Comparative Analysis of ^{210}Pb Dating Models: Evaluating Accuracy and Precision with Simulated Datasets

Marco A Aquino-López[*†] Nicole K. Sanderson[‡] Maarten Blaauw[§]
Joan-Albert Sanchez-Cabeza[¶] Ana Carolina Ruiz-Fernandez[‖] J Andrés Christen[**]

Abstract

The study of anthropogenic impacts on the environment can be addressed through dated sedimentary records for the last $\sim$ 100-200 years. During this period, radiocarbon (^{14}C) dating suffers from poor resolution and large uncertainties. To overcome this limitation, ^{210}Pb (lead-210) dating has been widely adopted as it provides absolute and continuous dates for this period. However, the Constant Rate of Supply (CRS, or Constant Flux - CF) model, which relies on the radioactive decay equation as an age-depth relationship, limits the accuracy of the model used to estimate older dates. In this work, we compare a classical approach to ^{210}Pb dating (CRS) with Plum, a Bayesian approach developed for analyzing sedimentary ^{210}Pb measurements.

Simulated ^{210}Pb profiles were generated according to three different sedimentation processes and constructed using the CRS model assumptions, which we analyzed using both methods. Our results indicate that the CRS model sometimes fails to capture true age values, and accuracy does not improve with more information. In contrast, Plum consistently provides more accurate results, even with relatively small sample sizes, and improves accuracy and precision with additional information.

Keywords: Plum, Age-depth models, Chronology, Constant Rate of Supply, Simulations, Comparison.

[*]Department of Geography, University of Cambridge, Cambridge, United Kingdom email: aquino@cimat.mx

[†]Corresponding author.

[‡]GEOTOP Research Centre, Université du Québec à Montréal, Montréal, Québec, H2X 3Y7, Canada. email: sanderson.nicole@uqam.ca

[§]School of Natural and Built Environment, Queen's University Belfast, Belfast, BT7-1NN, UK. email:maarten.blaauw@qub.ac.uk

[¶]Unidad Académica Mazatlán, Instituto de Ciencias del Mar y Limnología, Universidad Nacional Autónoma de Mexico, 82040 Mazatlán, México. email:jasanchez@cmarl.unam.mx

[‖]Unidad Académica Mazatlán, Instituto de Ciencias del Mar y Limnología, Universidad Nacional Autónoma de Mexico, 82040 Mazatlán, México. email:caro@ola.icmyl.unam.mx

[**]Centro de Investigación en Matemáticas (CIMAT), Jalisco s/n, Valenciana, 36023 Guanajuato, Gto, Mexico. email: jac@cimat.mx

1 Introduction

Lead-210 (^{210}Pb) is a naturally occurring radioactive isotope that is part of the ^{238}U decay chain and is formed in both the atmosphere and sediments. ^{210}Pb forms in the atmosphere through the decay of radon-222 (^{222}Rn), a gas that is released from soil and rocks. ^{222}Rn decays quickly to ^{210}Pb, attaches to aerosols and other atmospheric particles and can gets deposit onto the earth's surface through dry and wet fallout. The constant fallout of ^{210}Pb is called flux, and the atmospheric ^{210}Pb is called excess or unsupported ^{210}Pb. Once on the ground, excess ^{210}Pb can be transported, entering the sediment and mixing with in situ formed ^{210}Pb (supported ^{210}Pb). As sediment accumulates over time, ^{210}Pb is constantly buried in the sediment column, providing a chronological record of sediment deposition. This isotope, with a half-life of 22.23 ± 0.12 years, is commonly used to date recently accumulated sediments (< 150 years) and has become increasingly popular in recent decades for palaeoecological and pollution studies aimed at evaluating human impacts on the environment (e.g., Mustaphi et al., 2019).

The accuracy of chronologies is critical in environmental studies to correctly assign dates to physical, chemical, geological, biological, and ecological changes. Unlike other dating techniques, such as radiocarbon dating, dating a single sediment layer using a single ^{210}Pb measurement is not possible. Instead, ^{210}Pb activity is measured at different depths along a core (e.g., lake, peatland, marine sediments). A ^{210}Pb-chronology can only be established under certain assumptions regarding the sedimentation process and when a suitable portion of the excess ^{210}Pb decay curve is measured. The analysis of a complete series (data set) of ^{210}Pb measurements must be carried out to obtain meaningful dates Aquino-López et al. (2018).

Several traditional data analysis models are available for dating recent sediments using ^{210}Pb. These include the Constant Initial Concentration (CIC) model, also known as Constant Activity (CA) (Goldberg, 1963; Robbins and Edgington, 1975), the Constant Flux : Constant Sedimentation (CF:CS) model (Crozaz et al., 1964), and the Constant Rate of Supply (CRS) model, also known as the Constant Flux model (CF) (Appleby and Oldfield, 1978; Robbins, 1978; Sanchez-Cabeza and Ruiz-Fernández, 2012). The CIC model assumes that sediments have a constant initial ^{210}Pb concentration, while both the CF:CS and CRS models assume a constant flux of ^{210}Pb. The CF:CS model also assumes a constant sedimentation rate. Of these models, the CRS model is the most popular, as it allows for estimating of variable mass accumulation rates (see Figure 1). However, the flexibility of the CRS model in terms of its assumptions comes at a cost, as it requires the measurement of a sufficient portion of the excess ^{210}Pb inventory, which is the total amount of excess ^{210}Pb in the sediment, or the use of interpolation/extrapolation in order to estimate the complete ^{210}Pb inventory in the sediment.

In recent years, the CRS model has undergone several modifications to enhance its accuracy and practicality (Binford, 1990; Appleby, 2001; Sanchez-Cabeza et al., 2014). These modifications can be classified into two categories: (1) improvements in the model's uncertainty quantification, and (2) adjustments to the model's application when additional information is available, such as the presence of external dating mark-

Figure 1: Frequency of ^{210}Pb dating models used in papers between 1964 and 2017. Data gathered by Mustaphi et al. (2019) from a literature review of 271 papers. The models include the CF:CS (Constant Flux - Constant Sedimentation; Robbins, 1978), CIC (Constant Initial Concentration) (Goldberg, 1963; Crozaz et al., 1964; Robbins, 1978) and CRS (Constant Rate of Supply; Appleby and Oldfield, 1978; Robbins, 1978).Y-axis: Relative frequency (%).

ers like ^{137}Cs profiles, laminated sediments, tephras, and contaminated layers that correspond to known sedimentary events (Appleby, 1998, 2001, 2008).

A recent inter-laboratory model comparison experiment (Barsanti et al., 2020)presented poor consistency of chronologies. Two measured ^{210}Pbdatasets were sent to 14 laboratories worldwide, with varying degrees of expertise. Each laboratory was asked to provide a chronology, given the same dataset. Each laboratory used its preferred model, in most cases the CRS model. This experiment resulted in a wide range of chronologies, independently of the model used, providing different results even when the same model and dataset were used. The authors reinforced the need to use independent time markers (independent dating sources) to validate and "anchor" the chronologies, as suggested previously by Smith (2001). This comparison experiment clearly and critically shows the impact of user decisions and expert adaptations/revisions on the resulting chronologies. In order to replicate and update any given chronology, documenting such user

decisions becomes extremely important, such as providing the raw data. However, raw data sets and user decisions are rarely reported.

In recent years, Aquino-López et al. (2018) introduced a Bayesian approach to ^{210}Pb dating, named the Plummodel, an alternative to the classical models. In this approach, every data point is considered to originate from a forward model that takes into account both the sedimentation process and the radioactive decay process. The Plum model also assumes a constant flux of excess ^{210}Pb to the sediment, similar to the CRS model, although this assumption can be relaxed at the expense of computational power. One important difference between the two models is that Plum includes the supported ^{210}Pb, a naturally occurring component in sediments usually treated as a hindrance variable (i.e., the assumed supported 210Pb values are removed before further modelling).

Blaauw et al. (2018) presented a comparison between the construction of classical and Bayesian age-depth models construction, both for real and simulated ^{14}C-dated cores. They concluded that Bayesian age-depth models provide a more accurate result with a more realistic uncertainties under a wide range of scenarios. The objective of the present study was to test whether similar results are maintained in a more complex modelling situation, such as the construction of ^{210}Pb-based age-depth models. To do so, we compared ^{210}Pb dates (accuracy) and uncertainties (precision) from the widely applied CRS model against Plum using simulated cores, i.e. sedimentation "scenarios". We also aimed to observe the learning process of each model and estimate the amount of information needed to obtain a reasonable chronology for each model. This process is critical as the amount of information depends on the number of samples, which depends on the budget, time, and user decision of resources allocated to developing the age-depth model.

The paper is organized as follows: Section 2 briefly introduces the CRS model which is the most widely used, as well as Plum. Section 3 discusses the consideration and the experimental setup. Section 4 presents the tools we use for the model comparison, describing the three different sedimentation scenario simulations. Section 5 compares results for the overall chronologies and for single depths. Lastly, Section 6 presents the conclusions and discussion of the results obtained in section 5.

2 ^{210}Pb Data and models

Methods used to estimate ages from ^{210}Pb sedimentary profiles are based on a range of assumptions, which may result in differing chronologies. Except for Plum, it is necessary to distinguish between the supported and the excess ^{210}Pb. By estimating the excess vs. the supported ^{210}Pb along the core, and using the radioactive exponential decay law, one can estimate the depth-age relationship of the core. Under steady-state conditions, ^{226}Ra and supported (in situ) ^{210}Pb are assumed to be in secular equilibrium as they are part of the same decay chain. Therefore, by measuring ^{226}Ra in sediments using gamma spectrometry, an indirect measurement of the supported ^{210}Pb can be obtained. However, as the excess and supported ^{210}Pb inputs into sediments are otherwise indistinguishable by laboratory measurements, a proper estimation of

the excess ^{210}Pb is critical. In this section we present the data usually used for the derivation of sediment chronologies and how the supported and excess ^{210}Pb are handled by the CRS and Plum models.

2.1 Data

We will provide an example of the type of data used to create ^{210}Pb age-depth models using the dataset presented in Sanchez-Cabeza and Ruiz-Fernández (2012) and displayed in Table 1. This dataset was obtained by analyzing ^{210}Po (Pollonium-210) alpha decay through alpha-particle spectrometry, assuming that ^{210}Pb and ^{210}Po are in secular equilibrium (see Sanchez-Cabeza and Ruiz-Fernández, 2012, for details). The deepest three samples were utilized to estimate the supported ^{210}Pb activity. Alternatively, gamma-ray spectrometry can be utilized to measure ^{226}Ra, which can be used as a proxy for the supported ^{210}Pb activity. Both techniques have their benefits and disadvantages. Gamma spectrometry provides measurements of ^{226}Ra that can be used to infer the supported ^{210}Pb, whereas alpha spectrometry provides more accurate measurements. Depending on the study's budget and the laboratory's possibilities, each study may use either technique or a combination.

2.2 CRS

The Constant Rate of Supply model (Appleby and Oldfield, 1978; Appleby, 1998, 2001, 2008), assumes that the sediment experienced a constant input of ^{210}Pb (influx of ^{210}Pb). In order to account for the other material deposited into the sediment, the model uses the following relationship

$$P_0^U(t)r(t) = \Phi, \tag{1}$$

where P_0^U is the initial level of excess ^{210}Pb of the sediment of age t, Φ is the influx of ^{210}Pb to the sediment and $r(t)$ is the sedimentation rate at that moment (how quickly the sediment accumulates). Using this relationship and the decay equation one can reach the following relationship,

$$\begin{aligned} A_z &= \int_z^\infty \rho(w)P^U(w)dw \\ &= \int_z^\infty \Phi e^{-\lambda t(w)} \tfrac{dt(w)}{dw} dw \\ &= \int_{t(z)}^\infty \Phi e^{-\lambda y} dy, \end{aligned} \tag{2}$$

where z is a given depth in the sediment, $\rho(z)$ is the sediments density at depth z, P^U is the excess ^{210}Pb, Φ is the influx of ^{210}Pb to the sediment and λ the decay constant of ^{210}Pb. From equation 2 and by defining $A_0 = \int_0^\infty \Phi e^{-\lambda y} dy$, it is possible to derivate at the following expression:

$$t(z) = \frac{1}{\lambda} \log\left(\frac{A_0}{A_z}\right). \tag{3}$$

The CRS model, equation 3, is used to estimate the sediment ages by performing numerical integration. To do this, users () first calculate a supported ^{210}Pb by subtracting the previously defined supported ^{210}Pb

i, ID	x_i, Depth (cm)	d_i, Density ($\frac{g}{cm^3}$)	y_i, ^{210}Pb ($\frac{Bq}{kg}$)	σ_i, sd(^{210}Pb) ($\frac{Bq}{kg}$)	δ, Thickness (cm)
TehuaII-01	1	1.072	112.5	5.8	1
TehuaII-02	2	0.973	108.4	5.7	1
TehuaII-03	3	1.121	102.4	5.4	1
TehuaII-04	4	1.732	103.4	5.4	1
TehuaII-05	5	1.264	92.9	5.0	1
TehuaII-06	6	1.135	86.6	4.8	1
TehuaII-07	7	2.086	70.3	3.9	1
TehuaII-08	8	1.211	51.0	3.0	1
TehuaII-09	9	1.339	45.7	2.8	1
TehuaII-10	10	2.199	43.6	2.6	1
TehuaII-11	11	1.397	39.7	2.4	1
TehuaII-12	12	1.280	34.2	2.1	1
TehuaII-13	13	1.516	28.0	1.8	1
TehuaII-14	14	1.456	23.9	1.5	1
TehuaII-15	15	1.421	20.5	1.4	1
.....					
TehuaII-16	16	1.443	17.1	1.3	1
TehuaII-17	17	0.452	14.4	1.0	1
TehuaII-18	18	0.630	15.7	1.0	1

Table 1: Data from the Gulf of Tehuantepec, south-eastern Mexico (TEHUAII) reported in Sanchez-Cabeza and Ruiz-Fernández (2012). Depth represents the lower depth of each sample section, density is the sample's density which is used to correct for compaction, ^{210}Pb is the measured ^{210}Pb in the given section, sd(^{210}Pb) is the error reported by the laboratory and thickness is the section's thickness. Each column also includes the corresponding notation used in this work.

from the measured ^{210}Pb and multiplying the result by the density. This results in a vector of values, which can be used to estimate A_0 and A_{z_i} and infer the ages of the bottom of each section. The actual calculation of these ages under different conditions is outside the scope of this paper and Sanchez-Cabeza and Ruiz-Fernández (2012) provide details on the proper use of the CRS model.

A_0 is a crucial amount for the CRS method, as this represents the total excess ^{210}Pb in the sediment. Therefore, excess ^{210}Pb must reach $\sim$0 $\frac{Bq}{kg}$ ("background"; $\frac{Bq}{kg}$ is the SI unit for radioactive materials), i.e. total 210Pb and supported 210Pb must reach equilibrium. Sediments cores that have not reached the background excess ^{210}Pb are not suited for dating using this method. Regardless, some adaptations to the CRS model attempt to infer the missing ^{210}Pb excess of the sediment by forcing the age-depth model to

pass through a known dating marker (a depth where its age is known from other methods or with a constant sedimentation rate Sanchez-Cabeza and Ruiz-Fernández, 2012). However, in this work, we assume that the background is reached in all examples, so the standard CRS model may be used.

2.3 Plum

Plum (Aquino-López et al., 2018) is an Bayesian model for producing ^{210}Pb age-depth models. This model is the first Bayesian method for dating ^{210}Pb sediments and is receiving growing interest from the palaeoecological community.Plum assumes that there exists an (unknown) age-depth function $t(x)$ that relates depth x with calendar age $t(x)$. Conditional on $t(x)$, the model is assumed for the measured ^{210}Pb (y_i) between depths $x_i - \delta$ to x_i

$$y_i \mid P_i^S, \Phi_i, t \sim \mathcal{N}\left(P_i^S \rho_i + \frac{\Phi_i}{\lambda}\left(e^{-\lambda t(x_i-\delta)} - e^{-\lambda t(x_i)}\right), (\sigma_i \rho_i)^2\right). \tag{4}$$

Here, the supported ^{210}Pb (P_i^S) and the influx of excess ^{210}Pb (Φ_i) in section i are considered as unknowns. The age-depth model $t(x)$ is assumed to follow a flexible semiparametric piece-wise linear model, which is constrained by prior information on sediment accumulation rates and their variability using a Gamma autoregressive model (Blaauw and Christen, 2011). The relationship between the measured data y_i (the measured ^{210}Pb) and the unknowns, i.e., the supported ^{210}Pb, excess ^{210}Pb, and the actual chronology, is explicitly modelled. This approach results in a likelihood, which is used to obtain a posterior distribution of $t(x)$ and the other parameters through Bayesian inference using Markov Chain Monte Carlo (MCMC). The resulting posterior distribution provides date estimates at each core depth. More technical details regarding Plum may be found in (Aquino-López et al., 2018).

This data treatment allows for a formal statistical inference on a well-defined model and differs from the CRS model, which is not a likelihood-based approach. The CRS model employs the radioactive exponential decay equation to generate an age-depth function, leading to a limited age-depth model that solely considers the excess ^{210}Pb. In our notation, this corresponds to assuming that P_i^S is a known quantity.

Figure 2 displays the chronologies obtained from the CRS and Plum models applied to the dataset presented in Table 1. Previous studies have shown that, under ideal conditions, both models produce comparable results (Aquino-López et al., 2020), with Plum providing more realistic uncertainties and requiring minimal user input. In this study, we aim to compare the performance of the two ^{210}Pb dating models under realistic conditions by using synthetic data with known chronologies. Our analysis aims to provide insights into the accuracy, uncertainty quantification, and asymptotic behavior of the resulting chronologies under varying sample sizes, shedding light on the strengths and limitations of both approaches.

Figure 2: Comparison between ages resulting from applying the Plum, CIC and CRS models to the dataset in Table 1 and Sanchez-Cabeza and Ruiz-Fernández (2012)

3 Model considerations and experiment setup

Since the CRS model has had several revisions, and the CRS version choice considerably affect model outputs (Barsanti et al., 2020), we decided to apply the original version provided by Appleby (2001), with its suggested error propagation calculation. We will call this version of the CRS model the "classical implementation of the CRS" (CI-CRS). While this implementation may be less suitable in some particular cases and expert knowledge can greatly improve the precision and accuracy of the CRS model, it reduces the bias of any particular implementation on our results.

Since the late 1970s, when the CRS model was first introduced (Appleby and Oldfield, 1978; Robbins, 1978), it has undergone several improvements. For example, Barsanti et al. (2020) showed some modifications and improvements that can generate a range of age-depth models. Some improvements rely on independent dates, other isotopes, or techniques, and require user manipulation to "force" the method to agree with this information. One recent improvement, which requires little user manipulation and independent dates, is the comprehensive explanation, with expert notes, on the practical use of the CRS model by Sanchez-Cabeza and Ruiz-Fernández (2012). They also presented an improvement to the uncertainty quantification of the age estimates by using the Monte Carlo method (Sanchez-Cabeza et al., 2014) and released a publicly available Excel spreadsheet, which facilitates the calculation of their age estimates and Monte Carlo uncertainties.

Considering that this paper focuses on methods with minimal user manipulation, and given that these modifications are laboratory-specific and not made publicly available, we also present and compare results using an R implementation (provided by the authors) of the improved CRS by Sanchez-Cabeza et al. (2014), here labelled as revised CRS (R-CRS).

4 Simulations (experiment setup)

In order to quantify the accuracy and precision of any chronology, a known true age-depth function is required. Blaauw et al. (2018) presented a methodology for simulating radiocarbon dates and their uncertainties, and Aquino-López et al. (2018) presented an approach for simulating ^{210}Pb data given an age-depth function $t(x)$. These simulations follow the equations presented by Appleby and Oldfield (1978); Robbins (1978) guaranteeing that the CRS model assumptions are met. By using the approach presented by Aquino-López et al. (2018) for simulating ^{210}Pb data and the structure of uncertainty quantification presented by Blaauw et al. (2018), realistic simulated ^{210}Pb data may be obtained.

The simulation study was used to generate three different complete data sets, with ^{210}Pb measurements at every depth along the hypothetical core, which then were then sampled. This dataset sampling mimics the sample selection that each laboratory, or user, does on a real core. The quantity of samples is decided by the resources available to each project (budget, time) Blaauw et al. (2018). In some cases, very few samples are selected to create an age-depth model.

4.1 Simulation construction

Three scenarios were chosen to simulate sedimentation processes with their age-depth functions and parameters. These scenarios were selected as they provide three key challenges for the models:

- Scenario 1 presents an age-depth function resulting from increasing sedimentation and less compaction towards the present (surface), a quite common scenario for more recent sediments.
- Scenario 2 presents a challenging core structure since the age-depth function has a drastic and rapid shift in sediment accumulation around 15 cm depth, representing a change in environmental conditions (e.g. a change in local land use).
- Scenario 3 presents a cyclic and periodic change in accumulation rates, representing cyclic changes in environmental conditions (e.g. El Niño-Southern Oscillation (ENSO)).

Using the age-depth functions and parameters defined in Table 4.1, we obtain the ^{210}Pb activity, or concentration, at any given depth interval (a_1, a_2), by integrating the activity curve between $(t(a_1), t(a_2))$. In the real world, this process in the field is performed by cutting the sediment core at different depths and then measuring the ^{210}Pb within them, so this integration reflects this process. The integrated concentration

maybe be interpreted as observing the true ^{210}Pb concentration within the depth interval (section) (see Figure 3). Since there will always be measurement errors in any method used to measure these concentrations, it is crucial to correctly recreate these errors in the simulated data.

Uncertainty for measured ^{210}Pb activity is recreated using an modified version of Blaauw et al. (2018), developed for simulating radiocarbon dates under realistic working conditions. This methodology was chosen as it introduces different sources of uncertainty related to different steps of the measurement process. Other error simulation methodologies could be used, but the comparison remains valid as long as the same measurement errors are provided to both models.

Label	Age-depth function $t(x)$	Influx of ^{210}Pb (Φ) ($\frac{Bq}{m^2 yr}$)	Supported ^{210}Pb ($\frac{Bq}{kg}$)
Scenario 1	$\frac{x^2}{4} + \frac{x}{2}$	100	10
Scenario 2	$12x - 0.2x^2$	50	25
Scenario 3	$8x + 25\sin(\frac{x}{\pi})$	500	15

Table 2: Simulated age-depth function and parameters used in each scenario

Figure 3: Simulated sedimentation scenarios with their corresponding ^{210}Pb profiles. Left: Age-depth functions for the three different scenarios (Table 4.1). Right: Corresponding ^{210}Pb activity profiles with depth.

Let C_i be the true ^{210}Pb concentration in the interval $[x_i - \delta, x_i)$, given the age-depth function $t(x)$ and parameters Φ_i (influx of ^{210}Pb) and P_i^S (supported ^{210}Pb) in each scenario. To simulate disturbances in the

material, we can introduce dispersion around the true value, $y_i' \sim \mathcal{N}\left(C_i, s_{scat}^2\right)$, where s^2 is the amount of dispersion around the true value, in this case $s^2 = 10$, which is similar to the levels proposed by Blaauw et al. (2018).

The occurrence of outliers is a crucial consideration when using these measures. These outliers may be described as an error in the measuring processes, which results in a shift in the mean y_i'. In order to replicate outliers, a random variable δ_{shift} is defined. This variable will be 0 with probability $1 - p_{out}$ and $\mathcal{U}_a$ with probability p_{out}, where $\mathcal{U}_a$ is a uniform distribution between $(-a, a)$. This process can be simulated using the following variable $\delta_{shift} \mid \beta \sim \beta\ \mathcal{U}_{x_{shift}}$, where β is a Bernoulli random variable with parameter p_{out}, and x_{shift} is the level of the outlier.

Finally, to simulate the data provided by the laboratory, y can defined as:

$$y_i \sim \mathcal{N}\left(C_i + \delta_{shift}, \sigma_R^2\right), \tag{5}$$

where σ_R is the standard deviation reported by the laboratory. σ_R is defined as $\sigma_R = \max\left(\sigma_{min}, \kappa\ C_i\ \varepsilon\right)$, where σ_{min} is the minimum standard deviation assigned to a measurement. As this variable differs between laboratories; we used a default value of 1 Bq/kg (Becquerel over kilogram which an standard unit for radioactive material). Finally, ε is the analytical measuring uncertainty (default 0.01 $\frac{Bq}{kg}$), and κ an error multiplier (default 1.5). The default parameters were set in following Blaauw et al. (2018).

For this study, we created a dataset for each simulation by integrating in intervals by δ =1 cm intervals for depths 0 to 30 cm, where radioactive equilibrium was guaranteed (Aquino-López et al., 2018). The complete simulated ^{210}Pb data sets can be found at `https://github.com/maquinolopez/Paper_Simulations/tree/master/Code/Data`.

4.2 Model considerations

In order to create a comparison with minimal user interaction, each model was run automatically, with default settings. Default settings for Plum are; 1 cm model sections, 10 $\frac{cm}{yr}$ mean prior accumulation rate, 50 $\frac{yrBq}{m^2}$ mean prior influx and 10 $\frac{Bq}{kg}$ mean prior supported ^{210}Pb. As the CRS model (for both the CI-CRS and R-CRS) assumes that supported and excess ^{210}Pb have reached equilibrium, in order to reduce user input, we decided to fix the last sample (30 cm depth) for every case, as this allows every model to reach equilibrium. This step guarantees the consistent application of the CRS model and provides the model with a single bottom-most depth to be removed as required by the CRS model calculation process. Furthermore, as the CRS model only works with the excess ^{210}Pb, when certain excess activities reached negative values, the chronology was calculated below that depth. Plum deals with the supported ^{210}Pb variable automatically, as part of the inference. Consequently, Plum's resulting chronology always reaches 30 cm, as by default 1 cm sections are used for every simulation.

In the case of the supported ^{210}Pb and to reduce the influence of this variable, a constant level of supported ^{210}Pb was assumed for both models, in agreement with the simulations construction. For the

CRS model, the mean of the supported ^{210}Pb measurements was calculated and then subtracted from the total ^{210}Pb to obtain the excess ^{210}Pb. In contrast to certain implementations that directly subtract ^{226}Ra from the total ^{210}Pb, we opted for a different approach in this study. Our decision was driven by the desire to minimize the influence of individual outliers on the Constant Rate of Supply (CRS) model. Instead, we determined that calculating the mean value of the supported ^{210}Pb samples would yield a more accurate estimation of this variable. By employing this method, the CRS model benefits from an improved estimate of the supported ^{210}Pb, which is less susceptible to the influence of individual outliers.

5 Model comparison

To allow for a reasonable comparison between models, and to evaluate the effect that varying amounts of information may have on the accuracy and precision of ^{210}Pb models (reflected in this study as the bias and coverage, respectively), the three simulated data sets were used. As the sample size is strongly dependentstrongly depends on a project's project's budget and time, we considered the use of a varying sample size. Ssamples from the entire core or a portion of the core could be measured or only a portion of it. In order to analyze the effects of sample sizes, samples of size m were randomly generated to provided an percentage of information percentage, e.g., for a 20% information, a dataset with 6 random 1-cm samples sections -out of a possible total 30 (maximum) 1-cm samples- is was created. This sample was then used to create the chronology and calculate the bias, length of estimated intervals, and coverage. One hundred : 100 of these sub-datasets were created for different information percentages (from 10% to 95% at 5% intervals). The complete dataset was also used (i.e., 100% information percentage sample, namely, a fully analyzed core). Once a dataset was created, the CRS model and Plum were applied. Both outputs were then compared against the true known age value, see Figure 4.

Figure 4 shows a single "snapshot", as an example of the comparison between the ^{210}Pb models against the true value in every scenario. As we are dealing with a total of n=5,333 simulations, in order to evaluate the overall precision and accuracy of both models, we calculated the mean bias of the true age-depth model (in yr), the mean of length of the 95% intervals (in yr), and the mean coverage indicating the distance of modelled ages from the true value given the model's model's own uncertainty at each depth.

Figure 5 shows similar results to those presented by Blaauw et al. (2018). The classical model (CI-CRS and R-CRS) at first appears to provide a similar results to Plum, with significant greater biasses regarles of the amount of information provided to the method. On the other hand, the classical approaches appear to provide higher estimated precision, if we only consider the length of the 95% interval. It is important to note that the CI-CRS and R-CRS models' biases do improve if more information is available, but are always greater than those provided by Plum. To have a more realistic representation of how the models capture the true age-depth relationship, we consider the coverage. This variable shows the degree to which the models contain the truth within their uncertainty intervals (normalized to one standard deviation). Any model

Figure 4: Comparison between Plum, R-CRS and CI-CRS models against the true age-depth function (from scenario 3) using 95% of the information percentage (using 1-cm sections). Lines show the age estimates with the 95% credible intervals (Plum) and the 95% confidence interval (CRS). Dots show the coverage, i.e. the distance between the inferred age and the true age in relation to the standard error (the standard deviation in the case of the CI-CRS and the length of the confidence interval divided by 4 in the case of Plum). The vertical right-hand axis shows how many standard deviations each model is from the true age.

Figure 5: Bias between the modelled and true age of the CI-CRS (red), R-CRS (green), and Plum (blue) as a function of the amount of 210Pb data (% of information). Plum provides a small bias in almost every scenario, with both models improving their bias as more information is available. B) 95% confidence intervals and credible intervals in the case of Plum. The uncertainty provided by Plum is significantly larger for low percentages of information, and it constantly improves as more data are available. In contrast, the intervals provided by the CI-CRS and R-CRS appear to not decrease regardless of the available information (though maximum interval lengths tend to reduce as more information becomes available). C) Coverage, presenting the distance between the modelled age and the true age divided by the standard deviation (in the case of Plum, the length of the 95% interval divided by 4). The CI-CRS and R-CRS model's calculated standard deviation (on average) cannot capture the true age (running at distances over 2 standard deviations from the truths). On the other hand, Plum's credible intervals almost always capture the true age, even when little information is available.

with coverage larger than two (two standard deviations) cannot capture the true ages within its uncertainty intervals. While the CI-CRS and R-CRS estimate smaller uncertainties (confidence intervals) and the ages improve as sample size increases, it does so at the cost of accuracy as the improvements are insufficient to capture the true age. It also appears that the length of the 95% interval and bias are not affected by how much information is provided to the CRS model.

On the other hand, Plum seems to provide increasingly accurate results as more information is provided. Again, this agrees with the results outlined by Blaauw et al. (2018). When we observe the regular bias (not normalized), we find that Plum provides a smaller bias in comparison to the CI-CRS and R-CRS models; this, in combination with slightly larger (more realistic) modelled uncertainties, results in more consistently accurate age-depth models that can capture the true values within their uncertainty intervals. In fact, 87.86% (4686/5333) of Plum's runs remain within the 2 standard deviations, opposed to 7.48% (399/5333) of the CI-CRS models runs. Furthermore, only 0.54% (29/5333) of the CI-CRS model runs remain under one standard deviation, the most commonly reported interval when reporting CI-CRS results, with R-CRS providing very similar results. This result supports the claim that Plum provides more realistic uncertainties compared those obtained by the CI-CRS and R-CRS. Furthermore, it is evident that Plum exhibits rapid improvement in its credible interval length as additional information becomes available, with a mostly constant coverage, guaranteeing that the true age-depth model will be contained in the credible intervals. On the other hand, the CI-CRS and R-CRS models do not show a notable enhancement in their coverage until at least 75% of the information is available. Even then, the length of their intervals is insufficient to capture the true ages within their 2 standard deviation confidence intervals. Given the similarity of results between the CI-CRS and R-CRS models, the subsequent analysis will primarily focus on the CRS model, using the CI-CRS as the reference point for subsequent calculations.

In order to evaluate whether certain models better predict ages at certain sections of the sediment cores, we look at the coverage of every depth. These results are valid for the overall chronology (mean bias, interval, and coverage of the overall chronology).

The coverage of each simulation at different depths is illustrated in Figure 6 for both models. Plum demonstrates a noticeable learning pattern, influenced by the amount of available information, although an unusual inverse pattern is observed in scenario 2. On the other hand, the coverage of the CI-CRS model appears to be minimally affected by the percentage of information, in contrast to the findings of Plum. The inaccuracies of the CI-CRS model are not confined to specific sections of the chronology, suggesting that this is likely due to the small uncertainties estimated by the CI-CRS model.

6 Discussion and Conclusions

This research focuses on exploring the uncertainty and precision of the most commonly used ^{210}Pb dating method (CRS) in contrast to Plum. By using different scenarios, three different simulations were created.

Figure 6: Coverage (proportion of times that a credible interval or confidence interval successfully captures the true value) of every core section for the three simulated scenarios - CI-CRS age estimates at sample depths, and Plum's age estimates at 1 cm intervals. Dots go from lowest information percentage samples (few dated depths; red) to high percentage samples (nearly completely dated cores; purple). The CI-CRS's coverage shows no learning pattern at any particular depth, regardless of the available information (apart from narrowing coverage at deeper depths). This means that the CRS model can provide a reasonable chronology with low levels of information or an inaccurate age estimate with high levels of information. On the other hand, Plum demonstrates a systematic improvement in its age estimates as more data are available. As a likelihood-based approach, these results show that Plum consistently provides more reliable results.

These simulations were then sub-sampled at different percentages of information to observe the effects of different sample sizes on the resulting chronologies. This experiment was designed to objectively compare the accuracy and precision of both methods.

The experiment was conducted on two levels. First, we evaluated the overall accuracy and precision of each method. The mean of the bias, length of the 95% confidence and credible intervals, as well as the coverage were measured. Second, we quantified the ability of each model to capture the true value in their credible/confidence interval (coverage) of each scenario at every depth. These two comparisons provided a good picture of the difference in precision and accuracy between these methods.

From the overall accuracy (see Figure 5), both the CRS model and Plum reduce their bias as more data becomes available, with the Bayesian method providing, on average, a smaller bias regardless of the sample size. In terms of precision, the Bayesian method provides much larger uncertainties when small sample sizes are used. It is only with 60%, or more of information that the length of the intervals becomes comparable, meaning that it takes Plum much more information to achieve the same length of the credible intervals provided by the classical approach. This is a consequence of the linear/exponential interpolation between data points used by the CRS method, in contrast to the Bayesian approach (Plum).

The larger uncertainties provided by Plum are more realistic, as confirmed in this work. Further evidence that these uncertainties are more reasonable is that the widths of Plum's credible intervals become smaller as more data becomes available. On the other hand, the length of the confidence intervals provided by the classical model (CRS) remain almost constant at any sample size, with reduced biases as more data is available. Lastly, the coverage, which shows the ability of the model to capture the true values within their intervals, shows that the classical model (CRS), on average is incapable of capturing the true values within its 95% confidence interval. These results are of concern, as the classical ^{210}Pb dating community rarely reports 95% confidence intervals and instead uses only 65% confidence intervals (one standard deviation intervals). On the other hand, Plum's coverages almost always remain ≤ 2, guaranteeing that on average the true value is captured within its 95% credible intervals, even with small sample sizes. Furthermore, Plum's coverages are constantly improving and reaching stability with 50% or more of the information percentage. These experiments show that the Bayesian method, on average, provides more reliable results for both precision and accuracy, no matter the amount of information.

As the coverage demonstrates how well each model can estimate the real value within its intervals, this variable may be used to assess if a certain approach offers a more accurate estimate for various periods. Figure 6 presents the performance of both the CRS model and Plum for every simulated scenario by depth. It appears that the coverage of many of the CRS chronologies do not follow a clear estructure, meaning that the model does not have a period of time for which it is more precise. Moreover, the CRS does not exhibit a clear learning pattern, where the coverage is not much affected by the amount of information available. Even high levels of information percentage provide coverages > 2, in some cases closer to 4 for scenarios 2 and 3. Plum on the other hand, shows a structure where more data are reflected in improved models in

scenarios 1 and 3. It is only at low levels of information where Plum's coverage is > 2. Scenario 2, on the other hand, presents a case where Plum cannot capture the true value, for depths > 15 cm, and it appears that as more data becomes available, the model provides worse results. We recognize that this scenario is unrealistic as it presents an extreme change in the accumulation around 15 cm, which coincides with the depth at which the coverage becomes > 2. However, it is also important to acknowledge that this experiment was performed using default settings. In a real-world scenario, the user typically has some prior knowledge of the sedimentation process in about the site of interest, which could be incorporated as prior information to the model to improve the resulting chronology for both the CRS and Plum models.

In conclusion, the use of the Bayesian age-depth models is recommended for the consistent construction of sediment chronologies, not only on radiocarbon-based chronologies as presented by Blaauw et al. (2018) but also in the more complex case of ^{210}Pb shown here. While the classical approach provides a reduction of the bias as more data is available, the uncertainty quantification in these methods is insufficient, which may be a results of not being based on a proper statistical structure. In a real-world scenario, it is impossible to measure the true bias of a method, and therefore a proper uncertainty quantification becomes extremely important. These results support the recommendations presented by Smith (2001); Barsanti et al. (2020), where the CRS method, or any dating methodology, should be validated using independent dating markers.

Both Blaauw et al. (2018) and the present work show that Bayesian methods constantly improve as more data are added, and the uncertainty associated with the the method is realistic and coherent with the amount of information available. This leads to chronologies that can capture the true age in their credible intervals, especially with minimal expert input (unlike the CRS, which relies on expert modifications and modifications). The ability to capture the true value within the credible intervals becomes important when the problem is associated with decision making processes, as it provides a more realistic picture of the available knowledge of the process. Given that ^{210}Pb-dating is now widely used in studies of pollution, environmental dynamics, and climate change, which potentially have a high impact on both policy-making and public perception, realistic age estimates and uncertainties, as well as estimates of rates of change over time, become extremely important.

7 Acknowledgments

The authors are partially funded by CONACYT CB-2016-01-284451 and COVID19 312772 grants and an RDCOMM grant. The corresponding author was funded by CONACYT through the postdoctoral residence program with CVU 489201.

References

Appleby, P. (2008). Three decades of dating recent sediments by fallout radionuclides: a review. *The Holocene*, 18(1):83–93.

Appleby, P. and Oldfield, F. (1978). The calculation of lead-210 dates assuming a constant rate of supply of unsupported 210pb to the sediment. *Catena*, 5(1):1–8.

Appleby, P. G. (1998). Dating recent sediments by Pb-210: Problems and solutions. *Proc. 2nd NKS/EKO-1 Seminar, Helsinki, 2-4 April 1997, STUK, Helsinki*, pages 7–24.

Appleby, P. G. (2001). Chronostratigraphic techniques in recent sediments. *Tracking Environmental Change Using Lake Sediments: Basin Analysis, Coring, and Chronological Techniques*, pages "171–203".

Aquino-López, M. A., Blaauw, M., Christen, J. A., and Sanderson, N. K. (2018). Bayesian Analysis of ^{210}Pb Dating. *Journal of Agricultural, Biological and Environmental Statistics*, 23(3):317–333.

Aquino-López, M. A., Ruiz-Fernández, A. C., Blaauw, M., and Sanchez-Cabeza, J.-A. (2020). Comparing classical and Bayesian ^{210}Pb dating models in human-impacted aquatic environments. *Quaternary Geochronology*, 60:101106.

Barsanti, M., Garcia-Tenorio, R., Schirone, A., Rozmaric, M., Ruiz-Fernández, A., Sanchez-Cabeza, J., Delbono, I., Conte, F., Godoy, J. D. O., Heijnis, H., Eriksson, M., Hatje, V., Laissaoui, A., Nguyen, H., Okuku, E., Al-Rousan, S. A., Uddin, S., Yii, M., and Osvath, I. (2020). Challenges and limitations of the ^{210}Pb sediment dating method: Results from an IAEA modelling interlaboratory comparison exercise. *Quaternary Geochronology*, 59:101093.

Binford, M. W. (1990). Calculation and uncertainty analysis of ^{210}Pb dates for PIRLA project lake sediment cores. *Journal of Paleolimnology*, 3:253–267.

Blaauw, M. and Christen, J. A. (2011). Flexible paleoclimate age-depth models using an autoregressive gamma process. *Bayesian Analysis*, 6(3):457–474.

Blaauw, M., Christen, J. A., Bennett, K., and Reimer, P. J. (2018). Double the dates and go for Bayes — impacts of model choice, dating density and quality on chronologies. *Quaternary Science Reviews*, 188:58–66.

Crozaz, G., Picciotto, E., and de Breuck, W. (1964). Antarctic snow chronology with pb210. *Journal of Geophysical Research*, 69(12):2597–2604.

Goldberg, E. D. (1963). Geochronology with Pb-210. *Radioactive Dating*, pages 121–131.

Mustaphi, C. J. C., Brahney, J., Aquino-López, M. A., Goring, S., Orton, K., Noronha, A., Czaplewski, J., Asena, Q., Paton, S., and Brushworth, J. P. (2019). Guidelines for reporting and archiving ^{210}Pb sediment chronologies to improve fidelity and extend data lifecycle. *Quaternary Geochronology*, 52:77–87.

Robbins, J. (1978). Geochemical and geophysical applications of radioactive lead. *The biogeochemistry of lead in the environment*, pages 285–393.

Robbins, J. A. and Edgington, D. (1975). Determination of recent sedimentation rates in lake michigan using pb-210 and cs-137. *Geochimica et Cosmochimica Acta*, 39(3):285 – 304.

Sanchez-Cabeza, J. A. and Ruiz-Fernández, A. C. (2012). ^{210}Pb sediment radiochronology: An integrated formulation and classification of dating models. *Geochimica et Cosmochimica Acta*, 82:183–200.

Sanchez-Cabeza, J. A., Ruiz-Fernández, A. C., Ontiveros-Cuadras, J. F., Pérez Bernal, L. H., and Olid, C. (2014). Monte Carlo uncertainty calculation of ^{210}Pb chronologies and accumulation rates of sediments and peat bogs. *Quaternary Geochronology*, 23:80–93.

Smith, J. N. (2001). Why should we believe ^{210}Pb sediment geochronologies? *Journal of Environmental Radioactivity*, 55(2):121 – 123.