OPTIROYA

En lo que sigue, están las mejoras que se necesitaría para tener una herramienta flexible, ergonómica. Y potente.

Hay varias posibilidades de simplificación, que evidentemente reducen también la potencia de la herramienta:

* Podríamos simplificar lo señalado en los puntos 1 y 2, y definir que por defecto las variables que se van a usar para categorizar son “altura” y “tipo de variedad”. Los límites de las categorías, para la altura, y las modalidades, para la variedad también estarían predefinidas. Eventualmente, cada país podría definir qué variables quiere tener por defecto para categorizar, verificando previamente que sí explican la incidencia de roya. La variable “región” podría ser útil en algunos países.
* En el punto 3, los cambios son menores, pero necesarios.
* El punto 4 es importante. Una forma de simplificar drásticamente (pero no deseable), sería cargar los datos de un año solamente, el último que se tenga, por ejemplo, y basar el cálculo del número de parcelas en ese año únicamente.
* El punto 5 es clave.

Se describen secuencialmente las etapas, indicando, cada vez, cómo debería funcionar y cómo funciona en la actualidad.

1. **Pestaña « Variables más importantes – Datos de monitoreo »**
2. Cómo debería funcionar?

En la pestaña “Variables más importantes”, deben cargarse los datos históricos de la vigilancia (o monitoreo). Estos datos deben contener las variables en columna, con los nombres en la primera línea, y las parcelas observadas en las diferentes fechas en línea. Pueden utilizarse teóricamente todas las variables para correr el análisis “random forest” de esa pestaña. Hay dos variables indispensables, sin embargo: “Hojas.Evaluadas” y “Hojas.con.Roya”. Con esas dos variables se calcula la incidencia: 100\*Hojas.con.Roya / Hojas.Evaluadas. Evidentemente la variable de incidencia, si ésta está en la base de datos, no puede ser una variable explicativa, tampoco la fecha ni el nombre de los productores. Necesita un mensaje de advertencia.

Lo deseable sería tener un menú donde se le pida al usuario seleccionar las variables explicativas que él quiere utilizar para correr el análisis, y otro donde se le pida señalar donde están las variables que hemos llamado Hojas.Evaluadas y Hojas.con.Roya para calcular la incidencia.

Habría que poner un mensaje antes de correr el análisis de este tipo:

Este análisis debe hacerse una vez al año en el momento de la planificación anual. Tiene como objetivo identificar las variables que mejor explican la incidencia de roya, para usarlas en la construcción de categorías de fincas por monitorear en el año. Es importante incluir variables que permitan esta categorización (como, por ejemplo, altitud, susceptibilidad de la variedad, rendimiento esperado, cobertura de sombra, acidez del suelo etc…). Otras variables, particularmente las temporales, como el mes de la evaluación, el año, la fecha, o el nombre de los productores, aunque pueden entrar en el análisis, y explicar la roya, no pueden utilizarse en la categorización.

Después del análisis, sería bueno que salga un mensaje:

Las variables que explican mejor la incidencia de roya son *(y se indican las variables que salen en la figura por orden de explicación)*:

Para la categorización, se recomienda las dos variables más explicativas, considerando siempre que éstas permitan categorizar las parcelas por monitorear en el campo (excluyendo las variables temporales, por ejemplo)

1. Cómo trabaja ahora?

Hay un menú que nos pide seleccionar el año actual, o sea el año para el cual se quiere planificar el monitoreo. El modelo va a buscar en la base que se cargó los datos que no son del año seleccionado, para correr el análisis “random forest”.

En realidad, ese menú no se necesita. Pues el análisis debe correrse con los datos que se carguen, una vez al año, cuando se hace la planifiación anual.

1. **Pestaña « Variables más importantes – Elegir las variables »**
2. Cómo debería funcionar?

En el menú de las variables deberían salir las variables explicativas que el usuario identificó en el paso anterior. El usuario selecciona las variables que el “random forest” le sugiere u otras de acuerdo a su expertise o preferencia.

No deberían ser demasiadas categorías de fincas, para no complicar excesivamente el monitoreo: un máximo de 12 parece razonable. Esto significa que se pueden tener 4 clases en una variable y 3 en la otra como máximo, las cuales al cruzarse forman 12 categorías máximo.

Para formar las clases debería haber un paso que no existe en la actualidad. Cuando las variables iniciales son cuantitativas (por ejemplo, altura) o cualitativas con muchas modalidades (por ejemplo, variedad), debería haber un método que permita definir los límites de las clases de las variables cuantitativas y agrupar las modalidades en el caso de las variables cualitativas, de tal forma que no haya más de 4 clases por variable, y que esas clases difieran en cuanto a la incidencia de roya en los datos históricos.

Después de aplastar el botón “mostrar categorías”, debería haber un mensaje del tipo:

Recuerde estas categorías, e indíquelas en las bases de datos de los monitoreos mensuales para que OPTIROYA las reconozca.

1. Cómo trabaja ahora?

En la actualidad, sólo se puede seleccionar altura y variedad como variables que permiten la categorización. Se muestran las categorías de fincas por monitorear al aplastar el botón “mostrar categorías”, pero éstas están predefinidas en el código (4 clases para la altura, tres clases para la variedad). Las clases de variedad ya están en la base de datos: es la variable “cat\_Variedad”.

Las categorías de fincas no están en la base de datos.

1. **Pestaña « Número de parcelas – Parámetros »**
2. Cómo debería funcionar?

Esta pestaña sirve para definir algunos parámetros que permiten al modelo de funcionar. Debería indicarse en el cuadro azul, lo que representan los números: Número de parcelas existentes en cada categoría de parcelas.

Adicionalmente, deberían salir unos mensajes explicativos para que se sepa cómo eso afecta el cálculo del número de parcelas por monitorear:

Defina los parámetros siguientes.

Umbral de detectabilidad:

Es el umbral arriba del cual es fácil detectar la roya. Se recomienda usar 1% de incidencia. Cuando las incidencias estén debajo del umbral de detectabilidad, se incrementa el número de parcelas por monitorear para poder estimar adecuadamente la incidencia.

Número de plantas por parcela:

El aumentar el número de plantas monitoreadas por parcela facilita la detección de la roya y la estimación de incidencias bajas. Esto puede permitir reducir el número de parcelas por monitorear.

Número de parcelas mínimo:

Ese número se agrega al número de parcelas determinado por el modelo para evitar un número de parcelas por monitorear demasiado bajo. Se recomienda el número de 10, pero puede ser inferior.

1. Cómo trabaja ahora?

Defina los parámetros siguientes:

Número de parcelas existentes en cada categoría de parcelas:

Puede ser que algunas categorías de parcelas estén muy poco representadas en la realidad. No sería lógico monitorear muchas de estas parcelas si están poco representadas. El ingresar una estimación del número de parcelas existentes en cada categoría permite corregir los números de parcelas por monitorear que propone el modelo de acuerdo a esta información. Este dato es evidentemente muy aproximativo.

.

La herramienta trabaja así en la actualidad. Le faltan las explicaciones para entenderla, sin embargo.

1. **Pestaña « Número de parcelas – Tamaño a realizar »**

En este punto se hace la planificación anual del monitoreo.

1. Cómo debería funcionar?

Conociendo las variables que permiten categorizar las parcelas (punto 2), los parámetros (punto 3), y cargando los datos históricos de monitoreo de roya, se calcula el número de parcelas por monitorear en cada categoría de parcelas y por mes.

**Los datos de varios años se cargan todos juntos, pero se analiza un año a la vez**. El análisis podría hacerse automáticamente, identificando los años a través de la fecha.

El archivo debe contener las variables que definen la fecha (día, mes, año) y las variables “Hojas.Evaluadas” y “Hojas.con.Roya”. Igual que en el punto 1, el nombre de las variables debe estar en la primera línea. Adicionalmente, se deben conocer las categorías formadas. De acuerdo a lo indicado en el punto 2, se sugiere que la variable categoría esté en la base de datos.

Similarmente a lo señalado en el punto 1, sería útil tener un menú donde se puedan identificar las variables necesarias.

Hay un cálculo por año, y por lo tanto, para cada año, una propuesta diferente de número de parcelas por monitorear por cada categoría de parcelas y por mes. Estas propuestas deben poder guardarse para que el usuario tenga la posibilidad de escoger los números de parcelas por monitorear por categoría y por mes. Estos podrían ser:

\* los que se derivan de los datos del monitoreo de la roya de un año específico,

\* los promedios, medianas o máximos de los números de parcelas por monitorear por categoría y por mes, derivados de todos los años cargados.

Al final, el usuario podría descargar un archivo con esta información. Es importante que aparezca un mensaje donde se le señala que puede escoger la opción que más le conviene. Puede haber opciones más conservadoras que otras.

Estas son las opciones que tiene para la planificación anual del monitoreo de la roya en su país, según la categoría de parcela por monitorear y el mes. Seleccione la opción que más le conviene. Dentro de lo que hay que considerar para seleccionar la mejor opción están: su capacidad de monitoreo (cuantas parcelas puede efectivamente monitorear en términos de recursos humanos y económicos), su capacidad de re-monitorear al poco tiempo, en caso que el monitoreo inicial resulte sub-dimensionado (ver pestaña “datos monitoreo del mes”). Distribuya las parcelas en su país para que éstas representen lo mejor posible la realidad.

Distri.

1. Cómo trabaja ahora?

Actualmente, la herramienta sólo trabaja con las variables “altura” y “variedad” para categorizar las parcelas (punto 2). Las clases de “altura” están definidas en el código. Las clases de “variedad” son las de la variable “cat\_Variedad” en la base de datos.

La herramienta no trabaja si las categorías no pueden mostrarse en la pestaña “Variables más importantes – Elegir las variables”.

La herramienta utiliza la base de datos cargada en la pestaña “Variables más importantes – Datos de monitoreo”.

En la base, se utilizan los datos de las variables Fecha, “Hojas.Evaluadas” y “Hojas.con.Roya” de las líneas donde la variable “Tipo\_de\_datos” tiene el valor “a realizar”.

1. **Pestaña “Número de parcelas – Datos monitoreo del mes”**

En esta etapa se verifica que los monitoreos mensuales han sido eficientes o si requieren de complementarse con más parcelas.

1. Cómo debería funcionar?

Conociendo las variables que permiten categorizar las parcelas (punto 2), los parámetros (punto 3), y las variables “Hojas.Evaluadas” y “Hojas.con.Roya”, se verifica que el número de parcelas que se planificó monitorear (punto 4), ha sido suficiente para tener una buena precisión en la estimación de la incidencia.

Mensualmente, se cargan los datos del monitoreo mensual en curso. Igual que en los puntos 1 y 4, el nombre de las variables debe estar en la primera línea. Adicionalmente, se deben conocer las categorías formadas. De acuerdo a lo indicado en el punto 2, se sugiere que la variable categoría esté en la base de datos. Similarmente a lo señalado en los puntos 1 y 4, sería útil tener un menú donde se puedan identificar las variables necesarias.

Sería necesario un mensaje que indique que los datos por cargar son del mes en curso únicamente.

Cargue aquí los datos de monitoreo del mes en curso únicamente.

De los datos cargados, se deduce el número de parcelas monitoreadas en cada categoría.

El modelo vuelve a correr, y se comparan el número de parcelas monitoreadas con las que debieran hacerse para tener una buena estimación de la incidencia de roya en cada categoría, a luz de los datos cargados. Sale un mensaje OK, cuando el monitoreo ha sido suficiente, y sale un número de parcelas por hacer nuevamente, si el monitoreo no ha sido suficiente en ciertas categorías. En este último caso, hay que completar el monitoreo, y volver a cargar los datos, hasta que todas las categorías estén en OK.

Un mensaje podría salir explicándolo:

Si el resultado es OK para ciertas categorías de fincas, es que el número de parcelas que usted monitoreó, en el mes en curso para esas categorías es adecuado. Ya no tiene que monitorear más parcelas en estas categorías en el mes en curso.

Si en algunas categorías, el resultado es un número, ese número es la cantidad de parcelas que debe monitorear de nuevo para tener una buena estimación de la incidencia en el mes en curso. Monitoré ese número de parcelas en los próximos días y vuelva a cargar los datos para asegurarse que el monitoreo ahora es adecuado.

1. Cómo trabaja ahora?

Se carga una base de datos similar a la que se cargó en el punto 1, pero con datos adicionales del monitoreo del mes. Se vuelve a indicar el año actual (aunque no se utiliza; el año actual que se utiliza es el que se define en el punto 1). No es necesario, en todo caso, definir el año si sólo se cargan los datos del mes.

Se aplasta el botón “iniciar el análisis” y sale un mensaje que nos envía a la pestaña siguiente: “Tamaño efectuado y necesario”. No debería ser necesario cambiar de pestaña.

El número de datos monitoreados por categoría se saca de lo calculado en la pestaña “Número de parcelas – Tamaño a realizar”; no se calcula nuevamente a partir de la base de datos cargada. Se vuelve a correr el modelo con los datos cargados del mes para determinar el número de parcelas necesario por categoría. Para ese cálculo, se utilizan los datos para los cuales la variable “Tipo\_de\_datos” tiene el valor “efectuado”: básicamente porque la base contiene también datos históricos señalados con el valor “a realizar". Pero eso no es necesario si sólo se cargan los datos del mes.

Se obtiene un cuadro comparativo por categoría, donde se tiene el número de parcelas efectuado, el número de parcelas que debieran hacerse, la diferencia (lo que hay que completar) o un OK, si el monitoreo ha sido suficiente.

En esa misma pestaña, se pueden cargar los datos del mes que se han completado, y se repite el proceso.

La pestaña “Nueva ronda” es obsoleta.