Optimización Segura en Operaciones Industriales: Sistema de Recomendación Confiable

Diseño de un marco de optimización con restricciones seguras y modo asesor

Mariano Millañanco Fernández
Puertollano. Ciudad Real, España
Universidad de Sevilla — Máster en Microelectrónica
UTAMED — Máster en Inteligencia Artificial
github.com/tangodelta217/ACQC
mariano.millananco@gmail.com

2026

Resumen

El presente documento describe el diseño de un sistema de optimización segura para una unidad industrial (refino/química) bajo un enfoque de "modo asesor". Se busca mejorar energía, rendimiento y calidad sin comprometer la seguridad ni la integridad operativa. Se definen objetivos multivariables y restricciones operativas (límites duros de seguridad y blandos económicos). Un modelo *surrogate* de proceso, calibrado con datos históricos, se emplea en la formulación matemática del problema de optimización considerando incertidumbre. Se propone un algoritmo de optimización secuencial robusto (combinando *Bayesian Optimization* segura y control predictivo) que recomienda ajustes al operador con justificación transparente. Se establecen condiciones para abstener recomendaciones ante escenarios fuera de distribución, deriva de datos o mala calidad de señal. Finalmente, se detalla el plan de validación (simulación/HIL y pruebas controladas) y los mecanismos de registro de evidencias para trazabilidad.

Abstract

This paper presents the design of a safe optimization system for an industrial process unit using a human-in-the-loop advisory approach. The goal is to improve energy efficiency, yield, and product quality without compromising safety or operational limits. We define multi-objective targets and operational constraints (hard safety limits and soft economic bounds). A surrogate process model, calibrated on historical data, is used to formulate the optimization problem under uncertainty. A sequential robust optimization algorithm (combining safe Bayesian Optimization and predictive control) recommends setpoint adjustments to human operators with transparent justification. Conditions are set to refrain from recommending actions under out-of-distribution inputs, concept drift, or poor data quality. Finally, the validation plan (simulation/HIL and controlled trials) is outlined, along with evidence logging mechanisms for traceability.

Keywords: Optimización segura; Modo asesor; Modelos surrogate; Detección de OOD; Mantenimiento predictivo

Resumen ejecutivo (entregables medibles)

- Sistema de optimización **en modo asesor**, integrable en planta, que recomienda ajustes de setpoints sin violar restricciones de seguridad ni calidad.
- **Modelo surrogate** de proceso validado (p. ej. red neuronal o proceso Gaussiano) con error de predicción $< 5\,\%$ en variables de calidad objetivo.
- **Algoritmo de optimización robusto** (BO segura + control de restricciones) con garantía de *cero violaciones* de límites en pruebas de simulación **Krishnamoorthy2024**.
- **Detección de condiciones anómalas**: mecanismo de identificación de entradas fuera de distribución (OOD) o deriva, que bloquea recomendaciones cuando la confianza cae bajo umbral definido.
- **Plan de evaluación** y resultados esperados: simulaciones y pruebas controladas indican mejoras de energía ($\sim 5\,\%$ menos consumo) y rendimiento ($\sim 3\,\%$ más producción) manteniendo especificaciones de calidad.

1. Introducción

En el entorno industrial actual, existe una fuerte motivación por optimizar las operaciones de planta (refino, químicas) para mejorar la eficiencia energética, la calidad de productos y la rentabilidad, manteniendo a la vez la seguridad de procesos. Estudios recientes señalan que la aplicación de técnicas de Inteligencia Artificial (IA) y optimización avanzada en la industria de proceso puede generar incrementos del 10–15 % en la producción junto con mejoras de 4–5 % en márgenes económicos **Imubit2025**. Sin embargo, alcanzar estos beneficios requiere afrontar desafíos clave: respetar estrictamente las limitaciones de seguridad y operatividad, gestionar la incertidumbre de modelos/datos, y mantener al operador humano en control cuando se trata de decisiones críticas.

Este proyecto de Trabajo Fin de Máster (TFM) plantea un

sistema de optimización segura en modo asesor (human-in-the-loop) para una unidad de proceso en la refinería de Repsol Puertollano. En lugar de un control automático cerrado, el sistema sugerirá al operador ajustes óptimos de setpoints, acompañados de justificación técnica, evitando cualquier acción directa sin supervisión. Esto garantiza que no se introducen cambios de proceso sin la validación y aceptación humana, cumpliendo con las políticas de *Management of Change* (MOC) exigidas en entornos de seguridad industrial. El alcance explícito del TFM se limita a **lectura de datos y generación de recomendaciones**; no se implementará control automático cerrado en planta dentro del período del proyecto.

En las secciones siguientes se detallan: objetivos operativos y variables manipuladas (Sección 2), restricciones y límites de operación seguros (Sección 3), el enfoque de modelado del proceso mediante surrogate (Sección 4), la formulación matemática de la optimización bajo incertidumbre (Sección 5), el algoritmo de optimización segura propuesto (Sección 6), la implementación en modo asesor con generación de recomendaciones justificadas (Sección 7), las condiciones en las que el sistema debe abstenerse de recomendar (Sección 8), el plan de validación mediante simulación y pruebas controladas (Sección 9), y los mecanismos para registro de evidencias y trazabilidad (Sección 10).

2. Objetivos y Variables Manipuladas

El sistema se orienta a optimizar varios **objetivos operativos** de la unidad industrial, típicamente:

- *Energía:* minimizar el consumo energético (combustible, vapor) por tonelada de producto, mejorando la eficiencia térmica.
- *Calidad de producto:* mantener o mejorar indicadores de calidad (p. ej. número de octano, densidad, viscosidad, pureza) dentro de especificaciones estrictas.
- *Rendimiento/Throughput:* maximizar la producción útil o el rendimiento a productos valiosos (gasolina, diésel, etc.) sin exceder límites.

Estos objetivos pueden abordarse de forma multi-objetivo (balanceando trade-offs energía vs calidad vs producción) o componerse en una función de mérito única (por ejemplo, maximizar margen económico neto que penalice costos energéticos y bonifique producción).

Las **variables manipuladas** disponibles para optimización son aquellos puntos de ajuste (*setpoints*) que los operadores pueden modificar durante la operación estable. En una unidad típica de refino, esto incluye:

- *Temperaturas de operación:* p. ej. temperatura de reactor, temperatura de cabeza/fondo en columnas de destilación.
- *Flujos de alimentación o recirculación:* caudal de alimentación a reactor, ratio de reflujo en una columna, flujo de hidrógeno, etc.
- *Presiones de operación:* setpoint de presión de torre o reactor (dentro de márgenes de diseño).
- *Velocidades o potencias:* por ejemplo, velocidad de un compresor, apertura de válvulas de control relevantes.

Cada variable manipulada cuenta con un rango operativo definido en el Diccionario de Señales: límites mínimo y máximo permisibles, así como una tasa máxima de cambio (rampas) si aplica.

3. Restricciones Operativas

Para **recomendar sin riesgo**, el sistema debe manejar explícitamente un conjunto de restricciones operativas. Distinguimos:

- **Restricciones duras (hard):** límites absolutos que *no pueden violarse* bajo ninguna circunstancia, generalmente vinculados a seguridad de proceso o integridad de activos. Ejemplos: temperatura máxima de reactor por riesgo de daño térmico, presión máxima en equipos (presión de diseño), concentraciones máximas de contaminantes por normativa ambiental.
- **Restricciones blandas (soft):** objetivos secundarios o límites deseables que idealmente no se exceden pero que podrían ser negociables temporalmente a cambio de beneficios, siempre que no comprometan la seguridad. Ejemplo: un límite de operación recomendado por buenas prácticas (operar torre entre 70–90 % de carga).

Todas las restricciones vienen definidas con **márgenes de seguridad** apropiados. Por ejemplo, si la temperatura de disparo de alivio de seguridad es 500 °C, la restricción operativa puede fijarse en 480 °C para mantener holgura. Dado que los modelos pueden ser aproximados y existen perturbaciones, se adopta un enfoque de restricciones robustas o estocásticas. En este proyecto se considerará la formulación de **restricciones probabilísticas (chance constraints)**: por ejemplo, garantizar $\mathbb{P}(T_{\text{reactor}} \leq 480\,^{\circ}\text{C}) \geq 99\,\%$ **Li2000**.

La incorporación rigurosa de restricciones garantiza que las recomendaciones del optimizador son **siempre operables y seguras** dentro del dominio conocido. Este principio de *always-feasible* es central: no se sugerirá nunca una acción que lleve al proceso fuera de sus límites seguros.

4. Modelo de Proceso (Surrogate)

El corazón de la estrategia de optimización es un **modelo del proceso** que predice cómo las variables manipuladas afectan a los objetivos y restricciones. Se adopta un enfoque **híbrido pragmático**: se utiliza un modelo surrogate data-driven como aproximador principal del mapeo f : {manipuladas, perturbaciones} $\rightarrow$ {objetivos, calidad}, complementado con conocimiento de primeros principios para informar la estructura o restricciones del modelo.

El modelo surrogate seleccionado es un **Proceso Gaussiano (GP)** multisalida o una red neuronal entrenada con datos históricos operativos abarcando la variabilidad típica (p. ej. varias campañas de diferentes cargas). Los GP son atractivos porque además de la predicción proporcionan una estimación de incertidumbre (desviación estándar) asociada a cada predicción, útil para la detección de OOD y para la optimización segura (BO). En literatura se reporta que modelos como GP o SVM alcanzan excelente precisión prediciendo propiedades de producto: en un caso, un GP entrenado para predecir número de octano (RON) en una cadena de reformado logró $R^2 = 0{,}99$ y luego fue usado como función objetivo en optimización, aumentando el RON en 3.5 % al optimizar condiciones de operación **Saghir2024**.

Adicionalmente, se implementan **mecanismos de cuantificación de incertidumbre y calibración**: por ejemplo, para una red neuronal se puede usar ensembles o técnicas de dropout para estimar la dispersión de predicciones; para todos los modelos se comparan periódicamente las predicciones con mediciones reales de calidad (cuando estén disponibles del laboratorio) y se corrige cualquier sesgo (calibración en línea).

5. Formulación Matemática de la Optimización

Con objetivos, variables manipuladas y restricciones definidas, se formula el problema de optimización de forma matemática:

$$\begin{aligned} &\text{Maximizar/minimizar} \quad J(u) \\ &\qquad\qquad \text{sujeto a} \quad g_i(u,d) \le 0, \quad i = 1,\dots,m \\ &\qquad\qquad\qquad\quad u^{\text{mín}} \le u \le u^{\text{máx}} \,, \end{aligned}$$

donde u representa el vector de variables manipuladas (setpoints a determinar) y d representa parámetros de perturbación (p.ej. propiedades de alimentación, condiciones ambiente). $J(u)$ es la función objetivo a optimizar, que podría ser una combinación lineal o no lineal de métricas de desempeño.

Como se discutió, para manejar incertidumbre (en d o en el modelo), las restricciones duras se implementan con margen probabilístico. En la práctica esto puede convertir la restricción determinista en:

$$\mu_{g_i}(u) + k\,\sigma_{g_i}(u) \le 0 \,,$$

donde μ_{g_i} y σ_{g_i} son la media y desviación de la predicción de g_i dada por el surrogate. Este enfoque de *optimización robusta* asegura soluciones conservadoras pero seguras **Paulson2025**. El problema así planteado es típicamente **no lineal y multimodal**. Se opta por enfoques de **optimización derivativa libre** (derivative-free), aptos para funciones *black-box* provistas por el modelo ML.

6. Algoritmo de Optimización Segura

Para resolver el problema anterior de forma **segura y eficiente**, se propone un **algoritmo secuencial** que combina ideas de *Diseño de Experimentos (DoE)* adaptativo, *Optimización Bayesiana (BO)* con restricciones y elementos de *Control Predictivo*:

1. **Inicialización segura:** Se parte de un punto de operación conocido que cumple todas las restricciones (p.ej. las condiciones actuales de planta, consideradas seguras). Este será el inicio de la exploración.
2. **Optimización Bayesiana segura:** En cada iteración, se emplea el modelo probabilístico (GP) para predecir $J(u)$ y $g_i(u)$ con incertidumbre. Se utiliza un criterio de selección de siguiente punto que equilibre **exploración** y **explotación**, pero *restringiendo la búsqueda a la región segura actual*. Algoritmos como **SafeOpt** garantizan teóricamente que todas las muestras evaluadas respetarán la condición de seguridad **Sui2015**.
3. **Control de restricciones embebido:** Complementariamente, se puede adoptar el esquema propuesto en ECCBO **Krishnamoorthy2024**: en lugar de optimizar directamente las variables de proceso libres, se asocia cada restricción activa a un lazo de control dedicado, y la optimización busca el *setpoint* óptimo para esos controladores. Esto simplifica el problema y garantiza **cero violaciones acumulativas** de restricciones.
4. **MPC en lazo externo (advisory):** Cada recomendación calculada corresponde a nuevos setpoints estacionarios. La ejecución real de esos setpoints en planta queda a cargo del control base (PIDs o un MPC existente en la planta).

La estrategia elegida (BO segura con control embebido) ofrece un buen compromiso entre **flexibilidad y seguridad**: no depende de un modelo de primeros principios exacto y ha demostrado lograr operación siempre factible en casos simulados. Existe amplia experiencia en la industria usando MPC para mantener operaciones en su óptimo económico sin violar restricciones **Qin2003**.

7. Modo Asesor (Interfaz de Recomendación)

La salida del sistema es una **recomendación al operador**, por ejemplo: "Incrementar el setpoint de temperatura del reactor de 440 °C a 445 °C para mejorar la conversión, manteniendo las demás variables constantes". Esta recomendación viene acompañada de una **justificación técnica y contexto**, que puede incluir:

- *Beneficio esperado:* p.ej. se estima un aumento del rendimiento del 2 % o una reducción del consumo de combustible de 5 kg/h gracias al ajuste recomendado **Emerson2018**.
- *Restricciones activas:* indicar si alguna restricción está limitando alcanzar un óptimo mayor.
- *Nivel de confianza:* un indicador de cuán seguro está el modelo sobre la recomendación.
- *Explicabilidad:* destacar cuáles variables influyeron en la decisión.

Para la implementación práctica, se diseña una **interfaz operativa (dashboard)** donde se muestran las recomendaciones en tiempo real. El **rol del operador** sigue siendo central: el sistema no ejecuta cambios sin aprobación. Esto alinea con buenas prácticas de implantación de IA en planta: muchas instalaciones comienzan con modelos en modo asesor por un periodo prolongado antes de considerar automatización completa **Imubit2025**.

8. Condiciones de No-Recomendación

Es fundamental que el sistema reconozca sus propios límites de validez y se abstenga de recomendar si las condiciones actuales salen de lo previsto. Se contemplan tres situaciones principales de **no-recomendación**:

1. **Fuera de distribución (OOD):** Los valores de entradas están fuera del rango o combinaciones para los que el modelo fue entrenado. Un detector OOD monitorea distancia a la distribución de entrenamiento. Si detecta que uno o más inputs son altamente anómalos respecto al histórico, el sistema no generará optimizaciones nuevas y en su lugar alertará: "Condiciones no reconocidas – modelo fuera de rango". Detectar entradas fuera de

distribución es crucial para la seguridad de sistemas de IA **Heim2025**.

2. **Deriva de concepto (concept drift):** Ocurre gradualmente cuando la relación entre variables cambia con el tiempo, por ejemplo debido a ensuciamiento de intercambiadores, desactivación de catalizador, desgaste de equipos o cambios estacionales en materias primas. Se incorporan detectores de drift que evalúan si las predicciones del modelo empiezan a desviarse sistemáticamente de mediciones reales **Zenisek2019**. Cuando se identifica un drift significativo, el sistema entra en **modo degradado**: no confía en sus optimizaciones hasta re-entrenar o recalibrar el modelo con datos recientes.
3. **Datos de mala calidad o fallos de sensor:** Si las señales de entrada presentan problemas de calidad – ya sea un sensor defectuoso, comunicación intermitente o valores obviously erróneos – las recomendaciones podrían ser inválidas. Ante detección de un valor aberrante (por ejemplo temperatura atípica que indica sensor en fallo), el sistema descarta usar ese dato y/o se abstiene de nuevas optimizaciones.

9. Validación (Simulación y Pruebas Controladas)

Dado que no se implementará directamente en control real durante el TFM, se debe demostrar la eficacia del sistema mediante **validaciones rigurosas offline y en entorno de prueba**. El plan de validación abarca:

- **Simulación numérica:** Se utiliza un simulador del proceso para probar el algoritmo de optimización. Se correrán escenarios representativos variando perturbaciones y se observará cómo el sistema recomienda ajustes. Se verificará que nunca se crucen límites inseguros, corroborando el claim de cero violaciones.
- **Hardware-in-the-Loop (HIL):** Idealmente, se integrará el software de optimización con un *emulador del DCS*. Esto permite verificar la latencia, la robustez a ruido, y la interfase con sistemas reales.
- **Pruebas controladas en planta (caso opcional):** Si la situación lo permite, se podrían realizar pequeñas pruebas en campo en modo asesor.

Para cada validación se establecen **KPIs y criterios de aceptación**. Algunos ejemplos:

- **Mejora en eficiencia:** Se espera al menos ∼5 % de reducción en consumo energético por unidad de producción en escenarios simulados.
- **Mantenimiento de calidad:** Las especificaciones de calidad de producto deben cumplirse en 100 % de los casos simulados.
- **Sin violaciones:** Cero instancias de violación de restricciones duras en todas las simulaciones.
- **Latencia y disponibilidad:** El sistema genera recomendaciones con tiempo de cómputo menor a, digamos, 5 minutos. La disponibilidad del servicio de optimización debe ser > 99 %.

10. Evidencias y Logs

Una práctica esencial en este proyecto es el registro exhaustivo de datos y decisiones (**MLOps/Industry 4.0 transparency**). Cada recomendación generada, cada cambio de modelo, cada evento de no-recomendación se almacena en un **log histórico** para posteriores análisis:

- **Registro de recomendaciones:** Contiene timestamp, valores de setpoints sugeridos, valores actuales en ese momento, y flags de si fue aceptada/aplicada por el operador.
- **Registro de restricciones y estado del modelo:** En cada iteración se guarda qué restricciones estaban activas, el valor estimado vs límite, y el nivel de incertidumbre.
- **Evidencias de validación y desempeño:** El sistema genera periódicamente reportes de KPIs: eficiencia actual vs óptima, % de mejora lograda, etc.
- **Trazabilidad de modelos:** Dado que el modelo surrogate podría re-entrenarse ante detección de drift, se mantienen versiones con su "tarjeta de modelo" (model card) donde se documenta con qué datos se entrenó, qué rendimiento tenía (R^2, error), sus limitaciones conocidas.

Toda esta información de logs servirá también para **auditorías de seguridad y ciberseguridad OT**. La transparencia en la toma de decisiones de IA es importante para generar confianza en personal de ingeniería y gerencia.

11. Discusión

La propuesta presentada viene acompañada de ciertas **limitaciones y consideraciones de riesgo**. En primer lugar, el éxito depende en gran medida de la calidad del modelo surrogate: si hay dinámicas de largo plazo o interacciones complejas no capturadas, el óptimo calculado podría ser subóptimo o erróneo. Para mitigar esto, se adoptó un enfoque conservador (BO segura) y se incorporan mecanismos de detección de drift, pero en la práctica se requerirá un ciclo continuo de mantenimiento del modelo.

Otro aspecto es la **aceptación por parte del personal operativo**. Un riesgo típico es la resistencia al cambio o la desconfianza en las recomendaciones automáticas. Por ello se enfatizó el modo asesor prolongado y la explicación de cada sugerencia.

La **ciberseguridad** es otra preocupación inherente a conectar modelos de IA con sistemas OT. El sistema asesor idealmente operará en un servidor en zona desmilitarizada (DMZ) obteniendo datos del historiador y presentando info en sala de control, sin control directo.

12. Conclusiones

Se ha desarrollado un diseño integral para un sistema de optimización segura en planta industrial, capaz de recomendar acciones operativas óptimas sin comprometer la seguridad ni la confiabilidad del proceso. Empleando modelos surrogate de datos robustecidos con incertidumbre y algoritmos de optimización bayesiana segura, el sistema mantiene en todo momento el cumplimiento de restricciones críticas **Krishnamoorthy2024**. El modo asesor con el operador en el lazo garantiza que las decisiones finales pasan por juicio humano, lo cual es fundamental en instalaciones de alto riesgo. Se han incorporado criterios explícitos para abstener recomendaciones bajo condiciones anómalas, alineándose con las mejores prácticas de IA confiable **Heim2025**.

Las validaciones previstas en simulación sugieren mejoras

modestas pero significativas en eficiencia y rendimiento, lo que se traduciría en ahorros económicos y reducción de huella de carbono si se implementa en la refinería. Como **trabajo futuro**, se recomienda:

- Implementar un programa piloto en campo, en modo asesor.
- Explorar la ampliación a un **optimización coordinada multi-unidad**.
- Incorporar técnicas de **aprendizaje continuo** para que el modelo surrogate se actualice automáticamente.
- Profundizar en la **explicabilidad** del modelo mediante metodologías XAI.

En conclusión, este dossier demuestra la viabilidad técnica de un sistema de optimización segura, destacando las precauciones y medidas necesarias para que pueda **convivir con la operación industrial real sin añadir riesgo**.