Evaluation of Pb dating models using simulated datasets

Marco A Aquino-López[[1]](#footnote-1) [[2]](#footnote-2)

Nicole K. Sanderson[[3]](#footnote-3)

Maarten Blaauw[[4]](#footnote-4)

Joan-Albert Sanchez-Cabeza[[5]](#footnote-5)

Ana Carolina Ruiz-Fernandez[[6]](#footnote-6)

J Andrés Christen[[7]](#footnote-7)

Abstract

The study of anthropogenic impacts on the environment can be addressed through dated sedimentary records for the last 100-200 years. During this period, radiocarbon (C) dating suffers from poor resolution and large uncertainties. To overcome this limitation, 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 Pb dating (CRS) with *Plum*, a Bayesian approach developed for analyzing sedimentary Pb measurements. Simulated 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.

1

1

**Evaluation of Pb dating models using simulated datasets**

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

# Introduction

Lead-210 (Pb) is a naturally occurring radioactive isotope that is part of the U decay chain and is formed in both the atmosphere and sediments. Pb forms in the atmosphere through the decay of 222Rn, a gas released from soil and rocks. Radon-222 decays quickly to Pb, attaches to aerosols and other atmospheric particles and can deposit onto the earth's surface through dry and wet fallout. The constant fallout of Pb is called flux, and the atmospheric Pb is called excess Pb. Once on the ground, excess Pb can be transported, entering the sediment and mixing with in situ formed Pb (supported Pb). As sediment accumulates over time, Pb is continuosly buried into 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 ( years) and has become increasingly popular in recent decades for palaeoecological and pollution studies aimed at evaluating human impacts on the environment [e.g.,@Courtney2019].

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 Pb measurement is not possible. Instead, Pb activity is measured at different depths along a core (e.g., lake, peatland, marine sediments). A Pb-chronology can only be established under certain assumptions regarding the sedimentation process and when a suitable portion of the excess Pb decay curve is measured. The analysis of a complete series (data set) of Pb measurements must be carried out to obtain meaningful dates [@Aquino2018].

Several traditional data analysis models are available for dating recent sediments using Pb. These include the Constant Initial Concentration (CIC) model, also known as Constant Activity (CA) [@Goldberg1963; @Robbins1975], the Constant Flux : Constant Sedimentation (CF:CS) model [@Crozaz1964], and the Constant Rate of Supply (CRS) model, also known as the Constant Flux model (CF) [@Appleby1978; @Robbins1978; @Sanchez-Cabeza2012]. The CIC model assumes that sediments have a constant initial Pb concentration, while both the CF:CS and CRS models assume a constant flux of 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 variable mass accumulation rates (see Figure [1](#fig:210models)). 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 Pb inventory, which is the total amount of excess Pb in the sediment, or the use of interpolation/extrapolation in order to estimate the complete 210Pb inventory in the sediment.

|  |
| --- |
|  |

Figure 1: Frequency of Pb dating models used in papers between 1964 and 2017. Data gathered by [@Courtney2019] from a literature review of 271 papers. The models include the CF:CS [Constant Flux - Constant Sedimentation;@Robbins1978], CIC (Constant Initial Concentration) [@Goldberg1963; @Crozaz1964; @Robbins1978], and CRS [Constant Rate of Supply;@Appleby1978; @Robbins1978].

In recent years, the CRS model has undergone several modifications to enhance its accuracy and practicality [@Binford1990; @Appleby2001; @Sanchez-Cabeza2014]. 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 markers like Cs profiles, laminated sediments, tephras, and contaminated layers that correspond to known sedimentary events [@Appleby1998; @Appleby2001; @Appleby2008].

A recent inter-laboratory model comparison experiment [@Barsanti2020] presented poor consistency of chronologies. Two measured Pb datasets 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 [@Smith2001]. 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, [@Aquino2018] introduced a Bayesian approach to Pb dating, named the *Plum* model, an alternative to the classical models. In this approach, every data point is considered to originate from a forward model that considers both the sedimentation process and the radioactive decay process. The *Plum* model also assumes a constant flux of excess 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 the *Plum* model includes the supported 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).

[@Blaauw2018] presented a comparison between the construction of classical and Bayesian age-depth models, both for real and simulated C-dated cores. They concluded that Bayesian age-depth models provide a more accurate result and 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 Pb-based age-depth models. To do so, we compared 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.

# Pb Data and models

Methods used to estimate ages from 210Pb 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* Pb. By estimating the excess vs. the supported 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, Ra and supported (*in situ*) Pb are assumed to be in secular equilibrium as they are part of the same decay chain. Therefore, by measuring Ra in sediments using gamma spectrometry, an indirect measurement of the supported Pb can be obtained. However, as the excess and supported Pb inputs into sediments are otherwise indistinguishable by laboratory measurements, a proper estimation of the excess Pb is critical. In this section we present the data usually used for the derivation of sediment chronologies and how the supported and excess Pb are handled by the CRS and *Plum* models.

## Data

We provide an example of the type of data used to create Pb age-depth models using the dataset presented in [@Sanchez-Cabeza2012] and displayed in Table [1](#tab:tehuaii). This dataset was obtained by analyzing Po (Polonium-210) alpha decay through alpha-particle spectrometry, assuming that Pb and Po are in secular equilibrium [see@Sanchez-Cabeza2012 for details]. The deepest three samples were utilized to estimate the supported Pb activity. Alternatively, gamma-ray spectrometry can be utilized to measure Ra, which can be used as a proxy for the supported Pb activity. Both techniques have their benefits and disadvantages. Gamma spectrometry provides measurements of Ra that can be used to infer the supported 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.

Data from the Gulf of Tehuantepec, south-eastern Mexico (TEHUAII) reported in [@Sanchez-Cabeza2012]. Depth represents the lower depth of each sample section, density is the sample's density which is used to correct for compaction, Pb is the measured Pb in the given section, sd(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.

| , ID | , Depth () | , Density () | , Pb () | , sd(Pb) () | , Thickness () |
| --- | --- | --- | --- | --- | --- |
| TehuaII-01 | 1 | 1.071583866 | 112.5 | 5.8 | 1 |
| TehuaII-02 | 2 | 0.973213378 | 108.4 | 5.7 | 1 |
| TehuaII-03 | 3 | 1.121380264 | 102.4 | 5.4 | 1 |
| TehuaII-04 | 4 | 1.732484316 | 103.4 | 5.4 | 1 |
| TehuaII-05 | 5 | 1.263766643 | 92.9 | 5.0 | 1 |
| TehuaII-06 | 6 | 1.135424096 | 86.6 | 4.8 | 1 |
| TehuaII-07 | 7 | 2.085680966 | 70.3 | 3.9 | 1 |
| TehuaII-08 | 8 | 1.211092723 | 51.0 | 3.0 | 1 |
| TehuaII-09 | 9 | 1.339040564 | 45.7 | 2.8 | 1 |
| TehuaII-10 | 10 | 2.199381257 | 43.6 | 2.6 | 1 |
| TehuaII-11 | 11 | 1.397469527 | 39.7 | 2.4 | 1 |
| TehuaII-12 | 12 | 1.280204165 | 34.2 | 2.1 | 1 |
| TehuaII-13 | 13 | 1.516059058 | 28 | 1.8 | 1 |
| TehuaII-14 | 14 | 1.456445983 | 23.9 | 1.5 | 1 |
| TehuaII-15 | 15 | 1.42113905 | 20.5 | 1.4 | 1 |
| TehuaII-16 | 16 | 1.443497137 | 17.1 | 1.3 | 1 |
| TehuaII-17 | 17 | 0.451885447 | 14.4 | 1.0 | 1 |
| TehuaII-18 | 18 | 0.630431828 | 15.7 | 1.0 | 1 |

## CRS

The Constant Rate of Supply model [@Appleby1978; @Appleby1998; @Appleby2001; @Appleby2008] assumes that the sediment experienced a constant input of Pb (influx of Pb). In order to account for the other material deposited into the sediment, the model uses the following relationship:

where is the initial level of excess Pb of the sediment of age , is the influx of Pb to the sediment and 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:

where is a given depth in the sediment, is the sediment density at depth , is the excess Pb, is the influx of Pb to the sediment and the decay constant of Pb. From equation [[eq:sampleeqC2]](#eq:sampleeqC2) and by defining , it is possible to derive the following expression:

The CRS model [[eq:CRS]](#eq:CRS), is used to estimate the sediment ages by performing numerical integration. To do this, users first calculate a supported Pb by subtracting the previously defined supported Pb from the measured Pb and multiplying the result by the density. This results in a vector of values, which can be used to estimate and 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-Cabeza2012] provide details on the proper use of the CRS model.

is a crucial amount for the CRS method, as this represents the total excess Pb in the sediment. Therefore, excess 210Pb must reach ~0 Bq kg-1 (“background”; Bq kg-1 is the SI unit for radioactive materials), i.e. total 210Pb and supported 210Pb must reach equilibrium. Sediment cores that have not reached the background excess 210Pb are not suited for dating using this method. Some adaptations to the CRS model attempt to infer the excess Pb in the bottom missing portion 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-Cabeza2012]. In this work, we assume that the background is reached in all examples, so the standard CRS model may be used.

## *Plum*

Plum [@Aquino2018] is an alternative model for producing 210Pb age-depth models. This is the first Bayesian method for dating Pb sediments and is receiving growing interest from the palaeoecological community. *Plum* assumes that there exists an (unknown) age-depth function that relates depth with calendar age . Conditional on , the model is assumed for the measured Pb () between depths to :

Here, the supported Pb () and the influx of excess Pb () in section are considered as unknowns. The age-depth model 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 [@Blaauw2011]. The relationship between the measured data (measured Pb) and the unknowns, i.e., the supported and excess Pb, and the actual chronology, is explicitly modeled. This approach results in a likelihood, which is used to obtain a posterior distribution of 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 details regarding *Plum* may be found in [@Aquino2018].

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 Pb. In our notation, this corresponds to assuming that is a known quantity.

Figure 2: Comparison between ages resulting from applying the Plum, CIC, CRS, and CF:CS models to the dataset in Table 1 and [@Sanchez-Cabeza2012]

Figure [2](#fig:tehuaii) displays the chronologies obtained from the CRS and *Plum* models applied to the dataset presented in Table [1](#tab:tehuaii). Previous studies have shown that, under ideal conditions, both models produce comparable results [@Aquino2020], with *Plum* providing more realistic uncertainties and requiring minimal user input. In this study, we compare the performance of the two Pb dating models under realistic conditions by using synthetic data with known chronologies. Our analysis aim 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.

# Model considerations and experiment setup

Since the CRS model has had several revisions, and the version may considerably affect model outputs [@Barsanti2020], we decided to apply the original version provided by [@Appleby2001] with its suggested error propagation calculation. We call this version the "classical implementation of the CRS" (CI-CRS). While this implementation may be less suitable in some particular cases and expert knowledge can 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 [@Appleby1978; @Robbins1978], it has undergone several improvements. For example, [@Barsanti2020] 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-Cabeza2012]. They also presented an improvement to the uncertainty quantification of the age estimates using the Monte Carlo method [@Sanchez-Cabeza2014] 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-Cabeza2014], here labelled as revised CRS (R-CRS).

# Simulations (experiment setup)

In order to quantify the accuracy and precision of any chronology, a known true age-depth function is required. [@Blaauw2018] presented a methodology for simulating radiocarbon dates and their uncertainties, and [@Aquino2018] presented an approach for simulating Pb data given an age-depth function . These simulations follow the equations presented by [@Appleby1978; @Robbins1978], guaranteeing that the CRS model assumptions are met. By using the approach presented by [@Aquino2018] for simulating Pb data and the structure of uncertainty quantification presented by [@Blaauw2018], realistic simulated Pb data may be obtained.

The simulation study was used to generate three different complete data sets, with Pb measurements at every depth along the hypothetical core, which 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) [@Blaauw2018]. In some cases, very few samples are selected to create an age-depth model.

## 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., ENSO cycles).

Using the age-depth functions and parameters defined in Table [2](#tab:sim_param), we obtain the Pb activity, or concentration, at any given depth interval , by integrating the activity curve between . In the real world, this process in the field is performed by cutting the sediment core at different depths and then measuring the Pb within them, so integration reflects this process. The integrated concentration may be interpreted as observing the true Pb concentration in the section (see Figure [3](#fig:true_210)).

Uncertainty for measured Pb activity is recreated using a modified version of [@Blaauw2018], 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.

Simulated age-depth function and parameters used in each scenario

|  |  |  |  |
| --- | --- | --- | --- |
| Label | Age-depth | Influx of Pb | Supported Pb |
|  | function |  | () |
| Scenario 1 |  | 100 | 10 |
| Scenario 2 |  | 50 | 25 |
| Scenario 3 |  | 500 | 15 |

Figure 3: Simulated sedimentation scenarios with their corresponding Pb profiles. Left: Age-depth functions for the three different scenarios (Table [2](#tab:sim_param)). Right: Corresponding Pb activity profiles with depth.

Let be the true Pb concentration in the interval , given the age-depth function and parameters (influx of Pb) and (supported Pb) in each scenario. To simulate disturbances in the material, we can introduce dispersion around the true value, , where is the amount of dispersion around the true value, in this case , similar to the levels proposed by [@Blaauw2018]. Furthermore, we replicated the presence of measurement errors (a shift in the mean ) with a random variable . This variable will be with probability and with probability , where is a uniform distribution between . This process can be simulated using the variable , where is a Bernoulli random variable with parameter , and is the level of the error.

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

where is the standard deviation reported by the laboratory. is defined as , where is the minimum standard deviation assigned to a measurement. As this variable differs between laboratories, we used a default value of . Finally, is the analytical measuring uncertainty (default 0.01 Bq kg-1), and an error multiplier (default 1.5). The default parameters were set following [@Blaauw2018].

For this study, we created a dataset for each simulation by integrating by 1 cm intervals for depths from 0 to 30 cm, where radioactive equilibrium was guaranteed [@Aquino2018]. The complete simulated Pb data sets can be found at <https://github.com/maquinolopez/Paper_Simulations/tree/master/Code/Data>.

## 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 cm yr-1mean prior accumulation rate, 50 Bq yr-1 m-2 mean prior influx and 10 bq yr-1 mean prior supported Pb. As the CRS model assumes that supported and excess 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 Pb, when excess activities reached negative values the chronology was calculated below that depth. *Plum* deals with the supported 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 Pb and to reduce the influence of this variable, a constant level of supported Pb was assumed for both models, in agreement with the simulations construction. For the CRS model, the mean of the supported Pb measurements was calculated and then subtracted from the total Pb to obtain the excess Pb. This practice differs for some implementations where the Ra is directly subtracted from the total Pb. ; in this study, we decided that to reduce the impact of single outliers in the CRS model, the mean would provide a much better estimate of this variable.

# 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 Pb models (reflected in this study as the bias and coverage, respectively), the three simulated data sets were used. As the sample size strongly depends on a project's budget and time, we considered the use of a varying sample size. Samples of size were randomly generated to provide an information percentage, e.g., for a 20% information, a dataset with 6 random sections out of 30 (maximum) was created. This sample was then used to create the chronology and calculate the bias, length of estimated intervals, and coverage. One hundred 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](#fig:comparison1r).

Figure 4: Comparison between Plum, R-CRS, and CI-CRS models against the true age-depth function 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 [4](#fig:comparison1r) shows a single "snapshot" as an example of the comparison between the Pb models against the true value. As we are dealing with a total of simulations, to evaluate the overall precision and accuracy of both models we calculated the mean bias of the true age-depth model (yr), the mean length of the 95% intervals (yr), and the mean coverage X indicating the distance of modelled ages from the true value given the model's uncertainty at each depth.

Figure 5: A) 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 decrease slowly 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.

Figure [5](#fig:accpre) shows similar results to those presented by [@Blaauw2018]. Initially, the classical CRS model provides similar results (similar biases) to *Plum* but with higher estimated precision if we only consider the length of the 95% interval. Both the CI-CRS and R-CRS models' biases improve if more information is available. However, if we do not consider 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 relationship, we also 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 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 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 [@Blaauw2018]. When we observe the regular bias (not normalized), we find that *Plum* provides a smaller bias in comparison to the classical 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 model 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. These results confirm that *Plum* provides more realistic uncertainties than those obtained by the classical CRS models. We also observe that *Plum* increases its accuracy and precision to produce a better chronology as more information is available, whereas the CI-CRS and R-CRS models only improve slightly their coverage with additional data. As we obtained 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.

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).

|  |
| --- |
|  |

Figure 6: Coverage 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.

Figure [6](#fig:depths) shows the coverage of every simulation according to depth for both models. *Plum* shows a clear learning structure that depends on the model's available information. The information percentage appears to little affect the CI-CRS model coverage, contrary to the results obtained by *Plum*. 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.

# Discussion and Conclusions

This research focuses on exploring the uncertainty and precision of the most commonly used Pb dating methods (CRS, CIC, and CF:CS) in contrast to *Plum*. By using different scenarios, three different simulations were created. 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, and coverage (I think we need a more descriptive term, as the same term is used later for a different thing) were measured. Second, we quantified the ability of each model to capture the true value in their credible/confidence interval and the coverage of each scenario according. These two comparisons provided a good picture of the difference in precision and accuracy between these methods.

From the overall accuracy (see Figure [5](#fig:accpre)), 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. Only with 60% or more of information do the widths of the confidence intervals become comparable. 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 [@Aquino2020], 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) remains almost constant at any sample size. 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, cannot capture the true values within its 95% confidence interval. These results are of concern, as the classical Pb dating community rarely reports 95% confidence intervals and instead uses only 68% (one standard deviation interval). On the other hand, *Plum*'s coverages always remain , guaranteeing that 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](#fig:depths) presents the performance of both the CRS model and *Plum* for every simulated scenario. It appears that the bias (?) of many of the CRS chronologies is throughout the whole chronology, meaning that the model does not have a period 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 provide coverages , 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 bias is . Scenario 2, on the other hand, presents a case where *Plum* cannot capture the true value for depths 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 bias becomes . 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 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 [@Blaauw2018] but also in the more complex case of Pb shown here. While the classical approach provides good results regarding the bias, the uncertainty quantification in these methods needs improvement as it does 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 proper uncertainty quantification becomes extremely important. These results support the recommendations presented by [@Smith2001; @Barsanti2020], where the CRS method, or any dating methodology, should be validated using independent dating markers.

Both [@Blaauw2018] and the present work show that Bayesian methods constantly improve as more data are added, and the uncertainty associated with 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 decisions). 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 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.

# Acknowledgments

The authors are partially funded by CONACYT CB-2016-01-284451 and COVID-19 312772 grants and an RDCOMM grant. The corresponding author was funded by CONACYT through the postdoctoral residence program with CVU 489201.

1. Department of Geography, University of Cambridge, Cambridge, United Kingdom email: aquino@cimat.mx [↑](#footnote-ref-1)
2. Corresponding author. [↑](#footnote-ref-2)
3. GEOTOP Research Centre, Université du Québec à Montréal, Montréal, Québec, H2X 3Y7, Canada. email: sanderson.nicole@uqam.ca [↑](#footnote-ref-3)
4. School of Natural and Built Environment, Queen’s University Belfast, Belfast, BT7 1NN, UK. email:maarten.blaauw@qub.ac.uk [↑](#footnote-ref-4)
5. Unidad Académica Mazatlán, Instituto de Ciencias del Mar y Limnología, Universidad Nacional Autónoma de Mexico, 82040 Mazatlán, México. email:jasanchez@cmarl.unam.mx [↑](#footnote-ref-5)
6. Unidad Académica Mazatlán, Instituto de Ciencias del Mar y Limnología, Universidad Nacional Autónoma de Mexico, 82040 Mazatlán, México. email:caro@ola.icmyl.unam.mx [↑](#footnote-ref-6)
7. Centro de Investigación en Matemáticas (CIMAT), Jalisco s/n, Valenciana, 36023 Guanajuato, Gto, Mexico. email: jac@cimat.mx [↑](#footnote-ref-7)