A picture containing logo

Description automatically generated

Tesis para “Maestría en explotación de datos y descubrimiento del conocimiento”

Estudio de impactos de sequía en rendimientos de cultivos agrícolas mediante métodos de aprendizaje automático

Plan de trabajo

Alumno: Santiago Luis Rovere (Ingeniero en informática, UBA)

Director: Andrés Farall (Doctor en ciencias de la atmósfera, UBA)

8 de julio de 2021

1. Sinopsis

El presente documento constituye el Plan de Trabajo para la Tesis de Maestría a ser llevada a cabo por el autor. Este trabajo de tesis forma parte de un proyecto de investigación y desarrollo denominado “Diseño e implementación inicial de un Sistema de Información sobre Sequías para el Sur de América del Sur (SISSA)”, el cual es financiado por el programa de Bienes Públicos Regionales del Banco Interamericano de Desarrollo (Cooperación Técnica RG-T3308 [1]).

Por lo expuesto previamente, y a fin de mejorar la motivación del tema de estudio (Sección 2), se proseguirá con una descripción del proyecto SISSA, su misión y principales objetivos (Sección 2.1). A continuación, se presentará formalmente el tema que se desarrollará en este trabajo de tesis y la importancia de su abordaje (Sección 2.2). Se concluirá la Sección 2 con un listado de los principales objetivos buscados y una explicación acerca de la transferencia de los resultados encontrados (Sección 2.3).

Una vez presentado el tema de estudio, en la Sección 3 se describirá el plan de trabajo propuesto para llevar a cabo la investigación. Finalmente, en la Sección 4 se definirá un cronograma de tareas con el fin estructurar el plan de trabajo propuesto en tareas que tengan estipuladas una duración aproximada.

1. Tema de estudio

## ¿Qué es el SISSA?

El SISSA es el Sistema de Información sobre Sequías para el sur de Sudamérica. El SISSA provee herramientas e información (<https://sissa.crc-sas.org>) sobre las sequías y sus impactos a gobiernos, instituciones no gubernamentales y privadas, e individuos. Esta información permite:

* monitorear y predecir la ocurrencia de sequías;
* anticipar los impactos esperables en sectores económicos y comunidades; y
* fomentar la planificación y preparación anterior a la ocurrencia de sequías para mitigar sus daños, aumentar la resiliencia y reducir la vulnerabilidad.

El propósito último del SISSA es reemplazar las acciones *reactivas* (posteriores) a una sequía por un enfoque *proactivo* que permita gestionar los riesgos y reducir vulnerabilidades.

El SISSA es una institución virtual que funciona en el marco del Centro Regional del Clima para el sur de América del Sur (CRC-SAS, <https://www.crc-sas.org>). El CRC-SAS es una organización constituida en forma de red, según los principios definidos por la Organización Meteorológica Mundial (OMM, <https://public.wmo.int/es>). Se encuentra en fase operativa y ofrece servicios climáticos en apoyo a los Servicios Meteorológicos e Hidrológicos Nacionales (SMHN) y otros usuarios de los países situados en la región sur de América del Sur.

## Estudio de impactos de la sequía en el rendimiento de los cultivos

Los eventos de sequía constituyen el factor más importante asociado a la disminución del rendimiento de los cultivos del sector agrícola. A su vez, esta disminución en el rendimiento de los cultivos se traduce en millonarias pérdidas para el sector y la economía en general. Es por ello por lo que resulta de interés el estudio de este fenómeno con el objetivo de definir políticas de prevención y mitigación.

Para poder llevar a cabo acciones de prevención y mitigación, se requiere una comprensión de los vínculos entre las características de los eventos de sequía (por ejemplo, momento de ocurrencia, duración, intensidad, etc.) y sus impactos específicos sobre el rendimiento de los cultivos. El propósito de este trabajo será generar información cuantitativa que vincule condiciones de sequía con rendimientos obtenidos para cultivos comercialmente importantes.

Generar tal información será fundamental para diseñar un sistema de alerta temprana para el sector agrícola. Para emitir una alerta temprana es necesario, además, cuantificar o estimar los posibles impactos con cierto tiempo de antelación con el fin de gestionar los riesgos encontrados y llevar a cabo acciones de mitigación de dichos impactos. Para evaluar los impactos, se buscará vincular diferentes tipos de eventos secos (cortos pero intensos, largos y poco intensos, etc.) que ocurran en los distintos momentos del ciclo del cultivo con los rendimientos resultantes.

Un enfoque tradicional para abordar este estudio sería mediante la vinculación de variables meteorológicas históricas que definan condiciones de sequía y los resultados reales obtenidos. Sin embargo, abarcar todas las posibles combinaciones de tipos de eventos, momentos del cultivo y regiones geográficas (las cuales as su vez incluyen distintos tipos de suelo y de manejos de los cultivos) resulta una tarea imposible de ser realizada con series históricas, dado que no incluyen la diversidad condiciones que se quieren estudiar.

La alternativa propuesta se basa en la utilización del modelo de simulación DSSAT (<https://dssat.net>). Este modelo representa el crecimiento y rendimiento del cultivo, y además provee información sobre variables intermedias (ej. porcentaje de agua útil, estrés del cultivo, etc.). El uso de DSSAT además permite ignorar efectos de confusión que pueden agregar variación al rendimiento (enfermedades, plagas, innovaciones tecnológicas, etc.) causadas por factores no relacionados directamente con la sequía.

DSSAT es una herramienta de software que está compuesta por un conjunto de módulos, cada uno de los cuales tiene el propósito de modelar un determinado fenómeno. A su vez, cada módulo está compuesto por varios submódulos que representan procesos biológicos relativamente simples. Estos procesos están descriptos por modelos basados en ecuaciones diferenciales en derivadas parciales.

Un primer enfoque para vincular condiciones de sequía (datos de entrada) con rendimientos resultantes (datos de salida), sería mediante el estudio de las ecuaciones diferenciales que gobiernan cada uno de los procesos que componen el DSSAT. Sin embargo, esta tarea sería operacionalmente imposible debido a la dimensionalidad de los datos de entrada y la complejidad del modelo integrado resultante (que, a su vez, contiene un considerable número de parámetros asociados a manejos del cultivo, tipo de suelos, etc.).

El problema de la dimensionalidad radica en el gran número de variables de entrada, dado que es necesario definir valores de temperatura y precipitación para cada uno de los días del ciclo de vida del cultivo. Cada uno de estos valores representa un dato de entrada, por lo que el modelo resultante puede considerarse una composición de funciones dependientes de algunos cientos de variables. A su vez, esta composición de procesos (que son más de una decena, tal como se muestra en la Fig. 1) implica que el modelo, visto como un todo, resulta ser un sistema muy complejo como para ser estudiado por métodos analíticos tradicionales.

Por lo expuesto previamente, resulta más adecuado aplicar el método de Montecarlo para abordar el estudio de los efectos de la sequía en los rendimientos de los cultivos. Este método es utilizado para aproximar expresiones matemáticas complejas y costosas de evaluar con exactitud. Para ello se requiere la generación aleatoria de numerosos conjuntos de datos de entrada, los cuales puedan ser vinculados a datos de salida mediante procesos de simulación llevados a cabo mediante el uso de DSSAT.

Luego de obtener los resultados de las simulaciones, es necesario realizar tareas de análisis para vincular los datos de entrada con los rendimientos obtenidos. Considerando la dimensionalidad de los datos de entrada y la cantidad de condiciones que se quieren simular, resulta adecuado aplicar técnicas de análisis basadas en procesos de aprendizaje automático. Estas técnicas permitirán encontrar patrones que vinculen las condiciones de sequía con los rendimientos obtenidos.

## Objetivos buscados y transferencia de resultados

1. Plan de trabajo
2. Cronograma de tareas

Este documento resume las actividades realizadas por el autor en el marco de la Cooperación Técnica RG-T3308, “Diseño e Implementación de un Sistema de Información sobre Sequías para el Sur de América del Sur (SISSA)” financiada por el programa de Bienes Públicos Regionales del Banco Interamericano de Desarrollo (BID). En la Tabla 1 se listan los productos, módulos y tareas llevadas a cabo en el marco de cada uno de los cuatro entregables previos.

Tabla 1. Resumen de tareas ejecutadas en el marco de la Consultoría n° 2.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *#* | *Entregable* | *Producto* | *Módulo* | *Tareas* |
| 1 | Entregable I | Plan de trabajo | - | Desarrollo de plan de trabajo con la descripción de las tareas a realizar en el marco de la Consultoría n° 2 y la definición de un cronograma de entrega. |
| 2 | Entregable II | API | Relevamiento y análisis | Análisis de requerimientos mínimos y de servicios que se harán disponibles a través de la API. |
| 3 | Entregable II | API | Diseño conceptual | Elección de modelo de API y metodología de autenticación y control de permisos. |
| 4 | Entregable II | API | Implementación | Elección del lenguaje y paquetes para el desarrollo de la API y descripción de servicios implementados. |
| 5 | Entregable II | API | Documentación | Elaboración de manual de referencia con ejemplos en R y Python. |
| 6 | Entregable II | API | Difusión | Elaboración de material didáctico y realización de webinario llevado a cabo el día 30 de octubre de 2020. |
| 7 | Entregable III | CHIRPS-GEFS | Procesamiento | Descarga y actualización de datos de CHIRPS-GEFS. |
| 8 | Entregable III | CHIRPS-GEFS | Procesamiento | Selección de datos y construcción de series de precipitaciones observadas/pronosticadas. |
| 9 | Entregable III | CHIRPS-GEFS | Procesamiento | Cálculo de SPI en base a series de precipitaciones observadas/pronosticadas. |
| 10 | Entregable III | CHIRPS-GEFS | Visualización | Descripción de aplicación de visualización interactiva. |
| 11 | Entregable III | Estado actual de la sequía | Visualización | Descripción de aplicación de visualización interactiva. |
| 12 | Entregable IV | API | Series temporales | Implementación y documentación de nuevos servicios web para la extracción de series temporales de valores para productos satelitales. |
| 13 | Entregable IV | Aplicación sobre escenarios plausibles de SPI-3 | Procesamiento | Descripción del proceso de cálculo de escenario de SPI-3 a futuro en base a datos históricos de precipitaciones. |
| 14 | Entregable IV | Aplicación sobre escenarios plausibles de SPI-3 | Aplicación de visualización interactiva | Descripción de aplicación de visualización interactiva con distintos informes organizados en solapas. |
| 15 | Entregable IV | Validación de producto CHIRPS | Descripción de metodología y datos | Descripción de producto a validar, área y período de estudio, datos de verdad de campo y proceso de regionalización y selección de estaciones meteorológicas para validación. |
| 16 | Entregable IV | Validación de producto CHIRPS | Validación basada en montos de precipitación | Análisis de resultados de diagnósticos basados la comparación de montos de precipitación a nivel mensual y trimestral. |
| 17 | Entregable IV | Validación de producto CHIRPS | Validación basada en eventos de sequía | Análisis de resultados de diagnósticos basados la comparación eventos de sequía a nivel mensual y trimestral. |
| 18 | Entregable IV | Difusión de productos | Webinario | Descripción del temario a desarrollar durante el webinario a llevarse a cabo durante el mes de abril de 2021. |

Referencias

[1] RG-T3308: Diseño e Implementación Inicial de un Sistema de Información sobre Sequías para el sur de América del Sur, <https://www.iadb.org/es/project/RG-T3308>.