sólo se posee algún conocimiento hipotético, y por lo tanto gran incertidumbre. Supóngase que se modela la estimación de recuperación final de un pozo (EUR según su sigla en inglés) mediante una distribución lognormal  $LN(\mu_0, \sigma)$ . En el ejemplo desarrollado en este modelo los valores de la distribución lognormal truncada son una media de 5,7 millones de barriles y una desviación estándar de 3,4 millones de barriles, como muestra la **Figura 2**.

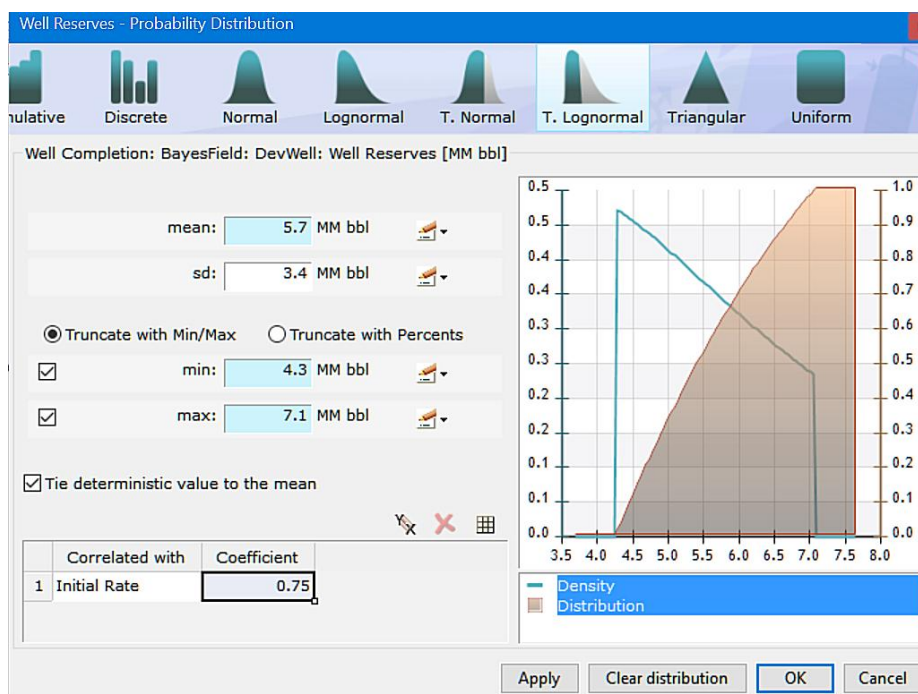


Fig. 2. Distribución lognormal del EUR (o reservas) de los pozos pilotos.

Para poder hacerlo se requiere un análisis que decida el valor medio  $\mu_0$  y la desviación estándar  $\sigma$ . Por supuesto, estos dos parámetros también son inciertos y por lo tanto es razonable reflejar esa incertidumbre en el modelo. En aras de la simplicidad, se supondrá que la incertidumbre en el valor medio es más crucial que la desviación estándar, y se concentrará la atención en ella. Por el teorema del Límite Central es razonable modelar la incertidumbre en torno al valor medio del EUR con una distribución normal  $N(\mu_0, \sigma_0)$ .

### 3.2 Nueva evidencia

Cada nuevo pozo que se perfora proporciona información fresca que uno desea combinar con la existente. El parámetro que se quiere mejorar a partir de ella es el valor medio del EUR de los pozos de desarrollo. Para ese valor medio se ha establecido una distribución normal  $N(\mu_0, \sigma_0)$ .

Supóngase que la nueva evidencia es el EUR de  $n$  pozos pilotos  $P_1, \dots, P_n$  y llámese a la nueva evidencia que ellos proveen  $e_1, \dots, e_n$ . La técnica bayesiana consiste en incorporar  $e_1, \dots, e_n$  en el modelo de tal modo que la probabilidad condicional  $prob(|e_1, \dots, e_n)$  se maximice. En otras palabras, la nueva estimación de la media del EUR será el valor más probable dada la evidencia  $e_1, \dots, e_n$ .

La literatura presenta la siguiente fórmula para el valor más probable:

$$\hat{\mu} = (\bar{\mu} + \varphi_n \mu_0) / (1 + \varphi_n),$$

Donde  $\varphi_n = \sigma^2 / (n\sigma_0^2)$  y  $\bar{\mu}$  es el promedio de los logaritmos naturales  $\ln(e_1), \dots, \ln(e_n)$ .

El estimador bayesiano  $\hat{\mu}$  tiene dos propiedades importantes. Primero, toma en consideración la hipótesis inicial sobre  $\mu_0$ ,  $\sigma_0$  y  $\sigma$ . Segundo, es *robusto*, es decir que no es fácilmente alterable por casos marginales. Esto significa que si uno de los pozos del piloto es demasiado bueno o demasiado pobre, su impacto no será tan extremo.

Nótese también que a medida que  $n$  se incrementa, el peso relativo de la hipótesis inicial se reduce comparado con la nueva información  $e_1, \dots, e_n$ . Esto es así porque todo conocimiento previo está capturado en  $\varphi_n$  y el valor de  $\varphi_n$  se aproxima a cero cuando  $n$  crece.

### 3.3 Intervalos de Confianza

La técnica de aprendizaje bayesiano descrita puede aplicarse para refinar el conocimiento previamente disponible a la luz de nueva información. Este procedimiento puede repetirse cada vez que un nuevo pozo o grupo de pozos se perforen.

Una de las propiedades interesantes del estimador bayesiano es que aun sin saber la media real que define certeramente la distribución lognormal  $\mu_A$  del EUR de los pozos, es posible estimar el peor error posible que lo separa del valor exacto. De modo que se puede responder a estas preguntas:

- 1) Dados  $n$  pozos pilotos, ¿cuál es la probabilidad de estimar  $\mu_A$  con un error que esté por debajo de cierto límite? En términos matemáticos, cuál es el valor de  $P(|\bar{\mu} - \mu_A| < \varepsilon)$ ?
- 2) ¿Cuántos pozos piloto se deberían perforar para tener una estimación con un error menor que  $\varepsilon$  y con una probabilidad mayor que p.ej. 90%?

La habilidad de responder esas preguntas es lo que hace convincente a la técnica bayesiana. No se pide que se confíe en ella, sino que se establezca el intervalo de confianza para el estimador, con lo cual la técnica es completamente objetiva. La aplicación de estos cálculos provee nuevas perspectivas que mejoran la calidad de las decisiones.

### 3.4 Estimando antes de los experimentos

El problema que se presenta a quienes deben modelar un proyecto mucho antes de que los pilotos se realicen es que los experimentos en el mundo real no estarán disponibles hasta mucho más tarde, así que la llegada de información nueva no puede ser parte del modelo. ¿Cómo aprovechar la técnica bayesiana entonces?

La respuesta es que sí se puede modelar la estimación previa de cuántos pozos pilotos serán necesarios, a través de un cálculo bayesiano. El software empleado provee una función **lognormalBayesianEvidence** que devuelve el número de experimentos necesarios dados una distribución de la media del EUR de los pozos, un intervalo de confianza, y una caracterización de la distribución normal a través de la razón  $p90/p10$ .

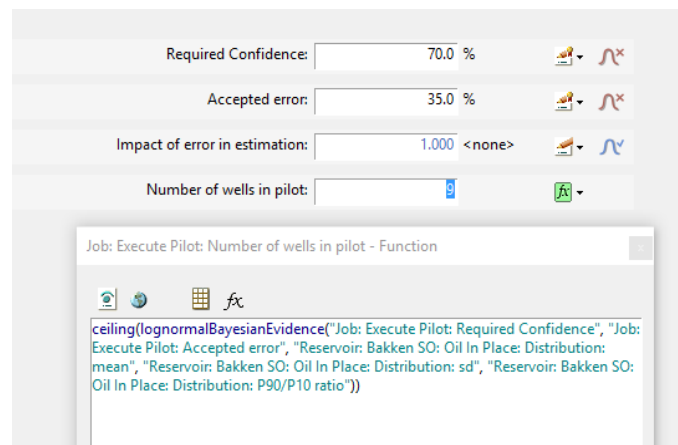
**lognormalBayesianEvidence**(confidence, error, mu0, sd0, P90/P10)

Returns the number of new experiments required to update the estimation of a Lognormal mean within the boundaries of the confidence interval given by **confidence** and **error**. The Lognormal distribution is assumed to have the percentile ratio **P90/P10** and a normally distributed mean with parameters  **$\mu_0$**  and  **$sd_0$** . The **error** is relative to the logarithmic width  $\ln(P90) - \ln(P10)$ .

En el modelo la expresión en cuestión quedó plasmada así:

```
lognormalBayesianEvidence("Job: Execute Pilot: Required Confidence", "Job: Execute Pilot: Accepted error", "Reservoir: Bakken SO: Oil In Place: Distribution: mean", "Reservoir: Bakken SO: Oil In Place: Distribution: sd", "Reservoir: Bakken SO: Oil In Place: Distribution: P90/P10 ratio").
```

El trabajo de perforación de pozos calcula entonces la cantidad de pozos del piloto a través de la expresión anterior. Eso arroja, según el intervalo de confianza ingresado, un número de nueve pozos a partir de estos valores (**Figura 3**).



**Fig. 3.** Valores de las variables que determinan el número de pozos en el piloto.

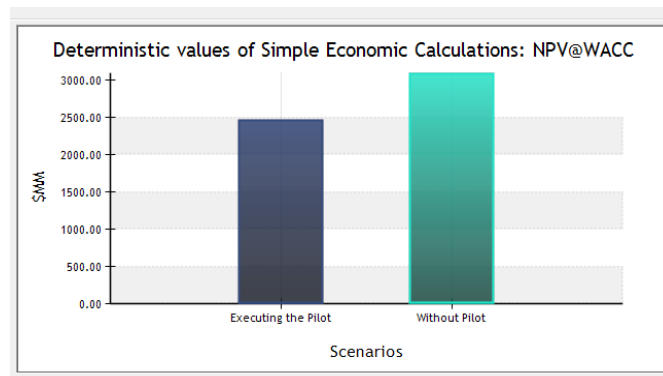
#### 4 El valor de la información

Una vez estipulado un número pozos para el piloto, están bien caracterizados su impacto en los costos y los tiempos del proyecto. Eso permite encarar la siguiente pregunta: ¿la información suministrada por el piloto vale realmente la pena? Para eso es necesario encuadrar el problema en el marco de la teoría del valor de la información. Este punto de vista ha sido estudiado anteriormente en la planificación de proyectos de exploración y desarrollo [4]. Básicamente el valor de la información puede definirse así:

**Valor de la Información = Valor del Proyecto con la Información – Valor del Proyecto sin la información**

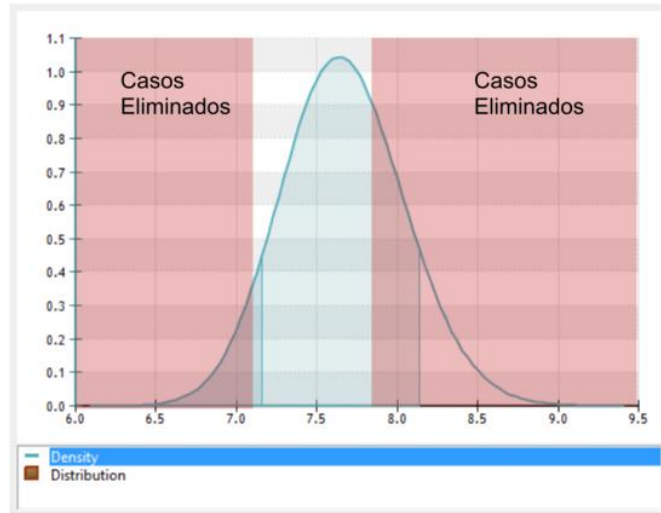
Esta ecuación obliga a construir dos escenarios: uno donde el proyecto se lleva adelante sin realizar el piloto, el otro donde sí se realiza el piloto. La comparación del Valor Actual Neto de cada uno de estos escenarios permitirá conocer el valor del proyecto piloto.

Para modelar las dos estrategias el equipo usó la herramienta Scenarios provista por el mismo sistema PetroVR. Sencillamente, en uno de los escenarios se activó la realización del piloto y en el otro no. La comparación determinística de escenarios antes de modelar los efectos beneficiosos de realizar el piloto sencillamente pone de manifiesto el costo del piloto, de acuerdo al número de pozos calculado por el cálculo bayesiano, como se observa en la **Figura 4**.



**Fig. 4.** Comparación determinística de escenarios.

Sin embargo, para poder estimar el valor del piloto se debería tener en claro en qué impacta la información que el piloto aporta. En el caso que se está estudiando, el principal impacto en el modelo es que el piloto reduce la incertidumbre sobre la distribución del EUR de los pozos. En la industria se suele estimar a priori esa reducción de incertidumbre como un truncamiento en la distribución original (**Figura 5**).



**Fig. 5.** Truncamiento de la distribución del EUR de los pozos como modo de modelar la reducción de incertidumbre aportada por el piloto.

En el caso del modelo estudiado, y debido a que todo el análisis se realizó antes de que los experimentos pudiesen llevarse a cabo en el mundo real, la medida de este truncamiento se basó exclusivamente en estimaciones aportadas por estudios previos sobre la efectividad del análisis de pilotos análogos.

Menor incertidumbre en la performance de los pozos implica mayor claridad para calcular la infraestructura del proyecto. Es allí donde el piloto puede hacer ahorrar dinero: si se conocen mejor los pozos se pueden construir instalaciones de tamaño adecuado. Sobredimensionar las instalaciones equivaldría a gastar más dinero del necesario. Por el contrario, subdimensionarlas haría que el proyecto se extendiera por más años y se demorara más en recuperar la inversión. Ésas son las penalidades que normalmente se pagan más caras en caso de no hacerse el piloto.

Pero es importante capturar en el modelo la circunstancia de que el piloto puede en ocasiones resultar engañoso, lo que se conoce como experimentos que arrojan falsos negativos o falsos positivos. Dependiendo de las características geológicas o la dinámica productiva del reservorio, el piloto podría llevar a creer que el campo es más productivo de lo que en realidad es –si los pozos han sido perforados en zonas especialmente productivas–, o viceversa, que es menos productivo de lo que en realidad es, si las zonas elegidas son de pobre rendimiento.

Para el análisis de esos falsos negativos y positivos lo que se hizo fue incorporar a la comparación de los escenarios el análisis de mundos posibles a través de la técnica de Monte Carlo. Por lo tanto la comparación de escenarios pasó a ser probabilística y no meramente determinística.



## 5 Análisis probabilístico

La toma de decisiones en proyectos con alta incertidumbre requiere asumir esa incertidumbre como un dato acerca del conocimiento imperfecto, un dato que debe ser modelado; en ese sentido las especificaciones determinísticas se revelan insuficientes ya que suponen que el modelador tiene un conocimiento exacto y completo de la información requerida por todos los parámetros. En el modelo estudiado aquí, era importante caracterizar la incertidumbre en cualquiera de los muchos parámetros existentes: reservas, tiempos de construcción, costos, etc. El modo de hacerlo es a través de distribuciones de probabilidades.

Las distribuciones de probabilidades posibilitan la aplicación de la herramienta matemática conocida como Simulación de Monte Carlo. La idea de esta simulación es simple: generar un muestreo aleatorio para cada variable del modelo que tenga definida una distribución de probabilidades, crear múltiples mundos posibles donde las variables asuman uno de los valores del muestreo generado, y correr una simulación determinística para cada mundo posible. Tras esas corridas será posible recolectar los resultados de esos mundos posibles y calcular sus indicadores estadísticos: valor medio, desviación estándar, Kurtosis, valor más probable, etc.

El uso de Monte Carlo en la planificación de proyectos de desarrollo petrolero está muy extendido [5], pero por lo general se limita a estudiar la incertidumbre geológica: en este trabajo el alcance de la técnica es más ambicioso: abarca todo tipo de incertidumbre, incluyendo incertidumbre operativa, de costos, y de precios.

La combinación de Monte Carlo con el modelado de Escenarios permite comparar el desempeño de dos (o más) estrategias en distintos mundos posibles: pone a prueba cada estrategia de desarrollo (en este caso realizar o no el piloto) en diferentes condiciones posibles. Desde el punto de vista del Valor de la Información, Monte Carlo aplicado a Escenarios permite poner a prueba el desempeño de la estrategia de construir un piloto cuando éste arroja un falso negativo y cuando arroja un falso positivo. Esos casos se combinan estadísticamente con los casos presumiblemente más frecuentes en que el piloto contribuye positivamente a entender las características productivas del campo, y juntos dan una idea más realista del valor del piloto.

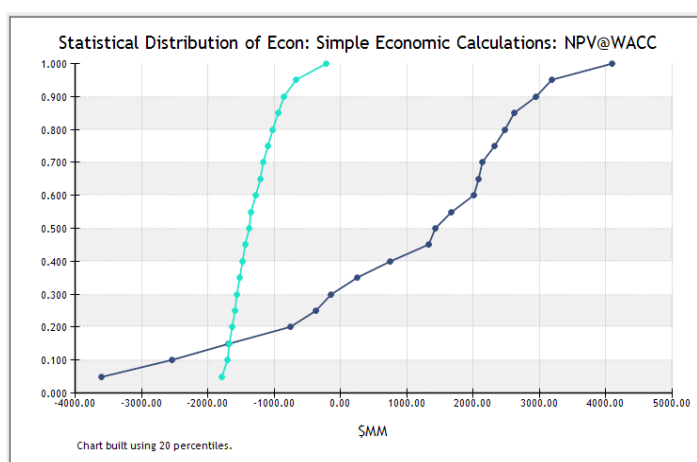
En el modelo estudiado, como se ha señalado anteriormente, el escenario en que se conduce el piloto tiene la distribución del EUR de los pozos reemplazada por una distribución con menor desviación estándar como modo de reflejar la mejora en el entendimiento en las características del campo productivo. La información que aporta el piloto no es perfecta (puede arrojar falsos positivos y falsos negativos) y ésta es la razón por la cual no se reemplaza la distribución por un valor exacto (el EUR real) sino por otra distribución.

Un error en la estimación del EUR, tanto en el caso de realizar el piloto como en el caso de omitirlo, impactará en el costo y tiempo de construcción de las instalaciones representadas por el objeto "Facility Network" cuya capacidad de procesamiento (e indirectamente sus tiempos de construcción y su costo) está definida con la siguiente función:

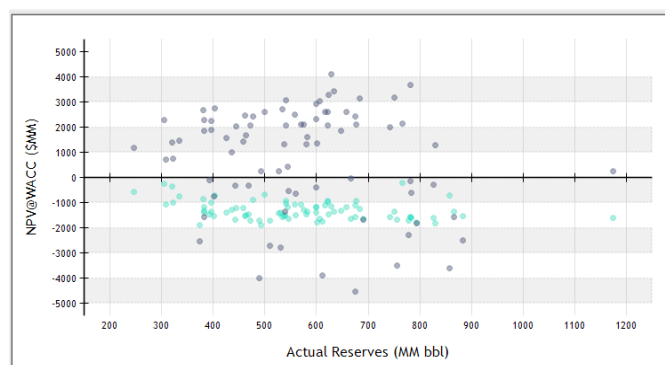
if("Job: Execute Pilot: Perform Task", "Pvr: Capacity based on actual reserves" \* "Job: Pilot: Impact of error in estimation", "Pvr: Capacity based on estimated reserves")

La función establece que en caso de ejecutarse el piloto la capacidad de procesamiento de la instalación estará basada en el EUR real de los pozos multiplicado por un factor de error de estimación que el piloto puede tener; de no hacerse el piloto, la capacidad estará basada en la estimación inicial del EUR.

Los resultados que arroja la comparación probabilística de los escenarios permite comparar las dos estrategias, como se observa en la **Figura 6** y la **Figura 7**.



**Fig. 6.** Comparación de la distribución acumulada del Valor Actual Neto de los dos escenarios. La línea celeste representa el escenario sin el piloto. La línea azul oscuro (el escenario con el piloto) permite mayores márgenes de ganancia en la mayor parte de los casos. Sólo en el 15% de los casos el escenario “sin piloto” es favorable: se trata de los casos en que el piloto provee información desacertada sobre las características productivas del reservorio.



**Fig. 7.** Cross Plot de las iteraciones de Monte Carlo para los dos escenarios: ante valores similares de reservas, los puntos oscuros (Execute Pilot) se ubican en su mayoría por encima de los celestes (Without Pilot), con lo cual denotan un mayor Valor Actual Neto. Unos pocos

puntos oscuros se ubican por debajo de los celestes: son los casos donde el piloto produjo información desacertada.

Por último, además de las dos estrategias básicas (desarrollar el campo con o sin piloto), se podrían construir escenarios adicionales donde el piloto se hace pero los intervalos de confianza elegidos varían. Esta comparación podría revelar que la pretensión de moverse en un margen de error muy pequeño cuesta desde un punto de vista estadístico demasiado dinero para el valor que proporciona, o viceversa, que un piloto de pocos pozos no provee la información suficientemente valiosa como para justificar su costo. De manera que estos escenarios adicionales podrían ser una herramienta útil para descubrir cuáles son los intervalos de confianza razonables.

## 5 Conclusiones

La experiencia presentada en este trabajo muestra que el concepto de aprendizaje bayesiano puede ser útil en el modelado de proyectos de exploración y desarrollo de petróleo y gas no convencionales. En ellos, como en muchos otros casos de la industria en que deben tomarse decisiones en presencia de una gran incertidumbre, ésta puede mitigarse con la adquisición de información; el modelo puede y debe capturar estas nociones: la incertidumbre, las estimaciones iniciales, y la forma en que la información nueva se combina con la previa. En este trabajo se ha observado cómo las técnicas de aprendizaje bayesiano pueden ser útiles para pensar correctamente el problema aún en casos como el presentado, donde el modelo se realiza antes de la existencia de experimentos (perforación de los pozos). Para lograrlo, es imprescindible contar con un sistema que facilite el cálculo del número de experimentos en un modelo que incluya comparación probabilística de estrategias. El sistema debe integrar el cálculo del piloto con los demás factores que impactan en el valor del proyecto: la infraestructura de superficie y el modelo económico.

## Referencias

1. Burkholder, M. K., Coopersmith, E. M., Schulze, J. H.: Appraisal Excellence in Unconventional Reservoirs, Paper SPE 162776, Presented at the SPE Canadian Unconventional Resources Conference, Calgary, Canada, October 2012.
2. Coopersmith, E. M., Burkholder, K. Mendoza, L.: Early Production Systems – When Do They Make Good Sense for your Project? Paper OTC-25440-MS, Presented in Offshore Technology Conference, Houston, Texas, USA, May 2014.
3. Stabell, Charles: Modeling Learning in Resource Play Assessment: An Integrated Approach for Decision Support, Paper SPE 162914, Presented at the SPE Hydrocarbon, Economics and Evaluation Symposium, Calgary, September 2012.

4. Bickel, J. E.: Decision Analytics, Value of Information, and Piloting in Unconventional Reservoirs, Paper SPE-169841, Presented at the SPE Hydrocarbon, Economics and Evaluation Symposium, Houston, Texas, USA, May 2014.
5. Arild, O. Lohne, H. P., Bratvold, R. A Monte Carlo Approach to Value of Information Evaluation, Paper IPCT 11969, Presented at the International Petroleum Technology Conference, Kuala Lumpur, Malaysia, December 2008.