Impacto de las Precipitaciones en el Costo de Vida en Colombia: Energía y Análisis de Alimentos

Seminario de Analítica y Ciencia de Datos

Resumen Descriptivo del Proyecto

El proyecto analiza cómo las precipitaciones afectan el mercado de alimentos y la energía eléctrica en Colombia. Se enfoca en la influencia de las lluvias en la producción agrícola y la generación hidroeléctrica, elementos clave en el costo de vida del país. Se busca entender cómo estas variaciones climáticas afectan la disponibilidad y los precios de los alimentos, así como el costo de la electricidad. Para alcanzar esta meta se debe aplicar modelos de análisis de datos preferiblemente ya utilizados en la literatura, que muestren resultados satisfactorios en la vida real. El análisis proporcionará información para una mejor planificación y gestión de recursos frente a los desafíos climáticos, utilizando datos gubernamentales disponibles públicamente.

Basado en el Data Product Canvas1 resuelve:

1. Problema.

a. ¿Cuál es mi problema alrededor de mi proyecto?

El problema es que no se tienen herramientas con los modelos adecuados y disponibles al público general que permitan pronosticar el impacto de las precipitaciones en dos aspectos críticos que afectan el costo de vida en Colombia: el mercado de alimentos y la energía eléctrica. Con la agricultura dependiente en un 90% en agua lluvia para labores de riego [1] y el suministro eléctrico en un 70% por hidroeléctricas [2]. La variabilidad en las precipitaciones puede tener consecuencias significativas en la producción agrícola y la generación hidroeléctrica, lo que a su vez influye en la disponibilidad y los precios de los alimentos, así como en el costo de la electricidad.

b. ¿Por qué se manifiesta el problema?

Este problema se manifiesta, debido a la importancia crítica que tienen la agricultura y la generación hidroeléctrica en la economía y la vida cotidiana de Colombia.

1. **Dependencia de la agricultura**: Colombia es un país donde la agricultura desempeña un papel fundamental en la seguridad alimentaria y la economía. La mayoría de la producción agrícola depende del agua de lluvia para su riego, por lo que cualquier variación en las precipitaciones puede tener un impacto significativo en la producción de alimentos. Esto puede llevar a escasez de alimentos, aumentos de precios y afectar la seguridad alimentaria de la población. En el anterior punto se resaltó un 90 % [1] de dependencia en agua lluvia, lo cual evidencia un atraso en la implementación de modelos predictivos o prescriptivos que permitan mitigar los riesgos relacionados a el vínculo tan estrecho entre el riego de los cultivos y el agua lluvia.
2. **Adherencia a la generación hidroeléctrica**: La generación hidroeléctrica es la principal fuente de energía en Colombia, proporcionando alrededor del 70% de la electricidad del país [2]. La cantidad de agua disponible para la generación hidroeléctrica está estrechamente ligada a las precipitaciones. Si las lluvias son insuficientes, puede haber una disminución en la capacidad de generación de energía hidroeléctrica, lo que lleva a cortes de energía, aumentos en los precios de la electricidad, afectando la estabilidad del suministro eléctrico. Lo anterior también muestra una necesidad urgente de diversificación energética, que permita abastecer la constante y creciente demanda de electricidad [3](https://www.elespectador.com/especiales/como-esta-colombia-en-transicion-energetica-oportunidades-y-desafios/).
3. **Mala planificación o aprovechamiento de los recursos**: En los anteriores puntos, se habla como no hay diversificación energética (sin mencionar los distintos problemas socio-económicos y de corrupción que han causado retrasos en el desarrollo Colombiano), además de la falta de soberanía alimentaria e implementación de estrategias que permitan reducir la dependencia de la agua lluvia en el entorno agrícola
4. **Modelos de proyección de la demanda y eventos climáticos:** Las implementaciones que se tienen planteadas para la predicción de la disponibilidad de energía o alimentos agrícolas no ha logrado ajustar o encontrar los parámetros y/o modelos (incluso la infraestructura) que permita integrar, correlacionar y predecir de manera adecuada los eventos de precipitación que causan las problemáticas antes mencionadas

Se puede afirmar que la manifestación de este problema se debe a la vulnerabilidad de Colombia a las variaciones climáticas, especialmente en lo que respecta a la disponibilidad de agua para la agricultura y la generación de energía hidroeléctrica, como también malas prácticas de planificación/aprovechamiento, proyección del consumo y eventos climáticos de precipitación que impactan directamente en el costo de vida de la población y en la estabilidad económica del país.

c. ¿De quién es el problema?

Es un problema compartido por múltiples partes interesadas, incluidos el gobierno, las empresas, los agricultores y la sociedad en general en Colombia.

1. **Gobierno**: El gobierno colombiano tiene la responsabilidad de establecer políticas y regulaciones que promuevan una gestión adecuada de los recursos naturales, así como de invertir en infraestructura y tecnología para mejorar la resiliencia del país frente a los impactos climáticos en la agricultura y la energía.

2. **Empresas**: Las empresas, especialmente en el sector energético, desempeñan un papel crucial en la planificación y optimización de la matriz hidroenergética. Mediante el uso de modelos predictivos respaldados por datos, estas empresas pueden mitigar el aumento de los indicadores de costo de vida. De manera similar, las empresas agrícolas tienen la responsabilidad de adoptar prácticas basadas en datos que impulsen una economía agrícola más sostenible y eficiente en el uso del agua.

3. **Agricultores**: Aunque los agricultores cuentan con un conocimiento arraigado y experiencia respecto a los ciclos naturales de precipitación en sus regiones, el impacto del cambio climático ha reducido significativamente el rendimiento de los cultivos[3]. Por consiguiente, se hace imperativo realizar una transición hacia nuevas prácticas agrícolas, fundamentadas en modelos y datos que se adapten a las condiciones climáticas y geográficas particulares de cada zona.

4. **Sociedad**: La sociedad en su conjunto también desempeña un papel crucial en la mitigación de los impactos de las fluctuaciones climáticas en la agricultura y la energía. Esto incluye el fomento de prácticas de consumo responsable, el apoyo a políticas ambientales y el impulso de la conciencia pública sobre la importancia de la gestión sostenible de los recursos naturales de manera colectiva.

En conclusión, es evidente que el problema es compartido y requiere la colaboración y el compromiso de múltiples actores para implementar soluciones efectivas y sostenibles. se podrían incluir otros actores, como el sector investigativo y tecnológico, pero implícitamente están incluidos en los 4 puntos mencionados anteriormente.

2. Datos.

a. Describa la fuente de datos (Sea más específico que en documento pasado)

1. En el ámbito de las precipitaciones se tienen 2 datasets obtenidos del sitio web de datos abiertos de Colombia ([www.datos.gov.co/](http://www.datos.gov.co/)). El primero es el dataset [Precipitación](https://www.datos.gov.co/en/Ambiente-y-Desarrollo-Sostenible/Precipitaci-n/s54a-sgyg/about_data) donde se tienen datos de precipitaciones extraídas del IDEAM que miden la cantidad de lluvia en milímetros en distintos departamentos y municipios diariamente (el dataset es actualizado diariamente también) el histórico viene desde 2019/08/27, además contiene datos geográficos de longitud y latitud y los distintos sensores que capturan las medidas. La segunda alternativa cuenta con medidas totales mensuales, pero es una opción delimitada sólo para el departamento de cundinamarca [Precipitaciones Totales Mensuales](https://www.datos.gov.co/en/Ambiente-y-Desarrollo-Sostenible/Precipitaciones-Totales-Mensuales/mb4n-6m2g/about_data) con columnas similares al anterior, la diferencia más significativa radica en que las mediciones de precipitaciones están divididas mensualmente en columnas, mientras que el primer dataset toma un registro diário específico. También el tiempo de los registros es diferente ya que el segundo dataset tiene registros desde el 2016/11/01 y el primero desde 2019/08/27, además, el tamaño de los datos es significativamente diferente, mientras el primer dataset posee 192 millones de registros, el segundo cuenta con aproximadamente 6200.
2. En cuanto a los temas de agricultura se han identificado 2 datasets de acceso libre que registran los precios y abastecimientos con fecha, tipo de alimento y lugar de destino de esos alimentos, con registros desde el 2013 al 2024. Estos datasets se encuentran disponibles públicamente en la página de la DIAN:

* [Sistema de Información de Precios y Abastecimiento del Sector Agropecuario Componente Abastecimiento de Alimentos - SIPSA - A - 2018 -2024](https://microdatos.dane.gov.co/index.php/catalog/697/get-microdata)
* [Sistema de Información de Precios y Abastecimiento del Sector Agropecuario Componente Abastecimiento de Alimentos - SIPSA - A - 2013 -2017](https://microdatos.dane.gov.co/index.php/catalog/645/get-microdata)

Además, se puede intentar correlacionar otro dataset que presenta información de áreas de cultivo en Antioquia [Áreas cultivadas y producción agrícola en Antioquia desde 1990-2022](https://www.datos.gov.co/en/Agricultura-y-Desarrollo-Rural/Areas-cultivadas-y-produccion-agr-cola-en-Antioqui/t2ca-uae5/about_data), con lo cual se podría obtener una relación entre la disponibilidad y los precios en Antioquia con base en las áreas de producción, este dataset tiene instancias con datos sobre el año, el municipio el área de siembra y el total de producción según el tipo de cultivo desde el año 1990 hasta el 2022.

1. Para el caso energético hay métricas disponibles desde el 2000 hasta la actualidad de la variación de los precios de oferta y demanda, además de exportaciones y demanda no atendida. Estos datos se encuentran en el sitio web de [XM](https://sinergox.xm.com.co/Paginas/Home.aspx) y [ido.xm.com.co/ido/SitePages/hidrologia.aspx?q=reservas](https://ido.xm.com.co/ido/SitePages/hidrologia.aspx?q=reservas).

b. ¿Cuál es la calidad que posee dicha fuente?

La calidad de los datos es buena, no se evidencian muchos datos nulos en ninguno de los datos de lluvias (no se han explorado todos ya que el volumen de datos es bastante grande, solo se han tomado algunas muestras porque incluso la API requiere de cierto nivel de conocimiento para consumirla ya que está paginada por eficiencia, al poseer 192 M de datos).

Con respecto a los datos de agricultura se observa un comportamiento igual sin nulos, lo que sí es evidente es un procesamiento de strings que están combinando distintas variables.

Con respecto a los datos energéticos una primera exploración muestra datos consistentes sin instancias nulas.

Todo esto es un análisis preliminar, ya que muchos otros retos de exploración, limpieza, modelación e integración pueden surgir en el camino

c. ¿Es accesible y disponible?

Como se expuso en el literal a, todos estos datos están disponibles en las fuentes referenciadas en el mismo literal (a).

3. Hipótesis.

a. ¿Qué es lo que pretende probar?

Se parte de la certeza que las lluvias afectan el suministro eléctrico y la agricultura en Colombia, traducido en el impacto monetario para los colombianos, es decir probar que las precipitaciones tienen una influencia importante en el precio de la energía y alimentos en Colombia

b. ¿Cuáles pueden ser las respuestas esperadas a lo que pretende probar?

Mediante el análisis de datos, se espera obtener información cuantitativa y procesable que permita comprender mejor el impacto de las precipitaciones en la agricultura y la generación de energía en Colombia, y desarrollar estrategias efectivas para abordar estos desafíos. Se pueden definir los siguientes puntos resultado de lo que se espera probar:

1. **Identificación de patrones climáticos**: Se espera que a través del análisis de datos climáticos históricos y actuales, los investigadores puedan identificar patrones y tendencias en las precipitaciones en diferentes regiones de Colombia. Esto permitiría comprender mejor cómo estas variaciones afectan la producción agrícola y la generación de energía hidroeléctrica.

2. **Correlación entre variables**: Se espera que el análisis de datos revele correlaciones significativas entre las fluctuaciones en las precipitaciones y la producción agrícola, así como la capacidad de generación hidroeléctrica. Esto proporcionaría evidencia cuantitativa de la influencia directa del clima en estos sectores críticos.

3. **Modelado predictivo**: Se espera que, a través del modelado predictivo y el análisis de datos, se puedan desarrollar modelos que pronostiquen cómo las variaciones futuras en las precipitaciones podrían impactar la producción agrícola y la generación de energía en Colombia. Estos modelos podrían ayudar a anticipar y mitigar los efectos adversos del clima en estos sectores.

4. **Optimización de recursos**: Se espera que el análisis de datos proporcione información valiosa para optimizar el uso de recursos, como el agua en la agricultura y la energía en la generación hidroeléctrica. Al entender mejor cómo las precipitaciones afectan la disponibilidad de recursos, se pueden desarrollar estrategias más eficientes para su gestión.

c. ¿Qué acciones debo tomar para cada respuesta anterior?

Para utilizar el análisis de datos como una herramienta efectiva para comprender y abordar los desafíos relacionados con las precipitaciones en la agricultura y la generación de energía en Colombia, planteamos los siguientes puntos:

1. **Recopilación y análisis de datos climáticos**:

* Identificar fuentes confiables de datos climáticos, como estaciones meteorológicas, bases de datos gubernamentales o instituciones de investigación (Lo cual se ha desarrollado en otros puntos anteriores).
* Recopila datos históricos de precipitaciones y otras variables climáticas relevantes para las regiones de interés en Colombia (Lo cual se ha desarrollado en otros puntos anteriores con las posibles bases de datos).
* Utiliza herramientas de análisis de datos, como software estadísticos preferiblemente Python o plataformas de ciencia de datos locales o en la nube, para analizar y visualizar los datos climáticos e identificar patrones y tendencias.

2. **Correlación y modelado predictivo**:

* Utilizar técnicas estadísticas para identificar correlaciones entre las variables climáticas y los indicadores de interés, como la producción agrícola o la generación hidroeléctrica.
* Desarrollar modelos predictivos utilizando técnicas de análisis de regresión u otras metodologías de modelado para predecir cómo las variaciones futuras en las precipitaciones podrían afectar estos indicadores.
* Validar y ajustar los modelos utilizando datos históricos y probar su capacidad predictiva utilizando conjuntos de datos independientes (se puede hacer una partición de los datos existentes).

3. **Desarrollo de estrategias de gestión de recursos**:

* Utilizar los resultados del análisis de datos para desarrollar estrategias de gestión de recursos que optimicen el uso de agua en la agricultura y la energía en la generación hidroeléctrica.
* Identifica áreas de intervención prioritarias basadas en las áreas más afectadas por las variaciones en las precipitaciones y las necesidades de mitigación.
* Colaborar con expertos en agricultura, energía y gestión de recursos hídricos para desarrollar e implementar políticas y prácticas que promuevan la resiliencia frente a los cambios climáticos.

Estos puntos son un planteamiento que podrían estar por fuera del alcance de la especialización.

1. **Despliegue del modelo, disponibilidad y consumo**:

* Desplegar los modelos, utilizando alguna herramienta actual, preferiblemente en la nube por facilidad.
* Disponibilizar algún consumo API para distintas organizaciones.
* Habilitar algún visualizador online de este modelo(s).

Estos puntos son un planteamiento que podrían estar por fuera del alcance de la especialización.

5. **Monitoreo y adaptación continua**:

* Establecer sistemas de monitoreo continuo para seguir de cerca los cambios en las variables climáticas y los indicadores de interés (en cuanto a las precipitaciones ya se tiene una base de datos actualizada diariamente).
* Ajustar las estrategias de gestión de recursos según sea necesario en respuesta a nuevas tendencias o eventos climáticos extremos.
* Fomenta la colaboración y el intercambio de información entre las partes interesadas para mejorar continuamente las prácticas de gestión y adaptación.

Igual que lo anterior estos puntos son un planteamiento complementario que podría estar por fuera del alcance de la especialización.

4. Solución.

a. Tipo. (¿Qué tipo de algoritmo de ML es? ¿Qué técnica es?

En la literatura se ha encontrado que tradicionalmente se han utilizado modelos conceptuales como el EBA4SUB para predecir fenómenos hidrológicos, investigaciones recientes indican que los modelos de Machine Learning, tales como ANFIS-PSO, ANFIS-FCM, MARS y M5Tree, tienen un desempeño superior en esta tarea [4], pero es importante tener otros modelos más simples como la regresión lineal múltiple en cuenta.

b. ¿Cómo debe ser implementado?

La idea sería encontrar el mejor algoritmo que permita integrar estos campos y ajustar los resultados a algo cercano a la realidad y proyectar el comportamiento en el futuro.

c. ¿Qué resultados se esperan?

Se busca desarrollar un modelo que tenga la capacidad de prever la variación en el precio de la energía y los alimentos en función de los cambios en la precipitación. Este modelo deberá contar con una interfaz gráfica que facilite la visualización de su comportamiento, e idealmente, una API que permita su utilización por parte de usuarios externos.

5. KPI’s

a. ¿Cómo espera evaluar el modelo?

Al utilizar una combinación de estos métodos de evaluación, se puede obtener una evaluación integral y confiable del modelo de análisis de datos, lo que permite determinar su utilidad y eficacia en la resolución de los problemas específicos relacionados con las precipitaciones en la agricultura y la generación de energía en Colombia.

**Validación Cruzada**: Dividir el conjunto de datos en subconjuntos de entrenamiento y prueba, ajustar el modelo a los datos de entrenamiento y evaluar su rendimiento en los datos de prueba. Este proceso se repite varias veces con diferentes divisiones de los datos para obtener una evaluación más robusta del modelo [5].

**Métricas de Desempeño**: Utilizar métricas específicas para evaluar el rendimiento del modelo, como el error cuadrático medio (ECM), el coeficiente de determinación (R cuadrado), la precisión, la sensibilidad y la especificidad, dependiendo del tipo de problema y del objetivo del modelo[6].

**Análisis de Residuos**: Examinar los residuos del modelo, es decir, la diferencia entre los valores observados y los valores predichos, para detectar patrones o anomalías que puedan indicar deficiencias en el modelo.

**Comparación con Modelos Alternativos**: Comparar el modelo desarrollado con otros modelos alternativos, como modelos de línea base simples o modelos más complejos, para determinar si el modelo propuesto ofrece mejoras significativas en el rendimiento[7].

**Validación Externa**: Evaluar el rendimiento del modelo en conjuntos de datos independientes o en entornos no observados para verificar su generalización y capacidad de aplicabilidad en situaciones del mundo real.

**Sensibilidad del Modelo**: Realizar análisis de sensibilidad para evaluar cómo pequeños cambios en los datos de entrada afectan las predicciones del modelo, lo que puede proporcionar información sobre la robustez y la estabilidad del modelo[7].

b. ¿Qué métricas debe usar?

Con los siguientes indicadores se podría tener una referencia de la efectividad de los modelos implementados:

**Indicadores de Resultados**: Estos indicadores miden los resultados o logros reales obtenidos como resultado del proyecto o iniciativa. Se centran en evaluar el impacto o efectividad del proyecto en términos de sus resultados previstos. Por ejemplo, observar si el modelo hace las proyecciones correctas en cuanto a los precios de energía y alimentos (provenientes de la agricultura colombiana). O si se implementaron algunas acciones basadas en el modelo, ver si tuvieron un efecto positivo o negativo en los precios de energía o alimentos (provenientes de la agricultura colombiana)

Como mapeo general de un indicador en la agricultura para medir la productividad luego de usar los modelos creados sería el *land Productivity* [8]

Como mapeo general de un indicador de energía para medir la productividad luego de usar los modelos creados sería el *Overall availability factor y Overall generation-forced outage rate*[9]*.*

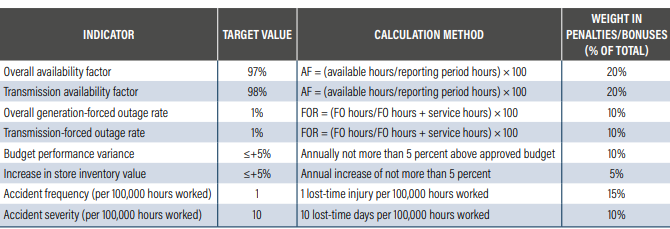


Figure 1. Hydroelectric indicators.

**Indicadores de Satisfacción de los Interesados (Actores)**: Estos indicadores miden los niveles de satisfacción de los principales interesados, como clientes, empleados o socios, con el proyecto o iniciativa. Proporcionan información sobre qué tan bien el proyecto está satisfaciendo las necesidades y expectativas de sus interesados y pueden ayudar a identificar áreas de mejora. Los indicadores de satisfacción de los interesados podrían incluir puntajes de retroalimentación del cliente, encuestas de compromiso de los empleados o calificaciones de satisfacción de los socios.

6. Actores.

a. ¿Quién es su cliente?

El cliente podría ser una entidad gubernamental, una organización sin fines de lucro (ONG), o incluso una empresa privada que esté involucrada en proyectos relacionados con la gestión de recursos hídricos, la agricultura sostenible o la generación de energía renovable en Colombia.

b. ¿Quiénes son los interesados (Stakeholders)?

Los interesados podrían incluir al gobierno colombiano a nivel nacional y local, instituciones agrícolas y de investigación, empresas agrícolas, empresas de generación de energía, consumidores de alimentos y electricidad, organizaciones ambientales, comunidades locales que dependen de la agricultura y la energía hidroeléctrica, entre otros.

c. ¿Quién usará la solución?

La solución desarrollada podría ser utilizada por agricultores y empresas agrícolas para mejorar la gestión del agua en la agricultura, por empresas de generación de energía para optimizar la producción de energía hidroeléctrica, por el gobierno para la planificación y políticas relacionadas con la seguridad alimentaria y la energía, así como por investigadores y organizaciones no gubernamentales para monitorear y analizar los impactos de las precipitaciones en la agricultura y la energía.

d. ¿A quién impactará?

La solución impactará a una amplia gama de partes interesadas, incluyendo a los agricultores y comunidades agrícolas que dependen de las lluvias para el riego de cultivos, a los consumidores de alimentos que pueden verse afectados por cambios en los precios de los alimentos, a las empresas y consumidores que dependen de la energía hidroeléctrica para la electricidad, así como al medio ambiente en general, que puede verse afectado por cambios en los patrones de precipitación y la gestión de recursos hídricos.

Referencias

[1] Semana, “Agricultura, el sector que más agua demanda en el país,” *Semana*, 2019. Accessed: Apr. 03, 2024. [Online]. Available: https://www.semana.com/medio-ambiente/articulo/agricultura-el-sector-que-mas-agua-demanda-en-el-pais/43450/

[2] Minenergía, “Colombia apuesta a la aplicación de un estándar mundial de sostenibilidad con el apoyo de la Cooperación Económica y Desarrollo (SECO) de Suiza y la Asociación Internacional de Energía hidroeléctrica (IHA),” *Minenergía*, 2023. https://www.minenergia.gov.co/es/sala-de-prensa/noticias-index/colombia-como-uno-de-los-líderes-latinoamericanos-en-energía-hidroeléctrica-le-apuesta-a-la-aplicación-de-un-estándar-mundial-de-sostenibilidad-con-el-apoyo-de-la-cooperación-económica-y-desar (accessed Apr. 03, 2024).

[3] C. Lau, A. Jarvis, and J. Ramirez, “Agricultura Colombiana : Adaptación al Cambio Climático,” *CIAT - Cent. Int. Agric. Trop.*, vol. 1, p. 4, 2011, [Online]. Available: http://dapa.ciat.cgiar.org/wp-content/uploads/2013/02/politica%7B\_%7Dsintesis1%7B\_%7Dcolombia%7B\_%7Dcambio%7B\_%7Dclimatico.pdf

[4] R. M. Adnan, A. Petroselli, S. Heddam, C. A. G. Santos, and O. Kisi, “Short term rainfall-runoff modelling using several machine learning methods and a conceptual event-based model,” *Stoch. Environ. Res. Risk Assess.*, vol. 35, no. 3, pp. 597–616, 2021, doi: 10.1007/s00477-020-01910-0.

[5] “Sets de Entrenamiento, Test y Validación | Aprende Machine Learning.” https://www.aprendemachinelearning.com/sets-de-entrenamiento-test-validacion-cruzada/ (accessed Apr. 28, 2024).

[6] “Cómo elegir y aplicar métricas de rendimiento para la ciencia de datos.” https://es.linkedin.com/advice/0/what-most-common-performance-metrics-data-scientists-pvwjf?lang=es (accessed Apr. 28, 2024).

[7] “3. Model selection and evaluation — scikit-learn 1.4.2 documentation.” https://scikit-learn.org/stable/model\_selection.html (accessed Apr. 28, 2024).

[8] FAO, “Productivity and Efficiency Measurement in Agriculture Literature Review and Gaps Analysis,” no. February, pp. 1–77, 2017, [Online]. Available: https://www.fao.org/3/ca6428en/ca6428en.pdf

[9] World Bank, “Operation and Maintenance Strategies for Hydropower: Handbook for Practitioners and Decision Makers,” *Oper. Maint. Strateg. Hydropower*, pp. 1–170, 2020, [Online]. Available: https://www.worldbank.org/en/topic/energy/publication/operation-and-maintenance-strategies-for-hydropower%0Afile:///C:/Users/bulfd/Downloads/Handbook-for-Practitioners-and-Decision-Makers.pdf