

A simulation study to compare ^{210}Pb dating analyses

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

A simulation study to compare ^{210}Pb dating analyses

Abstract

The increasing interest in understanding anthropogenic impacts on the environment have led to a considerable number of studies focusing on sedimentary records for the last $\sim 100 - 200$ years. Dating this period is often complicated by the poor resolution and large errors associated with radiocarbon (^{14}C) ages, which is the most popular dating technique. To improve age-depth model resolution for the recent period, sediment dating with lead-210 (^{210}Pb) is widely used as it provides absolute and continuous dates for the last $\sim 100 - 150$ years. The ^{210}Pb dating method has traditionally relied on the Constant Rate of Supply (CRS, also known as Constant Flux - CF) model which uses the radioactive decay equation as an age-depth relationship resulting in a restrictive model to approximate dates. In this work, we compare the classical approach to ^{210}Pb dating (CRS) and its Bayesian alternative (*Plum*). To do so, we created simulated ^{210}Pb profiles following three different sedimentation processes, complying with the assumptions imposed by the CRS model, and analysed them using both approaches. Results indicate that the CRS model does not capture the true values even with a high dating resolution for the sediment, nor improves does its accuracy improve as more information is available. On the other hand, the Bayesian alternative (*Plum*) provides consistently more accurate results even with few samples, and its accuracy and precision constantly improves as more information is available.

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

*Centro de Investigación en Matemáticas (CIMAT), Jalisco s/n, Valenciana, 36023 Guanajuato, Gto, Mexico. 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 de Mazatlán, Instituto de Ciencias del Mar y Limnología, Universidad Nacional Autónoma de Mexico, 82040 Mazatlán, México email: jasanchez@cmar1.unam.mx

^{||}Unidad Académica de 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 radionuclide, part of the ^{238}U decay chain, which forms naturally in the atmosphere as well as in sediments. This isotope, with a half-life of 22.23 ± 0.12 years, is commonly used to date recent recently accumulated sediments (< 150 years). In recent decades, increasing number of palaeoecological and pollution studies have focused on these recent sediments (e.g., Mustaphi et al., 2019) in order to evaluate human impacts on the environment. Unlike to other dating techniques such as ^{14}C (radiocarbon dating), dating single sediment layers is not possible from a single measurement of ^{210}Pb ; it is only when a suitable portion of the excess ^{210}Pb (atmospheric ^{210}Pb) decay curve is measured, representing the total inventory of ^{210}Pb from atmospheric deposition and runoff, and when certain assumptions about the sedimentation process are met that a chronology can be established. These studies strongly rely on the accuracy of their chronologies in order to correctly assign dates to chemical, biological and ecological changes. That is, unlike other dating techniques, an analysis of a series (data set) of ^{210}Pb measurements must be carried out in order to obtain meaningful dates. Samples are taken along a core (e.g., lake, peatland, marine sediments) at different depths, from which ^{210}Pb activity is measured. The whole series of ^{210}Pb measurements need to be analysed in order to attempt to produce a coherent chronology, see Aquino-López et al. (2018).

A range of traditional data analyses, or so called “models”, are available for dating recent sediments using ^{210}Pb ; e.g. the Constant Initial Concentration (CIC, Goldberg, 1963), also known as Constant Activity (CA, Robbins and Edgington, 1975), the Constant Flux : Constant sedimentation (CF:CS, Crozaz et al., 1964) and the Constant Rate of Supply (CRS, Appleby and Oldfield, 1978; Robbins, 1978; Sanchez-Cabeza and Ruiz-Fernández, 2012) also known as the Constant Flux model (CF). The CRS model is by far the most popular (see Figure 1) and has the most flexible assumptions. Other, more restrictive, models such as CF:CS and CIC also require the assumption of a constant flux of ^{210}Pb , as well as additional assumptions of the sedimentation process, as well as that of a constant supply of ^{210}Pb . The flexibility of the CRS model in terms of its assumptions, comes at the cost of needing to measure a sufficient portion of the inventory or to use of interpolation/extrapolation in order to properly estimate the complete inventory of ^{210}Pb in the sediment. For details see the following section.

The CRS model has undergone several revisions in the last decade in order to improve its applicability and precision. There are two types of revisions to this model: (1) revisions to its uncertainty quantification (eg. Binford, 1990; Appleby, 2001; Sanchez-Cabeza et al., 2014) and (2) to its application where extra information is available, such as external independent dating markers (e.g. ^{137}Cs profiles), laminated sediments, tephras, contaminated layers (known sedimentary events) (eg. Appleby, 1998, 2001, 2008).

A recent inter-laboratory model comparison experiment (Barsanti et al., 2020) presented concerning results. Two measured ^{210}Pb data sets were sent to 14 laboratories around the world with varying degrees of expertise in the ^{210}Pb dating method. Each laboratory was asked to provide a chronology, given the same data. It is important to note that each laboratory applied their preferred model; in most cases the

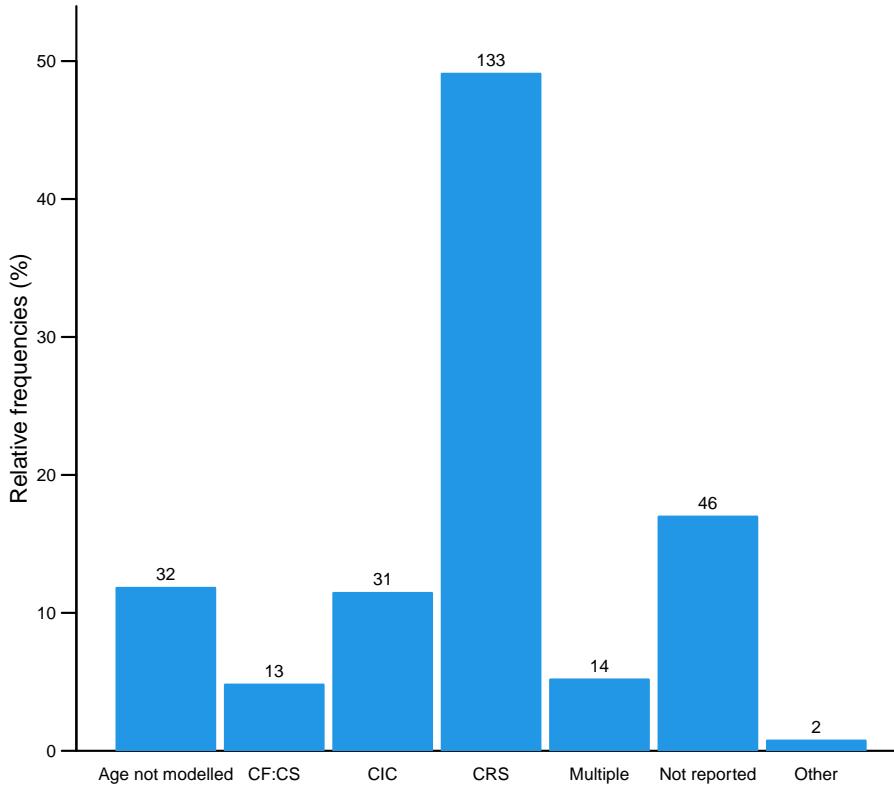


Figure 1: Frequency of ^{210}Pb dating models used in papers between 1964 and 2017. Data gathered by Mustaphi et al. (2019) from a literature review of 271 papers. The models include CF:CS model (The Constant Flux - Constant Sedimentation; Robbins, 1978), CIC (Constant Initial Concentration) (Goldberg,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

CRS model was calculated. This experiment resulted in a wide range of chronologies, independently of the model used, providing different chronologies even when the same model and dataset was used. The authors reinforced the need to use of independent time markers (independent dating sources) to validate and “anchor” of the chronologies, as suggested previously by (Smith, 2001). This comparison experiment clearly and critically shows the impact that user decisions and applying expert adaptations/revisions have on the resulting chronologies. In order to replicate and/or update any given chronology, such user decisions becomes extremely important. In addition, raw data sets are also required; unfortunately, both the raw data sets and/or user’s decisions are rarely reported.

Recently Aquino-López et al. (2018) presented an alternative to these classical models, by introducing *Plum*, a Bayesian approach to ^{210}Pb dating. This model treats every data point as originating from a forward model that includes both the sedimentation process and the radioactive decay process. *Plum* also assumes

a constant rate of supply to the sediment, similar to the CRS model (this assumption can be relaxed at the cost of computational power). Another important difference between the CRS and *Plum* is that the latter incorporates the supported ^{210}Pb , which naturally forms in the sediment and is normally threaded as a hindrance variable.

Blaauw et al. (2018) presented a comparison between 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 and more realistic uncertainties under a wide range of scenarios. The objective of the present study is to test whether the results obtained by Blaauw et al. (2018), concerning the accuracy and precision of the Bayesian approach, are maintained in a more complex modelling situation, such as the construction of ^{210}Pb -based age-depth models. To do so, we compare ^{210}Pb dates and uncertainties from the widely applied CRS model (by far the most popular age-depth model for ^{210}Pb) against *Plum* using simulated cores, i.e. sedimentation “scenarios”. We also aim to observe the learning process of each of the models and estimate the amount of information is needed to obtain a reasonable chronology for each model. This process is of crucial importance as the amount of information depends on the number of samples; which in turn depends on the budget and user decision of how much of such budget gets used to developing the age-depth model.

The paper is organized as follows: second section sets the tools we use for the model comparison, describing the simulations of the three different scenarios. Section 3 describes the comparison for both the overall chronologies and for single depths. Section 4 shows the impact of expert revisions. Lastly, Section 5 presents the conclusions and discussion of the results obtained in section 3.

2 ^{210}Pb Models

As previously mentioned there exist a series of methods to estimate ages from sediments containing ^{210}Pb . These methods are based on different assumptions to achieve different chronologies. It is important to note that all these models, excluding *Plum*, only deal with unsupported ^{210}Pb , which is atmospheric ^{210}Pb (also known as excess ^{210}Pb). Considering that almost any sediment contains some level of local ^{210}Pb (call supported ^{210}Pb), which is constantly replenished, and it is indistinguishable from the unsupported ^{210}Pb the proper calculation of the unsupported ^{210}Pb is critical. In this chapter, we will go into some details of each model.

2.1 CIC

The Constant Initial Concentration model (CIC) (Goldberg, 1963; Crozaz et al., 1964; Robbins, 1978), this model assumes that the initial concentration at any layer of the sediment is the same. With this assumption, we can define the initial concentration of any layer as $P_0^U = \text{Constant}$, by directly using the decay equation

the age of any depth as,

$$t_i = \frac{1}{\lambda} \log \left(\frac{P_0^U}{P_i^U} \right) \quad (1)$$

where λ the decay constant of the ^{210}Pb $\approx 0.03118 \pm 0.00017 \text{ yr}^{-1}$, P_0^U and P_i^U are the initial ^{210}Pb concentration and measured concentration of layer i respectively.

As we can observe this model does not take into account any fluctuation in the influx of ^{210}Pb to the sediment. This assumption is particularly strong as ^{210}Pb influx (Buscar una cita). As Figure 1 shows, this model is rarely used and its uses are mostly limited to early uses or uses where other models' requirements are not met.

An example of the use of this model is shown at the end of this chapter with a comparison to the other models. For details on the application of the model refer to Sanchez-Cabeza and Ruiz-Fernández (2012).

2.2 CF:CS

This model assumes both a constant initial concentration of ^{210}Pb at any layer but also a constant accumulation rate over the accumulated mass in the sediment.

$$t_i = -\frac{m_i b}{\lambda} \quad (2)$$

Where m_i is the accumulated dry mass of the sediment up to layer i , and b is approximated using linear regression over the following model,

$$\log(P_i^U) = \log(P_0^U) - \frac{\lambda}{r} m_i \quad (3)$$

$$= \hat{a} + \hat{b} m_i \quad (4)$$

It is important to note that this models even if more flexible it is still very restrictive, as it can be seen this model is even more rarely use in the practice and again, it only use when the CRS model cant not be used. For the proper implementation of this model, refer to Sanchez-Cabeza and Ruiz-Fernández (2012).

2.3 CRS

Up to this point we have deal with models which have extremely restrictive assumptions and simple extensions of the decay equation. The CRS (Appleby, 1998, 2001, 2008), in contrast, assumes a constant supply of atmospheric ^{210}Pb (also known as excess ^{210}Pb) to the sediment and allows for changes in the sedimentation rate. In order to allow to flexibility in the sedimentation rate the CRS model uses remaining activity. This variable is obtained by multiplying ^{210}Pb concentration and the density of the sediment.

This model also requires that the sediment is measured to the point where excess ^{210}Pb is no longer found, this is called reach equilibrium or background. In order to build a chronology, the CRS model uses a ratio between the complete “inventory” (the excess ^{210}Pb activity accumulated in the sediment column, between the surface and the equilibrium depth, where excess ^{210}Pb can no longer be found) and the remaining inventory from depth x to the previously defined equilibrium depth,

$$t(x) = \frac{1}{\lambda} \log \left(\frac{A_0}{A_x} \right), \quad (5)$$

where A_0 is the complete inventory, and A_x is the inventory up to depth x .

Once more details on the proper use of this and the previous models can be found in Sanchez-Cabeza and Ruiz-Fernández (2012).

2.4 Plum

Lastly, *Plum* is the most recent age depth model (Aquino2018). This model is the first Bayesian method for dating ^{210}Pb sediments and it has received a moderate interest by the palaeoecological community with 18 citations up to the moment of writing this manuscript. This is more than the citations obtained by the CF:CS model and in the literature review by Mustaphi et al. (2019).

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 following model is assumed for the measured ^{210}Pb y_i between depths $x_i - \delta$ to x_i

$$y_i | P_i^S, \Phi_i, \bar{t} \sim \mathcal{N} \left(A_i^S + \frac{\Phi_i}{\lambda} \left(e^{-\lambda t(x_i - \delta)} - e^{-\lambda t(x_i)} \right), (\sigma_i \rho_i)^2 \right). \quad (6)$$

Here A_i^S is the supported ^{210}Pb in the sample and Φ_i the supply of excess ^{210}Pb to the sediment, the age-depth model $t(x)$ is based on a piece-wise linear model constrained by prior information on the sediment’s accumulation rates (Blaauw and Christen, 2011), see Aquino-López et al. (2018) for details.

This treatment of the data allows for a formal statistical inference on a well-defined model with specific parameters. In order to infer the parameters of the model, a Bayesian approach is used. This differs from the CRS model, which does not provide a formal statistical inference. The CRS model uses the decay equation to obtain an age-depth function, resulting in a more restrictive age-depth model. It only deals with the excess ^{210}Pb , the estimated supported ^{210}Pb having been previously removed before modelling. *Plum* has shown to provide accurate results with a realistic precision using different case scenarios (Aquino-López et al., 2018, 2020) - both in simulations as well as for real cores. Under optimal dating conditions, *Plum* and the CRS model have been shown to provide similar results (Aquino-López et al., 2020), with *Plum* providing more realistic uncertainties, with minimal user interaction.

3 Model considerations and experiment setup

Given that the CRS model has had several revisions, the choice of which can considerably affect model outputs as shown by Barsanti et al. (2020), we decided to apply the original version of the equations provided by Appleby (2001), with its suggested error propagation calculation; we will call this version of the CRS the “classical implementation of the CRS” (CI-CRS). While we acknowledge that this implementation may be the less suitable in some particular cases and then expert knowledge can greatly improve the precision and accuracy of the model, but this will reduce the bias of any particular implementation has on our results.

Since the late 1970’s, when the CRS method was first introduced (Appleby and Oldfield, 1978; Robbins, 1978), the CRS has undergone several improvements. Some of these improvements rely on independent dates, other isotopes or techniques, and/or require user manipulation to “force” the method to agree with these independent dates. One recent improvement, which does not require user manipulation and/or independent dates, is the comprehensive explanation, with expert notes, on the practical used of the CRS model by Sanchez-Cabeza and Ruiz-Fernández (2012). The same authors presented an improvement to the uncertainty quantification of the age estimates by using the Monte Carlo method (Sanchez-Cabeza et al., 2014) and release a publicly available Excel spreadsheet, which facilitate the calculation of their age estimates and Monte Carlo uncertainties. Barsanti et al. (2020) showed that there exist several modifications and improvements to the CRS, and that the choice of modifications can generate a range of age-depth models. Considering that this research focuses on the methods with minimal user manipulation, and given that these modifications are laboratory-specific and not made publicly available, an R implementation (provided by the authors) of the improved CRS by Sanchez-Cabeza et al. (2014), here labelled as revised CRS (R-CRS)

The other models (CIC and CF:CS) were also used but the results show a big bias, as well as, an insufficient uncertainties to capture the true age-depth functions, as expected from such restrictive models. These models were performed using the serac R package (Bruel and Sabatier, 2020). These results can be observed in the appendix A of this paper.

4 Simulations (experiment setup)

In order to observe 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, while Aquino-López et al. (2018) presented an approach for simulating ^{210}Pb data given an age-depth function $t(x)$. It is important to note that these simulations follow the equations presented by Appleby and Oldfield (1978); Robbins (1978) guaranteeing that the CRS 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), reliable simulated ^{210}Pb data can be obtained.

This simulation will be used to generate three different data sets, which then later will be used to do

sampling, which will mimic the act of sampling from a real core, which each laboratory or ^{210}Pb user has to do. The quantity of samples is decided by the budget of each project, as explained by Blaauw et al. (2018). In some cases very few samples are selected creating an age-depth model.

4.1 Simulation Construction

Three different scenarios (see Table 4.1) were chosen to simulate sedimentation processes, with their own 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 which is the result of increasing sedimentation and less compaction towards the present (surface); this is quite common for recent sediments. Scenario 2 presents a challenging core structure as the function has a drastic and rapid shift in sediment accumulation around depth 15 cm depth. Lastly Scenario 3 presents a cyclic and periodic change in accumulation rates. Using the age-depth functions and parameters defined in Table 4.1, we obtain the ^{210}Pb activity, or concentration, at any given depth or interval, by integrating the age-depth curve for that interval. Although these concentrations may be interpreted as error-free measurements (see Figure 2), we replicated the ^{210}Pb activity uncertainty, following a similar methodology to Blaauw et al. (2018). This methodology was chosen as it introduces different sources of uncertainty related to different steps of the measurement process. Other uncertainty quantification methodologies could be used, but as long as the same methodology and uncertainty is provided to both models the comparison remains valid.

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

Table 1: Simulated age-depth function and parameters used in each scenario

Let $C_{\hat{x}}$ be the true ^{210}Pb concentration in the interval $\hat{x} = [x - \delta, x]$, given the age-depth function $t(x)$ and parameters Φ and A^S in each scenario. To simulate disturbances in the material, we can introduce scatter centred around the true value, $\theta \sim \mathcal{N}(C_{\hat{x}}, y_{scat}^2)$, where y_{scat}^2 is the amount of scatter for this variable (in this case $y_{scat}^2 = 10$). Now, to replicate outliers, a shift from the true value (x_{shift}) is defined, which occurs with a probability p_{out} . This results in a new variable θ' which is defined as

$$\theta' = \begin{cases} \mathcal{U}(\theta - x_{shift}, \theta + x_{shift}), & p_{out} \\ \theta, & 1 - p_{out} \end{cases}. \quad (7)$$

Finally, to simulate the uncertainty provided by the laboratory, we can define the simulated measurements as $y(\theta') \sim \mathcal{N}(\theta', \sigma_R^2)$, where σ_R is the standard deviation reported by the laboratory. σ_R is defined as

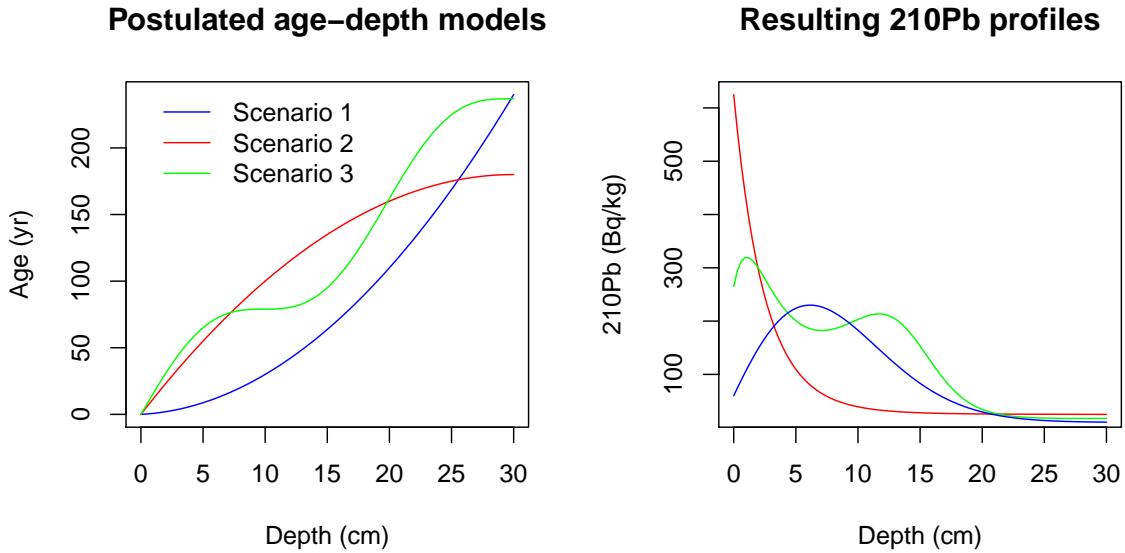


Figure 2: Simulated sedimentation scenarios with their corresponding ^{210}Pb profiles. Left: Age-depth functions for the three different scenarios (Table 4.1). Right: Corresponding ^{210}Pb activity profiles in relation to depth.

$\sigma_R = \max(\sigma_{min}, \mu(\theta') \varepsilon y_{scat})$, where σ_{min} is the minimum standard deviation assigned to a measurement. This variable differs between laboratories, we use a default value of 1 Bq/kg. Finally, ε is the analytical uncertainty (default .01) and y_{scat} an error multiplier (default 1.5). The default parameters were set in accordance with Blaauw et al. (2018).

For this study we created a data set for each of the three simulations by integrating in intervals of $\delta = 1$ cm, for depths from 0 to 30 cm where radioactive equilibrium was guaranteed (Aquino-López et al., 2018). The complete simulated ^{210}Pb data sets can be found in 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. In the case of *Plum*, default settings were used in order to minimize user interaction. As for the CRS model (for both the CI-CRS and R-CRS) assumes that background (supported) ^{210}Pb has been reached, in order to reduce user innervation, we decided to fix the last sample (30 cm depth) for every case. This step not only guarantees the consistent application of the CRS model, it also provides the model with a single bottom-most depth to be removed as it is common practice when using the CRS model. Conversely, as CRS model only models the excess ^{210}Pb (the total ^{210}Pb minus

the supported ^{210}Pb), when certain excess activities at depth fall below zero, the chronology will only be calculated up to that depth. *Plum* deals with this variable supported ^{210}Pb variable automatically, as part of the inference. Because of this, *Plum*'s resulting chronology will always reaches 30 cm, as by default 1 cm sections are used for every simulation.

In the case of the supported ^{210}Pb and in order to reduce the influence of this variable in the bias of the models, a constant level of supported ^{210}Pb was assumed for both models, which coincides with the way the simulation were constructed. 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 , as it is common practice when using the CRS model.

In order to provide an objective comparison, the bias of the true age-depth model (in yr), length of the 95% intervals (in yr) and coverage (bias over the standard deviation) were calculated (the coverage indicates the distance of modelled ages from the true value given the model's own uncertainty). The main discussion will revolve around the coverage as it provide an intuitive measure of the accuracy a model by taking into account the levels of uncertainty provided by each model.

5 Model Comparison

To allow for a reasonable comparison between models, and to evaluate the effect that different amount of information may have on the accuracy and precision of ^{210}Pb models, reflected in this study as the bias and coverage, the three simulated data sets previously described will be used. In order to reflect the effect of sample sizes, which depending on the budget of the project, the core could entirely sample or only a portion of it. A samples of size m is randomly generated provided a percentage of information, e.g. for a 20% information a dataset with 6 random 1-cm samples -out of a possible total 30 1-cm samples- is created. This sample was then used to create the chronology and calculate the bias, length of interval and coverage: 100 of these sub-datasets were created for different information percentages (from 10% to 95% at 5% intervals - i.e., 10%, 15%, 20%,...,95%). The complete dataset was also used (i.e 100% percentage of information sample, or fully sampled core). Once a dataset was created, both CRS models and *Plum* were applied (CIC and CF:CS were also applied and results shown on the appendix). Both sets of outputs were then compared against the true known age value, see Figure 3.

Figure 3 shows a single “snapshot” an example of the comparison between the ^{210}Pb models against the true value. As we are dealing with a total of $n = 5333$ simulations, in order to evaluate the overall precision and accuracy of both models, we decided to calculate the mean bias to the true age-depth model (in yr), the mean of length of the 95% intervals (in yr), as well as the mean coverage indicating the distance of modelled ages from the true value given the model's own uncertainty at each depth.

Figure 4 show results similar to those presented by Blaauw et al. (2018). The classical model (CI-CRS and R-CRS) at first appears to provide a similar results (similar biasses) to the Bayesian alternative (*Plum*),

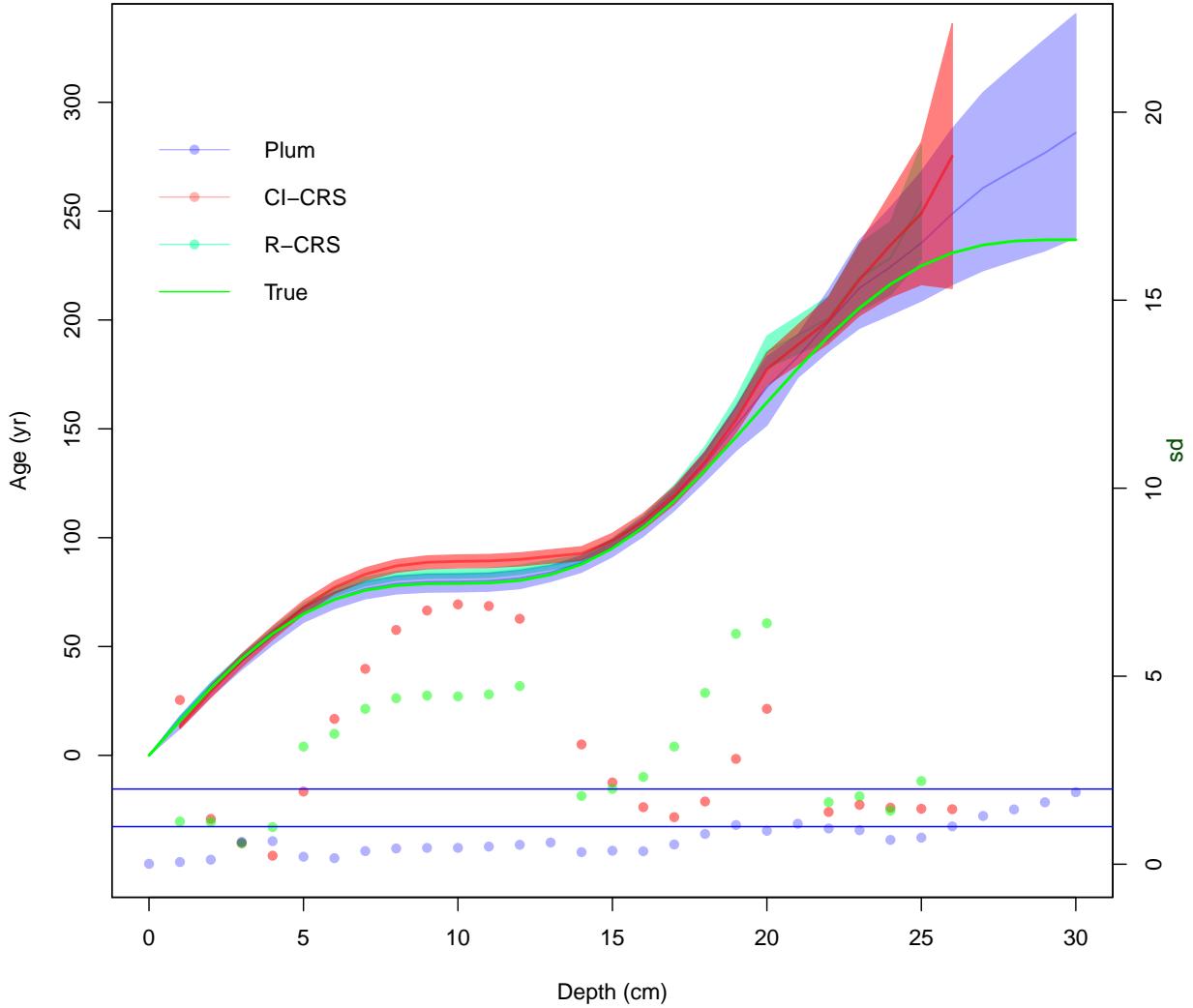


Figure 3: Comparison between *Plum*, R-CRS and CI-CRS model against the true age-depth model using 95% of the information percentage (using 1-cm samples). Lines show the age estimates with the 95% credible intervals (*Plum*) and the 95% confidence interval (CI-CRS). Dots show the coverage, i.e. the distance 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 axis shows how many standard deviations is each model from the true age.

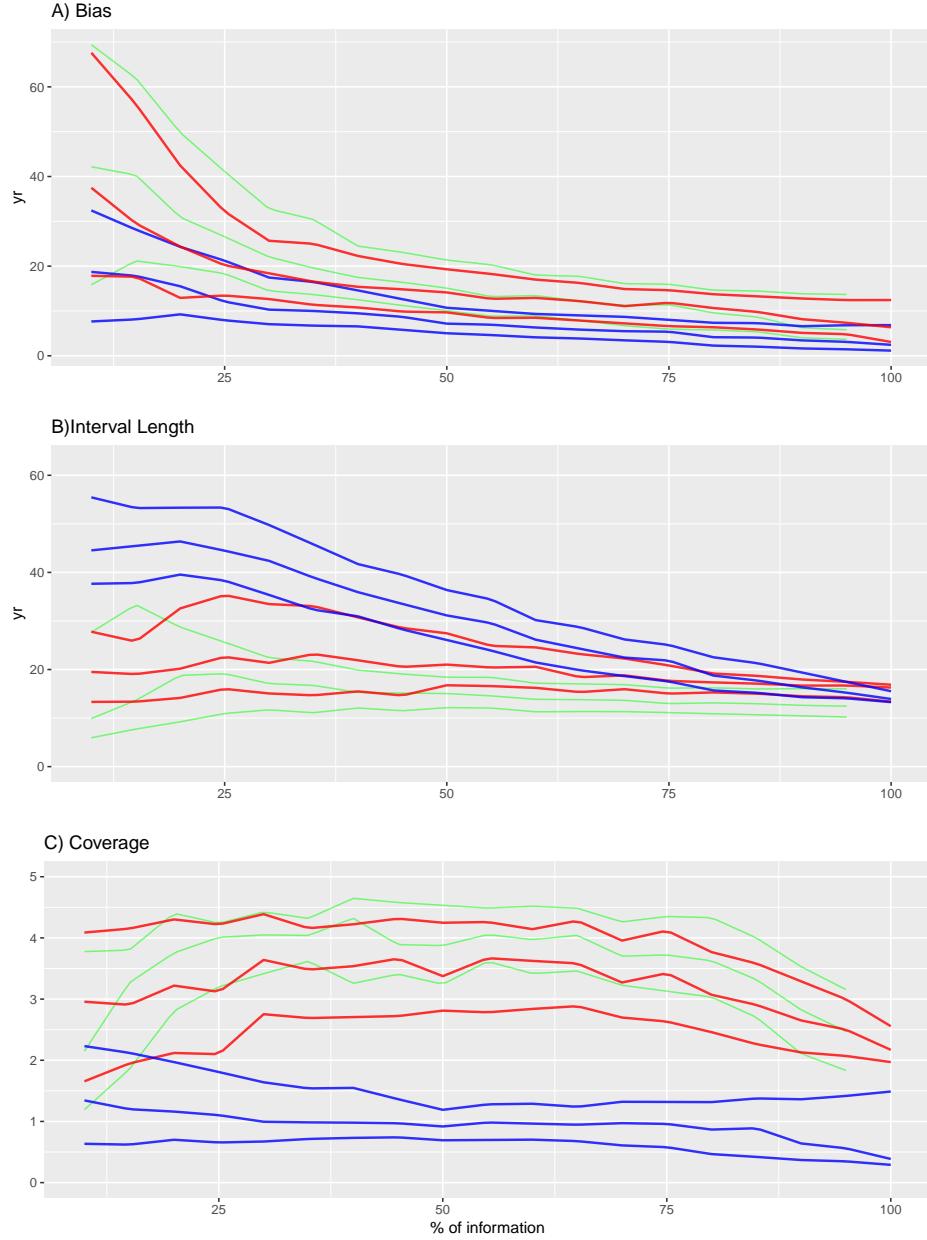


Figure 4: Top panel A) shows the bias between the modelled and true age of the CI-CRS (red), R-CRS (green) and *Plum* (blue). This panel shows how *Plum* provides a small bias in almost every scenario with both models improving their bias as more information is available. Middle panel B) shows the 95% confidence intervals. It is clear, from this panel, than the uncertainty provided by *Plum* is a lot bigger for low percentage of information and it constantly improves as more data is available, whereas the length of the intervals provided by the CI-CRS and R-CRS appear to stay constant regardless of the available information. Bottom panel C) shows the coverage, presenting the distance between the modelled age and the true age normalized divided by the standard deviation (in the case of *Plum*, the length of the 95% interval divided by 4). This panel presents a worrying situation where the CI-CRS and R-CRS model's calculated standard deviation (on average) is incapable of capturing the true age. On the other hand, *Plum*'s credible intervals almost always capture the true age even when little information is available.

but at higher estimated precision, if we only consider at the length of the 95% interval. It is important to note that the CI-CRS and R-CRS models' bias improves as more information is available. However, if we do not consider both the effects of both the bias and length of the interval together, the results are not favourable to the CI-CRS. To have a more realistic representation of how the models capture the true age-depth models relationship, we should observe the coverage. This variable shows the degree to which the average models contain the truth within their uncertainty intervals (normalized to one standard deviation). Any model with a coverage larger than two (two standard deviations) is incapable of capturing the true ages within its uncertainty intervals. This means that, while the CI-CRS and R-CRS estimates smaller uncertainties and its ages improve as more data is available, it does so at the cost of its accuracy and the improvements are not sufficient 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 added to the model. This again coincides 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 are capable of capturing the true values within their uncertainty intervals. This result supports the claim that *Plum* provides more realistic uncertainties compared those obtained by the CI-CRS and R-CRS. Another important statistic to take into account is that 87.86% (4686/5333) of *Plum*'s runs remain within the 2 standard deviations, opposed to 7.48% (399/5333) for the CI-CRS models. Furthermore, only 0.54% (29/5333) of the CI-CRS models runs remain under the 1 standard deviation, which is the most commonly reported interval when reporting CI-CRS results, with R-CRS providing very similar results. We can also observe a clear structure in the way that *Plum* increases its accuracy and precision to obtain a better chronology as more information is available, whereas the CI-CRS and R-CRS model does not appear to improve its ability to capture the true value from additional data, i.e it does not improve its coverage as more data is available. Since we obtained very similar results from both the CI-CRS and R-CRS, the discussion will now focus on the CRS in general taking the CI-CRS as its base for the following calculations.

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

Figure 5 shows the coverage of every simulation according to depth for both models. *Plum* shows a clear learning structure which depends on the information available to the model. The information percentage appears to be irrelevant to the coverage of the CI-CRS model, contrary to the results obtained by *Plum*. It is important to note that the inaccuracies of the CI-CRS model are not exclusive to any particular sections of the chronology; this is most likely driven by the small uncertainties estimated by the CI-CRS model. See below for a discussion of how *Plum* behaved in sedimentation simulation 2.

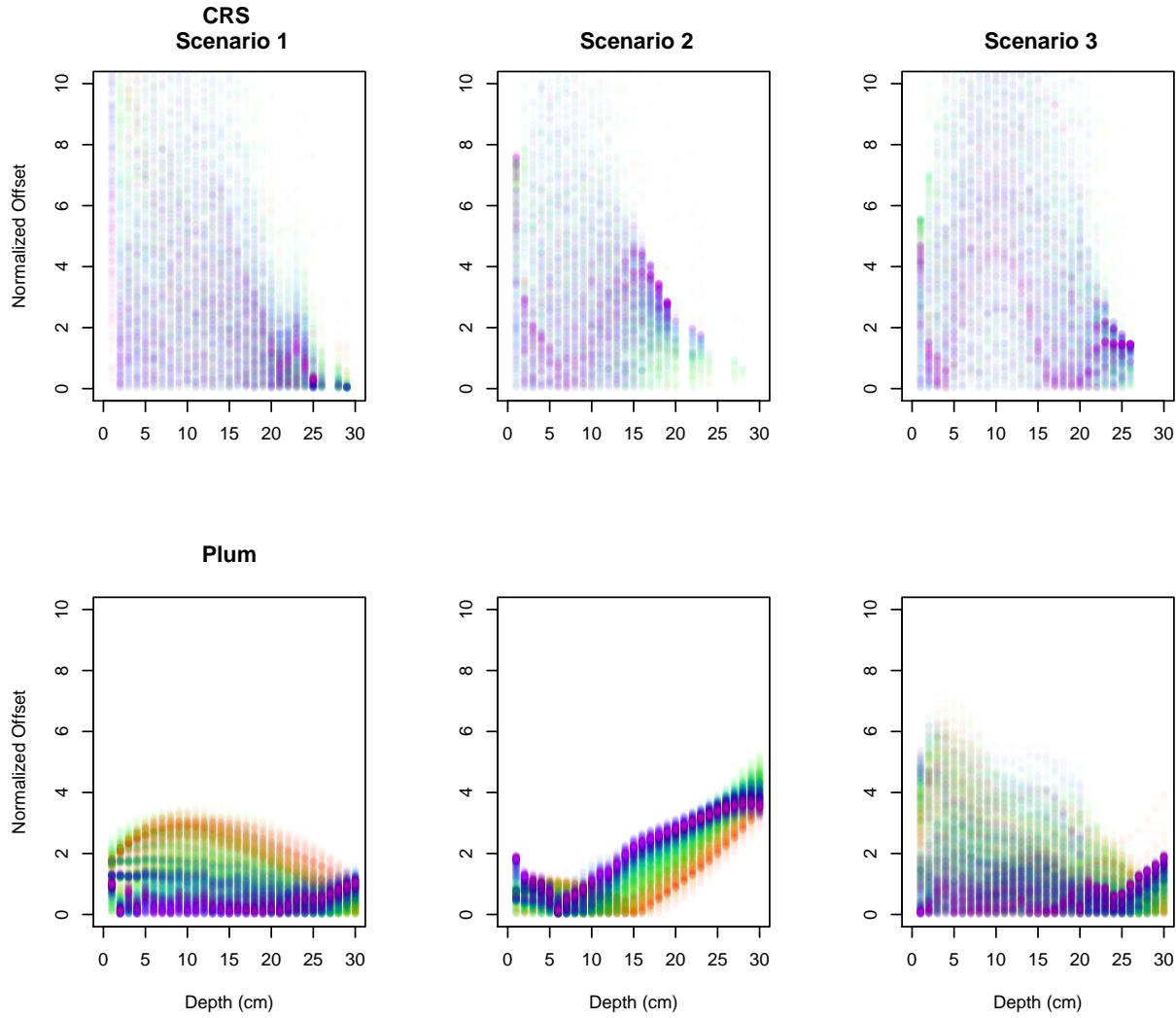


Figure 5: coverage of every sampling sample at every depth for the three simulated scenarios - CI-CRS age estimates at samples depths and *Plum*'s age estimates at 1 cm intervals. Dots go from lowest information percentage samples (few dated depths; red) to high percentage samples (nearly completely dated cores; purple). The CI-CRS's normalized bias shows no learning pattern at any particular depth regardless of the available information. This means that the model can provide a reasonable chronology with low levels of information or a very inaccurate age estimate with high levels of information at any given depth resulting in a distrustful age-depth model. On the other hand, *Plum* demonstrates a systematic improvement in its age estimates as more data is available. This results assures that a Bayesian approach would consistently provide more reliable results.

6 Discussion and Conclusions

This research focuses on exploring the uncertainty and precision of the most commonly used ^{210}Pb dating methods (CRS and *Plum*). By using different scenarios, three different simulations were created. These simulations were then sub-sampled at different percentages of information in order to observe the effects that different sample sizes have on the resulting chronology. This experiment provided an objective comparison of the accuracy and precision of both methods.

The experiment was measured in two different level. For the first level, the overall accuracy and precision of the method were evaluated. The mean of the bias, length of the 95% confidence and credible intervals, as well as the coverage were measured. The second level focused on the ability of each model to capture the true value in their credible/confidence interval, and the coverage of every scenario per depth was calculated. These two comparisons provided a good picture of the difference in precision and accuracy between these methods.

From the overall accuracy (see Figure 4) it is clear that 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. On the subject of precision, the Bayesian method is providing much larger uncertainties when small sample sizes are used. It is only with 60%, or more, of the information that the length of the intervals becomes comparable. This is a consequence of the linear interpolation, between data points, used by the CRS method, in contrast to the Bayesian approach (*Plum*) using a proper statistical inference. As has been previously discussed by Aquino-López et al. (2020), the larger uncertainties provided by *Plum* are more realistic, and this experiment confirms the latter. Further evidence that these uncertainties are more sensible is that the length of the credible intervals becomes 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. Lastly, the coverage, which shows the capability 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 especially worrisome considering that the ^{210}Pb dating community rarely report 95% confidence intervals and instead tend to use only 65% confidence intervals (one standard deviation intervals) are reported. On the other hand, *Plum*'s coverages always remain ≤ 2 , therefore guaranteeing that on average the true value is capture within its 95% credible intervals, even with small sample sizes. *Plum*'s coverages are constantly improving and reaching stability with 50%, or more, of information percentage. These experiments show that the Bayesian method, on average, provides more reliable results.

Because the coverage shows the capability of capturing the true value within its intervals, this variable can be used to conclude if any given method is better at estimating certain time period. Figure 5 presents the performance of both the CRS model and *Plum* for every simulated scenario. It appears that, the coverage of many of the CRS chronologies are > 2 throughout the whole chronology, meaning that the model does

not have a period of time for which it is more precise. Moreover, the CRS does not exhibits a clear learning pattern, where the coverage appears to be indifferent to the amount of information available. It appears that 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 is 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* is both incapable of capturing the true value, for depths deeper than 15 cm, and it appears that as more data becomes available the model provides worse results. This may be of concern if we do not 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 . 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, about the site of interest, which could be incorporated as prior information to the model to improve the resulting chronology.

In conclusion, the use of the Bayesian age-depth models is preferred 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 as demonstrated by this research. While the classical approach provides reasonable results, regarding the bias, unfortunately the uncertainty quantification in these methods needs improvements as they do not rely 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 methods markers.

Lastly, it is important to highlight the benefits of the Bayesian methods. From both Blaauw et al. (2018) and the present work, it is shown that Bayesian methods constantly improve as more data are added, the uncertainty associated to the method is realistic and coherent with the amount of information available. This leads to chronologies that are capable of capturing the true age in their credible intervals. The ability to capture the true value in 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 pollution, environmental and climate change studies, which potentially have a high impact on both policy making and public perception, realistic age estimates and uncertainties become extremely important.

7 Acknowledgments

The authors are partially founded by CONACYT CB-2016-01-284451 and COVID19 312772 grants and a RDCOMM grant. The corresponding author is founded 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 ^{210}pb 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.
- Bruel, R. and Sabatier, P. (2020). serac: an r package for ShortlivEd RAdionuclide chronology of recent sediment cores. *Journal of Environmental Radioactivity*, 225:106449.
- 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 210pb 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). 210Pb 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 210Pb chronologies and accumulation rates of sediments and peat bogs. *Quaternary Geochronology*, 23:80–93.
- Smith, J. N. (2001). Why should we believe 210pb sediment geochronologies? *Journal of Environmental Radioactivity*, 55(2):121 – 123.

8 Supplementary Material

Data for each simulation and code used are hosted at: https://github.com/maquinolopez/Paper_Simulations

9 Appendix A

Bias, length of 95% confidence intervals and coverage of the CIC and CF:CS models. As it can be observe and as expected the bias of the CIC model (which can be consider the simples model) is much bigger when compare to any other alternative and it does not decrease as more data is available. The confidence interval remains regarles of the information available, and coverage is almost always above two, which means that the uncertainties are not sufficient to capture the true age-depth model. On the other hand the CF:CS model does decrease its bias as more data is available and after 60% to 70% this decrease does not seen improve. The length of the interval remains the same regarles of the information available and the coverage just like the CIC model

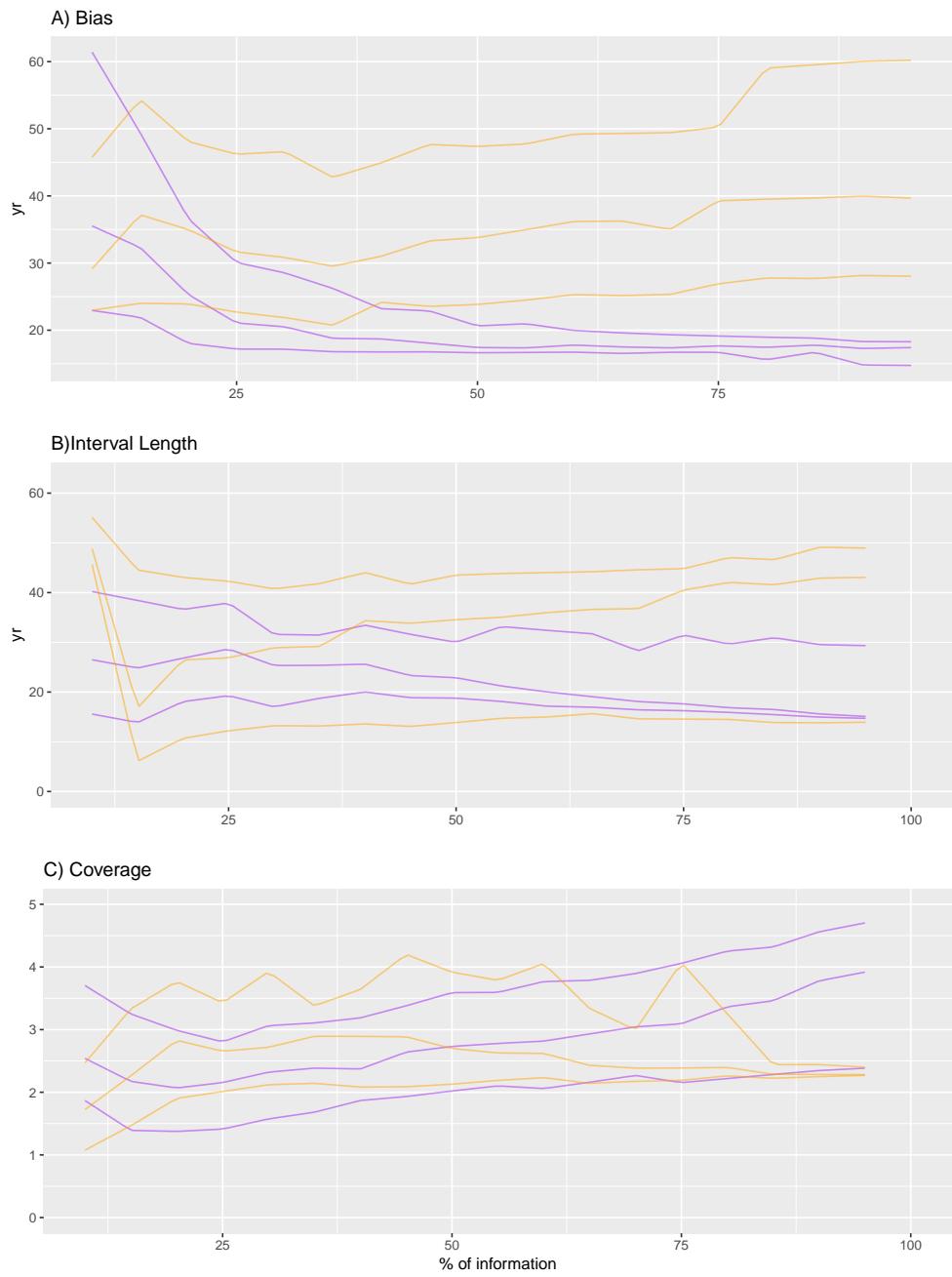


Figure 6: Top panel A) shows the bias between the modelled and true age of the CF:CS (purple) and CIC (Orange). Middle panel B) shows the 95% confidence intervals. Bottom panel C) shows the coverage, presenting the distance between the modelled age and the true age normalized divided by the standard deviation.