

ISC-315-T. Proyecto de soporte de lenguaje para la manipulación de un modelo de una investigación.

Objetivo general del proyecto

El objetivo consiste en elaborar un lenguaje de dominio específico (*DSL*) para la manipulación y ajuste de un modelo en una investigación en modo interactivo. *No se solicita la realización del lenguaje propuesto.*

El diseño de cualquier lenguaje, especialmente los *DSL*, exige un conocimiento muy detallado del dominio a que debe responder. La descripción que sigue es para dar una idea de las características de una situación más o menos típica.

La situación que se describe es extremadamente detallada. Se advierte que es una situación *típica*, aunque en cada caso pudieran presentarse otras diferentes aunque similares.

Descripción general de una investigación típica.

La investigación tiene los siguientes objetivos:

- Obtención de data experimental sobre la interceptación foliar de un cultivo de caña de azúcar.
- Análisis de la data experimental, orientada hacia el descubrimiento de patrones de comportamiento de la data acopiada.
Este análisis probablemente envuelva la exploración/identificación de patrones a través de análisis de regresión, eventualmente con influencias factoriales.
- Contrastación de un modelo teórico del comportamiento, tanto en ajuste como en validación.

El proyecto que se solicita se enfoca en abordar este último punto. Se supone que los restantes han sido resueltos de alguna manera.

La caña de azúcar se caracteriza por tener una gran masa foliar. Esto hace que las plantaciones de caña de azúcar formen una tupida masa de hojas. La investigación se propone indagar sobre la influencia de esta masa foliar en la cantidad de agua de lluvia que resulta obstaculizada por las hojas y que entorpece su llegada al suelo.

A este fin se concibe una investigación que arroje información sobre ese comportamiento.

Términos empleados

A continuación se detallan los términos utilizados en este documento.

Término	Descripción
Investigación	Período de tiempo corto o largo en que se obtienen datos para ser custodiados y procesados. Habitualmente está compuesto por una o más campañas.
Campaña	Se entenderá por ella un conjunto de experimentos cercanos en el tiempo. Se caracterizan por un inicio y final (fechas)
Sesión	Conjunto de mediciones realizadas en un lugar e intervalo de tiempo especificado. Supondremos que los experimentos son homogéneos y tienen características equivalente.
Medición	La sesión consta de la información y mediciones asociados a él.
Cañaveral	Plantío de cañas, especialmente caña de azúcar.
Guardarraya	Camino estrecho entre dos espacios cultivados, especialmente en un cañaveral, que permite el paso de personas, animales y vehículos y que sirve de línea divisoria en campos de cultivo.

Término	Descripción
Lote	Espacio de terreno con caña de azúcar sembrada y mantenida con las técnicas agronómicas usuales. El lote es de dimensiones notablemente menores que una plantación de explotación. Está delimitada por un borde no cultivado, que se corresponde con su límite (guardarraya).
Surco	Hendidura que se hace en la tierra con el arado.
Camellón	Lomo de tierra, como el que queda entre surco y surco al arar un terreno o el que se levanta con la azada.
Plantón	Agrupación típica en un lote de caña de azúcar que ha sido cosechado varias veces. En ese caso, los brotes sucesivos da lugar que varios tallos compartan un mismo rizoma, del cual parten a menara de racimo.
Radio efectivo de interceptación	Distancia de la masa foliar más lejana respecto al centro del plantón.
Pluviómetro	Dispositivo de medición de la lámina de lluvia [mm] captada en un punto dado a una altura determinada en un momento dado. Su diseño y funcionamiento no tienen importancia para este trabajo.
Pluviómetro de referencia	<i>Pluviómetro</i> situado fuera del lote pero cercano a él. Se asume que su captación no es obstaculizada por nada. Su diseño y funcionamiento no tienen importancia para este trabajo.
GPS	Dispositivo que permite ubicar la posición absoluta de cualquier elemento del experimento. Su salida será en términos de latitud/longitud.
Borde	Plantones del lote ubicados en su frontera: son aquellos plantones que están inmediatos a la guardarraya. Se considerará que el borde del lote está especificado dando las posiciones de los plantones del borde.

Todos estos términos serán *italizados* en su uso para recordar que siempre deberá entenderse el significado dado antes.

Desarrollo de una sesión

Asumiendo que los pluviómetros se encuentran instalados en el lote, la secuencia típica de acciones de captación de la información sería:

1. Anotación de los datos identificadores del *lote*.
2. Anotación de la *campaña, sesión*, fecha de inicio.
3. Anotación de las posiciones de los *plantones* del *borde* del *lote*.
4. Anotación de la posición del *pluviómetro* de referencia y su altura.
5. Anotación de la posición de los *pluviómetros* de captación a diferentes alturas.
6. Anotación de la posición de los *plantones* en torno a los *pluviómetros* de captación.
7. Anotación de la *hora de inicio*.
8. Anotación de las lecturas de *lámina* de los *pluviómetros* cada cierto tiempo.
9. Anotación de la *fecha de terminación* y *hora de terminación* de la *sesión*.
10. Anotaciones suplementarias **libres** sobre la *sesión* desarrollada.

En caso de que la *sesión* durara más de un día y se tropezara con un incidente (caída de un *pluviómetro*, por ejemplo), el investigador cerraría la sesión iniciada llevando a cabo los pasos 8 y 9 e inmediatamente abriría una nueva realizando anotaciones del incidente.

Detalles adicionales de las informaciones registradas asociadas a los elementos de la investigación

Elemento	Informaciones registradas
Lote	Identificador del <i>lote</i> , posición del <i>lote</i> (dado por <i>GPS</i> de un punto cercano al centro del <i>lote</i>).
Campaña	Fechas de inicio y terminación de la campaña.
Sesión	Fecha de inicio, hora de inicio, fecha de terminación, hora de terminación, intensidad y velocidad del viento, temperatura.
Pluviómetro	<i>Lote</i> , <i>campaña</i> , <i>sesión</i> , <i>posición</i> , altura respecto al suelo, si es interno, si se encuentra sobre un <i>surco</i> o un <i>camellón</i> . Parejas de datos (momento, lámina) donde el momento incluye fecha y hora (hh:mm:ss).
Plantón	<i>Lote</i> , <i>campaña</i> , <i>sesión</i> , <i>posición</i> , altura máxima de las hojas, posición del <i>pluviómetro</i> más cercano. Debe registrarse si se trata o no de un plantón de <i>borde</i> .

Observe que el momento de registro de la lámina debe darse *incluyendo la fecha*, puesto que puede haber *sesiones* que duren varios días.

Fórmula de cálculo de la distancia entre dos puntos dadas su latitud ϕ y longitud λ

Se ofrece un método de cálculo para la distancia entre dos puntos de posiciones tomadas de *GPS*.

$$a = \sin^2 \left(\frac{\Delta\phi}{2} \right) + \cos(\phi_1) \cdot \cos(\phi_2) \cdot \sin^2 \left(\frac{\Delta\lambda}{2} \right)$$

Usando la definición de *arctan2*,

$$\text{atan2}(y, x) = \begin{cases} 2 \arctan \left(\frac{y}{\sqrt{x^2 + y^2} + x} \right) & \text{si } x > 0, \\ 2 \arctan \left(\frac{\sqrt{x^2 + y^2} - x}{y} \right) & \text{si } x \leq 0 \text{ y } y \neq 0, \\ \pi & \text{si } x < 0 \text{ y } y = 0, \\ \text{indefinida} & \text{si } x = 0 \text{ y } y = 0. \end{cases}$$

calculando

$$c = 2 \cdot \text{arctan2}(\sqrt{a}, \sqrt{1-a})$$

entonces, la distancia d entre ellos será

$$d = R \cdot c$$

considerando

$R = 6371$ (radio medio de la Tierra en km)

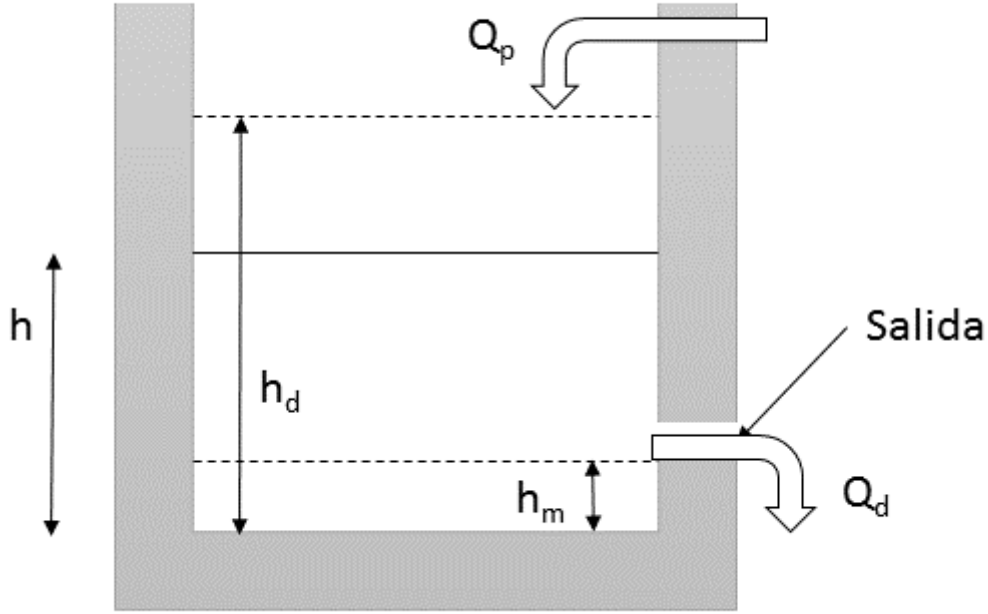


Figura 1: Esquema del modelo

Descripción cualitativa del modelo

El esquema muestra un modelo de interceptación muy simple, compuesto por una entrada externa H , un compartimiento de acumulación y una salida de descarga. El volumen acumulado (lámina) en un instante dado t se describe con la variable h . El modelo supone que el agua recibida se va acumulando (sin producir descarga) hasta un valor acumulado determinado h_d . En ese momento, se dispara la salida (siguiendo una ley de descarga especificada). Esa descarga se mantiene hasta el momento en que se alcanza una altura de lámina correspondiente al “volumen muerto” h_m . Si continúa la entrada externa, el proceso se repite.

Debe desarrollarse para este modelo un conjunto de funciones en R que permita experimentar con diferentes valores de sus parámetros y, eventualmente, facilidades para su expresión gráfica.

Estructura general del modelo y significado de las variables y parámetros

Variable/parámetro	Tipo	Descripción
$Q_p(t) = \frac{dH}{dt}$	variable externa	Intensidad de la precipitación recibida.
h	variable	Lámina acumulada.
h_d	parámetro	Lámina máxima posible de acumular.
h_m	parámetro	Lámina mínima (lámina muerta). Esta lámina no es evacuada nunca, aún cuando inicialmente esté vacía.
A_{int}	parámetro	Área efectiva de interceptación.
A_{ef}	parámetro	Área efectiva de descarga.

Variable/parámetro	Tipo	Descripción
$\phi(h) = A_{ef}\sqrt{2gh}$	función	Especifica la ley que gobierna el proceso de descarga del compartimiento.
$Q_d = \frac{dh}{dt} = \phi(h)$	variable	Intensidad de la descarga.

Los parámetros son constantes durante todo el proceso. La variable externa es el input del modelo. Las variables varían en el tiempo durante el proceso debido a la entrada de la precipitación y la influencia de los parámetros.

Dinámica del modelo

A continuación se describe la dinámica del modelo junto con las relaciones matemáticas correspondientes.

1. Al inicio de la sesión, $ht(t) = 0$.
2. Se produce salida cuando $h(t) > h_m$.
3. Cuando $h(t) = h_d$, se produce una descarga que la lleva a alcanzar el valor h_m . Esa descarga viene dada por $Q_d = A_{ef}\sqrt{2gh}$, donde A_{ef} es la sección efectiva de la vía de descarga.
4. Como el agua no puede desaparecer, tenemos que siempre $A_{int}\frac{dh}{dt} = Q_p - Q_d$, donde A_{int} es área de interceptación efectiva.

Entradas y salidas del modelo

Aparte de los parámetros h_m , h_d , A_{ef} , A_{int} , que se suponen constantes durante el proceso, la entrada es la precipitación como lámina acumulada en el tiempo $Q_p(t)$.

Las salidas o productos del modelo serán:

- $Q_d(t)$: dependencia funcional de la descarga hacia el suelo en el tiempo.
- $h(t)$: lámina en compartimiento en el tiempo t .

Capacidades de manipulación del modelo.

Debe entenderse que las siguientes capacidades de manipulación interactiva por parte del lenguaje propuesto son *mínimas*. Se detallan después aquellas que equieren una descripción adicional.

- Especificación de selección sobre la data de entrada / salida.
- Especificación de consolidación de la data de entrada / salida.
- Ajuste del modelo a la data experimental.
- Ejecución del modelo para valores dados de sus parámetros y entradas / salidas especificadas.
- Graficado de entradas y salidas.

Ajuste del modelo a la data experimental. Ejemplo del caso típico.

Se debe ser capaz de ajustar los parámetros constantes del modelo a partir de los datos experimentales correspondientes a partir de $Q_p(t)$ y la lámina obtenida en los captadores.

Debe tenerse en cuenta que la descarga hacia el suelo Q_d se corresponderá (aproximadamente) con la lámina del captador a una altura determinada.

“Ajustar” significa que deberán obtenerse valores de los parámetros que minimicen las desviaciones de la predicción del modelo respecto a la data medida. La minimización debe realizarse respecto a la suma de las desviaciones cuadráticas *de los volúmenes*.

Esto significa que la función objetivo deberá expresarse como

$$\Delta(h_m, h_d, A_{ef}, A_{int}) = \sum_i [Q_{ob}(t_i) - Q_{est}(h_m, h_d, A_{ef}, A_{int}, t_i)]^2$$

O, alternativamente, en términos de las láminas

$$\Delta(h_m, h_d, A_{ef}, A_{int}) = \sum_i [h_{ob}(t_i) - h_{est}(h_m, h_d, A_{ef}, A_{int}, t_i)]^2$$

Perfil del usuario final.

- Investigador con capacidad y experiencia en el uso de lenguajes de programación a nivel de “power user” pero *no* como desarrollador para otros.
- Amplia experiencia en uso de ambientes interactivos de consola como *R* y/o *Python*.
- Puede realizar pequeños *scripts* en lenguajes de scripting como *R* y/o *Python*.
- Ocasionalmente ha escrito programas “stand-alone” en lenguajes compilados como *Fortran* cuando se ha visto casi forzado a hacerlo, pero prefiere evitarlo.
- Sus capacidades computacionales, aunque amplias, se limitan a lo que le ha resultado necesario para su trabajo investigativo.
- Ha sido capaz de llevar a cabo tareas de complejidad al nivel de la expuesta en la situación típica.

Productos detallados del proyecto.

Los productos del proyecto deben satisfacer las capacidades de manipulación del modelo expuestas antes.

La secuencia que sigue es la que parece más razonable, aunque algunas de las etapas pudieran tener una diferente según el criterio del diseñador del lenguaje.

1. Descripción *detallada* del posible entorno de ejecución del intérprete.
2. Selección y fundamentación del paradigma de programación.
3. Especificación detallada de los objetivos del lenguaje propuesto.
4. Especificación sintáctica del lenguaje propuesto.
5. Especificación semántica del lenguaje propuesto.
6. Indicaciones generales del procedimiento de realización del intérprete para el lenguaje propuesto. *No se solicita su realización.*
7. Ejemplo de una sesión de trabajo con el lenguaje. El ejemplo que se ofrezca deberá ser cuidadosamente documentado.

Formato de los documentos a presentar del proyecto

Los documentos serán presentados como documentos de Word o pdf. Todos los archivos serán remitidos en un archivo compactado de nombre “Proyecto.zip”.

Aclaraciones sobre el proyecto.

Las aclaraciones que sean requeridas serán solicitadas a través de la comunidad del curso. Eventualmente, las aclaraciones del profesor pudieran derivar en modificaciones de este documento pero nunca en una extensión de sus propósitos.