



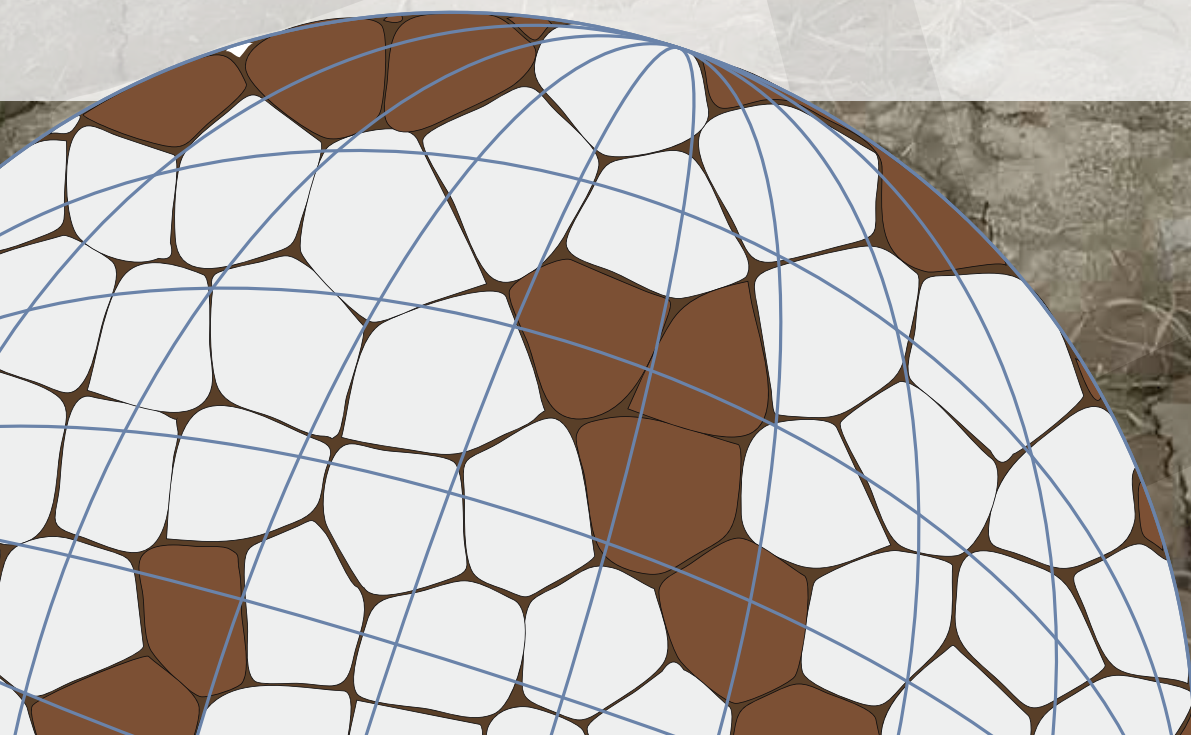
Organización de las Naciones
Unidas para la Alimentación
y la Agricultura

Mapeo de suelos afectados por salinidad

Mapa GSS

LECCIÓN 3

Modelado espacial de
indicadores (propiedades) de
suelos afectados por salinidad



ALIANZA MUNDIAL
POR EL SUELO

Mapeo de suelos afectados por salinidad Mapa GSS

LECCIÓN 3

Modelado espacial de indicadores (propiedades) de suelos afectados por salinidad

Christian Thine Omuto
Ronald Vargas
Kostiantyn Viatkin
Yusuf Yigini

Cita recomendada

Omuto CT, Vargas R, Viatkin K, Yigini Y. 2021. *Mapeo de suelos afectados por salinidad: Lección 3- Modelado espacial de indicadores (propiedades) de suelos afectados por salinidad*. FAO, Roma.

Las denominaciones empleadas en este producto informativo y la forma en que aparecen presentados los datos que contiene no implican, por parte de la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO), juicio alguno sobre la condición jurídica o nivel de desarrollo de países, territorios, ciudades o zonas, ni sobre sus autoridades, ni respecto de la demarcación de sus fronteras o límites. La mención de empresas o productos de fabricantes en particular, estén o no patentados, no implica que la FAO los apruebe o recomiende de preferencia a otros de naturaleza similar que no se mencionan.

Las opiniones expresadas en este producto informativo son las de su(s) autor(es), y no reflejan necesariamente los puntos de vista o políticas de la FAO.

Primera edición: 2020 (versión en inglés)

© FAO, 2021



Algunos derechos reservados. Esta obra se distribuye bajo licencia Creative Commons Atribución-NoComercial-CompartirIgual 3.0 Organizaciones intergubernamentales (CC BY-NC-SA 3.0 IGO; <https://creativecommons.org/licenses/by-nc-sa/3.0/igo/deed.es>).

De acuerdo con las condiciones de la licencia, se permite copiar, redistribuir y adaptar la obra para fines no comerciales, siempre que se cite correctamente, como se indica a continuación. En ningún uso que se haga de esta obra debe darse a entender que la FAO refrenda una organización, productos o servicios específicos. No está permitido utilizar el logotipo de la FAO. En caso de adaptación, debe concederse a la obra resultante la misma licencia o una licencia equivalente de Creative Commons. Si la obra se traduce, debe añadirse el siguiente descargo de responsabilidad junto a la referencia requerida: “La presente traducción no es obra de la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO). La FAO no se hace responsable del contenido ni de la exactitud de la traducción. La edición original en [idioma] será el texto autorizado”.

Todo litigio que surja en el marco de la licencia y no pueda resolverse de forma amistosa se resolverá a través de mediación y arbitraje según lo dispuesto en el artículo 8 de la licencia, a no ser que se disponga lo contrario en el presente documento. Las reglas de mediación vigentes serán el reglamento de mediación de la Organización Mundial de la Propiedad Intelectual <http://www.wipo.int/amc/en/mediation/rules> y todo arbitraje se llevará a cabo de manera conforme al reglamento de arbitraje de la Comisión de las Naciones Unidas para el Derecho Mercantil Internacional (CNUDMI).

Materiales de terceros. Si se desea reutilizar material contenido en esta obra que sea propiedad de terceros, por ejemplo, cuadros, gráficos o imágenes, corresponde al usuario determinar si se necesita autorización para tal reutilización y obtener la autorización del titular del derecho de autor. El riesgo de que se deriven reclamaciones de la infracción de los derechos de uso de un elemento que sea propiedad de terceros recae exclusivamente sobre el usuario.

Ventas, derechos y licencias. Los productos informativos de la FAO están disponibles en la página web de la Organización <http://www.fao.org/publications/es> y pueden adquirirse dirigiéndose a publications-sales@fao.org. Las solicitudes de uso comercial deben enviarse a través de la siguiente página web: www.fao.org/contact-us/licence-request. Las consultas sobre derechos y licencias deben remitirse a: copyright@fao.org.

Fotografías de las cubiertas anterior y posterior: ©FAO/Christian Thine Omuto

Índice

Resumen ejecutivo.....	iv
1. Introducción	1
1.1 Perspectiva	1
1.2 Objetivo.....	1
1.3 Resultados esperados	1
2. Requisitos para evaluar los suelos afectados por salinidad.....	2
2.1 Requisitos de datos.....	2
2.2 Requisitos de programas de computación	3
3 Recursos	4
4 Actividades.....	5
4.1 Cargando de datos y paquetes R	5
4.2 Verificación y armonización de la distribución estadística de las capas SIG	8
4.3 Armonización de los datos del suelo	11
4.4 Modelado espacial de indicadores	13
5. Resultados	19

Figuras

Figura 1: Ubicación caso de prueba en el Estado del Norte de Sudán	2
Figura 2: Creando la carpeta de trabajo	5
Figura 3: Cargando archivos del Proyecto R	6
Figura 4: Configuración del directorio de trabajo para los datos de prueba.....	6
Figura 5: Distribución estadística empírica de los índices de imágenes.....	9
Figura 6: Correlación de índices de imágenes y gráfico de sus componentes principales	10
Figura 7: Ejemplo de armonización profunda para CE_{se}	12
Figura 8: Representación gráfica de la distribución de frecuencia con una de predicción a un intervalo de confianza del 95%.....	15
Figura 9: Representación gráfica de CE prededidos Vs.los CE medidos.....	16
Figura 10: Representatividad de los rangos de validación de CE (puntos de muestreo) en el mapa de predicción (mapa de características)	17
Figura 11: Ancho de la predicción espacial al 95% del intervalo de confianza y sobreposición de los puntos de validación	18

Cuadro

Cuadro: Resumen de resultados del mapa digital de las propiedades de suelos para el establecimiento de la información nacional de los suelos afectados por salinidad	vi
Cuadro 1: Organización de los datos del caso de prueba	2
Cuadro 2: Imágenes de sensores remotos MODIS	3

Resumen ejecutivo

Esta Lección constituye el tercer paso del programa de la FAO diseñado para desarrollar las capacidades nacionales en mapeo digital de suelos enfocado al problema de salinidad de suelos. Igualmente, este programa para es útil para armonizar procedimientos y desarrollar nueva información y nuevos conocimientos sobre los suelos afectados por salinidad en los ámbitos nacional y global. El objetivo general de esta Lección es proveer a los participantes, herramientas de conocimiento para el desarrollo de mapas raster de los indicadores (propiedades) de suelos afectados por salinidad a escala nacional. Se espera que, al término de esta lección, los participantes cuenten con la capacidad técnica para generar, con una perspectiva nacional, información espacial respecto a los indicadores del suelo para la identificación de áreas geográficas afectados por salinidad de suelos en sus países.

Cuadro: Resumen de resultados del mapa digital de las propiedades de suelos para el establecimiento de la información nacional de los suelos afectados por salinidad

Concepto	Descripción	Formato
Mapas digitales de propiedades del suelo	Suelo superficial (0-30 cm) CE, pH, y PSI	archivos raster geoTiff
	Subsuelo (30-100 cm) CE, pH y PSI	archivo raster geoTiff
Mapas digitales de incertidumbre	Suelo superficial (0-30 cm) CE, pH, y PSI	archivos raster geoTiff
	Subsuelo (30-100 cm) CE, pH y PSI	archivo raster geoTiff
Archivo de Texto	Suelo superficial (0-30 cm) índices de precisión (ME, RMSE, r2, NSE)	Hoja de Cálculo
	Subsuelo (0-30 cm) índices de precisión (ME, RMSE, r2, NSE)	Hoja de Cálculo



1 Introducción

1.1 Perspectiva

Existen muchos métodos en la literatura para el mapeo de los suelos afectados por salinidad. Entre los cuales se encuentran aquellos basados en mapas por tipo de suelo, imágenes de sensores remotos, opinión especializada y mapeo digital de las propiedades del suelo (MDS) afectados por salinidad. El enfoque que utiliza el mapeo digital de indicadores del suelo tiene el potencial de cuantificar la incertidumbre y la precisión cartográfica además de desarrollar información espacial (mapas) de las propiedades del suelo afectados por salinidad. Esta Lección se centra en el mapeo digital de indicadores de suelos afectados por salinidad. La Lección también describe los pasos para evaluar la precisión e incertidumbres asociadas con el modelado espacial de indicadores de suelos afectados por salinidad. Está dirigida a expertos nacionales de instituciones con el mandato de generar y actualizar información de suelos a nivel nacional. Esta lección pretende el desarrollo de capacidades para el uso de estos datos sobre indicadores de suelos afectados por salinidad en las instituciones de los países participantes. Se espera que sus resultados contribuyan al desarrollo de la información espacial de suelos afectados por salinidad a nivel nacional y global.

1.2 Objetivo

El objetivo general de esta Lección es el modelado espacial de indicadores de suelo de problemas salinos utilizando el enfoque de la cartografía digital de suelos.

1.3 Resultados esperados

Se espera que, al finalizar esta Lección, los participantes puedan:

- i. Modelar espacialmente las propiedades de suelos afectados por salinidad, utilizando el enfoque del mapeo digital de suelos.
- ii. Evaluar la precisión e incertidumbre del mapeo de suelos.
- iii. Producir mapas digitales de los indicadores (propiedades) de suelos afectados por salinidad

2 Requisitos para evaluar los suelos afectados por salinidad

2.1 Requisitos de datos

Esta Lección utiliza datos de prueba que fueron recolectados del Estado del Norte de Sudán (<http://susis.sd>) (Figura 1). Los datos incluyen:

- Datos de suelo (conductividad eléctrica (CE), pH y Porcentaje de Sodio Intercambiable (PSI)).
- Covariables espaciales como precipitación media anual, cobertura de la tierra, geología, hidrogeología, imágenes de sensores remotos MODIS, altitud (MED).

Los datos del suelo se distribuyen como se muestra en el Cuadro 1.

Cuadro 1: Organización de los datos del caso de prueba

Muestra	Pozos	Longitud	Latitud Superior	Inferior	Horizonte	CE	pH	PSI	
5	2	29.81	20.62	0	10	1.000	1.900	8.600	10.000
6	2	29.81	20.62	10	30	2.000	0.700	7.800	5.000
7	2	29.81	20.62	30	100	3.000			
8	3	31.57	17.15	0	35	1.000	0.900	7.600	5.000
9	3	31.57	17.15	35	60	2.000	0.400	7.800	2.000
10	3	31.57	17.15	60	100	3.000	0.400	7.900	2.000

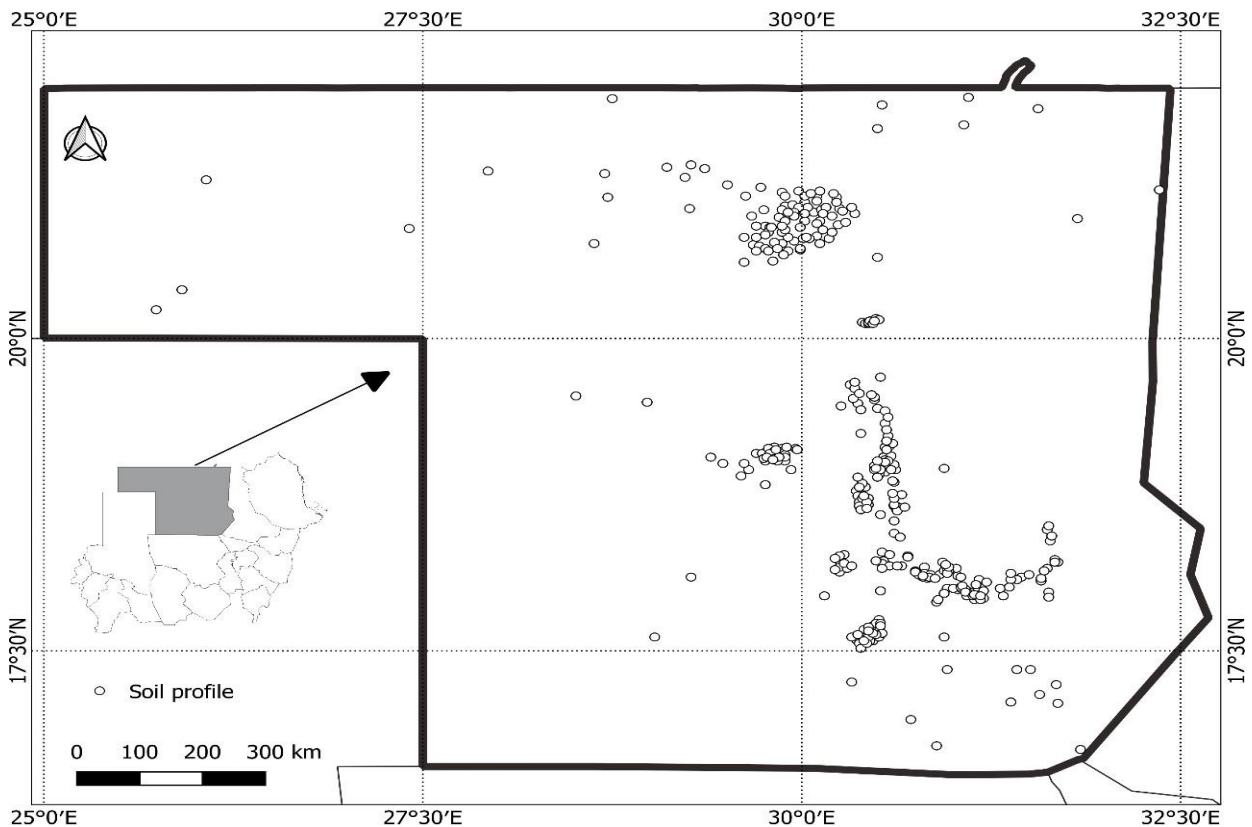


Figura 1: Ubicación caso de prueba en el Estado del Norte de Sudán

Otros archivos SIG son archivos raster a una resolución espacial de 1 km y proyectados como WGS 84 (UTM 36N). Las imágenes de sensores remotos son bandas de reflectancia corregidas MODIS (Cuadro 2).

Cuadro 2: Imágenes de sensores remotos MODIS

Imagen	Bandas espectrales	
MODIS MOD009GA V6	Banda 3	Azul: 0.459-0.479 μm
	Banda 4	Verde: 0.545-0.565 μm
	Banda 1	Roja: 0.62-0.67 μm
	Banda 2	NIR: 0.841-0.876 μm
	Banda 6	SWIR1: 1.628-1.652 μm
	Banda 7	SWIR2: 2.105-2.13 μm

Los datos del caso de prueba (*soil.RData*, *predictors.RData*) han sido almacenados en formato de archivo de datos R. *Soil.RData* es el conjunto de datos de calibración del suelo mientras que *predictors.RData* es una pila de archivos SIG raster.

2.2 Requisitos de programas de computación

La última versión de cada programa debería haberse instalado (Lección 2)

- i. R (<https://www.r-project.org/>)
- ii. QSIG (<https://qgis.org/en/site/forusers/download.html>)
- iii. RStudio (<https://rstudio.com/products/rstudio/download/#download>)
- iv. ILWIS (<https://www.itc.nl/ilwis/download/ilwis33/>)
- v. Programas de Hoja de Cálculo (Excel, Access) y programas para procesamiento de documentos (Word, Notepad)

También es necesario instalar los siguientes paquetes de R requeridos en esta lección para el modelado espacial en R: *soilassessment*, *sp*, *foreign*, *rgdal*, *car*, *carData*, *spacetime*, *gstat*, *automap*, *randomForest*, *e1071*, *caret*, *raster*, *soiltexture*, *GSIF*, *aqp*, *plyr*, *Hmisc*, *corrplot*, *factoextra*, *spup*, *purrr*, *lattice*, *ncf*, *ranger*.

3 Recursos

Los siguientes recursos serán de utilidad durante la recolección de datos:

- Referencias
 - Guías técnicas y manual para el mapeo de suelos afectados por salinidad¹ (GSP-Secretariat@fao.org).
 - Guías y especificaciones nacionales para el mapeo global de suelos afectados por salinidad².

¹ Omuto CT, Vargas R, El-Mobarak AM, Mohammed N, Viatkin, K. 2020. Global Mapping of Salt-affected Soils: Technical Guideline and Cookbook. FAO. Rome

² FAO. 2020. Mapping of salt-affected soils: Technical specifications and country guidelines. FAO. Rome

4 Actividades

4.1 Cargando de datos y paquetes R

#Paso 1: Cargue los datos y configure el directorio de trabajo

Cree una carpeta en C y nómbrala *Salinidad* (C:/Salinidad) dando clic en el botón derecho en *New Folder* en C, utilizando el explorador de Windows (Figura 2). Descomprima el archivo comprimido y descargado (*DSM_saltaffected.zip*) en C/Salinidad.

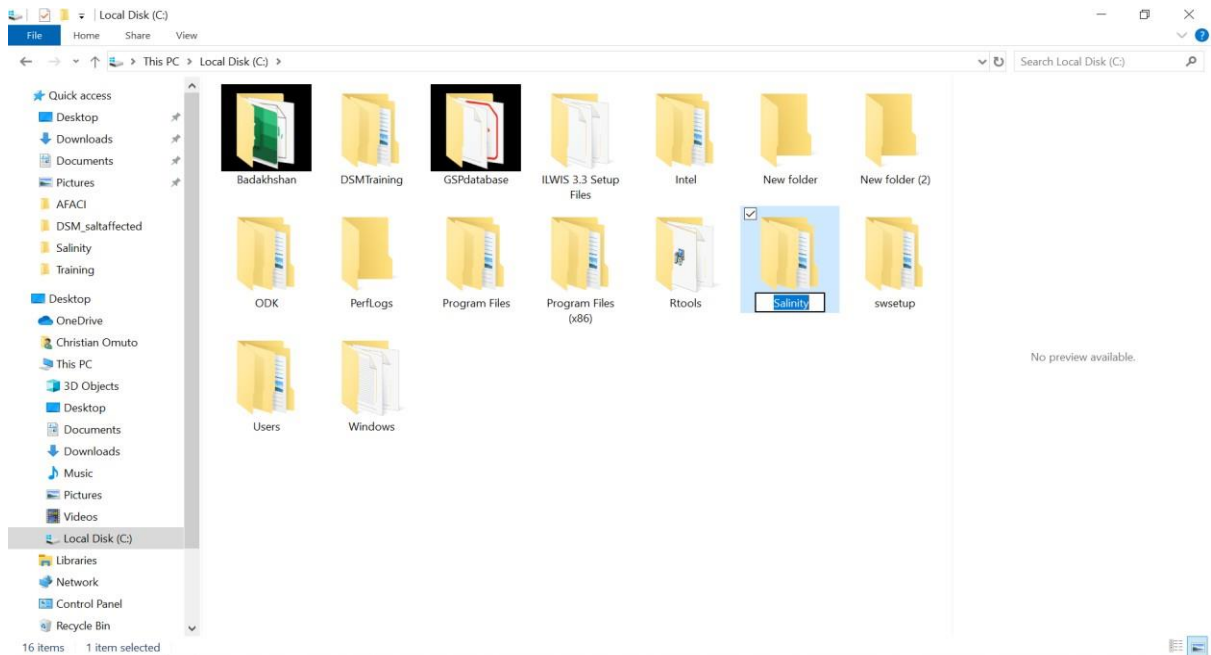


Figura 2: Creando la carpeta de trabajo

En RStudio, en la esquina superior derecha de la ventana *Espacio de Trabajo*, haga clic en *proyecto* y desplácese hacia abajo a *"Open Project"* (Abrir Proyecto). Navegue en *C:/Salinity/DSM_saltaffected* para localizar el archivo *"DSM_salaffected.Rproj"*. Seleccione para cargar el Proyecto. En la ventana de *Ficheros/Gráficos/Paquetes* de R, seleccione *File* (Archivo) (siendo las demás pestañas: *Plots* (Gráficos), *Packages* (Paquetes), *Help* (Ayuda), *Viewer* (Visor)). *File* mostrará un grupo de archivos en el *DSM_salaffected.Rproj* (Figura 3). Haga doble clic en los archivos (uno a la vez) y acepte el diálogo siguiente (*Digital_mapping_of_saltaffected_soils.R*, *predictors.RData*, *soil.RData*, *soilvalid.RData*).

Los paquetes de R anteriormente instalados contienen funciones útiles para el mapeo digital de suelos afectados por salinidad. Si los paquetes no fueron instalados durante la Lección 2, entonces deben instalarse primero antes de cargar las librerías.

```
install.packages(c("raster", "sp", "rgdal", "car", "carData", "dplyr", "spacetime", "gstat", "automap", "randomForest", "fitdistrplus", "e1071", "caret", "soilasses", "soiltexture", "GSIF", "aqp", "plyr", "Hmisc", "corrplot", "factoextra", "spup", "purrr", "lattice", "ncf", "npsurv", "lsei", "qrnn", "nnet", "mda", "RColorBrewer", "vcd", "readxl", "mapproj", "neuralnet", "psych", "ggrepel", "plotly"))
```

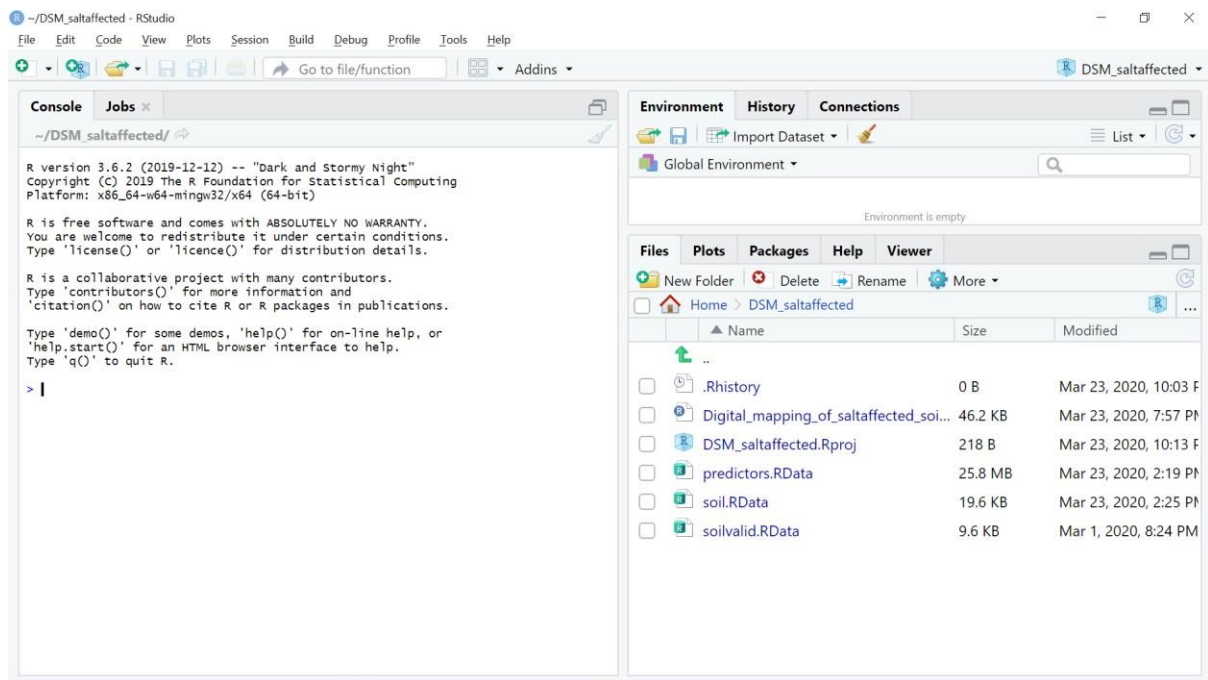


Figura 3: Cargando archivos del Proyecto R

En ventana *Espacio de Trabajo*, seleccione *More* (Más), y desplácese para elegir *Set as Working Directory* (Configurar como Directorio de Trabajo) (Figura 4). Este paso configura el directorio de trabajo de los datos de prueba.

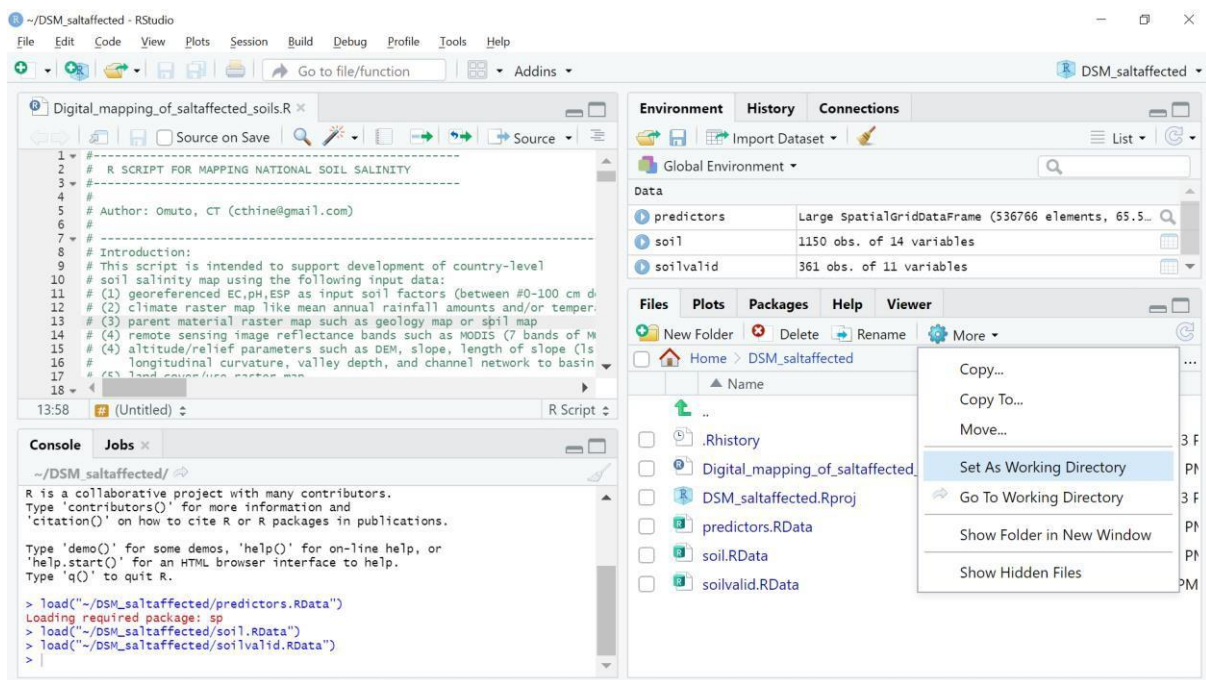


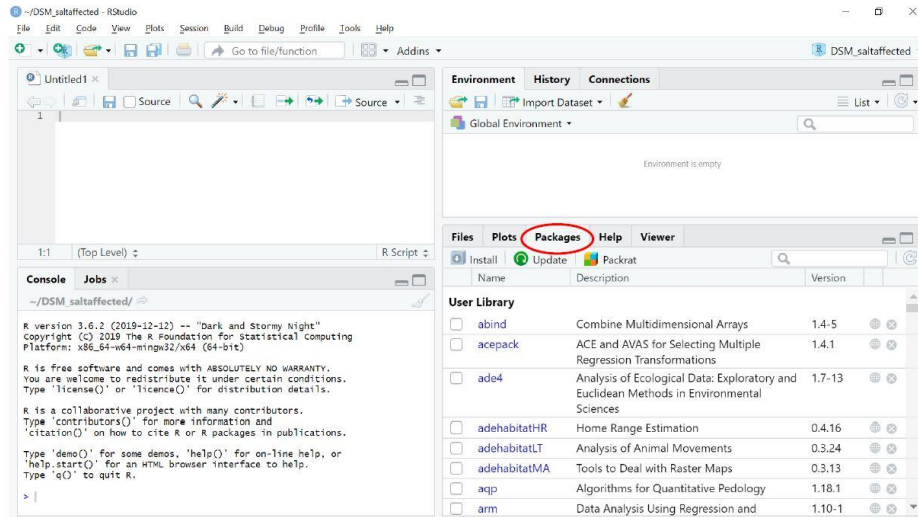
Figura 4: Configuración del directorio de trabajo para los datos de prueba

#Cargando paquetes R

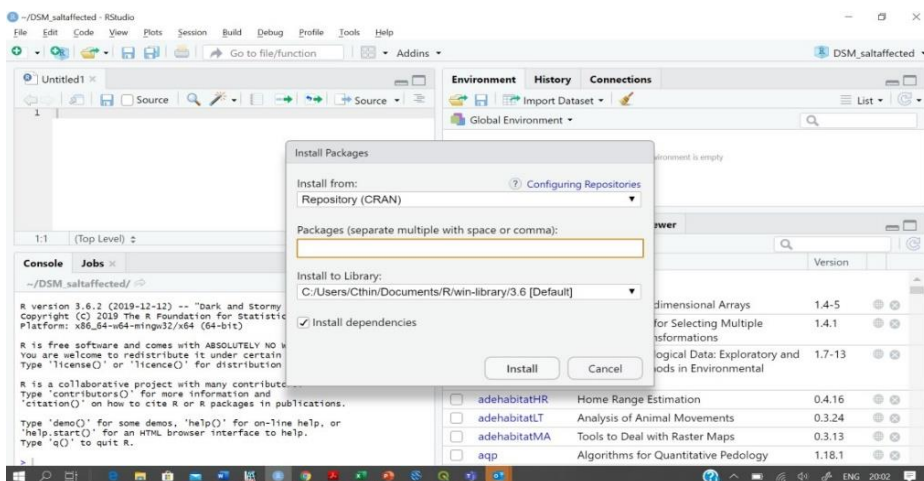
En caso de existir un error en la instalación:

1. Verifique si la conectividad de Internet es adecuada.

2. Si la conectividad de Internet es buena, entonces localice los paquetes que no estén instalados e instálelos manualmente.
3. En ventana *Espacio de Trabajo*, seleccione *Packages* (Paquetes) e *Install* (Instalar) en la vista emergente.



4. Se desplegará una ventana para instalar el paquete. En ese momento la conectividad de Internet debe estar activa.



5. Escriba el nombre del paquete en el espacio de *Packages* (si son varios paquetes, sepárelos con un espacio o coma): (Nota: También puede copiar y pegar cada viñeta debajo en el mismo espacio).
 - a. raster, sp, rgdal, car, carData, dplyr, spacetime, gstat, automap, randomForest, fitdistrplus, e1071;
 - b. caret, soilassessment, soiltexture, GSIF, aqp, plyr, Hmisc, corrplot, factoextra, spup, purrr, lattice;
 - c. ncf, npsurv, lsei, qrnn, nnet, mda, RColorBrewer, vcd, readxls, maptools, neuralnet, psych, ggrepel, plotly.

Las librerías deben cargarse después de instalar los paquetes:

```
>library(sp);library(foreign);library(rgdal);library(car);library(carData)
);library( maptools)
>library(spacetime);library(gstat);library(automap);library(randomForest)
);library(fitdistrplus);
>library(e1071);library(caret);library(raster);library(soilassessment);li
brary(soil texture);
>library(GSIF);library(aqp);library(plyr);library(Hmisc);library(corrplot
);library( factoextra)
>library(spup);library(purrr);library(lattice);library(ncf);library(npsur
v);library( lsei);
>library(nnet);library(class);library(mda);library(RColorBrewer);library(
vcd);library(grid);
>library(neuralnet);library(readxl);library(psych);library(qrnn);library(dplyr)
```

4.2 Verificación y armonización de la distribución estadística de las capas SIG

Antes de que se realice la verificación y armonización de la distribución estadística de las capas SIG, es importante verificar y eliminar *pixeles No Asignados* (en adelante NA).

#Verifique y elimine los NA

```
> summary (predictors)
Objeto de clase Coordinadas Marco de
Datos de Cuadrícula Espacial:
Min,

Máx
x      -356126.8
      4
65873.2
y      1825343.5
      2478343.5

Está proyectado:VERDADERO
proj4string :
[+proj=utm +zone=36 +datum=WGS84 +units=m +no_defs +ellps=WGS84 +towgs84=0,0,0]
Atributos de Cuadrícula:
 cellcentre.offset cellsize cells.dim x -355626.8 1000 822
y 1825843.5 1000 653
Atributos de datos:
..
 lcover      geology      pgeology      Rain      swir1
Min. : 2.0    Min. : 1.00    Min. : 0.9977  Min. : 0.1938  Min. : 0.01449
Q1 : 178.0    Q1 : 31.00    Q1 : 3.0000    Q1 : 2.0000    Q1 : 0.58141
Mediana :178.0  Mediana :32.00  Mediana : 3.0000  Mediana : 4.0000  Mediana :0.65651
Media :177.4    Media :47.55    Media : 3.8039    Media : 7.2899    Media :0.63076
Q3 :178.0    Q3 :66.00    Q3 : 3.0000    Q3 : 7.8400    Q3 :0.70472
Máx. :188.0    Máx. :88.00    Máx. :10.0000  Máx. :70.5665  Máx. :0.93891
..
```

Se mostrarán los NAs en las capas en donde ocurran. Primero debe investigarse donde y cuando ocurren. Si se encuentran predominantemente fuera del área de estudio y ocurrieron debido al recorte espacial de los mapas en formato raster, entonces pueden ser eliminados, por ejemplo, reemplazándolos con la media de los datos.

```
> predictors$slope=ifelse(is.na(predictors$slope),mean(!is.na(predictors$slope)),pr
edictors$slope)
```

#Derive los índices de sensores remotos

Derive los índices de sensores remotos de problemas de salinidad y adjúntelos en la pila de *predictors* (predictores) de capas SIG. Los índices se derivan utilizando la función *image Indices* en la librería *soil assessment* (Omuto, 2020³).

```
> predictors$SI1=imageIndices(predictors$BBlue,predictors$BGreen,predictors$BRed,predictors$BIRed,predictors$swir1,predictors$swir2,"SI1");summary(predictors$SI1)
  Min.  Q1      Mediana Media Q3      Máx.  0.04471  0.31981  0.35173
0.34386 0.37887 0.54641
```

```
> predictors$SI2=imageIndices(predictors$BBlue,predictors$BGreen,predictors$BRed,predictors$BIRed,predictors$swir1,predictors$swir2,"SI2");summary(predictors$SI2)
  Min.  Q1      Mediana Media Q3      Máx.  0.0389  0.2297 0.2540 0.2490 0.2749
0.4108
```

-

```
> predictors$BI=imageIndices(predictors$BBlue,predictors$BGreen,predictors$BRed,predictors$BIRed,predictors$swir1,predictors$swir2,"BI");summary(predictors$BI)
  Min.  Q1      Mediana Media Q3      Máx.  0.07189  0.73497  0.83005
0.80074 0.89241 1.22795
```

Cualquier NAs que resulte del cálculo de los índices de imágenes debe ser eliminado donde sea necesario.

#Verifique la asimetría utilizando la distribución del histograma empírico

```
> hist(predictors@data[,27:29]) # Figura 5.2
> summary(predictors$SI6)
#      Min.  Q1      Mediana Media Q3      Máx.  #      0.003647
0.940943 1.129692 1.068303 1.232106 1.663694
```

```
predictors$BI=sqrt(predictors$BI)
```

>

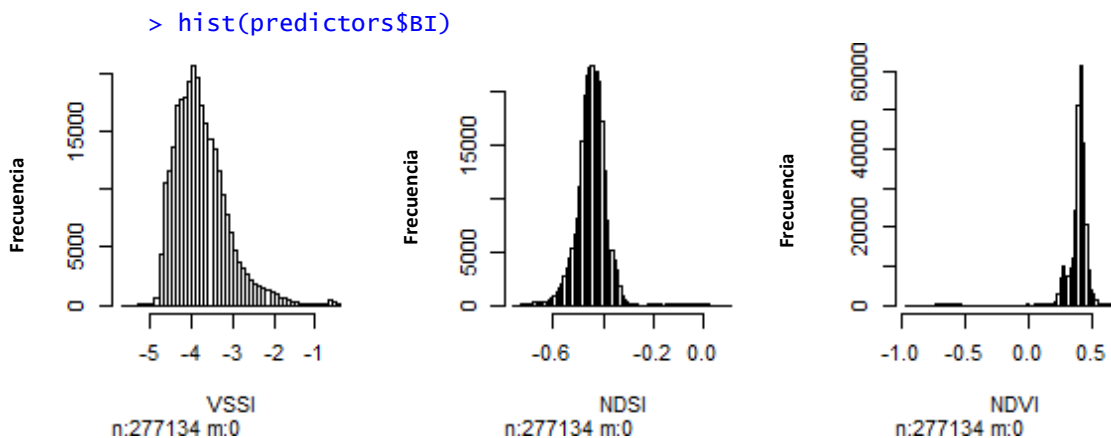


Figura 5: Distribución estadística empírica de los índices de imágenes

Por el momento, puede probarse la raíz cuadrada o la transformación logarítmica para la normalización de datos.

```
> hist(predictors@data[,"rain"])
> summary(predictors$rain)
  Min.  Q1      Mediana Media Q3Max.
0.1938      2.0000  4.0000
```

³ CT. 2020. Soil assessment: Assessment Models for Agriculture Soil Conditions and Crop Suitability <https://cran.r-project.org/web/packages/soilassessment/index.html>

7.2899 7.8400

70.5665

```
> predictors$rain=log(predictors$rain)
```

#Realice el PCA [análisis de componentes principales, por sus siglas en inglés] y seleccione los primeros PCs [componentes principales] que representen más del 95% de la variación de los índices de imágenes.

Después de normalizar los índices de imágenes, se las seleccionan y convierten en un *dataframe* (marco de datos) para permitir la determinación de la correlación y el análisis de los componentes principales. Posteriormente, los PCs seleccionados se convierten de nuevo a la pila raster.

Extraiga las capas de imágenes

```
> predictors=predictors@data[,c("SI1", "SI2", "SI3", "SI4", "SI5", "SI6", "SAVI", "VSSI", "NDSI", "NDVI", "SR", "CRSI", "BI")]
> soil.cor=cor(predictors)
> corrplot(soil.cor,method="number",number.cex = 0.8) # Figura 6a
> pca<-prcomp(predictors[, scale=TRUE])
> fviz_eig(pca) # Figura 6b
```

El gráfico de correlación (Figura 6a), muestra la correlación entre los índices de imágenes. Por ejemplo, SI1 y SI2 tienen un coeficiente de correlación de Pearson equivalente a 86%. El PCA examina estas correlaciones y determina los ejes principales en donde los datos están altamente correlacionados. Estos ejes también se conocen como componentes principales (o dimensiones en la Figura 6b). La Figura 6 es importante para orientar la elección de los PCs para representar las (13) capas completas de índices de imágenes.

Estos análisis (correlaciones y PCs) nos permiten hacer hipótesis sobre la variabilidad espacial de suelos en el área de interés a partir de los datos disponibles. Estas hipótesis nos ayudan a entender la importancia de las variables explicativas de las propiedades de suelo que son modeladas, y nos ayuda a simplificar el gasto computacional para hacer predicciones detalladas (e.g., al nivel de parcela).

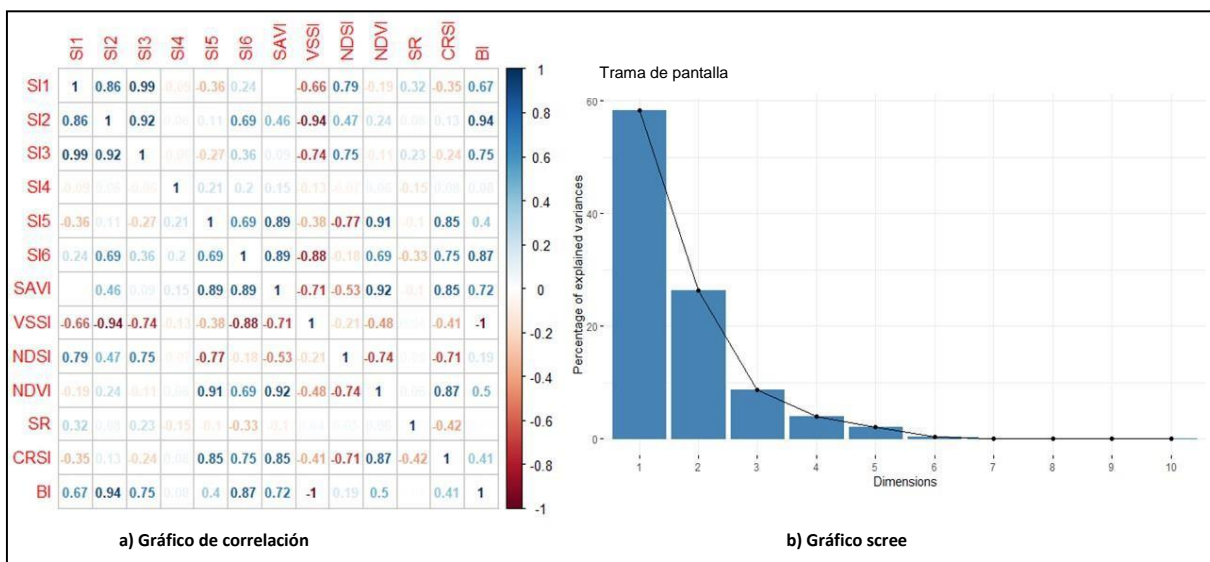


Figura 6: Correlación de índices de imágenes y gráfico de sus componentes principales

En la Figura 6b, la suma acumulada de los primeros 5 PCs (Dimensiones) suma más del 95% de la variación explicada en el total de las 13 capas de los índices de imágenes. Por consiguiente, los primeros 4 PCs pueden representar adecuadamente los índices de las 13 imágenes. “Este enfoque puede utilizarse en cualquier grupo de índices de imágenes para seleccionar el número apropiado de PCs para reducir los datos en bruto”.

Regrese los PCs seleccionados a la pila raster para completar el proceso de armonización

```
> Pred.pcs<-predict(pca,predictors[])
> predictors@data$PCA1=Pred.pcs[,1]
> predictors@data$PCA2=Pred.pcs[,2]
> predictors@data$PCA3=Pred.pcs[,3]
> predictors@data$PCA4=Pred.pcs[,4]
```

4.3 Armonización de los datos del suelo

Armonice las mediciones del indicador de entrada con aquellas para el extracto de pasta saturada de suelo

Se pueden utilizar muchos métodos para determinar la CE. Entre los que se incluyen: (1) el uso de extracto de pasta saturada del suelo; (2) usando otros extractos; (3) uso de modelos de pedotransferencia de otras propiedades del suelo, o (4) inducción electromagnética. La armonización busca estandarizar los métodos 2, 3 y 4 con valores equivalentes al método 1 pues los esquemas de clasificación popular utilizan valores obtenidos por el método 1. Los datos del caso de prueba se determinaron utilizando el extracto de pasta saturada del suelo. Por consiguiente, no requerirán armonización de la CE. Omuto et al. (2020)⁴ señala los pasos para la armonización de la CE para los métodos 2, 3 y 4.

La armonización de la profundidad del suelo tiene como objetivo desarrollar información del suelo para lograr una profundidad uniforme en todos los datos del suelo. Esta armonización se logra utilizando el enfoque integrador de profundidad *spline* (Bishop et al., 1999)⁵. La herramienta para implementar el enfoque está incluida en el paquete *GSIF* (Hengl, 2019)⁶.

```
> lon=soil1$Longitude
> lat=soil1$Latitude
> id=soil1$Pits
> top=soil1$Upper
> bottom=soil1$Lower
> horizon=soil1$Horizon
> CEdp=soil1$CE
> prof1=join(data.frame(id,top,bottom, CEdp, horizon),data.frame(id,lon,lat),type="
inner")
Joining by: id
> dept
hs(prof1)=id~top+bot
tom Warning message:
converting IDs from factor to character
> site(prof1)=~lon+lat
> coordinates(prof1) = ~lon+lat
> proj4string(prof1)=CRS("+proj=longlat +datum=WGS84 +no_defs")
> depth.s = mpspline(prof1, var.name= "CEdp", lam=0.8,d =
t(c(0,30,100,150))) Fitting mass preserving splines per profile...
```

⁴ Omuto CT, Vargas RR, El Mobarak, AM, Mohamed N, Viatkin K, Yigini Y. (Eds). 2020. Global mapping of salt-affected soils: A technical guideline and cookbook. Rome.

⁵ Bishop, T.F.A, McBratney, A.B., Laslett, G.M., 1999. Modelling soil attribute depth functions with equal-area quadratic smoothing splines.

⁶ Hengl T. 2019. GSIF: Global Soil Information Facilities [GSIF: Instalaciones mundiales de información sobre el suelo]. <https://cran.r-project.org/web/packages/GSIF/index.html>


```

=====| 100%
> plot(prof1, color= "CEdp", name="horizon",color.palette =
rev(brewer.pal(8, 'Accent')),par=c(cex.lab=2.0)) #Figura 7

```

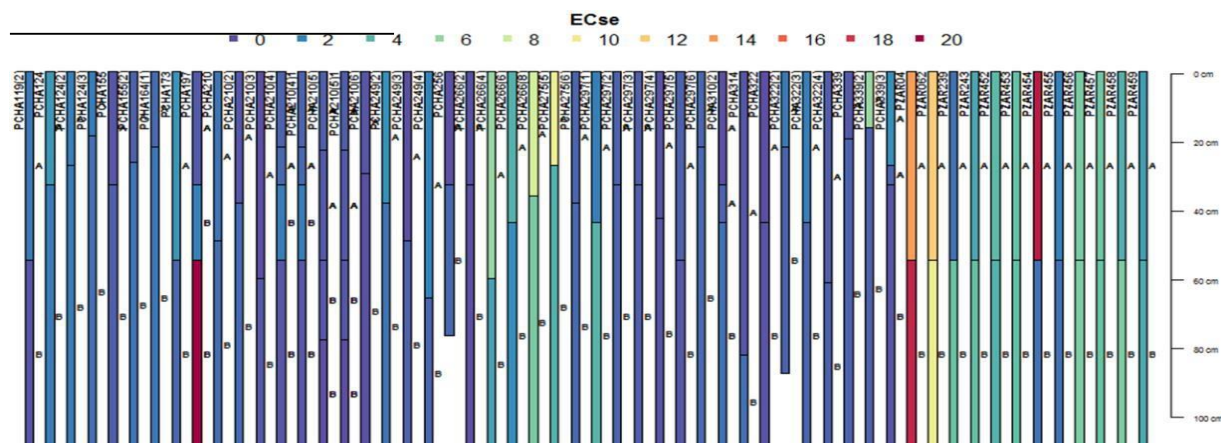


Figura 7: Ejemplo de armonización profunda para CE_{se}

Extraiga los datos del suelo de profundidad armonizada y reprojecte

```

> soilhrmdepths=data.frame(depth.s$idcol, depth.s$var.std, check.names = TRUE)
> soil2=merge(soil1,soilhrmdepths,by=intersect(names(soil1),names(soilhrmdepths)),b
y.x="Pits",by.y="depth.s.idcol",all=TRUE)
> coordinates(soil2)=~Longitud+Latitude

```

#Agregue los CRS (Sistemas de Coordenadas de Referencia, en adelante CRS) a los datos

```

> proj4string(soil2)=CRS("+proj=longlat +datum=WGS84")

```

#Armonice los CRS y asegúrese de usar +proj y +zone correctos para el área de estudio

```

> soil1=spTransform(soil2,CRS("+proj=utm +zone=36 +ellps=WGS84 +units=m +no_defs"))
> soil1=soil2
> hist(soil1$CE)
> soil1=subset(soil1,!is.na(soil1$CE))

```

#Armonización de distribución estadística

Esta armonización se realiza para transformar la distribución de frecuencia a una distribución normal. Esta transformación es opcional para los algoritmos de modelado espacial. En caso de realizarse, entonces se establece primero la distribución empírica a través del análisis del histograma y se implementa la transformación en caso de encontrar una distribución asimétrica. La función *hist* se utiliza para extraer y graficar el histograma. Se prefiere la transformación de Box-Cox (1964)⁷. Los siguientes *scripts* ilustran los pasos para transformar la distribución estadística. Primero se obtiene un resumen de la distribución para establecer si existen ceros, *NAs*, o valores negativos. Es deseable eliminarlos antes de implementar la transformación Box-Cox.

```

> Summary(soil1$X0.30.cm)
  Min.   Q1     Mediana  Media   Q3     Máx.
0.0007 0.6291  1.8709  6.6812  5.3121 154.2463

```

```

> soil1$dummy=(soil1$CE) # (usted puede) agregar"+0.001" si mínimo X0.30.cm es cero

```

⁷ G. E. P. Box and D. R. Cox. 1964. An Analysis of Transformations. Journal of the Royal Statistical Society. Series B (Methodological) Vol. 26, No. 2 (), pp. 211-252

```
> hist(soil1$dummy, main="Frequency distribution (before transformation)", xlab="Harmonized CE (dS/m)")
> soil1$Tran=(soil1$dummy^(as.numeric(car::powerTransform(soil1$dummy, family="bcPower")["lambda"]))-1)/(as.numeric(car::powerTransform(soil1$dummy, family="bcPower")["lambda"])))
```

4.4 Modelado espacial de indicadores

El modelado espacial de indicadores de suelos afectados por salinidad se basa en el concepto de mapeo digital del suelo (MDS). En este concepto, se construye una relación entre los indicadores de suelos afectados por salinidad y los factores formadores de suelo (capas SIG que funcionan como variables explicativas de los indicadores de suelos afectados por salinidad). Este enfoque permite la cuantificación de:

1. Información espacial de los indicadores de suelos afectados por salinidad (CE, pH, PSI) y distintas profundidades del suelo.
2. Incertidumbres y precisión de mapeo.
3. Información espacial de clases e intensidad de problemas de salinidad.

Los modelos populares utilizados con frecuencia para representar la función f que relaciona los datos de suelos con sus predictores, son modelos *lineales*, *random-forest*, *support-vector machine*, *mixedeffects*, *regression kriging*, etc. El paquete *soilassessment* proporciona la función *regmodelSuit* para orientar la elección del modelo adecuado para cartografiar las variables del suelo; prueba diferentes modelos y regresa a los nueve modelos principales utilizando *RMSE* (error cuadrático medio), *ME* (error medio), *NSE* (coeficiente de eficiencia de Nash-Sutcliffe) y r^2 . El RMSE más bajo, ME más bajo, NSE más alto y r^2 más alto, son entonces utilizados como los criterios orientadores para elegir el modelo apropiado. Un paso inicial para el modelado espacial es construir el modelo en un conjunto de datos de calibración y después probar el modelo utilizando un conjunto de datos independiente. Esto requiere del establecimiento de conjuntos de datos de calibración y validación. Estos conjuntos de datos deben tener propiedades del suelo (indicadores) y predictores espaciales bien alineados en cada punto de muestreo georeferenciado. La extracción del valor de pixel de las capas SIG (predictores) utilizando datos puntuales (*soil1*) es un método adecuado para desarrollar ya sea la calibración o validación del conjunto de datos.

Primero verifique si hay similitudes en el CRS y después extraiga los predictores

```
> crs(predictors); crs(soil1) argumentos CRS:
+proj=utm +zone=36 +datum=WGS84 +units=m +no_defs +ellps=WGS84 +towgs84=0,0,0
argumentos CRS:
+proj=utm +zone=36 +datum=WGS84 +units=m +no_defs +ellps=WGS84 +towgs84=0,0,0
```

Es importante garantizar que los CRS para los predictores y la base de datos del suelo sean los mismos antes de empezar la extracción de pixeles.

Después extraiga los valores de los pixeles para todos los predictores en el dataframe de los datos del suelo

```
> {predictors.ov=over(soil1, predictors)
+ soil1$dem=predictors.ov$dem
+ soil1$slope=predictors.ov$slope
+ soil1$cnbl=predictors.ov$cnbl
+ soil1$ls=predictors.ov$ls
+ soil1$valley=predictors.ov$valley
+ soil1$loncurve=predictors.ov$loncurve
+ soil1$lcover=predictors.ov$lcover
+ soil1$rain=predictors.ov$rain
+ soil1$pgeology=predictors.ov$pgeology
```

```

+   soil1$geology=predictors.ov$geology
+   soil1$PCA1=predictors.ov$PCA1
+   soil1$PCA2=predictors.ov$PCA2
+   soil1$PCA3=predictors.ov$PCA3
+   soil1$PCA4=predictors.ov$PCA4
+ }

```

#Paso 2-2: Establezca el modelo MDS adecuado

```

> soil1=subset(soil1,!is.na(soil1$dem))
> soil11a=soil1@data[,c("Tran","dem","slope","ls","cnb1","loncurve","valley","rain",
,"lcover","pgeology","geology","PCA1","PCA2","PCA3","PCA4")]
> regmodelSUIT(soil11a,Tran,dem,geology,pgeology,slope,rain,loncurve,cnb1,valley,lcover,ls,PCA1,PCA2,PCA3,PCA4)
|=====| 100%

```

	MERMSE		R2NSE	
Linear	1.37034834	1.8129133	0.1320264	-4.45424486
RandomForest	0.24614749	0.4291176	0.9623631	0.99707891
SVM	1.34745209	1.8212570	0.1357457	-4.44014767
BayesianGLM	1.36669809	1.8051662	0.1399516	-4.55079779
BaggedCART	0.88676091	1.1705841	0.7018074	0.44302759
Cubist	0.07851255	0.2744213	0.9753726	1.00000000
CART	1.40147986	1.8274332	0.1320797	-4.56273851
Ranger	0.26852953	0.4205875	0.9655103	0.99702690
QuantRandForest	0.04923343	0.2855093	0.9761419	1.00000000
QuantNeuralNT	1.16162791	1.7075411	0.2582558	0.07778314

Los resultados anteriores muestran los modelos *quantile regression*, *random forest* y *cubist* como adecuados para el modelado de CE_{se} 0-30 cm, utilizando los predictores espaciales dados en los datos de prueba del estudio de caso.

La construcción de modelos estadísticos y las estrategias de prueba, recomiendan conjuntos de datos independientes para la construcción del modelo y la prueba del modelo. Lo ideal es que estos conjuntos de datos sean muestreados con enfoque para la construcción y prueba del modelo. En ausencia del conjunto de datos muestreados independientemente para cualquiera de los dos modelos, el de construcción (calibración) o el de prueba (validación), se utiliza con frecuencia la estrategia de división de datos. La estrategia de división de datos, divide aleatoriamente (o aleatoriamente estratifica) los datos en dos partes. Una parte es tomada como calibración, la otra como validación. El conjunto de datos de validación se utiliza para evaluar la precisión. Los índices para reportar la precisión del modelo incluyen RMSE, ME, r^2 , NSE y la representación gráfica de los valores modelados contra los armonizados. La división de datos puede elegirse arbitrariamente de acuerdo con, o dependiendo del tamaño de los datos. Sin embargo, una vez seleccionado el modelo apropiado, el mapa final debe ser generado con todos los datos. En este caso, debemos usar el modelo con mayor capacidad predictiva con todos los datos y generar una predicción para el mapa final.

#Paso 2-3: Construya y pruebe del modelo

```

> {soil4=as.data.frame(soil1)
+   bound <- floor((nrow(soil4)/4)*3)
+   soil3 <- soil4[sample(nrow(soil4)), ]
+   df.traina <- soil3[1:bound, ]
+   df.testa <- soil3[(bound+1):nrow(soil3), ]}

> rf.ec=train(Tran~(slope+rain+loncurve+ls+cnb1+valley+lcover+dem+PCA1+PCA2+PCA3+PCA4+PCA5),
data = df.traina, method = "qrf", trControl=trainControl( method = "cv"
,number=5,returnResamp = "all",savePredictions = TRUE, search = "random",verboseIter = FALSE))

```

Muestre el intervalo de predicción

```

> df.testa$Strain=predict(rf.ec,newdata=df.testa)
> hist(df.testa$Strain,xlab="Box-Cox Transformed CEse (0-30cm)", main=NULL)

```

```
> abline(v = quantile(df.testa$Strain, probs = c(0.05, 0.95)), lty = 5, col = "red")
```

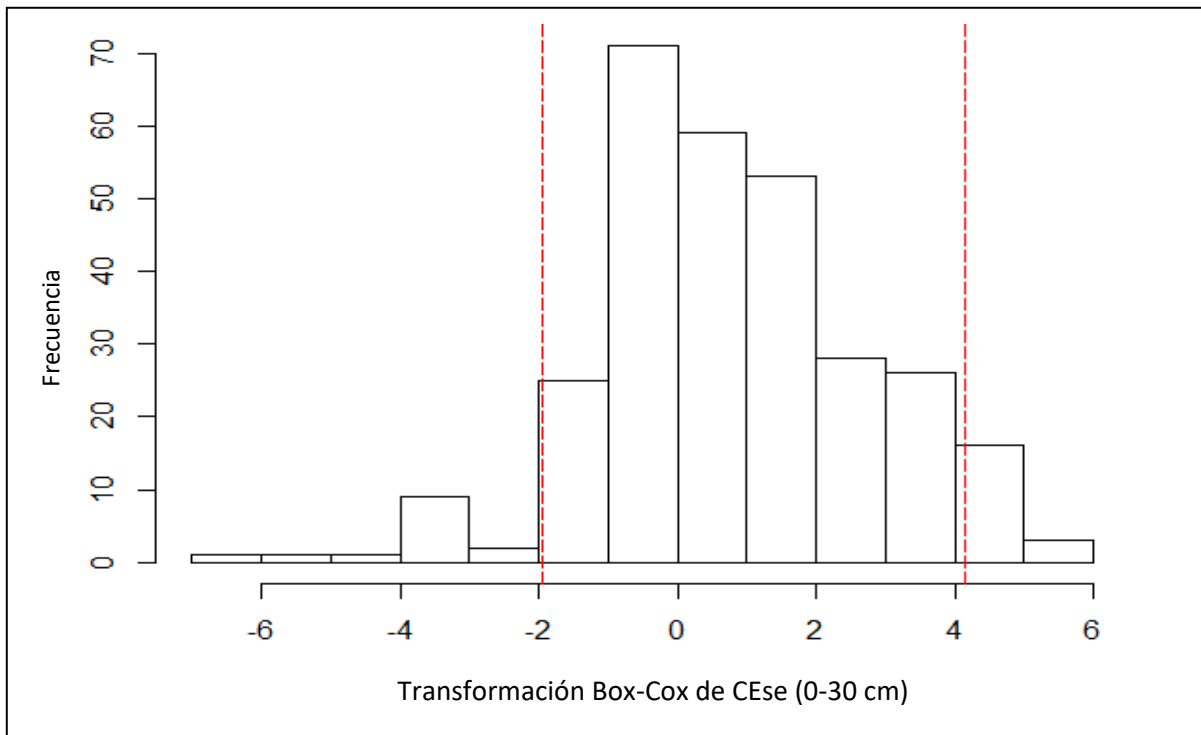


Figura 8: Representación gráfica de la distribución de frecuencia con una de predicción a un intervalo de confianza del 95%

Los límites de predicción de los valores transformados via Box-Cox al 95 % de confianza muestra el intervalo alrededor de la media de 0.77 como [-2, 4.1].

#Evaluación de la precisión

```
> cor(df.testa$Strain,df.testa$dummy)^2 [1] 0.9950319
> {plot(df.testa$Strain~df.testa$dummy, xlab="Measured CEse",ylab="Modelled CEse",
main="Accuracy assessment on hold-out samples")
+ abline(a=0,b=1,lty=20, col="blue")} # Figure 9
```

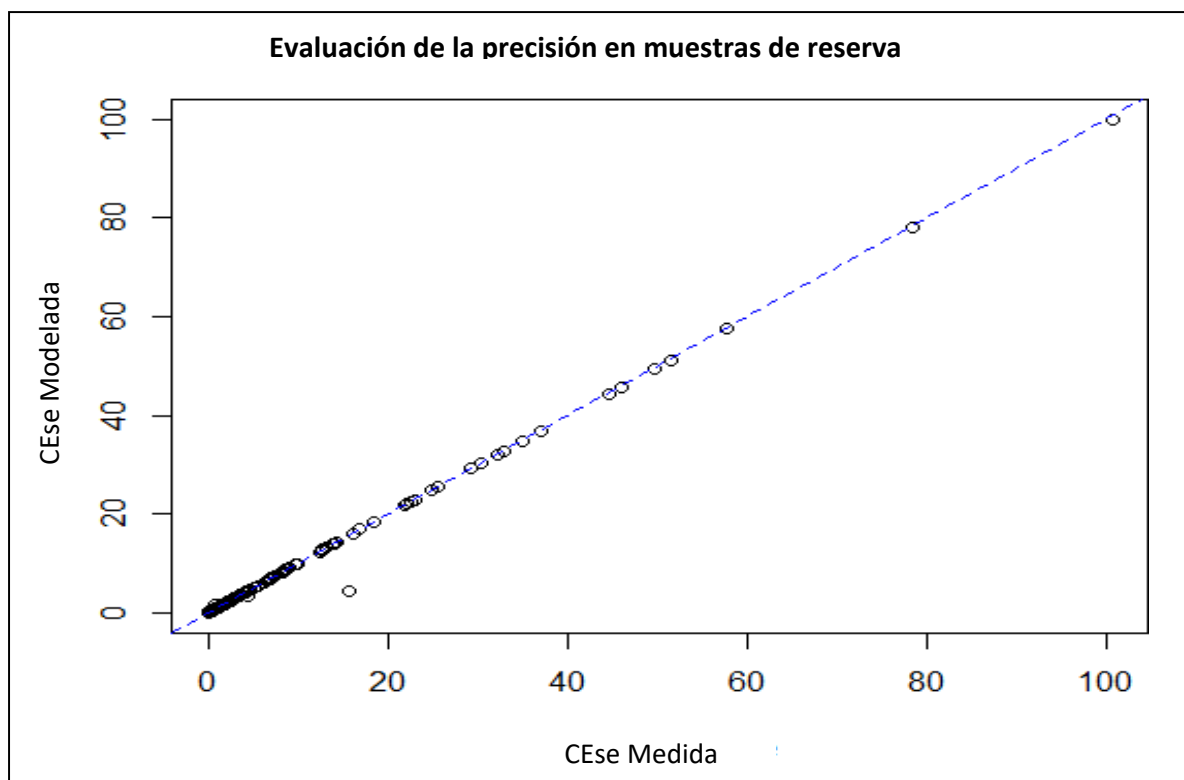



Figura 9: Representación gráfica de CE predcidos Vs.los CE medidos

```
> Bias=mean(df.testa$Strain-df.testa$dummy,na.rm=TRUE)
> RMSE=sqrt(sum(df.testa$Strain-df.testa$dummy,na.rm=TRUE)^2/length((df.testa$Strain -
df.testa$dummy)))
> Rsquared=cor(df.testa$Strain,df.testa$dummy)^2
> NSE=1-sum(df.testa$Strain-df.testa$dummy,na.rm=TRUE)^2/sum((df.testa$Strain-mean(
df.testa$dummy,na.rm=TRUE))^2,na.rm=TRUE)
> statia=data.frame(Bias,RMSE,Rsquared,NSE);View(statia)
> write.csv(statia,file = "CE0_30_validmodel_stats.csv")
> statia
```

	Bias	RMSE	Rsquared	NSE
1	-0.1019564	1.751158	0.9950319	0.982046

#Use el modelo desarrollado para predecir el mapa de CE

```
> lmbda1=(as.numeric(powerTransform(soil1$dummy, family = "bcPower")["lambda"]))
> predictors$CEte=predicta(rf.ec,predictors)
> coordinates(df.testa)=~Longitude+Latitude
> proj4string(df.testa)=CRS("+proj=utm +zone=36 +datum=WGS84 +units=m +no_defs +ell
ps=WGS84 +towgs84=0,0,0") # Asegúrese de utilizar el CRS correcto
> predictors.ov1=over(df.testa, predictors)
> df.testa$Predre=predictors.ov1$CEse
> cor(df.testa$dummy,df.testa$Predre)^2 [1] 0.9978655
```

#Compare la predicción espacial y validación del conjunto de datos

```
> featureRep(predictors["CEse"],df.testa) #Figura 10
```

Prueba de dos muestras *Kolmogorov-Smirnov*

datos: dist.histbb\$left y dist.histbb\$right D = 0.52174, p-valor = 0.003819
hipótesis alternativa: dos colas

```
> summary(predictors$CEse);summary(df.testa$dummy)
```

Min.	Q1	Mediana	Media	Q3	Máx.
0.00007	0.48810	1.17487	1.51685	1.61781	112.74435
Min.	Q1	Mediana	Media	Q3	Máx.
0.00048	0.59755	1.71126	6.60388	5.05220	113.50941

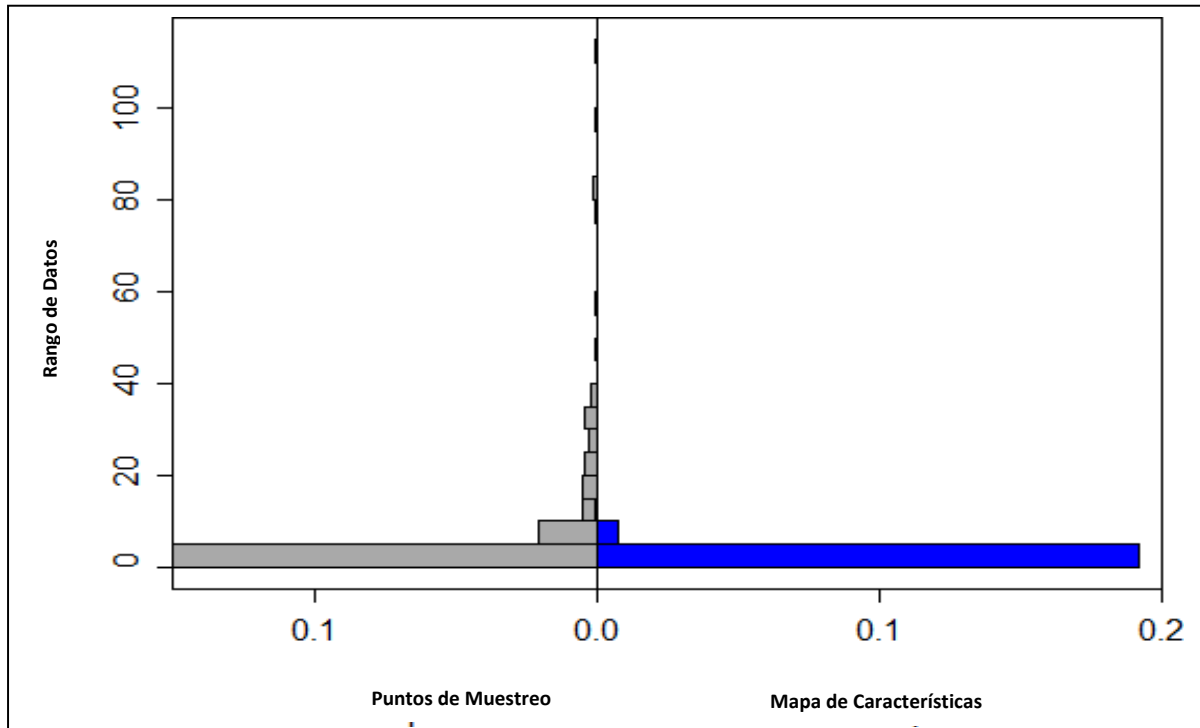


Figura 10: Representatividad de los rangos de validación de CE (puntos de muestreo) en el mapa de predicción (mapa de características)

La representación de características muestra qué tan bien está contenido el rango de mediciones CE (con respecto a CE de validación) en el mapa de predicción. En el muestreo del caso de estudio, los valores de CE altos (>40 dS/m) están contenidos deficientemente en el mapa de predicción. El eje de las x muestra la frecuencia (densidad de probabilidad) de ocurrencia de los valores de datos (CE) en el eje de las y. La representación deficiente de altos valores de CE (CE > 40) implica la incertidumbre del modelo para valores altos de CE. Esto se investigará con mayor detalle cuando se produzcan las incertidumbres.

#Exporte los resultados

```
writeGDAL(predictors["CEse"], drivername = "GTiff", "Top0_30CEse.tif")
```

#Evaluación de la incertidumbre

```
> soil6a=soil1[,c("Tran")]
> predictors6a=predictors[c("dem","slope","cnb1","lcover","loncurve","rain","pgeology",
"geology","ls","valley","PCA1","PCA2","PCA3","PCA4","PCA5")]

> pred_uncerta=predUncertain(soil6a,predictors6a,3,95,"qrandomforest")
|=====| 100%

> spplot(pred_uncerta, "pred_width", scales = list(draw = TRUE),col.regions=heat.colors(20,rev = TRUE)) + spplot(df.testa,"dummy",pch=3,cex=0.4) #Figura 11
```

#Paso 2-7: Exporte los mapas de incertidumbre

```
> CE0_30_uncertain=(pred_uncertain$pred_width*lambda+1)^(1/lambda)
> writeRaster(CE0_30_uncertain, filename="CE0_30_uncertain.tif",format="GTiff")
```

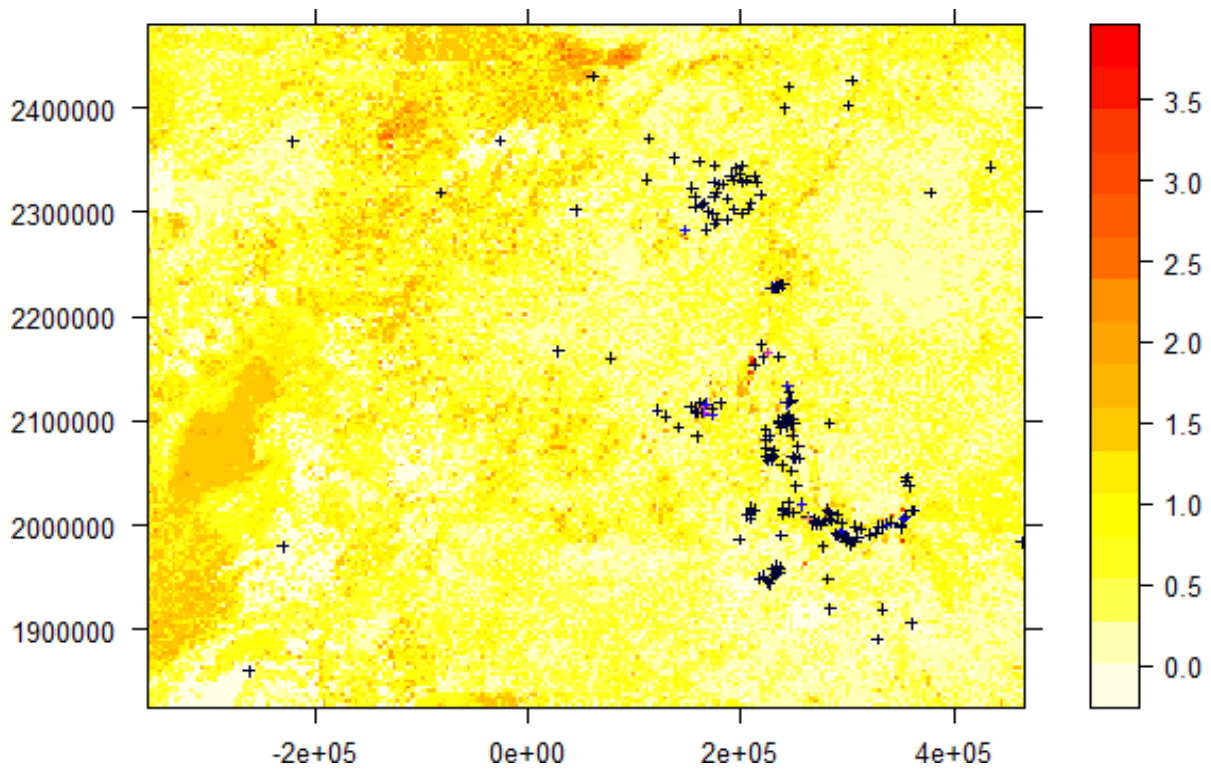


Figura 11: Ancho de la predicción espacial al 95% del intervalo de confianza y sobreposición de los puntos de validación

Los pasos anteriores para el modelado espacial de CE, deben repetirse para pH y PSI, además de la profundidad de 0-30 cm, para profundidad de suelos de 30-100 cm.

5 Resultados

Al finalizar la lección, se espera que cada participante utilice los conocimientos aprendidos, para que produzca los siguientes mapas de indicadores de suelos afectados por salinidad:

1. Mapas raster GeoTiff de indicadores de suelos afectados por salinidad (CE, pH, y PSI), para 0-30 cm y 30-100 cm (6 mapas raster en total).
2. Mapas raster GeoTiff de evaluación de la incertidumbre para cada propiedad del suelo (CE, pH, y PSI), para 0- 30 cm y 30-100 cm (6 mapas raster en total).
3. Archivo de texto de la evaluación de la precisión para cada propiedad del suelo (CE, pH, y PSI) para 0-30 cm y 30- 100 cm (6 mapas raster en total).



Gracias al apoyo financiero de



Ministerio de Finanzas de la
Federación de Rusia



Comisión
Europea



Rural Development
Administration

