

16 de diciembre de 2022

**ASUNTO:** Reporte de evaluación — SISLAC, producto 3.2.1.2

**M.Sc. Ana Luisa Posas Guevara**

Oficial de Agricultura – FAO

PRESENTE

Con referencia al proyecto “*Evaluaciones para favorecer la adopción de soluciones digitales en intervenciones en ALC*” y en el marco de las actividades de la “**Asesoría técnica para las bases de datos de suelos en el marco de la iniciativa del Mapa Global de Nutrientes de los Suelos (GSNmap) en una nueva versión de SISLAC – Producto 3.2.1.2**”, hago envío del reporte de evaluación de los datos del SISLAC y los siguientes documentos adjuntos:

- Base de datos del SISLAC versión 2.0 en formatos para PostgreSQL (.backup) y valores separados por comas (.csv).
- Diccionario de datos con la descripción de cada atributo de la base de datos del SISLAC versión 2.0
- Archivo excel “AnálisisBaseDatosSISLAC” en el cual se encuentra el detalle de los perfiles que presentan algún cambio de la base de datos original.
- Bases de datos originales agregadas al SISLAC (6 repositorios de 5 países)

Atentamente,



M.Sc. Sergio Díaz Guadarrama

Consultor México

C.c.p. Carolina Olivera ([carolina.oliverasanchez@fao.org](mailto:carolina.oliverasanchez@fao.org))

C.c.p. Souroush Parsa ([souroush.parsa@fao.org](mailto:souroush.parsa@fao.org))



## Contenido

<b>Resumen ejecutivo.</b>	3
<b>Introducción</b>	5
<b>1.- SISLAC</b>	5
1.1 Estructura	5
1.2 Integridad de los datos	6
Resultados	11
<b>2.- Nuevas bases de datos a incorporar al SISLAC</b>	12
<b>3.- Base de datos SISLAC versión 2.0</b>	17
<b>4.- Conclusiones.</b>	20
<b>5.- Recomendaciones.</b>	20

## Resumen ejecutivo.

El **objetivo** de esta consultoría consistió en “**Integrar bases de datos nacionales de suelos para incorporar al SISLAC**”. En ese contexto, en este documento se describen cinco fases: la primera describe el proceso de análisis y depuración de la base de datos original con 49,084 perfiles de suelo que luego de revisar pasa a tener 41,202 perfiles con una correcta descripción de sus horizontes. La segunda corresponde al análisis, depuración e incorporación de nuevas bases de datos disponibles (de Argentina, Brasil, Chile, Ecuador y México) a la base de datos del SISLAC. La tercera detalla la nueva versión de los datos, denominada SISLAC 2.0 con 66,746 perfiles de suelo, lo cual representa un aumento del 61% de perfiles respecto a la base de datos del SISLAC depurada, además, se muestra un comparativo de las gráficas de variabilidad vertical de 5 atributos (carbono orgánico, pH, arcilla, limo y arena) entre la base de datos original y la nueva versión. La cuarta lista las conclusiones de este trabajo y la quinta describe una serie de recomendaciones a tener en cuenta para los países que aporten nuevos datos y para recabar más datos desde otras fuentes. El producto obtenido en esta consultoría es la base de datos de perfiles de suelo con 66,746 perfiles y 192,568 horizontes, todos con una correcta descripción, se entrega en dos formatos: i) *backup* para PostgreSQL (PostGIS) denominado **bdSISLAC\_v2** con una tabla para los perfiles, otra para los horizontes y la tabla de referencia espacial, la columna clave entre ambas tablas (perfiles y horizontes) es el identificador del perfil “*profile\_identifier*” y ii) archivo de valores separados por comas **sislac\_v2.csv** con toda la información contenida en la base de datos. También se ponen a disposición los códigos (para el *software* R) con las reglas de validación y depuración de nuevos datos de perfiles de suelo, estos se encuentran en el repositorio: <https://github.com/FAO-GSP/SISLAC-DB>.

## Executive summary

The aim of this consultancy was to **"Integrate national soil databases for incorporation into SISLAC"**. In this context, this document describes five stages: the first one describes the process of analysis and improvement of the original database with 49,084 soil profiles, which after review were obtained 41,202 profiles with a correct description of their horizons. The second corresponds to the analysis, processing and incorporation of new available databases (from Argentina, Brazil, Chile, Ecuador and Mexico) to the SISLAC database. The third details the new version of the data, called SISLAC 2.0 with 66,746 soil profiles, which represents an increase of 61% of profiles with respect to the revised SISLAC database. In addition, a comparison of the vertical variability graphs of 5 attributes (organic carbon, pH, clay, silt and sand) between the original database and new version is shown. The fourth lists the conclusions of this work and the fifth describes a list of recommendations to be taken in consideration for countries that provide new data and to collect more data from other sources. The product obtained in this consultancy is the soil profile database with 66,746 profiles and 192,568 horizons, all with a correct description, it is delivered in two formats: i) backup for PostgreSQL (PostGIS) called bdSISLAC\_v2 with a table for the profiles, another for the horizons and the spatial reference table, the key column between both tables (profiles and horizons) is the profile identifier "profile\_identifier" and ii) comma separated values file `sislac_v2.csv` with all the information contained in the database. The codes (for R software) with the validation and correction rules for new soil profile data are also made available and can be located at <https://github.com/FAO-GSP/SISLAC-DB>.

## Introducción

En el marco de la consultoría “Asesoría técnica para las bases de datos de suelos en el marco de la iniciativa del Mapa Global de Nutrientes de los Suelos (GSNmap) en una nueva versión de SISLAC”, producto 3.2.1.2, el **objetivo** de este trabajo consistió en “**Integrar las bases de datos nacionales de suelos** y dar soporte técnico a la implementación de la iniciativa del Mapa Global de Nutrientes de los suelos, **para integrar al SISLAC**”. En ese aspecto, a continuación se detallan los trabajos realizados y que están comprendidos en cinco grandes apartados. El primero abarca la descripción de la base de datos original del SISLAC, su análisis y depuración para lograr que los perfiles cuenten con un nivel mínimo de integridad que permita que estos puedan ser utilizados. El segundo lista las bases de datos de perfiles de suelos disponibles en la región y su integración a la base de datos del SISLAC, logrando así, una nueva versión con más datos y una mejor integridad. El tercero muestra la comparación de las gráficas de variabilidad vertical del carbono orgánico, pH, arcilla, limo y arena con la versión original y la nueva versión. Finalmente, se listan las conclusiones y recomendaciones resultantes de este trabajo para los países con miras a futuros esfuerzos por aumentar el volumen de la base de datos del SISLAC.

## 1.- SISLAC

### 1.1 Estructura

La base de datos del SISLAC está compuesta por 49,084 perfiles de suelo y un total de 139,777 horizontes. Los atributos del sitio se describen en la Tabla 1, en la primera columna se indica el nombre del atributo, en la segunda su descripción y finalmente se indica el tipo de dato que se almacena. Como se observa los primeros atributos describen el sitio y su ubicación espacial, posteriormente se encuentran atributos referentes a los datos del proveedor y su licencia de uso.

Tabla 1: Atributos de los perfiles

Columna	Descripción	Tipo de dato
profile_identifier	Código único del perfil	Texto
latitude	Latitud de la ubicación del perfil (grados decimales)	Numérico
longitude	Longitud de la ubicación del perfil (grados decimales)	Numérico
country_code	Código del país. formato ISO 3166-1	Texto
date	Fecha de captura de los datos	Fecha
source	Origen de los datos	Texto
contact	Correo electrónico de contacto	Texto
order	Orden se suelo	Texto
type	Tipo (perfil o barreno)	Texto
license	Código de licencia <sup>1</sup>	Texto

<sup>1</sup> Public Domain Dedication and License: PDDL; Attribution License: ODC-By; Open Database License: ODC-ODbL; Creative Commons Attribution 4.0 International: CC-BY; Creative Commons Attribution - Non-Commercial

Por otra parte, los atributos de los horizontes se detallan en la Tabla 2, Al igual que la tabla anterior, las tres primeras columnas corresponden al nombre de columna, descripción y tipo. La cuarta columna indica el porcentaje de horizontes con datos válidos. Como se puede observar de las 11 propiedades medibles, tres de estas (agua útil, porcentaje de fragmentos gruesos y densidad aparente) no superan el 20% de registros con datos; tres más (densidad aparente, carbono inorgánico y conductividad eléctrica) tienen entre el 25% y 50% de datos; dos tienen entre el 40 y 60% de datos (carbono orgánico y limo) y finalmente, solo tres (arena, arcilla y Ph) superan el 82% de registros con valores. Lo anterior se indica sin considerar que los valores contenidos sean válidos ya que pueden ser valores atípicos, fuera de rango o valores 0 (cero).

Tabla 2: Atributos de los horizontes.

Columna	Descripción	Tipo de dato	% de horizontes con datos
<b>layer_identifier</b>	Código único de cada horizonte <sup>2</sup>	texto	100
<b>designation</b>	Nomenclatura del horizonte (por ej. A, Bw, C2)	texto	29
<b>top</b>	El límite superior del horizonte	numérico	100
<b>bottom</b>	El límite inferior del horizonte	numérico	91
<b>bulk_density</b>	Densidad aparente	numérico	17.1
<b>ca_co3</b>	Carbono inorgánico	numérico	41.7
<b>coarse_fragments</b>	Porcentaje de fragmentos gruesos	numérico	9.4
<b>eccec</b>	Capacidad de intercambio catiónico	numérico	48.7
<b>conductivity</b>	Conductividad eléctrica	numérico	26.9
<b>organic_carbon</b>	Carbono orgánico	numérico	56.8
<b>ph</b>	Ph	numérico	82.7
<b>clay</b>	Porcentaje de arcilla	numérico	82.6
<b>silt</b>	Porcentaje de limo	numérico	58.3
<b>sand</b>	Porcentaje de arena	numérico	81.1
<b>water_retention</b>	Agua útil	numérico	8.5

## 1.2 Integridad de los datos

En un primer análisis exploratorio de los datos se identificó que estos no cuentan con validaciones previas que garanticen su consistencia, por ejemplo, existen perfiles fuera de

4.0 International: CC-BY-NC; Creative Commons Attribution - Non-Commercial No Derivatives 4.0 International: CC-BY-NC-ND.

<sup>2</sup> Este debe ser distinto entre elementos del mismo perfil, pero puede existir un mismo código en distintos perfiles.

zona, perfiles duplicados espacialmente, horizontes con una incorrecta descripción de sus límites o atributos con valores no válidos como valores de pH fuera de rango (0-14) o valores atípicos de porcentaje de carbono orgánico (*organic\_carbon*). Para poder determinar el nivel de integridad, se realizaron cuatro tipos de revisión enfocados en verificar que los perfiles cumplan con una serie de requisitos mínimos de integridad que posibiliten su uso, a continuación se detallan estas revisiones de integridad:

### 1.2.1 Integridad espacial

Al realizar una revisión para comprobar que los perfiles están donde dicen estar (con base en el atributo *country\_code*), se identificaron 4 variantes de localización de perfiles: i) 46358 que se encuentran donde corresponde; ii) 223 que se encuentran en los límites exteriores del país correspondiente, como se ejemplifica con el perfil [IGAC-AMA-PA-071](#) (Figura 1, recuadro superior izquierda), una posible causa de este desfase puede ser el método con que se tomaron las coordenadas ya que estos pueden tener un margen de error en la posición; iii) 2,471 que tienen una ubicación incorrecta a causa de que las coordenadas están invertidas, tal como se observa, por ejemplo, en el perfil [P-1284](#) (Figura 1, recuadro superior derecho); ii) 32 que no se corresponden con la ubicación, un ejemplo de esto es el [perfil 50726](#) que corresponde a Paraguay (Figura 1, recuadro inferior), sin embargo, se localiza en Brasil. Estos últimos perfiles al no estar ubicados correctamente fueron excluidos del conjunto de datos.

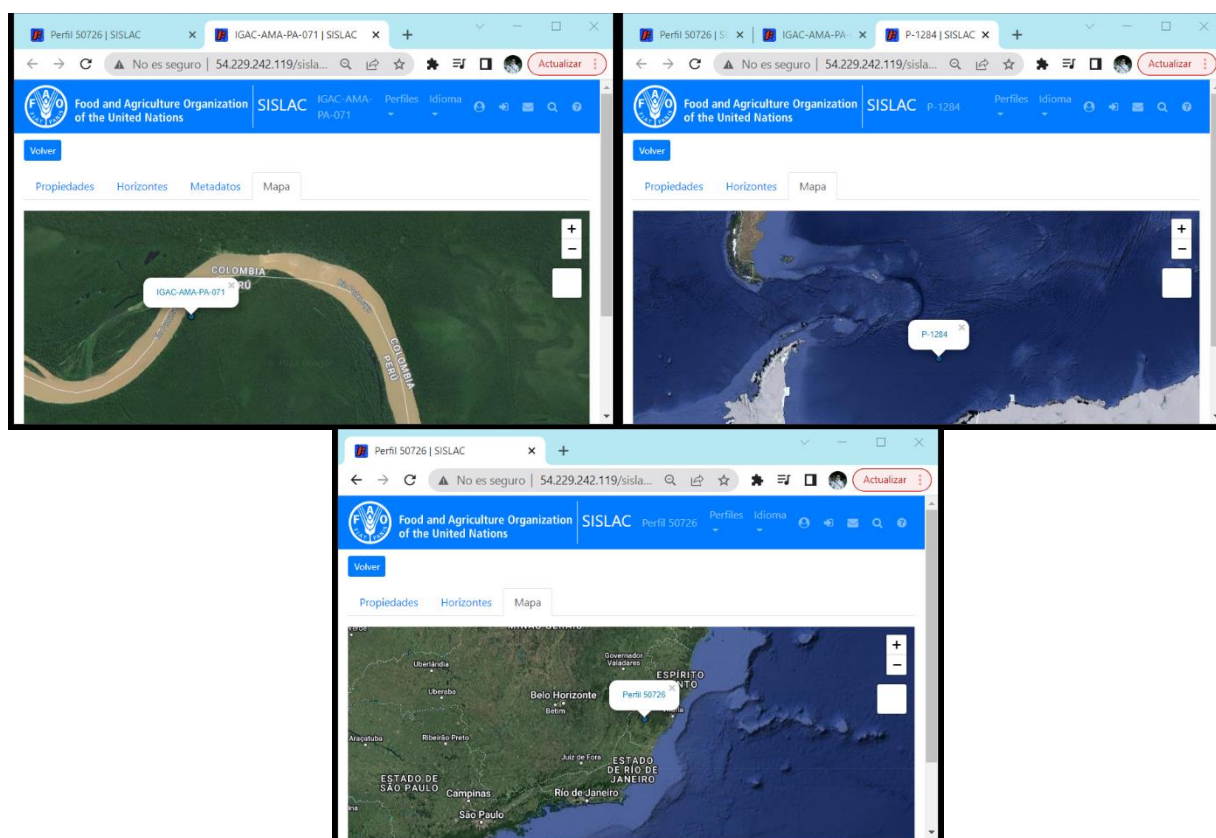


Figura 1, Verificación de localización de los perfiles.

### 1.2.2 Duplicidad espacial de perfiles

Esta revisión consistió en identificar si existen perfiles con las mismas coordenadas, se identificaron 1,989 casos, la ocurrencia de estos por país se detalla en la Tabla 3, la mayoría de estos pertenecen a Brasil con 1,680 casos, un par de ejemplos de la duplicidad se muestra con los perfiles [201724](#) y [201725](#) de Brasil y los perfiles [68150](#) y [68151](#) de Jamaica. Estos perfiles fueron “marcados” en la base de datos, el atributo perfil\_duplicado con valor TRUE indica que ese perfil tiene otro(s) en su misma ubicación.

Tabla 3: Ocurrencia de duplicidad de perfiles por país.

País	Perfiles
Brasil	1680
Argentina	94
Colombia	50
Jamaica	40
Venezuela	28
Uruguay	16
Surinam	11
Guatemala	9
Bolivia, Ecuador, Honduras y México	7
El Salvador, Guyana y Nicaragua	6
Panamá	5
Costa Rica y Perú	4
Cuba	2
<b>Total</b>	<b>1989</b>

### 1.2.3 Perfiles vacíos

Existen casos en los que los perfiles se han cargado sin datos de horizontes o con un único horizonte el cual solo incluye el límite superior y el resto de los atributos están vacíos, tal como se ilustra en la Figura 4. Como se observa en la Tabla 4, este caso se presentó en un total de 7,781 perfiles, la mayoría de estos se encuentran en México y Paraguay con 4,942 y 2830 perfiles respectivamente. Estos perfiles vacíos fueron también excluidos de la base de datos.

Tabla 4, Cantidad de perfiles vacíos por país.

País	Perfiles
México	4942
Paraguay	2830
Colombia	3
Argentina	2
Venezuela	2
Brasil	1
Chile	1
<b>Total</b>	<b>7781</b>



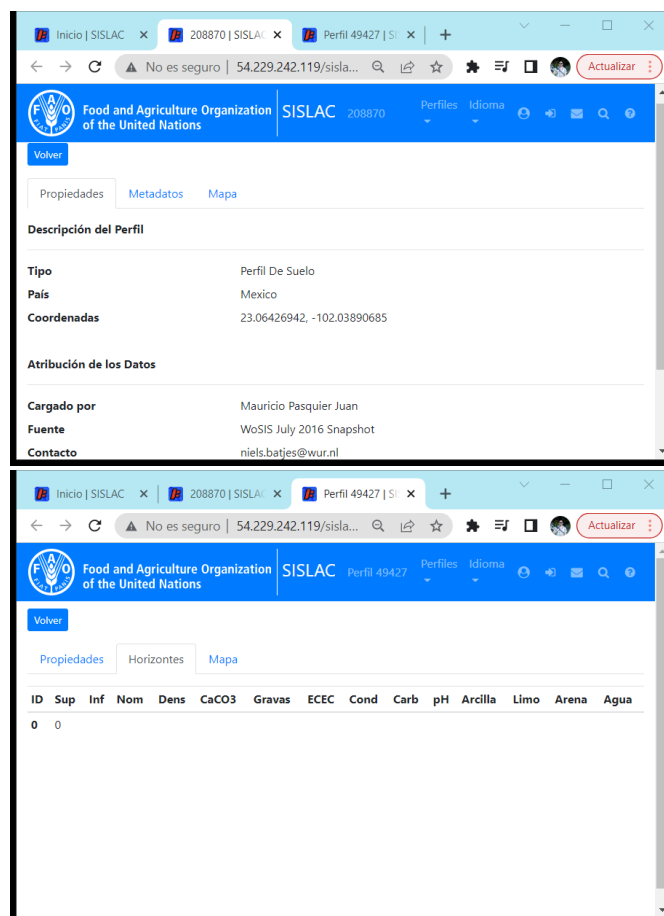


Figura 2, Perfiles sin horizontes (arriba) o con un único horizonte vacío (abajo)

#### 1.2.4 Superposición de horizontes:

Otro de los problemas que se presenta en la descripción de los horizontes es la superposición de estos, lo cual representa un error. Todos los perfiles con sobreposición fueron revisados para tratar de minimizar los perfiles a excluir corrigiendo aquellos en que existía evidencia de cómo solucionar y/o reparar el perfil. Se analizaron 719 perfiles, de los cuáles 398 fueron excluidos, 271 fueron divididos en dos conjuntos de datos (subset1 y subset2), un ejemplo de esta división se muestra en la Figura 3 que corresponde al [perfil 202383](#) de Brasil (arriba) y al [perfil 69514](#) de Cuba (abajo).

**Screenshot 1: Profile 202383**

ID	Sup	Inf	Nom	Dens	CaCO3	Gravas	ECEC	Cond	Carb	pH	Arcilla	Limo	Arena	Agua
1006399	0	15					16.0		22.8	5.7				
1006400	0	15					13.8		22.8	5.7				

**Screenshot 2: Profile 69514**

ID	Sup	Inf	Nom	Dens	CaCO3	Gravas	ECEC	Cond	Carb	pH	Arcilla	Limo	Arena	Agua
2098	0	22			37.0		0.2	15.9		78.9	9.1	12.1		
2099	3	8												54.68
2100	22	55			37.0		0.3	4.5		70.8	17.6	11.6		
2101	23	28												57.08
2102	55	80			313.0		0.2	2.4		57.4	28.0	14.6		
2103	65	70												44.98
2104	80	85												43.95
2105	80	90			411.0		0.2	1.5		43.2	51.2	5.6		
2106	100	105												39.8
2107	100	130			339.0		0.2	1.1		30.1	57.3	12.5		
2108	140	170			304.0		0.2	0.1		40.1	48.8	10.9		

Figura 3, división de horizontes de un perfil en dos subconjuntos de datos.

## 1.2.5 Correcciones menores a los horizontes

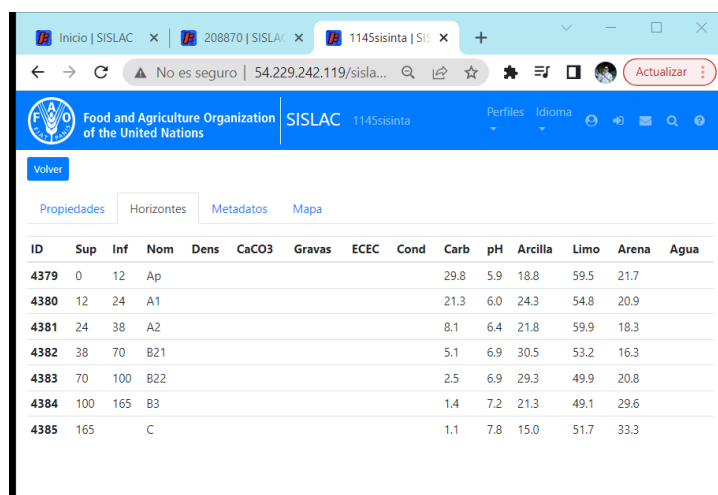
Finalmente, se corrigieron otras inconsistencias menores para garantizar que los perfiles estuvieran descritos correctamente, estas fueron:

- Perfiles con horizonte orgánico: Esta inconsistencia consiste en que el primer horizonte de un perfil está descrito de manera inversa y a partir del segundo comienza la descripción “normal”, para corregirlo, los valores del primer horizonte se invierten y los siguientes se reescalan con base en la diferencia del primero. Se presentaron 40 casos y en la Figura 4 se ejemplifica el proceso de reasignación de valores de los límites de los horizontes.

ID Perfil	ID Horizonte	Top	Bottom	SOC %	ID Perfil	ID Horizonte	Top	Bottom	SOC %
C-03	C-03-1	5	0		C-03	C-03-1	0	5	
C-03	C-03-2	0	5	3.9	C-03	C-03-2	5	10	3.9
C-03	C-03-3	5	25	1.1	C-03	C-03-3	10	30	1.1
C-03	C-03-4	25	40	0.7	C-03	C-03-4	30	45	0.7
C-03	C-03-5	40	77	0.3	C-03	C-03-5	45	82	0.3
C-03	C-03-6	77	115	0.3	C-03	C-03-6	82	120	0.3
C-03	C-03-7	115	180	0.2	C-03	C-03-7	120	185	0.2

Figura 4, Ejemplo de reescalamiento de los valores de los límites de un perfil por horizonte orgánico. A la izquierda el perfil con los valores originales y a la derecha los nuevos valores.

- Perfiles con el límite inferior del último horizonte vacío: como su nombre lo indica, el último horizonte no tiene un valor del límite inferior asignado, por convención se le asigna el valor de su límite superior +10. Se presentaron 992 casos, todos correspondientes a Argentina, 959 de esos valores fueron obtenidos de la última versión de los datos de [SISINTA](#) y con esos valores fueron actualizados esos registros, mientras que los otros 33 sí se aplicó la regla mencionada anteriormente. Un ejemplo de esta inconsistencia se muestra en la Figura 5.



ID	Sup	Inf	Nom	Dens	CaCO3	Gravass	ECEC	Cond	Carb	pH	Arcilla	Limo	Arena	Agua
4379	0	12	Ap						29.8	5.9	18.8	59.5	21.7	
4380	12	24	A1						21.3	6.0	24.3	54.8	20.9	
4381	24	38	A2						8.1	6.4	21.8	59.9	18.3	
4382	38	70	B21						5.1	6.9	30.5	53.2	16.3	
4383	70	100	B22						2.5	6.9	29.3	49.9	20.8	
4384	100	165	B3						1.4	7.2	21.3	49.1	29.6	
4385	165		C						1.1	7.8	15.0	51.7	33.3	

Figura 5, Perfil con el último horizonte sin valor de límite inferior.

### 1.2.6 Integridad de los atributos de los horizontes

En esta fase se analizaron dos aspectos, el primero referente a la cantidad de registros no vacíos de los atributos medibles de los horizontes, tal como se observa en la Tabla 2, se tienen valores en el 47% de los registros, sin embargo, excluyendo valores en 0 (cero), el porcentaje disminuye al 38%. Lo anterior es un punto crítico ya que esa falta de datos puede limitar la generación de información a partir de esos atributos. En futuras versiones de los datos proporcionados por los países se debe pedir mayores esfuerzos para que los datos estén lo más completos posibles. Finalmente, se identificó que para el caso del pH y el carbono orgánico existe una cantidad considerable de datos atípicos, para el primero se tiene un 10% de valores fuera de rango (0 a 14) y para el segundo el 2.9% de valores que podemos considerar altos (mayores a 30), una posible causa para este último puede obedecer a que se han proporcionado en otra unidad de medida.

## Resultados

Luego de los seis procesos de análisis descritos se logró consolidar una nueva versión de la base de datos en la que todos los perfiles cuentan con una correcta descripción de sus horizontes ya que se han corregido las inconsistencias que estos pudieron presentar. El detalle de los perfiles editados se encuentra en el archivo adjunto denominado AnalisisBaseDatosSISLAC.xlsx el cual contiene 5 hojas que se corresponde con las 5 primeras validaciones listadas anteriormente y son:

- 1-PerfilesFueraDeZona
- 2-PerfilesDuplicados
- 3-HorizontesSuperpuestos
- 4-PerfilesSinDatosHorizontes
- PerfilesInconsistenciasLeves

El resultado final de esta etapa son 41,202 perfiles validados, además, a la base de datos se han agregado dos atributos a los perfiles que son:

- **perfil\_duplicado:** Aquellos que tienen un valor TRUE son aquellos que comparten coordenadas con otro y se sugiere una revisión de los usuarios de esos perfiles.
- **tier:** o nivel de confianza de los datos, este puede tener tres valores numéricos (1, 2 y 3) y cada valor representa lo siguiente:
  1. Perfiles que están correctamente descritos
  2. Perfiles que fueron modificados para que estén correctamente descritos. Del archivo AnalisisBaseDatosSISLAC.xlsx adjunto aquellos perfiles que fueron corregidos tienen este valor.
  3. Perfiles que contienen inconsistencias graves y que limitan su uso.

## 2.- Nuevas bases de datos a incorporar al SISLAC

Luego de haber depurado la base de datos del SISLAC, el trabajo se centró en identificar otras fuentes de datos de perfiles de suelo en la región, las que se lograron obtener se listan en la Tabla 5, en esta se indica el proveedor de los datos, el link de acceso a estos, el número de perfiles, la licencia de uso y el sistema de referencia espacial en el cual se encuentran, este último es importante ya que todo debe armonizarse a coordenadas geográficas, datum WGS84, EPSG 4326.

Tabla 5, Descripción de las bases de datos para incorporar a la nueva versión del SISLAC.

País	Proveedor	Link a los datos	No. de perfiles	Licencia de uso	EPSG	Contacto
Argentina	<a href="http://sisinta.inta.gov.ar/">http://sisinta.inta.gov.ar/</a>	<a href="http://sisinta.inta.gov.ar/">http://sisinta.inta.gov.ar/</a>	6180	No se indica	4326	Marcos Angelini <a href="mailto:Marcos.Angelini@fao.org">Marcos.Angelini@fao.org</a>
Brasil	<a href="https://www.pedometria.org/">https://www.pedometria.org/</a>	<a href="https://www.pedometria.org/febr/ctb0003/">https://www.pedometria.org/febr/ctb0003/</a>	400	Attribution 4.0 International (CC BY 4.0)	4326	Alessandro Samuel-Rosa <a href="mailto:alessandrosamuelrosa@gmail.com">alessandrosamuelrosa@gmail.com</a>
Chile	<a href="https://www.earth-system-science-data.net/home.html">https://www.earth-system-science-data.net/home.html</a>	<a href="https://doi.org/10.17605/OSF.IO/NMYS3">https://doi.org/10.17605/OSF.IO/NMYS3</a>	13612	Attribution 4.0 International (CC BY 4.0)	4326	Marco Pfeiffer <a href="mailto:mpfeiffer@uchile.cl">mpfeiffer@uchile.cl</a>
Ecuador	<a href="https://www.earth-system-science-data.net/home.html">https://www.earth-system-science-data.net/home.html</a>	<a href="https://doi.org/doi:10.6073/pasta/1560e803953c839e7aedef78ff7d3f6c">https://doi.org/doi:10.6073/pasta/1560e803953c839e7aedef78ff7d3f6c</a>	13542	Attribution 4.0 International (CC BY 4.0)	32717	Daphne Armas <a href="mailto:daphne.arms@gmail.com">daphne.arms@gmail.com</a>
México Series I y II	<a href="https://www.inegi.org.mx/">https://www.inegi.org.mx/</a>	<a href="https://www.inegi.org.mx/app/biblioteca/ficha.html?upc=702825266707">https://www.inegi.org.mx/app/biblioteca/ficha.html?upc=702825266707</a>	13965	No se indica	6362	Rocio Hernández Reyes Sin correo

Las bases de datos anteriores fueron sometidas al mismo procedimiento de validación de integridad para garantizar que los perfiles cumplan los requisitos mínimos de integridad. Una vez validados, se comprobó si esos perfiles existían en la base de datos del SISLAC mediante el identificador único del perfil y/o las coordenadas de estos (en caso de ser posible). A continuación en la Tabla 6 se detallan el país, los perfiles, cuántos de estos ya existían en SISLAC, cuántos se excluyen por inconsistencia grave y finalmente, cuántos perfiles fueron agregados al SISLAC.

Tabla 6, Detalle de perfiles validados para su ingreso a la base de datos del SISLAC

País	Perfiles	En SISLAC	Con inconsistencia grave	Agregados
Argentina	6180	3374	61	2745
Brasil	400	0	0	400
Chile	13612	0	13	13359
Ecuador	13542	4633	67	8842
México	13965	7274*	6493**	7472
<b>Total</b>	<b>47699</b>	<b>8007</b>	<b>6634</b>	<b>32818</b>

\*No es posible determinar con exactitud si existen ya que no hay campos en común para hacer la comprobación, pero se comprobó que son de la misma fuente original (Series 1 y 2 de INEGI)

\*\*Corresponden a perfiles vacíos, sin datos de horizontes por lo que no se incluyen.

Como indica en la Tabla 6, para Argentina, Brasil, Chile y Ecuador fue factible comprobar si los perfiles ya existían en SISLAC y así no duplicar información. Para el caso de México no existen campos clave entre los datos del SISLAC y los nuevos datos, tampoco las coordenadas se corresponden, por lo que no existe forma de verificar qué perfiles ya existen, sin embargo, se comprobó mediante revisión aleatoria de perfiles que ambos conjuntos de datos provienen de la misma fuente, es decir, de las series 1 y 2 de INEGI. Por lo anterior, los nuevos datos se marcaron como *tier 1* (confiables) y los antiguos como *tier 3* (no confiables), así se tiene identificado el origen de cada perfil y su identificador único se corresponde con los datos originales al cual se le ha agregado el prefijo MexS1 o MexS2 según la serie a la que corresponde. Además, los datos se corresponden en cantidad ya que los que provienen de SISLAC son 7,274 y los que se ingresaron a partir del análisis son 7,472 perfiles.

A continuación en la Tabla 7 se muestran unos ejemplos del caso mencionado de México, como se puede observar en tres pares de perfiles, el identificador del perfil (*perfil\_ID*) y el identificador del horizonte (*layer\_ID*) no coinciden. Se observa que entre esos pares de perfiles sí se corresponde los límites de los horizontes y los valores de los atributos disponibles en el SISLAC, pero los nuevos datos de las series 1 y 2 tienen más atributos con valores por lo que representan un mejor insumo. Adicionalmente, se observa en la Figura 7 que espacialmente los perfiles se encuentran a una distancia de aproximadamente 4 kilómetros. Los perfiles de las series 1 y 2 se encontraban en sistema de referencia espacial EPSG 6362 y fueron transformados a EPSG 4326 que es en el que se encuentran los datos del SISLAC.

Evaluaciones para favorecer la adopción de soluciones digitales en intervenciones en ALC  
Asesoría técnica para las bases de datos de suelos en el marco de la iniciativa del Mapa Global de  
Nutrientes de los Suelos (GSNmap) en una nueva versión de SISLAC – Producto 3.2.1.2

Tabla 7, Casos de perfiles de México en los que los identificadores de perfil y horizonte no coinciden pero sí los datos de los horizontes.

FUENTE	Perfil_ID	Layer_ID	Límite Superior	Límite Inferior	Designación	CIC	Conductividad	Carbono Orgánico	pH	Arcilla	Limo	Arena
SISLAC	217228	1065725	0	20	S/D		0.1	4	7.3	6	23	71
SISLAC	217228	1065726	20	32	S/D		0	0	7.6	2	9	89
México Serie 2	MexS2-154107	MexS2-154107-1	0	20	A1	20.5	0.1	0.4	7.3	6	23	71
México Serie 2	MexS2-154107	MexS2-154107-2	20	32	C1	15	0	0	7.6	2	9	89
SISLAC	217298	1065931	0	13	S/D		0			0.1	9	6
SISLAC	217298	1065932	13	39	S/D		0			0	6	6.7
SISLAC	217298	1065930	39	61	S/D					0	4	7.1
SISLAC	217298	1065933	61	90	S/D					0	4	7.1
México Serie 2	MexS2-156151	MexS2-156151-1	0	13	A1		0		9.6	0.1	0.9	6
México Serie 2	MexS2-156151	MexS2-156151-2	13	39	A2		0		8.3	0	0.6	6.7
México Serie 2	MexS2-156151	MexS2-156151-3	39	61	A3		0		10.2	0	0.4	7.1
México Serie 2	MexS2-156151	MexS2-156151-4	61	90	B1		0		10.1	0	0.4	7.1
SISLAC	208456	1062103	0	20	S/D					1		8.3
México Serie 2	MexS1-62028	MexS1-62028-1	0	20	C1				7.5	1	0.1	8.3

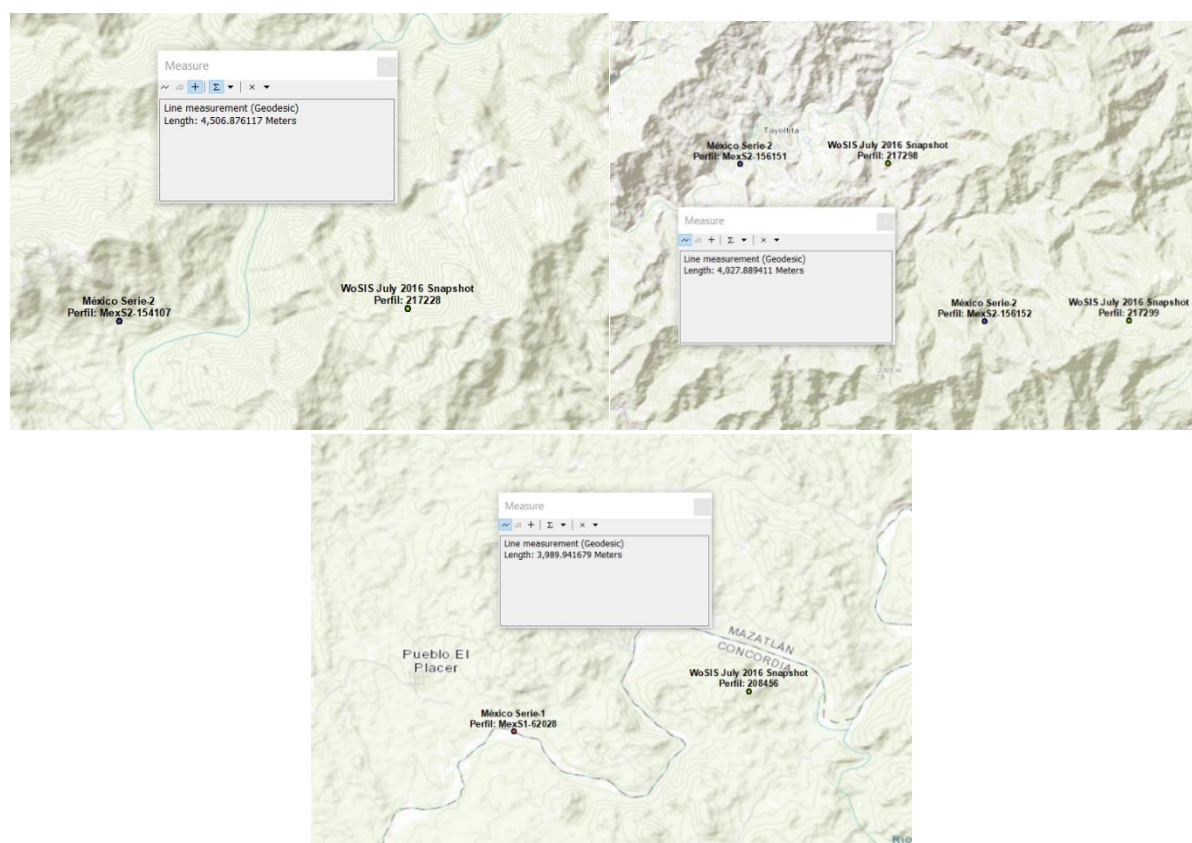


Figura 6, Ejemplo de perfiles idénticos son diferente identificador de perfil y de horizontes.

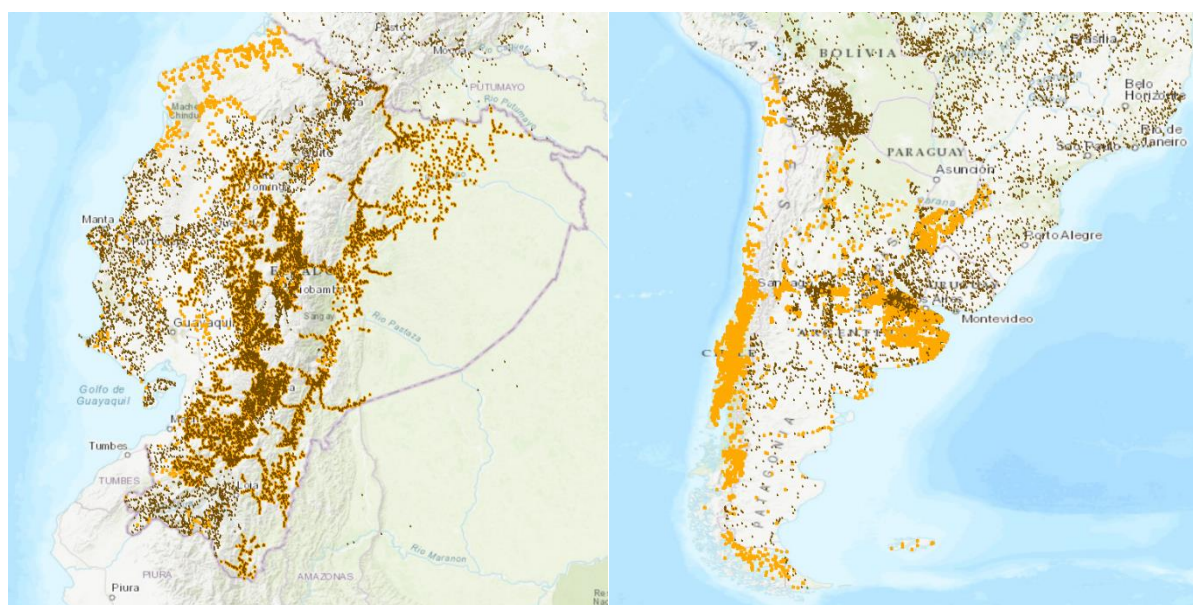
Luego de agregar los perfiles citados en la Tabla 6, la nueva base de datos del SISLAC pasa de tener 41,202 perfiles válidos (excluyendo los existentes en México) a 66,746, todos estos validados y con una correcta descripción de sus horizontes, lo anterior significa un aumento del 61% de perfiles. A continuación en las siguientes figuras se muestra las zonas en que se agregaron perfiles, los puntos color café corresponden a los perfiles originales y los puntos naranja a nuevos perfiles: en la Figura 7 se muestra la distribución espacial de los nuevos perfiles. La Figura 8 muestra la nueva cobertura de perfiles para Ecuador (izquierda), y Chile y Argentina (derecha). La Figura 9 corresponde a los perfiles de México, la diferencia en este



caso no es perceptible ya que como se comentó, los perfiles provienen de la misma fuente por lo que casi todos tienen su par con la misma ubicación. Finalmente, en la Figura 10 una vista general de la región con los puntos originales y los nuevos perfiles agregados.



*Figura 7, Localización de los nuevos perfiles agregados al SISLAC.*



*Figura 8, Nueva cobertura de perfiles en Ecuador, Argentina y Chile. En naranja los nuevos perfiles y en café los perfiles del SISLAC original.*



*Figura 9, Total de perfiles para México.*



*Figura 10, Visualización de la distribución de los perfiles originales y nuevos en la región.*

A continuación en la Tabla 8 se muestra un comparativo de los perfiles iniciales y actuales de los países en los que se adicionaron datos:



Tabla 8. Detalle de los perfiles iniciales, agregados y totales en los países con nuevos datos.

País	Perfiles SISLAC	Perfiles nuevos	Total por país
Argentina	3770	2745	6515
Brasil	7714	400	8114
Chile	44	13359	13403
Ecuador	13070	8842	21912
México	7274*	7472	7472
TOTAL	24598	32818	57416

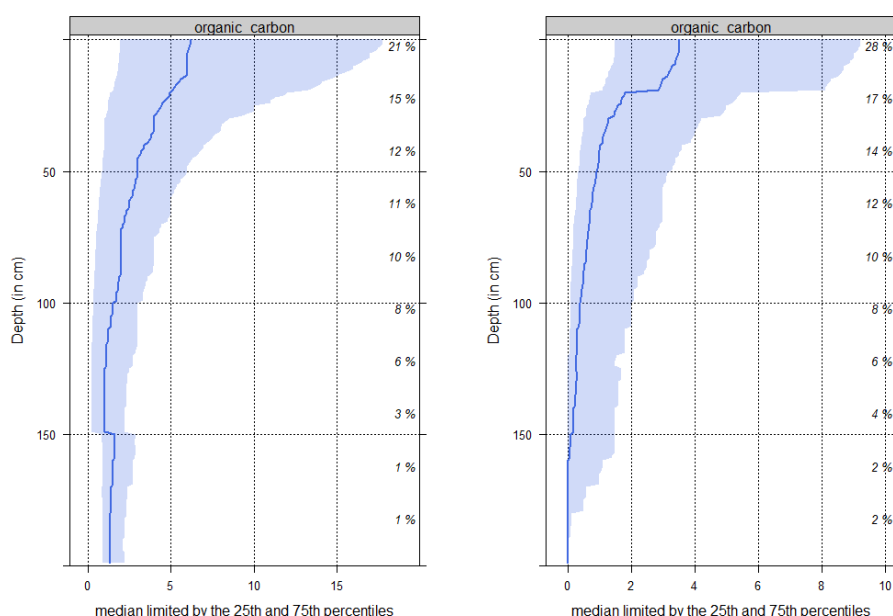
\*Como se mencionó estos fueron marcados como tier 3 por lo que no deben considerarse como válidos.

### 3.- Base de datos SISLAC versión 2.0

Como se describió en el apartado anterior, la base de datos del SISLAC versión 2.0 cuenta con 66,746 perfiles de suelo con una correcta descripción de sus horizontes, los cuales suman un total de 192,568 registros. La nueva base de datos ha sido complementada con nuevos atributos en los perfiles y en los horizontes para almacenar otras propiedades como las de los nutrientes con miras a la producción de los nuevos Mapas Globales de Nutrientes de la FAO. En el documento adjunto número 2 “Diccionario de datos” se listan los nombres de los atributos para la tabla de perfiles y horizontes, su descripción, el tipo de dato que almacena, observaciones y el valor por default que almacena. Adicionalmente, se proporcionan los códigos para el *software* R para el análisis y depuración de los datos para que estos puedan ser usados por los usuarios que así lo requieran. Esos códigos y la documentación descriptiva estarán alojados en el repositorio oficial de la FAO <https://github.com/FAO-GSP/SISLAC-DB>.

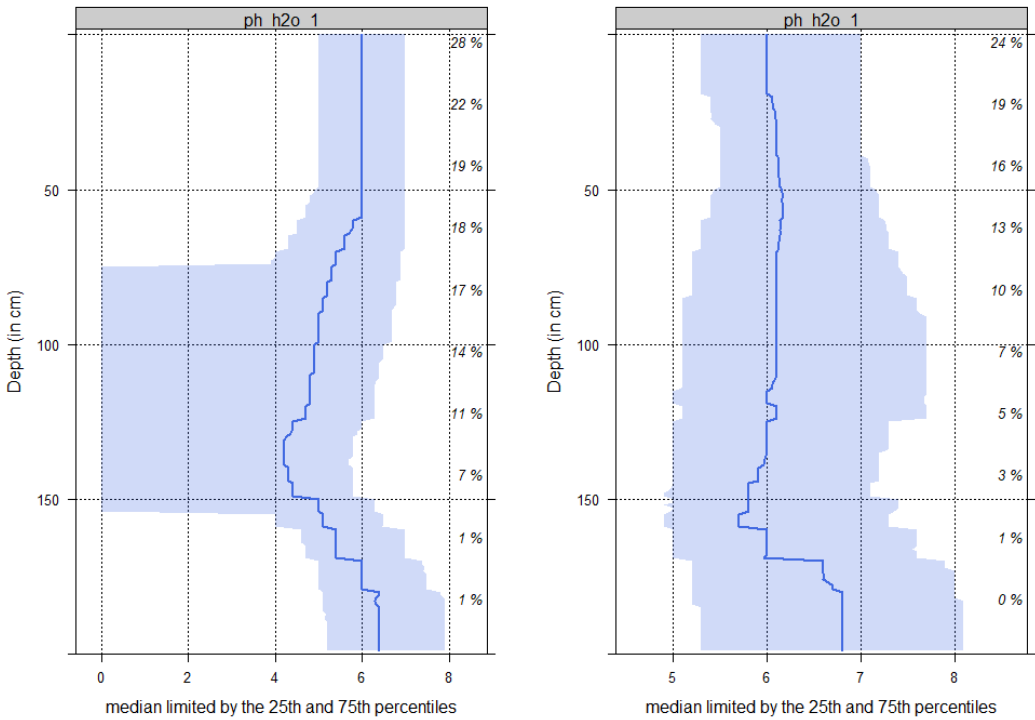
Enseguida se muestran las gráficas de la variabilidad vertical del carbono orgánico, pH, arcilla, limo y arena con los datos originales y con la nueva versión del SISLAC. Dichas variables cuentan con la mayor cantidad de registros con valores, lo que permite evidenciar el cambio presentado con el nuevo volumen de datos. En las gráficas siguientes la de la izquierda corresponde a los datos originales y la de la derecha a la nueva base de datos.

#### Carbono orgánico

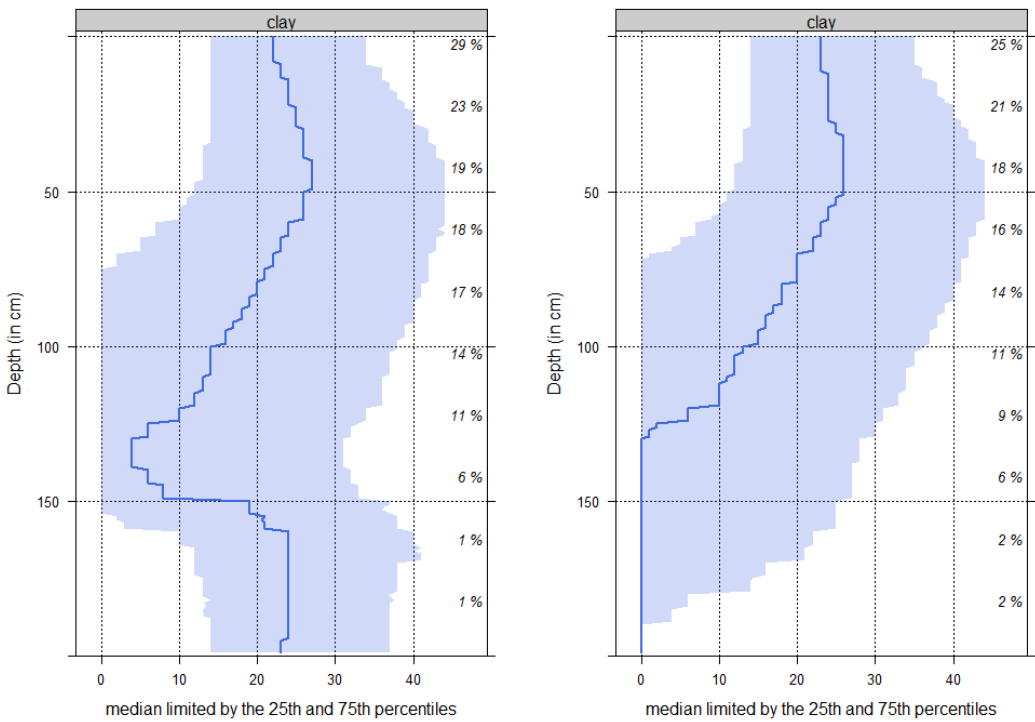




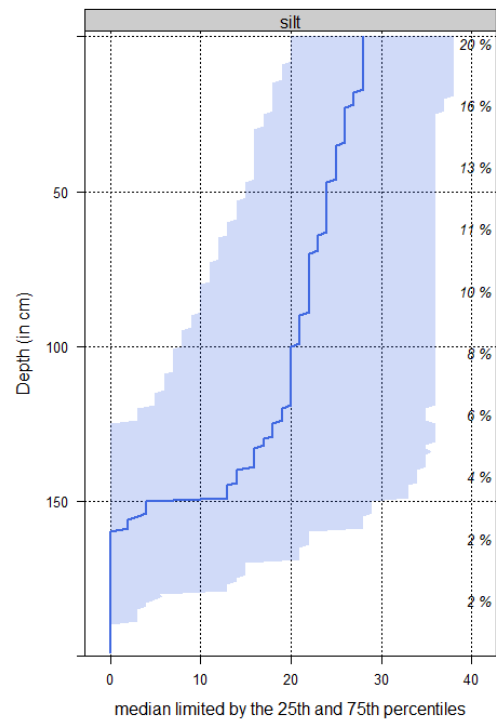
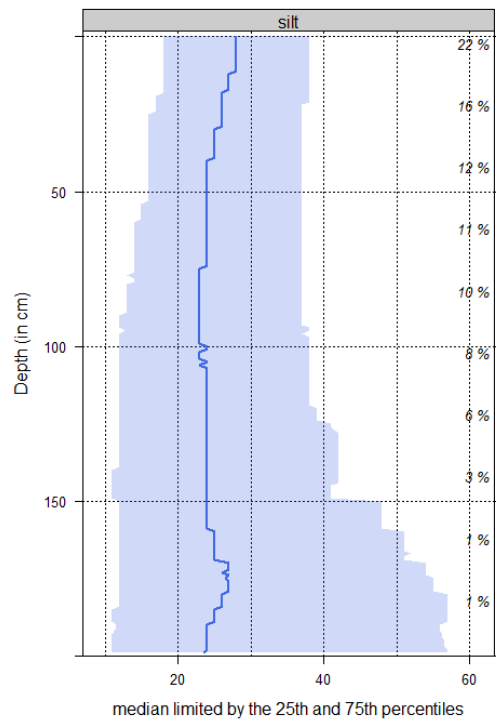
**pH**



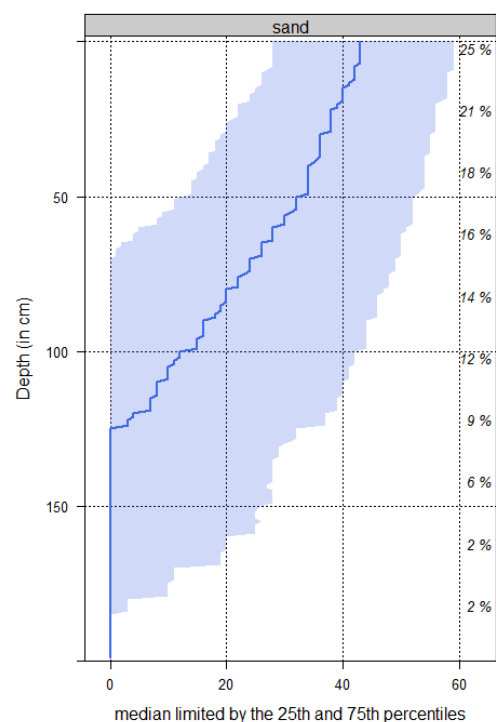
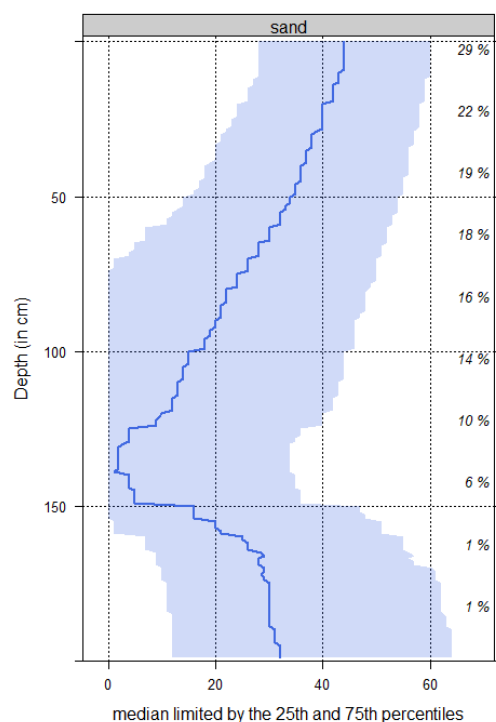
**Arcilla**



## Limo



## Arena



Como se puede observar, mientras que en las gráficas de carbono orgánico el porcentaje de datos aumenta, en el resto disminuye, esto se debe en parte a que, una de las bases de datos agregadas (Chile, con aproximadamente 13,000 perfiles nuevos) solo presenta datos de carbono, por lo que este valor aumenta, pero los otros restantes no, lo que hace que el porcentaje disminuya.

#### 4.- Conclusiones.

La nueva versión de la base de datos, que para efectos prácticos llamamos SISLAC 2.0 pasó de tener 49,084 perfiles originales y 41,202 perfiles válidos a 66,746 registros, todos con una correcta descripción de los horizontes de los perfiles, siendo esto clave para facilitar su procesamiento y análisis. Lo anterior se evidencia en la generación óptima de las gráficas de variabilidad vertical del apartado 3. Se identificó que las inconsistencias en la descripción de los horizontes ocurren de manera sistemática, lo cual ha servido para informar a los países sobre cómo subsanar problemas puntuales con el objetivo de que los perfiles cuenten con un mejor nivel de integridad previo a su envío. Un ejemplo de corrección de los datos es el de Argentina, ya que en los datos obtenidos ya incluyen los valores del límite inferior del último horizonte, ese valor fue tomado y actualizado en los datos originales del SISLAC que presentaban dicha inconsistencia. En la revisión de la integridad de los horizontes tanto de la base original como de las nuevas bases de datos a agregar se procuró al máximo subsanar las inconsistencias de los perfiles, esto se puede identificar en los perfiles que en la columna *tier* indican el valor 2.

#### 5.- Recomendaciones.

Finalmente, se hacen algunas recomendaciones que pueden ayudar a que nuevos datos proporcionados por los países cuenten con niveles mínimos de integridad previo a su envío e incorporación al SISLAC.

- Es importante que los países o proveedores de datos traten en la medida de lo posible garantizar que los valores de los atributos: identificador del perfil, identificador del horizonte, límite superior y límite inferior siempre contengan datos ya que este inconveniente es persistente.
- En la medida de lo posible, se debe garantizar que los horizontes tengan un identificador secuencial que facilite el ordenamiento de estos dentro del perfil ya que sin este, el ordenamiento de los datos para su análisis, por ejemplo, utilizando la librería para el *software* R “AQP” se dificulta, esto puede generar un orden incorrecto de los horizontes que a su vez ocasiona que los análisis no se puedan realizar hasta que no se ordenen correctamente.
- Un atributo vacío informáticamente puede representar que el valor se “perdió” o no se ingresó, etc., para los atributos numéricos se recomienda que cuando estos no contengan valor se coloque *NULL* y S/D para los atributos de texto, así, se tiene control sobre los valores ya que, por ejemplo, en los atributos numéricos el valor 0 (cero) representa un *valor válido* que puede sesgar los análisis realizados. En este trabajo

se garantizó que todos los datos ingresados llevaran un valor *NULL* o S/D según corresponda cuando no existían valores en los campos.

- Los perfiles contenidos en el SISLAC han sido aportados por las instituciones oficiales relacionadas con el manejo del recurso suelo de cada país, sin embargo, existe otra fuente de datos con una potencial cantidad de perfiles: las instituciones académicas y de investigación (por ejemplo universidades, centros de investigación o incluso empresas privadas), se recomienda contemplar la posibilidad de hacer esfuerzos por recabar esos datos que sin duda harán la base de datos más robusta, lo cual puede derivar en una mejor calidad de la información generada.