Comments by anonymous referee on: CHLSOC: The Chilean Soil Organic Carbon database, a multi-institutional collaborative effort

Data:

1. I would prefer unabbreviated column headings with units or an explanation table for column headings and units in the data paper.
2. There are 0 values in columns oc and crf in the data – is this possible?

Discussion paper:

Abstract. ~~One of the~~A critical aspect~~s~~ in ~~modelling~~ predicting soil organic carbon (SOC) concentrations ~~predictions~~ is the lack of ~~access~~ available~~to~~ soil information; where information on soil characteristics is available, this is ~~which is~~ usually concentrated ~~o~~in regions of high agricultural interest. To date i~~I~~n Chile, ~~most~~ a large proportion of the soil and SOC data has been collected in a quarter of ~~to date is highly concentrated in 25% of~~ the territory that has intensive agricultural or forestry use, however, ~~. V~~vast areas beyond the~~o~~se forms of land use have few or no soil data available. Here~~,~~ we present a new database of SOC for the country, which is the result of an unprecedented national effort under the framework of the Global Soil Partnership. This partnership has helped ~~that help to~~ build the largest database on SOC to date in Chile ~~- named~~ the Chilean Soil Organic Carbon database ("CHLSOC)", comprising 13,612 data points~~. This dataset is the product of the compilation from~~ compiled from numerous sources including unpublished and difficult to access data. The database will allow users to ~~, allowing to~~ fill ~~numerous~~ spatial gaps where no SOC estimates were publicly available~~ before~~. ~~The~~ Presented values of SOC ~~compiled in CHLSOC~~ range from 6×10−5 to 83.3 percent, reflecting the variety of ecosystems that exist~~s~~ in Chile. The dataset has the potential to inform and test current models that predict SOC stocks and dynamics at larger spatial scales thus enabling benefit ~~Profiting~~ from the richness of geochemical, topographic and climatic variability in Chile~~,~~. ~~the 10 dataset has the potential to inform and test models trying to predict SOC stocks and dynamics at larger spatial scales.~~

The d~~D~~ataset is freely available to registered users at https://www.doi.org/10.17605/OSF.IO/NMYS3 (Pfeiffer et al., 2019b) under the terms of the Creative Commons Attribution 4.0 International Public Licence.

Introduction

Soil organic carbon (SOC) stocks play a ~~crucial~~ vital role in the global Carbon (C) cycle~~,~~ and equal nearly two thirds of the total terrestrial carbon ~~stocks~~ pool (Eswaran, 2000; Sarmiento and Gruber, 2002). Therefore, knowledge of the contents and dynamics of the SOC stock is ~~pool is 2 https://doi.org/10.5194/essd-2019-161 Open AccessEarth System Science Data Discussions Preprint. Discussion started: 14 October 2019 c Author(s) 2019. CC BY 4.0 License. are~~ essential to estimate trends in the evolution of atmospheric carbon dioxide (CO_2), ~~content s~~ to be used as an input and applied ~~in~~ to models of global climate change (Jones et al., 2005; Davidson and Janssens, 2006). However, predictions of the SOC stock vary widely due to limited availability of soil data for remote regions and existing soil datasets being biased towards highly managed forest and agroecosystems (Duarte-Guardia et al., 2018). Chile is not exempt ~~of~~from the~~o~~se difficulties, having publicly available soil and SOC data focused ~~o~~in ~~the~~ intensively cultivated ~~areas of the~~ central regions (Padarian et al., 2012, 2017). In fact, v~~V~~ast areas of the country~~, however,~~ are situated in the high Andean mountains, the hyperarid Atacama Desert or the inaccessible Magellanic moorlands ~~in~~ of the

Patagonian fjords for which very few soil data is available. These areas are of particular interest for SOC dynamics and stock predictions as they represent the extreme ends of a huge latitudinal climate gradient from Earth's driest extreme in the north (Atacama Desert) to the very humid conditions of Patagonian pacific margin, all flanked by the second highest mountain range in the world (Garreaud et al., 2009; Ewing et al., 2008; Loisel and Yu, 2013).

A~~The a~~ccess to spatially explicit, consistent and reliable soil data is essential to model and map the status of soil resources globally at increasing detailed resolution in order to respond and assess world global issues (Arrouays et al., 2014; FAO, 2015; Hengl et al., 2014; Omuto et al., 2013). Furthermore, soil datasets are ~~also~~ one of the most important inputs for Earth System Models (ESM), to address, for example, the importance of terrestrial~~s~~ sinks and sources ~~for~~ of greenhouse gases (Dai et al., 2018; Luo et al., 2016). At the same time, soils in ESM are one of the largest sources of uncertainty (Dai et al., 2018). ~~This is why~~Hence, in recent years there has ~~being~~ been a growing effort to improve access to and quality of soil datasets, ~~being one of the~~a key goal~~s~~ of ~~pillar 4 of~~ the Global Soil Partnership Pillar 4 Implementation Plan~~global soil partnership~~ sponsored by the Food and Agriculture Organization of the United Nations (Batjes et al., 2017; Omuto et al., 2013). ~~In this sense, e~~Efforts to increase access to harmonized soil products~~,~~ containing comparable and consistent datasets including soil carbon are highly valuable and appreciated by an increasing number of users (Arora et al., 2013; Baritz et al., 2014; Batjes et al., 2017; Hendriks et al., 2016; Jones and Thornton, 2015; Luo et al., 2016; Maire et al., 2015).

In an unprecedented national effort, between May 2018 and April 2019, a group of professionals from 39 public and private institutions ~~joint~~ joined together to build the largest (to date) Chilean SOC database (CHLSOC). The dataset was ~~We built this dataset based on the compilation of~~compiled from varied data ~~different~~ sources including~~corresponding to~~ soil surveys, publications, private reports, unpublished research data and ~~cryptic~~ obscure documents unknown to the public and often ~~or~~ difficult to access.

The entire resultant CHLSOC dataset ~~All data in CHLSOC~~ (13,612 data points from 25 sources; ~~25~~ summarised in Table 1) is ~~publicly~~ freely available to registered users for~~and can be~~ download~~ed for free~~ at https://www.doi.org/10.17605/OSF.IO/NMYS3 (Pfeiffer et al., 2019b). This joint effort has resulted in a ~~consistent~~ comprehensive Chilean soil dataset ~~of Chile to be~~that is available for the international community for analysis, exchange and interpretation.

2 Soil Data harmonization

2.1 Database sources

In order to fill the gaps in the current data, and t~~T~~o build CHLSOC ~~(Chilean Soil Organic Carbon database), we~~ 889 soil profiles and 12,723 topsoil samples from all over Chile (Table 2) were gathered, curated and harmonized. ~~889 soil profiles and 12,723 topsoil samples from all over Chile (Table 2)~~. Eighty nine percent of this information was previously unpublished or un~~-and not~~ available to the national and global scientific community. ~~This~~ The resultant soil information was from ~~cover~~ all of the administrative regions and 16 out of 17 ecological zones of Chile (Figure~~.~~ 1, Table 3).

Data~~The data~~ compiled from the literature is ~~properly~~ referenced in Table 1. Sources include legacy soil surveys, environmental assessment reports, research papers, private reports, theses and unpublished data ~~facilitated~~ provided by researchers. A m~~-M~~inimum requirement for inclusion to ~~s for studies to enter~~ the database were ~~to have~~ geographic coordinate information~~s~~, records of soil horizon depth~~s~~ and soil organic carbon content (or organic matter content). ~~If available, o~~Other soil variables such as bulk density, texture and/or coarse fragments, sampling depth, sampling year and measurement methods, were ~~also~~ included where available. Approximately 20% of ~~the total~~ horizon samples included information on bulk density (BLD)~~. Where available, this information was~~ measured using the clod ~~method~~ or the core (cylinder) method and~~.~~ ~~Only~~ 382 horizons (<3~~2.8~~%) included information about coarse fragments (CRF). The resulting database (summarised in Table 1) includes datasets of variable size, source and composition. ~~All u~~Unpublished data sources are referenced in the database to the ~~coauthor~~co-author and group who provided the data. Examples of unpublished data sources are shown in Table 1 and include those of Oficina de Estudios y Políticas Agrarias (ODEPA) with 782 points provided by J. Ramirez, METHANOBASE (Table 1), corresponding ~~10~~ to surface samples (0 - 25 cm) from the Magallanes Region collected during 2016 and provided by L. Cabrol and M. Barret. (Table 3). A f~~F~~urther~~,~~ 51 data points from the E~~e~~nvironmental Impact A~~a~~ssessment ~~database~~ System (SEIA) were includ~~ed from under represented~~ ~~ing focusing on~~ areas ~~with low representation,~~such as the Andean Cordillera and the Atacama Desert.

The largest contributor to CHLSOC (~~in terms of number of~~ 9,935 data points (~~9,935~~) ~~as well as~~and a wide geographic extent)~~ical extension were~~ was the SOC dataset of the Agricultural and Livestock Service (SAG by its Spanish acronym). ~~This~~ The data comprsie~~dataset only includes~~ SOC~~,~~ obtained from the first 20 cm of soil by auger or excavation methods ~~15~~ sampled ~~taken~~ by beneficiaries (farmers) of the SAG subsidy program.

Another ~~relevant~~ important data contribut~~or~~ion was ~~taken~~ the~~from~~ legacy soil ~~legacy~~ survey data compiled by Centro de Información de Recursos Naturales (CIREN) (~~Centro de Información de Recursos Naturales~~) ~~and~~ reported as regional soil surveys~~, based on soil surveys~~ that were carried from the 1960s ~~up~~ to 2007. ~~The database compiled by CIREN is distributed over 177,500 km2 , equal to about a 24.5 % of the total Chilean territory~~. In total, CIREN compiled 37 soil surveys, ~~including~~ totalling 540 data points over 177,500 km2 (equal to almost 25% of the total Chilean territory), much of which are already compilations of former studies originally not referenced by CIREN (CIREN, 1996a, b, 1997a, b, 20 1999, 2002, 2003, 2005a, b, 2007).

Commented [CJ1]: Sorry, what do you mean here?

2.2 Data harmonization processing and caveats

The assembled data ~~has been~~was sampled over several decades and for some data points it was ~~no longer~~ not possible to find ~~and~~ or verify the original data source ~~for some of the data points~~. A potential source of ~~dditional~~ uncertainty may be the ~~is~~analytical method employed for analysis; ~~introduced as~~ SOC content for most samples (97%) was analyzed using the wet oxidation method and a~~. Only~~ small number~~a minor fraction of the SOC data~~ were analyzed by total combustion (CN elemental analyzer). ~~The difference in the analytical methods could be a source of uncertainty for future modelling initiatives which is not properly addressed in Chile on a national level.~~ Discrepancies in ~~the results of~~ SOC ~~analyses~~ results between combustion methods have identified wet combustion ~~as an unreliable~~as a less reliable assessment method for SOC (Kumar et al., 2019). This ~~is not yet~~issue has not been ~~properly~~ addressed in Chile to date. The~~, as the~~ recommended methods for SOC determination are currently ~~are still~~ wet oxidation and loss on ignition; ~~, not mentioning~~ however, dry combustion i~~a~~s a more accurate alternative (Sadzawka et al., 2006). Future data collection initiatives of ~~data collection~~ should stress for consistent analytical procedures ~~while~~ and a revision of local standards is urgently required. A possible source of bias in data from SAG~~'s data~~ is that samples we~~a~~re taken by farmers following SAG guidelines where ~~provided by SAG~~, ~~taking~~ a composite sampling is taken for each parcel.

3 Spatio-temporal distribution of the Chilean SOC dataset

3.1 Spatial distribution

To date, CHLSOC is the most complete data compilation for mainland Chile, ~~CHLSOC currently is~~ comprising~~ed of~~ 13,612 points, ~~which is~~ a great improvement in comparison with ~~compared to~~ former databases used in Chile for SOC assessments. For example, national SOC mapping studies (Padarian et al., 2017; Reyes Rojas et al., 2018) we~~a~~re based almost exclusively on CIREN data~~ of CIREN~~ (540 points). ~~To date, CHLSOC is the most complete data compilation for mainland Chile~~. CHLSOC~~ and~~ can be used to ~~reflect~~ show the influence of soil, vegetation and climatic conditions on SOC concentrations. Table 3 shows the ~~amount of points~~number of data compiled in this work, by vegetation formation. It is important to not~~ic~~e that this scheme corresponds to the potential vegetation belt~~s~~ that originally occupied the territory and does not necessarily reflect~~s~~ ~~the~~ current land use (Luebert and Pliscoff~~, (~~2006). ~~In this sense, we will~~We refer to vegetation formations as "ecosystems" ~~as~~ since this is a more common term and to avoid further specific disciplinary discussion, which is outside~~ of~~ the ~~reach~~ scope of this work. In order to represent~~Representativeness of~~ each ecosystem (by vegetation formation) ~~in~~ the CHLSOC database is based on the number of data points divided by the total coverage of the ecosystem in Chile.

Commented [CJ2]: Which scheme?

Commented [CJ3]: Sorry not sure what you mean here

~~About~~ More than two-thirds (85.73%) of ~~our total~~the data are sampled from a concentrated area~~ in~~ (25% of the total country area) ~~and located~~ found in the following four ecosystems: deciduous forest, broad-leaved forest, sclerophyllous forest and thorny forest. T~~; t~~he first two ecosystems are located in the northern section of the temperate macro bioclimate zone, and the second two in the southern section of the mediterranean macro bioclimate zone (Moreira-Muñoz, 2011) respectively. These ecosystems are characterized by ~~its~~ a combination of benign climate, high quality soils and water availability~~le~~ (for irrigation), ~~which~~ resulting~~ed~~ in a long history of agricultural activity and human settlement~~s~~ (Armesto et al., 2010). For this reason, these areas ~~Because of this, they comprise the area that~~ historically experienced the highest land-use conversion ~~for~~ to agriculture, forestry and urban use in the country (Echeverría et al., 2006; Schulz et al., 2010; Arroyo et al., 2008). Deciduous forests (~~— which comprise~~ 14.7% of the country)~~ — is~~ are the most represented, with 52.14% of the data points collected in CHLSOC located between latitudes 35°S to 41°S (Figure 1).

The second ~~biggest~~ largest pool of data~~ points, with~~ (8.6% of the total data compiled in this work)~~,~~ ~~covers~~ is for evergreen forest, steppe and grassland (Table 3). ~~These two ecosystems, that~~ which comprise~~e~~ 10.3% of the country area. These ecosystems~~,~~ are located between 41°S and 53°S (Figure 1) in the Temperate macrobioclimate zone (Moreira-Muñoz, 2011), a thermally homogeneous territory with a considerable precipitation gradient that can reach several meters of mean annual precipitation o~~i~~n its western section, along the pacific coast (Garreaud et al., 2009). ~~It~~ The areas contain~~s~~ ~~big~~ vast sections of pristine forest, with only 8% of the land being converted to other land use (Pliscoff 20 and Fuentes-Castillo, 2011). Most of the data collected here correspond to the eastern section of the administrative region of Aysén in Patagonia. The relatively high representation of these ecosystems in the database can be attributed to (i) the intense agricultural use of the northern section of the evergreen forest, and (ii) an unprecedented effort in soil sampling in the Aysén Region (43.5°S – 49°S) (Figure 1) by SAG and the Agricultural Resarch Institute (INIA by its Spanish acronym) (Table 1) (Hepp and Stolpe, 2014).

Arguably t~~T~~he most ~~relevant~~ important ecosystem, in terms of SOC stocks for Chile, corresponds to the moorlands, which comprise a large area located o~~i~~n the Pacific coast of Patagonia where the landscape is fragmented into fjords and ~~little~~ small islands (between 44°S and 55°S (Figure 1)). ~~This ecosystem, which~~The moorlands cover~~s~~ a significant section (9.1%) of the country area and are~~, is likely~~ probably the ~~biggest~~ largest soil carbon reservoir ~~of~~ in Chile, with an almost continuous carpet of thick peat bog~~s~~ to a depth of ~~that can reach up to~~ 5 m in some ~~sections~~ places (Loisel and Yu, 2013; Minasny et al., 2019). Despite ~~its relevance~~the importance of moorland soils, most of our knowledge ~~on soils from~~of this ecosystem comes from ~~its~~ the northern and eastern borders ~~where there is some accessibility,~~ whereas there is limited information about peat soils in remote areas of the western fjords (20 observations in this database).

The ~~ecosystems that comprise the~~ Atacama Desert section of Chile (Table 3; desert, low desert scrub and desertic scrub) comprise 2.18% of the CHLSOC database ~~presented in CHLSOC~~ but correspond to 6% of the country area. ~~Despite this low percentage~~However~~,~~ the number of data points compiled for this region (298) ~~still~~ constitute a great improvement compared ~~to previous works, as the only~~ previous national work on SOC ~~that includes data points~~ for the Atacama Desert ~~considered~~ which only includes~~only~~ 3 points (Padarian et al., 2017).

The scarce SOC information of this region ~~to date is mainly because of the~~may be due to the extreme aridity of the region, ~~and~~ low biological activity and low ~~amount of~~ SOC accumulation (McKay et al., 2003). Vegetation is restricted to a narrow belt along the coast that receives water from fog, ~~to the~~ deep valleys that cross the desert and ~~toward~~ the western flank of the Andes (Moreira-Muñoz, 2011).

Regions of h~~H~~igh altitude and mountainous ~~regions~~ areas comprise 102 data points (0.74% of the database) representing 16.2% of the country area. Two characteristic alpine vegetation formations exist in the Andean Cordillera of Chile between 18°S and 38°S (Figure 1) that comprise ~~,~~ herbaceous alpine vegetation and alpine dwarf scrub. Most of the data is concentrated on the lower part (alpine dwarf scrub), while virtually no soil data is available for the higher section of the Andes (above 3000 m a.s.l.). The scarcity of soil data ~~in~~ for this region means that ~~poses a high uncertainty to~~ assessments of the impact of climate change on soil C stocks are uncertain, because large quantities of SOC are stored ~~is~~ in this~~type of~~ ecosystem (Bockheim and Munroe, 2014).

~~Other vegetation formations that have f~~Few data ~~correspond to~~ are available for ~~the~~ coniferous forest, deciduous shrubland, thorny shrubland and arborescent shrubland areas of vegetation (Table 3~~). Those formations are~~ located in areas of low forestry or agricultural interest but these areas comprise ~~together with a small surface area (~~less than 2.5% of the country~~)~~.

In summary, ~~our~~ the data we have ~~compilation~~ compiled demonstrate the imbalance ~~shows that there is an unbalance~~ between areas of agricultural and forestry interest and ~~those~~ areas beyond ~~that~~ those land uses. Three areas of high value in terms of ecological, scientific and ecosystem service ~~their surface and significance, not only~~ nationwide (and~~but also~~ worldwide), are under-represented in terms of soil data: High Andean Cordillera, Atacama Desert and Western Patagonia. ~~These areas are of high interest for ecological, scientific and ecosystem service purposes.~~ Government efforts ~~oriented~~ to develop soil surveys in these regions should be ~~are~~ promoted urgently~~. and should be promoted.~~ In particular, a SOC inventory of western Patagonia is essential to properly assess the national stock ~~on~~ of SOC, and the potential to include this area in carbon offset~~s~~ programs.

3.2 Temporal distribution

~~In CHLSOC we provide the date~~The date of sample collection is provided in more ~~were this was available, so users can constrain the use of the data to a certain time frame, or only to data were the time of sampling is available. More~~ than 90% of the included data (12,318 data points)~~ reported the date of soil sampling~~. The majority of points were ~~Most of the data was taken~~ sampled between 2006 and 2018 (Figure~~.~~ 2). The high number of data from the last decade enables users to ~~the last decade reflects that an important proportion of this database consist of recent data that can be used~~

~~in doing~~ estimat~~ions of~~e modern carbon ~~contents~~ stocks in Chilean soils. Most of the data with years of sampling are concentrated in a short timeframe and mainly corresponds to the SAG database and to sampling efforts related to research projects such as ODEPA in 2006; INIA and NAMA in 2015; NAMA (2016), Methanobase in 2016, and a compilation of data from INIA La Platina that was compiled recently by Corradini et al. (2019). Data from CIREN (Table 1) did not report sampling date. However, as it consist of compilations from known former soil surveys, we can limit the period in which samples were collected and analyzed to 30 the period between 1970 and 2007. The oldest data points correspond to those collected by Holdgate (1961) in the Western Patagonian fjords in 1959.

Commented [CJ4]: Suggest rewriting this. Doesn't quite make sense.