Article:
Selzer, S, Annett, AL and Homoky, WB orcid.org/0000-0002-9562-8591 (2021) RaDeCC reader: Fast, accurate and automated data processing for Radium Delayed Coincidence Counting systems. Computers & Geosciences, 149. 104699. ISSN 0098-3004

https://doi.org/10.1016/j.cageo.2021.104699

RaDeCC Reader: Fast, accurate and automated data processing for Radium Delayed Coincidence Counting systems

Sean Selzer[1]*, Amber L. Annett[2], William B. Homoky[1,3]

1. Department of Earth Sciences, University of Oxford, South Parks Road, Oxford, OX1 3AN, UK
2. Ocean and Earth Science, University of Southampton, Waterfront Campus, National Oceanography Centre, European Way, Southampton, SO14 3ZH, UK
3. Present address: School of Earth and Environment, University of Leeds, Leeds, LS2 9JT, UK

*sean.selzer@earth.ox.ac.uk

Code Availability:

RaDeCC Reader program and supporting files can be found on GitHub

Github Repository: (https://github.com/oxradreader/RaDeCC_Reader/releases)

Authorship Statement:

S.S. conceived and wrote the RaDeCC Reader program and A.L.A. and W.B.H. contributed to its design, testing and implementation. A.L.A. provided real sample data files and S.S. carried out the validation experiments. S.S. prepared the manuscript, with edits and contributions throughout from A.L.A. and W.B.H.

Abstract

A Python program is presented to expedite the process of correcting raw data and propagating the related uncertainties from Radium Delayed Coincidence Counting (RaDeCC) instruments. The performance of the program was validated against an established method with real data. Excellent agreement between determinations of excess radium-223, actinium-227, excess radium-224, thorium-228 and radium-226 was achieved, with minor discrepancies in the

results attributed to logical improvements in our implementation. The RaDeCC Reader program is able to process one thousand data files in only a few minutes, and thereby offer distinct advantages in the processing speed combined with reliable accuracy of data processing implementations.

Keywords: Data Processing; Software Engineering; Data Assimilation; Environmental Science; Hydrogeology;

1. Introduction

Radium is a valuable tracer for environmental geochemistry due to the conservative nature of radium in seawater and the predictable rates of decay of its isotopes. Disequilibria between these isotopes can allow the quantification of rates of exchange between natural reservoirs (Cochran, 1982). The development of Radium Delayed Coincidence Counting (RaDeCC) systems has made radium-based studies in aqueous environments more feasible (Moore and Arnold, 1996). For example, radium isotopes are increasingly used to trace, quantify and advance understanding of many fundamental ocean processes in coastal (Moore, 2000; Tamborski et al., 2020), shelf sea (Hendry et al., 2019), open ocean surface (Charette et al., 2007) and deep water settings (Kipp et al., 2018).

To measure the activities of radium-223 and radium-224 in aqueous environments, sample water is commonly pumped through manganese oxide impregnated acrylic or polyethylene fibres (Moore, 1976). These fibres extract radium, its parent isotopes thorium and actinium, and other species with high affinity for MnO_2, from the water via binding to the MnO_2 functional groups present on the fibres. The precise activities of radium isotopes on these fibres can be determined by counting their daughter isotopes radon and polonium using a scintillation counting technique that is optimally performed by the RaDeCC apparatus (https://www.radecc.com) (Moore and Arnold, 1996).

The RaDeCC system of delayed coincidence counting was originally devised by Giffin *et al.*, (1963) and forms the basis of the RaDeCC apparatus devised by Moore and Arnold, (1996). The RaDeCC apparatus measures the activities of radon isotopes – the nuclides produced from radium decay - emanating from sample fibres over the course of a counting period, herein termed "read". The flow of helium through a closed loop carries this radon between the sample fibre container and the scintillation cell. Radon decay in the scintillation cell produces an alpha particle which is detected, generating a signal which is routed to three channels: total counts, radon-219 and radon-220. The total counts channel records a count when any signal is received. In the radon-219 and radon-220 channels the system looks for a second count, corresponding to the subsequent decay of daughters polonium-215 and polonium-216 (respectively) after a short delay for the signal to stabilise: 0.01 ms for the radon-219 channel and 5.61 ms for the radon-220 channel (Moore and Cai, 2013). After these delays a gate is opened in each channel (5.6 ms and 600 ms for radon-219 and radon-220, respectively; Moore and Arnold, 1996) in which an additional signal of alpha decay is required in order to register a count. During a read the RaDeCC software logs the counts and accumulated counts per minute for each channel at regular user-defined intervals to a text file.

Factors that need to be corrected for in the raw output include: interference between the detector channels for radon-219 and -220 and chance coincidence events; the counting efficiency and background (blank) of each detector; decay that occurred between sampling and measurement; rescaling sample activities to their original sample volumes (Giffin et al., 1963; Moore and Arnold, 1996). The expressions used to propagate uncertainties associated with these corrections were derived by Garcia-Solsona *et al.*, (2008).

The amount of radium parent isotope on the MnO_2 coated fibres determines the rate of production of the radon isotope daughter, and therefore the activity sustained in the flow of helium through a closed loop between sample and the RaDeCC system. The decay of actinium and thorium on the fibres supplements the amount of 'excess' radium-223 and radium-224 that is initially present (creating supported activity) leading to the activities of radon-219 and radon-220 initially measured by the RaDeCC system. These supported activities must be accounted for to accurately determine the excess, or unsupported, activities of radium-223 and radium-224. Finally, there is the ingrowth of radon-222 from its long-lived parent isotope, radium-226, recorded by the total channel. Determining the rate of radon-222 ingrowth can be used to estimate the activity of radium-226 (Geibert et al., 2013).

To perform the necessary raw data correction and uncertainty propagation calculations, many workers construct large Excel spread sheets and individually import their saved read file outputs from RaDeCC apparatus. Although this process allows a very granular view of the raw data and can serve its purpose well, it remains time intensive and large sets of data are susceptible to user-error. A faster, user-defined automation that preserves details of the calculation processes could therefore offer significant improvements to data processing speed and the reliability of outputs. Herein, we present our approach to expedite the process of correcting raw data and propagating the related uncertainties from Radium Delayed Coincidence Counting (RaDeCC) instruments using a newly designed program, RaDeCC Reader. We prove the validity of our new method by comparing results obtained with RaDeCC Reader to those we obtained by a previously established method using real data collected from karstic spring-, coastal- and open-ocean water samples.

2. Theory

2.1 Calculation of excess radium-223 and radium-224 activities

To convert raw decay counting statistics into the activity of radium-223 or radium-224 of a sample, a number of factors must be considered and corrected for. A table of variables and their units is included for reference (Table 1). Uncertainties in the raw counts must also be propagated through each of these corrections to determine uncertainties in final calculated activities.

An erroneously registered count due to chance coincidence events (*Y CC*, in cpm) is the first correction to be made. An erroneous count can be made when a decay event that is unrelated to the isotope of interest occurs while the detector-gate for that isotope's channel is open. These can originate from the background activity in the detectors or the decay of radon-222 while the 219 or 220 channels are open. The counts per minute (cpm) attributed to chance coincidence events are subtracted from the count rate in the relevant channel. Expressions to calculate the fraction of chance coincidence events in each channel (Equations 1,2) were derived by Giffin *et al.* (1963) and were included by Garcia-Solsona *et al.* (2008), where *cpm total, cpm219* and *cpm220* are the counts per minute in the total, 220 and 219 counting channels respectively.

$$Y\ 220\ CC\ \frac{(cpm\ total - cpm220 - cpm219)^2 \times 0.01}{1 - [(cpm\ total - cpm220 - cpm219) \times 0.01]} \qquad (1)$$

$$Y\ 219\ CC = \frac{(cpm\ total - corr220 - cpm219)^2 \times 0.000093}{1 - [(cpm\ total - corr220 - cpm219) \times 0.000093]} \qquad (2)$$

These chance coincidence events are then subtracted from the counts per minute in the relevant channel to determine the coincidence corrected counts (*corr220*, *corr219*).

$$corr220 = cpm220 - Y\ 220\ CC \qquad (3)$$

$$corr219 = cpm219 - Y\ 219\ CC \qquad (4)$$

In certain circumstances the decays associated with radon-219 can be erroneously registered in the 220-channel. This can happen if two atoms of radon-219 decay within the time that the 220 channel is open. Radon-220 can also cause interference in the radon-219 channel since the gate for this channel is open for enough time that 2.55% of radon-220 decay events occur while the gate is open. Expressions to account for these cross-channel interferences were devised by Giffin et al. (1963) and adapted by Moore and Arnold (1996).

$$final\ 220 = corr220 - \frac{(1.6 \times corr219)^2 \times 0.01}{1+[(1.6 \times corr219) \times 0.01]} \quad (5)$$

$$final\ 219 = corr219 - (corr220\ \times\ 0.0255) \quad (6)$$

In addition, background measurements may be run with MnO_2-coated fibres that were not used for sampling, assessing any counts due to contamination on fibre or within the RaDeCC apparatus itself, although the need for a background correction varies with sample type and application. Where required, the background count rate (in cpm) in each channel is averaged over multiple reads for each detector. The averaged background count rate from the applicable detector and channel is then subtracted from *final220* and *final219* before detector efficiencies are accounted for (*bkgcorr224, bkgcorr223* respectively; Equations 7,8).

$$bkgcorr224 = final220 - \ Average\_bkg\_220 \quad (7)$$

$$bkgcorr223 = final219 - \ Average\_bkg\_219 \quad (8)$$

The detection efficiencies, *E219* and *E220*, are evaluated by measuring the activities of standards with a known amount of radium-223 or radium-224 adsorbed to their fibres and comparing these measured activities (in cpm) to their known activities (in dpm) after corrections for decay since manufacture (Equations 9,10). These standards are made by adsorbing known activities of thorium-232 or actinium-227 in secular equilibrium with their daughter isotopes, radium-224 and radium-223 respectively.

$$E220 = \frac{final220\ (standard)}{thorium-232} \quad (9)$$

$$E219 = \frac{final219\ (standard)}{actinium-227} \quad (10)$$

Alternatively *E219* can be determined from *E220* using equations 11 and 12 (Moore and Cai, 2013).

$$Ratio_{E219/E220} = \frac{P_{219} \times (1 - L_{219})}{P_{220} \times (1 - L_{220})} \quad (11)$$

$$E219 = E220 \times Ratio_{E219/E220} \quad (12)$$

In which *E220* is the 220-channel system efficiency, *P* is the probability of the radon isotope decaying in the counting cell and *L* is the fractional loss due to delay and window settings. Fraction loss (*L*) will depend on the default RaDeCC apparatus time constants or those set by the operator as described by Moore and Cai (2013). The ratio *E219/E220* for the RaDeCC with default settings and normal configuration is 0.91 (Moore and Cai, 2013).

The counts per minute due to radon-219 and radon-220 are converted to disintegrations per minute (dpm) by dividing *final220* or *final219* by the detection efficiency of the channel, (*E219* or *E220*, respectively; Equations 13,14) (Giffin et al., 1963; Moore and Arnold, 1996).

$$dpm224 = \frac{bkgcorr224}{E220} \quad (13)$$

$$dpm223 = \frac{bkgcorr223}{E219} \quad (14)$$

Finally, the *dpm* values are divided by the sample *volume* (or mass) to produce the volume-corrected radium-223 (*vdpm223*) and radium-224 (*vdpm224*) sample activities (both in dpm m^{-3}) for each read (Equations 15,16).

$$\mathrm{vdpm224} = \frac{dpm224}{Volume} \times 1000 \quad (15)$$

$$vdpm223 = \frac{dpm223}{Volume} \times 1000 \quad (16)$$

To obtain the excess radium-224 and radium-223 activities of the samples at the time of sampling, two further factors must be accounted for: decay of the isotope between sampling and measurement and any activity supported by the parent isotope. The respective parent or supporting isotopes of radium-223 and radium-224 are actinium-227 and thorium-228.

In order to distinguish the activities of parent and daughter isotopes, each sample must be analysed multiple times at different intervals relative to the time of collection. The 1st interval read, performed as soon after sampling as possible, is a measurement of radium-223 and radium-224 activity, this will be a combination of excess and supported activities. A 2nd interval, 7-10 days after sampling, can provide a more accurate radium-223 activity due to reduced interference from radium-224 and radon-220 decay (Moore, 2008), and is essential in instances where the 220/219 count rate is greater than 10, or greater than 4 and the 220 channel exceeds 5 cpm (Diego Feliu et al. 2020). Eventually, >99% of measured radium-224 and radium-223 activities will be supported by their parent isotopes. This occurs after 25 days for radium-224 and after 80 days for radium-223, and dictates the timing of 3rd and 4th intervals. In effect, 3rd and 4th interval reads provide an indirect measurement of these parent isotope activities, thorium-228 and actinium-227 respectively.

$$ {}^{223}_{xs}Ra = \frac{{}^{223}_{i}Ra - {}^{223}_{s}Ra}{e^{-\lambda_{223} t}} \qquad (17) $$

$$ {}^{224}_{xs}Ra = \frac{{}^{224}_{i}Ra - {}^{224}_{s}Ra}{e^{-\lambda_{224} t}} \qquad (18) $$

Excess radium-224 and excess radium-223 at the time of sampling is then calculated via equations 17 and 18, where ${}^{223}_{i}Ra$ and ${}^{224}_{i}Ra$ are the radium-223 activity of the 1st or 2nd interval read and the radium-224 activity of the 1st interval read, ${}^{223}_{s}Ra$ the activity supported by actinium-227 decay (4th interval) and ${}^{224}_{s}Ra$ the radium-224 activity supported by thorium-228 decay (3rd interval). The time between sampling (in days) and the first measurement of each isotope is denoted by *t* and the respective decay constants of radium-223 and radium-224 by λ_{223} and λ_{224}. For all calculations, including detector efficiencies, error propagation follows the equations presented in Garcia-Solsona et al. (2008).

2.2 Calculation of radium-226 activity

The activity of long-lived radium-226 is measured indirectly via the rate of ingrowth of its decay product, radon-222. The half-life of radon-222 is 3.8 days, so as radium-226 in the sample decays over the course of a read, radon-222 accumulates in the system. This accumulation is seen in the total channel, with counts in the total channel increasing throughout the read in proportion to the radium-226 activity of the sample (Geibert et al. 2013). The *slope of cpm total* versus time during a run thus provides a measure of the radium-226 activity of the sample, based on the conversion factor *'m'*, which has a theoretical value of $1.80 \pm 0.07 \cdot 10^{-4}$ min^{-1} (Diego-Feliu et al. 2020).

Each RaDeCC detector must also be calibrated by measuring a standard with known radium-226 activity, calculated as for *E220* in equation 7. Volume-corrected radium-226 activity of the sample (*vdpm*226, in dpm/m^3) is then calculated using equation 20, where '$vdpm226_{initial}$' is the initial volume corrected radium-226 activity (in cpm/m^3; equation 19) and '*E226*' is the efficiency of system in determining radium-226 activity. This method was devised by Geibert et al., (2013) and modified by Diego-Feliu et al., (2020).

$$vdpm226_{initial} = \frac{slope\ of\ cpm\ total}{m} \div Volume \times 1000 \quad (19)$$

$$vdpm226 = \frac{vdpm226_{initial}}{E226} \quad (20)$$

3. Implementation: The RaDeCC Reader Program

The RaDeCC Reader program is a collection of python scripts that quickly processes RaDeCC output files. The program works from a single folder containing all read files including those of standards and backgrounds (or blanks), sample log sheets and a small amount of user input via a graphical user interface (GUI; Figure 1). From this folder, it creates an organised directory of read files, a table of calculated detector efficiencies with propagated uncertainties and a table of corrected excess radium-223, excess radium-224, thorium-228, actinium-227 and radium-

226 activities (in dpm/m^3) (Figure 2). The tabulated outputs also detail each correction and its propagated uncertainty for each read of each sample. Additional transparency is provided by plots of counts-per-minute vs. time for the 219, 220 and total channels, produced for each read (Figure 3) as well as plots depicting any anomalous spikes that have been automatically removed. Data quality warnings and errors are also flagged alongside calculated results in output tables as outlined in Section 3.2.2.

3.1 Essential information for the program

The RaDeCC Reader program receives information in three ways: text files output by the RaDeCC apparatus, the sampling log-sheet and the graphical user interface (GUI). Information entered into the GUI entry fields are used to aid the program in file-handling and provide standard and instrument specific parameters required for the data corrections and uncertainty propagations. Once completed these GUI entries can be saved by the user and reloaded for subsequent runs of the program.

3.1.1 Directories

To start, the user sets up the following folder and contents:

C:/.../Main_folder/Read_files_and_logsheet/

C:/.../Main_folder/RaDeCC_Reader_Scripts/

The first entry fields in the GUI are the input and output directories (Figure 1a) and the logsheet. The input directory is where the program will find all the input read files and the logsheet. The output directory is where the program will place the organised read files, the logsheet file and output files.

3.1.2 Logsheets: Linear and Branched sample sets

Logsheets form the basis of the eventual output files, in which all the metadata contained within a logsheet will be included. A logsheet must contain information that is essential to data correction calculations: sample names, sample volumes and mid-point sampling times; as well as any sub-sample names (for herein so-called ‘branched’ datasets) if applicable. There should be a column displaying each of these variables in a logsheet. Any additional information contained in a logsheet (e.g. the latitude, longitude and depth of individual samples) is preserved in the output files and will not interfere with the calculations but may prove useful for later analysis. An example logsheet file is included in the Supplementary Information. The date format convention for read files and the logsheet must be consistent and can be indicated via a tickbox in the GUI.

Data outputs can be organised differently to assist the user. How data outputs will be organised depends on whether or not the user indicates a sample set includes sub-samples. Herein sample sets that do not contain sub-samples (e.g. multiple locations sampled once, or time series at a single location) are termed ‘linear’. Sample sets with sub-samples (e.g. multiple locations each sampled at multiple times, or a series of depth profiles) are termed ‘branched’. In the case of a sample set where some samples have sub-samples, this could be processed using the branched setting. In this case, samples without subsamples would be seen as samples with one subsample each. The distinction between linear and branched can be indicated via a tickbox in the GUI.

3.1.3 File naming and identifiers

In order to acquire raw data, the program requires the text files generated by the RaDeCC systems for sample, standard and (if required) background reads. The formatting of these filenames needs to be consistent and must include information on the sample (and sub-sample) name and the detector used. For example, ‘1-StnX001-A001-010220-det1.txt’, contains the sample name ‘StnX001’, subsample name ‘A001’ and the detector name ‘det1’. The number ‘1’ at the start of the file name designates the read interval (e.g. 1 for radium-224 quantification), although this is recorded by the program it is not used in excess calculations. Instead, the program assigns a read interval automatically by calculating the elapsed time between sampling and RaDeCC analysis. It is important to note that sample and sub-sample names must be distinct from each other, no sample name should contain another sample name within it (StnX1 and StnX10, for example). Once the first panel of entries is completed in the GUI (Figure 1a), these entries are checked by the program, and if verified, the user can proceed to the second panel in the GUI to assign details of the standards and backgrounds.

3.1.4 Information on detectors, standards and backgrounds

Upon verified completion of the first panel of entries in the GUI (Figure 1a), a second panel will appear requesting inputs for individual detectors (names, *E219/E220* ratios, radium-226 slope calibration values and radium-226 system efficiency values) and details specific to individual standards (names, dates of manufacture and initial activity) (Figure 1b). Only an identifying name is requested for background runs. If background measurements are not required then the 'No. of Background Standards' field can be set to '0' in panel 1 of the GUI (Figure 1a). These inputs are all required for the calculation of detector efficiencies and the resulting corrections to the raw data.

3.1.5 Assigning variables

The final GUI entries are the titles of log-sheet columns containing sample name, sub-sample name, sample volume, sample volume error, sampling date and sampling time. These column titles should not contain spaces and must be selected via the drop-down lists that appear in the second panel of the GUI after a log-sheet file is selected in the first (Figure 1b).

Once these details are completed and verified, the user can then proceed to run the RaDeCC Reader. A step-by-step explanation of information input and program setup is also provided in the *Instructions.md* or *Instructions.txt* files in the GitHub repository along with example data to check that the program is functioning properly.

3.2 How it works

3.2.1 Data, directories and detector efficiencies

Upon clicking the ‘Run RaDeCC Reader’ button, the directory building function will create an organised directory of read and logsheet files using input from the GUI as well as sample and sub-sample names in the logsheet.

The directory_filler function will then use each folder/sub-folder name as a search criteria and search through the main folder of reads for files that match each folder name and then subsample. When a match occurs, the file is copied to the folder it was matched with. Any files not matching sample/sub-sample folder names or standard or background folder names will be copied to the miscellaneous (*misc.*) folder.

Once the directory is built and populated with reads, a dataframe of detector efficiencies is produced. The efficiencies calculation function searches through the appropriate standard and background subfolders for each detector specified by the user in the GUI. The program creates a dataframe of corrected reads for each standard with the appropriate channel efficiency for each read calculated as well as a dataframe of background reads. These offer the user a more granular view of read results when validating the average efficiencies displayed in the summary efficiencies dataframe. These four dataframes are automatically exported as .csv files. The detector efficiency of the 219-channel for each detector is calculated using the actinium-227 standard as well as the method devised by Moore and Cai (2013), based on system volume and 220-channel efficiency using the thorium-232 standard. In parallel, the two separate 219-channel efficiencies are used to calculate two separate final corrected radium-223 values. Use of radium-223 values based on the Moore and Cai (2013) method requires verification of an *E219/E220* ratio (Section 3.1.4): the Reader includes the value determined by Moore and Cai for the standard RaDeCC configuration as a default.

The program uses the sample name (and sub-sample name) in each row of the logsheet as search criteria, finding the corresponding read files to scan. Using the data scanned from the read files the program performs the appropriate corrections and related propagation of uncertainties. For each read, these new corrected values along with their uncertainties are combined with the sample's corresponding metadata from the logsheet and entered as a new row in the read results dataframe.

3.2.2 First level corrections

Every read file for each sample/sub-sample is scanned, the interval logged data is extracted from the text file, and the first level of corrections are performed (Garcia-Solsona et al., 2008). These include:

- Chance coincidence counts per minute in the 219 and 220 channels (*Y 219 CC, Y 220 CC*)
- Corrections for 220 interference in the 219 channel (to give *final*219)
- Corrections for 219 interference in the 220 channel (to give *final*220)
- Total counts corrected for counts due to 219 or 220 (to give *corr total*)

As the program scans through a read, each interval is evaluated using the guidelines outlined by Diego-Feliu et al., (2020) for the measurement and quantification of radium-223 and radium-224 (Figure 4). For each read, the program records the percentage of intervals for which quantification of radium-223 or radium-224 is not recommended and logs these percentages in an error column of the read results dataframe. This allows the user to quickly establish whether an anomalous result might be due to cross-talk or other interferences. The scanning of read files is not obstructed by files with lines enclosed by quotation marks or extra lines added by a pause function.

Spike removal: If the number of counts in either the total, 219- or 220-channel during one time interval is higher than the next interval by more than the *'Spike sensitivity'* constant the program removes this time interval as it is considered to contain a counting anomaly - likely due to a spike in the electrical supply to the RaDeCC apparatus. If an anomaly is removed, the value of the anomaly is recorded in the *Spike_Value* column of the results dataframe. The calculated counts per minute values of each interval are then averaged over the whole read.

The *'Spike sensitivity'* constant is set at 10^6 counts by default, meaning that spike removal is inactive, but may be activated via a change in the *'Spike sensitivity'* constant value by the user to allow for higher or lower sample activities.

Radium-226 estimates from radon-222 ingrowth and Raw Data Plots: The rate of radon-222 ingrowth seen in the Total channel (*cpm total*) is calculated in order to estimate the activity of radium-226 in the sample. The equilibration time variable (0 minutes as default) allows the user to set the time required for the radon-222 activity throughout the RaDeCC circuit to accumulate sufficiently to be detectable in the total channel. The time interval is set by the user in the GUI prior to initiating a read and is the number of minutes between the software logging each line of the output file (Figure 1a). The time interval is used by the RaDeCC Reader program here to decide how many lines to miss at the start of the read file before calculating radon-222 ingrowth and therefore the radium-226 activity estimate. A plot of read-time vs. total counts per minute (*cpm total*), 219 channel counts per minute (*cpm*219) and 220 channel counts per minute (*cpm*220) for each read is saved in the '*Read Plots*' folder (Figure 3.). These plots provide a graphical view for raw data quality assessment by the user, for instance to evaluate system stability as well as the build-up of radon-222 during each read. Estimated radium-226 activity will only be calculated from reads with durations >600 minutes, shorter reads may be less reliable due to the short period for ingrowth of radon-222 to occur. In the event of a short read, the '*Err226_short_read*' error is logged in the read results dataframe.

3.2.3 Second level corrections and output

After the first level of corrections is complete, generating values for *final*219, *final*220 and *cpm total* for radium-226 estimation, the read results dataframe containing these new values is passed on to the second level of calculations. Second level corrections expand the read results dataframe with the calculated values and propagated uncertainties as described by Garcia-Solsona *et al.* (2008). Second level corrections include:

- Detector background corrections in all channels
- Corrections for detector channel efficiencies
- Corrections for sample volume, producing volume corrected activity (*vdpm*)

These final calculations complete the series of corrections and uncertainty propagations providing disintegrations per minute per 1000 L (dpm/m^3) for radium-223 and radium-224 as well as an estimate of radium-226 activity (dpm/m^3) for each read. These individual read results are saved as a table in comma-separated-value (.csv) format before being combined to calculate sample activities.

3.2.4 Sample activity calculations and outputs

The final stage of calculations is the combination of read-specific values calculated in the results dataframe to calculate excess radium-223, excess radium-224, radium-226, actinium-227 and thorium-228 activities for each sample.

The 2nd and 4th interval reads of each sample/sub-sample are combined using equation 13 to calculate excess radium-223. For the calculation of excess radium-224, 1st and 3rd interval reads are combined using equation 14. In many circumstances 1st interval reads are sufficient to accurately quantify radium-223 activity, so if 2nd interval reads are unavailable, excess radium-223 is calculated using 1st reads. Similarly, if 3rd reads are unavailable, excess radium-224 is calculated using 4th interval reads. Actinium-227 activity is essentially the supported radium-223 activity calculated for the 4th (or 3rd) interval read of a sample/sub-sample while thorium-228 is the supported radium-224 activity calculated for the 3rd (or 4th) interval read of a sample/sub-sample. If the results dataframe contains more than one read of a particular sample/sub-sample for a given interval (1st-4th), the average activity of the relevant reads will be used in the calculation. The radium-226 activity of a sample is determined by averaging the radium-226 activity of reads >600 minutes in duration. Any radium-226 activities that are more than one standard deviation from the mean are then removed and a new average is calculated.

The results of these final calculations are tabulated in a summary dataframe and exported as a comma-separated value (.csv) file. Any read-interval substitutions in the calculation of excess activities are logged in the error column of this summary dataframe alongside any errors raised using the logic outlined by Diego-Feliu et al. (2020) for all read results used.

4. Validation

4.1 Experimental Design

To evaluate the performance and accuracy of the RaDeCC Reader program, the processing time and corrected data outputs from real sample, standard and background determinations by RaDeCC instruments were compared to those derived from a Microsoft Excel implementation of the calculations outlined by Garcia-Solsona et al. (2008) and Geibert et al. (2013).

A total of 208 raw data files from 44 samples were used for the purpose of this evaluation. Open ocean samples (106 raw data files, 19 seawater samples) were collected from 60-100 litres of seawater using MnO_2 impregnated fibres, during the along southwest Greenland during the ICY-LAB expedition aboard RRS *Discovery* in 2017 (Hendry et al., 2019). Coastal surface seawater samples (~0.5 m depth) (40 raw data files, 9 surface samples) and karstic spring-water samples (62 raw data files, 16 samples) were collected offshore of the Calanques of Marseille-Cassis on 27–28 March 2018 aboard the R/V Antédon II, by trace-metal clean submersible pump and scuba-divers respectively (Tamborski et al., 2020).

The range of 219, 220 and total count rates (219: 0 - 6.3 cpm, 220: 0 – 16.8 cpm, total: 0 – 35 cpm) and counting times (60-4002 minutes) tested here, are realistic ranges encountered in submarine aquifer and open ocean fieldwork and 35% of the maximum quantification limit of the RaDeCC apparatus (Diego-Feliu et al., 2020). These samples, previously published in Hendry *et al*. (2019) and Tamborski et al. (2020), were calibrated using standards prepared at LEGOS, OMP (Toulouse, France) with solutions of ^{228}Th (in equilibrium with ^{232}Th) and ^{227}Ac obtained from the International Atomic Energy Agency (Monte Carlo, Monaco). Here we repeat their raw data processing using our standardised Excel-based methodology and compare the outputs to those obtained using the RaDeCC Reader. This approach allows any disparity in results to be attributed to differences in implementation. Nine variables were compared: the corrected activities of excess radium-223, actinium-227, excess radium-224, thorium-228 and radium-226, and the propagated uncertainties for excess radium-223, actinium-227, excess radium-224 and thorium-228.

4.2 Results and Discussion:

Implementation time of either method is certain to vary between users. For new users of the RaDeCC Reader time will be needed to name and organise files and prepare logsheets. In this exercise, however, the implementation of the Excel-based methodology took an experienced user over 2 hours to process the outputs from standards, backgrounds and 30 samples; amounting to a total of 233 raw data files. This compared to a processing time of 2 minutes to perform the equivalent functions using the RaDeCC Reader, a time saving that would be magnified with larger datasets or familiarity with the required file naming conventions.

Excellent agreement ($R^2 > 0.99$, Standard Error < 0.02) was seen for the corrected activities and propagated uncertainties of excess radium-223 and excess radium-224 (Figure 5) as well as actinium-227, thorium-228 (Figure 6). The small amount of variance seen, possibly due to a difference in the treatment of background measurements, is an order of magnitude smaller than any propagated uncertainties associated with the activities determined in this study.

Radium-226 activity determined by our Excel method and the RaDeCC Reader also displayed very strong agreement, with greater variance than was seen for the short-lived radium isotopes or their supporting isotopes ($R^2 = 0.99$, Standard Error = 0.02, Figure 6). We attribute this greater variance between methods to the fact that radium-226 activities determined by RaDeCC apparatus are inherently less precise than those determined for excess radium-223, actinium-

227, excess radium-224 and thorium-228. The activity of radium-226 is measured via the ingrowth of its daughter-isotope radon-222 and therefore the slope of the activity in the total channel with time. Many workers may choose not to include a portion of measurements at the start of a read to allow for the partial pressure of radon-222 in the system to accumulate above background. This equilibrium time may not be applied uniformly, whereas the RaDeCC Reader's user defined equilibration time is applied to all reads consistently. The slope in total activity with time is also sensitive to system leaks as well as the length of time a sample is measured for, particularly for samples with low activity, and therefore should be evaluated separately for samples with markedly different total activities.

5. Conclusions

We have developed a program that simplifies and expedites the process of correcting raw RaDeCC data, propagating related uncertainties and calculating the activities of excess radium-223, actinium-227, excess radium-224, thorium-228 and radium-226. With a logsheet and read file names in the required format, the RaDeCC Reader program is capable of processing a substantial real data set in a matter of minutes, and is therefore able to save users considerable time and effort in data processing when compared to previous and widely used Excel-based methodologies. By letting users evaluate their sampling methods and analytical performance more efficiently, the RaDeCC Reader has potential to enhance experimental design, for example, during maritime research expeditions. RaDeCC Reader maintained the accuracy of results attributed to previous methods, and preserved transparency of data processing by displaying the values of each stage of calculation, providing a view of the original raw data via saved plots and flagging results with data quality warnings. We attribute minor discrepancies in calculated excess radium-223, actinium-227, excess radium-224 and thorium-228 activities between methods to a difference in background treatment by the RaDeCC Reader's implementation. This provided no significant changes to the results from samples used in our test, however the implementation used by RaDeCC Reader mitigated the risk of greater inaccuracies that might have arisen from raw data files containing larger or more frequent counting anomalies.

Acknowledgments:

The authors gratefully acknowledge support from the UK's Natural Environment Research Council who funded S.S. through the Environmental Research Doctoral Training Partnership with University of Oxford, and A.L.A. and W.B.H. through Independent Research Fellowships (NE/P017630/1 and NE/K009532/1). We would like to thank J. Tamborski and P. van Beek who shared their RaDeCC output and sample data to validate the RaDeCC Reader across a wider range of environmental activities.

Computer Code Availability:

RaDeCC Reader, developed by Sean Selzer, Department of Earth Sciences, South Parks Rd, OX1 3AN, Oxford (01865 272000, sean.selzer@earth.ox.ac.uk). First available in 2019.

Hardware Requirements: 2 x 64-bit 2.8 GHz 8.00 GT/s CPUs, 32 GB RAM (or 16 GB of 1600 MHz DDR3 RAM), 300 GB Storage.

Written in Python 3.6, RaDeCC Reader (179 KB) is available on GitHub (https://github.com/oxradreader/RaDeCC_Reader)

References:

Charette, M.A., Gonneea, M.E., Morris, P.J., Fones, G., Planquette, H., Salter, I., Garabato, A.N., 2007. Radium isotopes as tracers of iron sources fueling a Southern Ocean phytoplankton bloom. Deep Sea Res. Part II Top. Stud. Oceanogr. 54, 1989–1998. https://doi.org/10.1016/J.DSR2.2007.06.003

Cochran, J., 1982. The oceanic chemistry of the U- and Th-series nuclides, in: Ivanovich, M., Harmon, R. (Eds.), Uranium Series Disequilibrium: Applications to Environmental

Problems. Clarendon Press, Oxford, pp. 384–431.

Diego-Feliu, M., Rodellas, V., Alorda-Kleinglass, A., Tamborski, J., van Beek, P., Heins, L., Bruach, J.M., Arnold, R., Garcia-Orellana, J., 2020. Guidelines and Limits for the Quantification of Ra Isotopes and Related Radionuclides With the Radium Delayed Coincidence Counter (RaDeCC). J. Geophys. Res. Ocean. 125. https://doi.org/10.1029/2019JC015544

Garcia-Solsona, E., Garcia-Orellana, J., Masqué, P., Dulaiova, H., 2008. Uncertainties associated with 223Ra and 224Ra measurements in water via a Delayed Coincidence Counter (RaDeCC). Mar. Chem. 109, 198–219. https://doi.org/10.1016/j.marchem.2007.11.006

Geibert, W., Rodellas, V., Annett, A., Beek, P. van, Jordi Garcia-Orellana, Hsieh, Y. Te, Masque, P., 2013. 226Ra determination via the rate of 222rn ingrowth with the radium delayed coincidence counter (RaDeCC). Limnol. Oceanogr. Methods 11, 594–603. https://doi.org/10.4319/lom.2013.11.594

Giffin, C., Kaufman, A., Broeker, W., 1963. Delayed Coincidence Counter for the Assay of Actinon and Thoron. J. Geophys. Res. 68, 1749–1757.

Hendry, K.R., Huvenne, V.A.I., Robinson, L.F., Annett, A., Badger, M., Jacobel, A.W., Ng, H.C., Opher, J., Pickering, R.A., Taylor, M.L., Bates, S.L., Cooper, A., Cushman, G.G., Goodwin, C., Hoy, S., Rowland, G., Samperiz, A., Williams, J.A., Achterberg, E.P., Arrowsmith, C., Alexander Brearley, J., Henley, S.F., Krause, J.W., Leng, M.J., Li, T., McManus, J.F., Meredith, M.P., Perkins, R., Woodward, E.M.S., 2019. The biogeochemical impact of glacial meltwater from Southwest Greenland. Prog. Oceanogr. 176, 102126. https://doi.org/10.1016/j.pocean.2019.102126

Kipp, L.E., Sanial, V., Henderson, P.B., van Beek, P., Reyss, J.-L., Hammond, D.E., Moore, W.S., Charette, M.A., 2018. Radium isotopes as tracers of hydrothermal inputs and neutrally buoyant plume dynamics in the deep ocean. Mar. Chem. 201, 51–65. https://doi.org/10.1016/J.MARCHEM.2017.06.011

Moore, W.S., 2008. Fifteen years experience in measuring 224Ra and 223Ra by delayed-coincidence counting. Mar. Chem. 109, 188–197. https://doi.org/10.1016/j.marchem.2007.06.015

Moore, W.S., 2000. Determining coastal mixing rates using radium isotopes. Cont. Shelf Res. 20, 1993–2007. https://doi.org/10.1016/S0278-4343(00)00054-6

Moore, W.S., 1976. Sampling 228Ra in the deep ocean. Deep. Res. Oceanogr. Abstr. 23, 647–651. https://doi.org/10.1016/0011-7471(76)90007-3

Moore, W.S., Arnold, R., 1996. Measurement of 223Ra and 224 Ra in coastal waters using a delayed coincidence counter. Measurement 101, 1321–1329.

Moore, W.S., Cai, P., 2013. Calibration of RaDeCC systems for 223Ra measurements. Mar. Chem. 156, 130–137. https://doi.org/10.1016/j.marchem.2013.03.002

Tamborski, J., van Beek, P., Conan, P., Pujo-Pay, M., Odobel, C., Ghiglione, J.F., Seidel, J.L., Arfib, B., Diego-Feliu, M., Garcia-Orellana, J., Szafran, A., Souhaut, M., 2020. Submarine karstic springs as a source of nutrients and bioactive trace metals for the oligotrophic Northwest Mediterranean Sea. Sci. Total Environ. 732, 1–14. https://doi.org/10.1016/j.scitotenv.2020.139106

Figures and Tables:

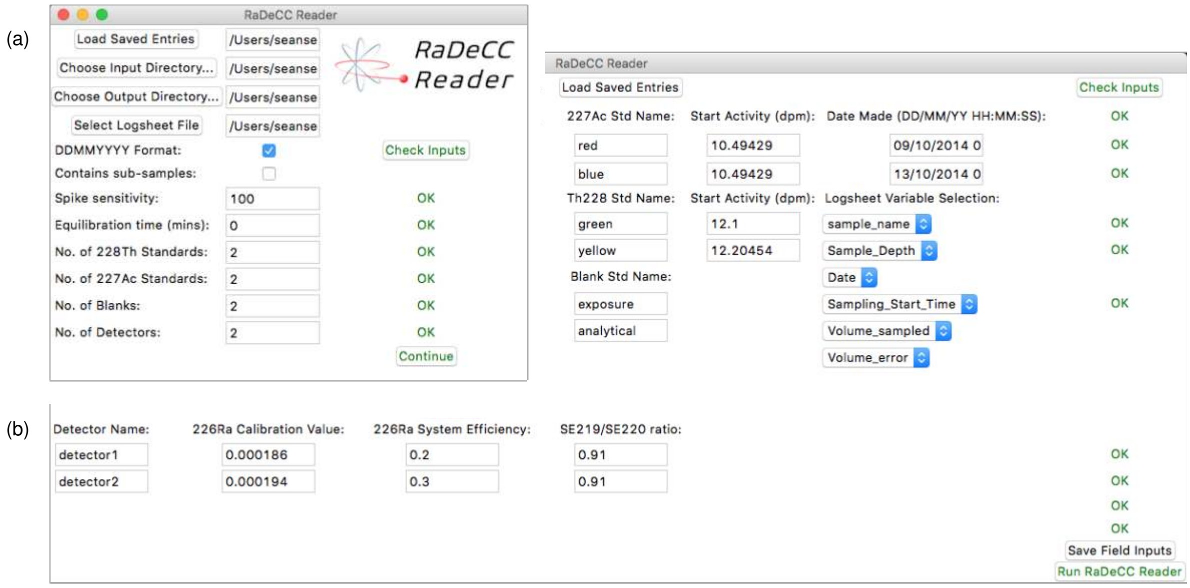
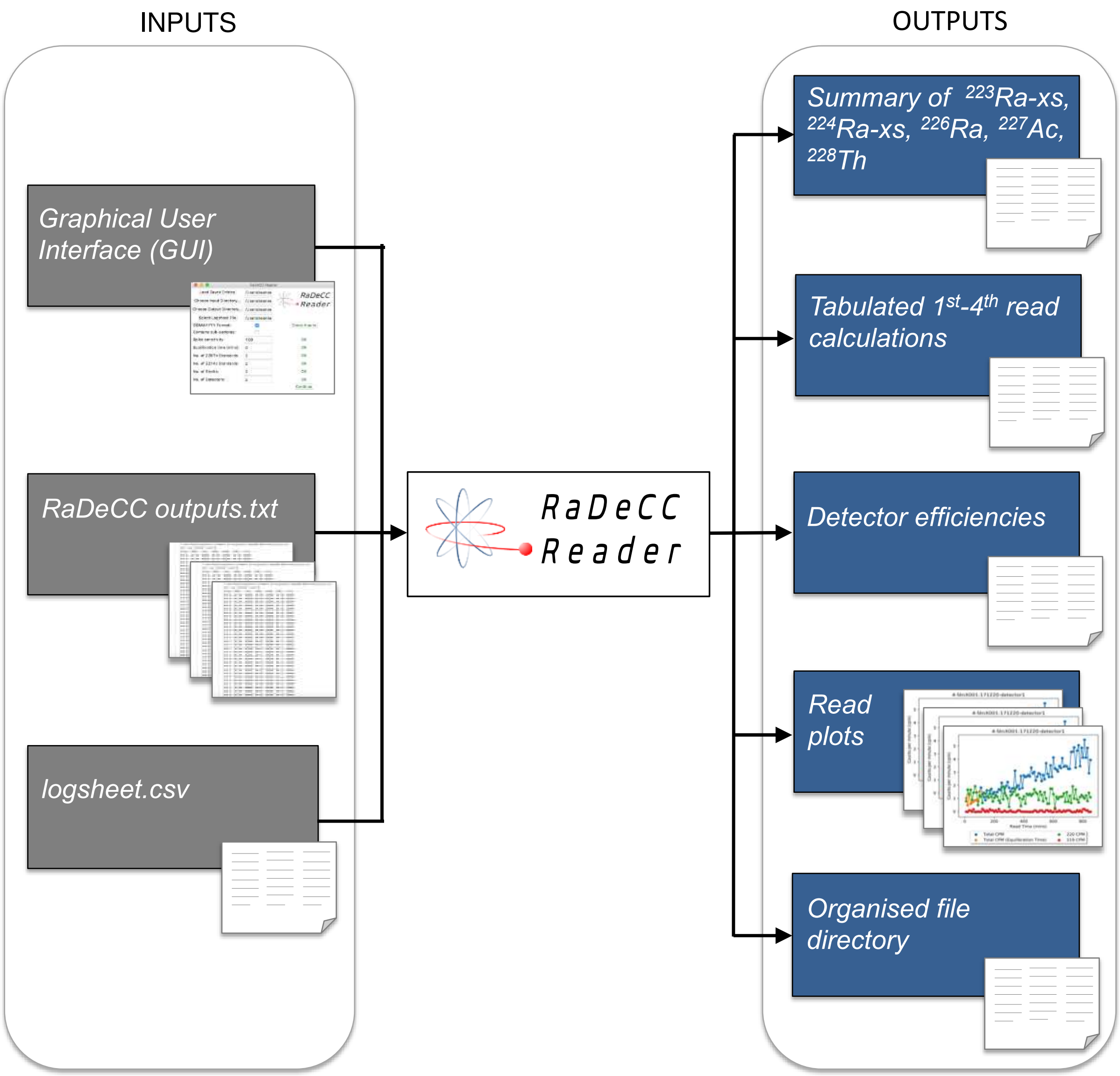


Figure 1. Details of the Graphical User Interface (GUI) used to operate RaDeCC Reader. This provides a verifiable summary of editable and necessarily user defined input parameters. Including (a) input and output file directories, date and data formats, calculation preferences, and the inventory of standards, backgrounds and detectors, and (b) details of individual detector names and efficiencies, standard names and activities, background names, and logsheet variables. In all fields of the GUI users may save and load previous inputs and check inputs before running the RaDeCC Reader programme.



550

551 **Figure 2.** Summarised inputs and outputs of the RaDeCC Reader program.

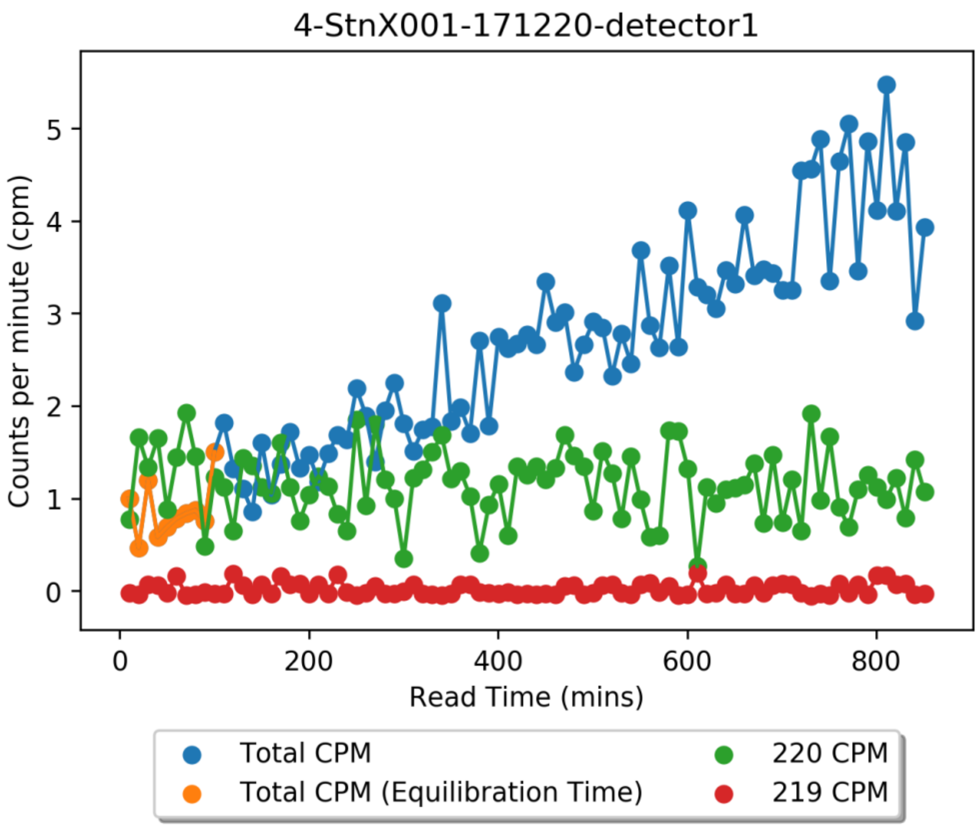


Figure 3. Example of a read plot produced by RaDeCC Reader of counts per minute for the total, radon-219 and radon-220 channels over the course of a sample read. Spikes in counts per minute (any counts that exceeded the default *'Spike sensitivity'* constant) have been removed. Counts in the total channel that are used in the estimation of radium-226 activity are shown in blue. Counts in the total channel that are ignored in the estimation of radium-226 activity during a user-defined period of detector equilibration are shown in orange.

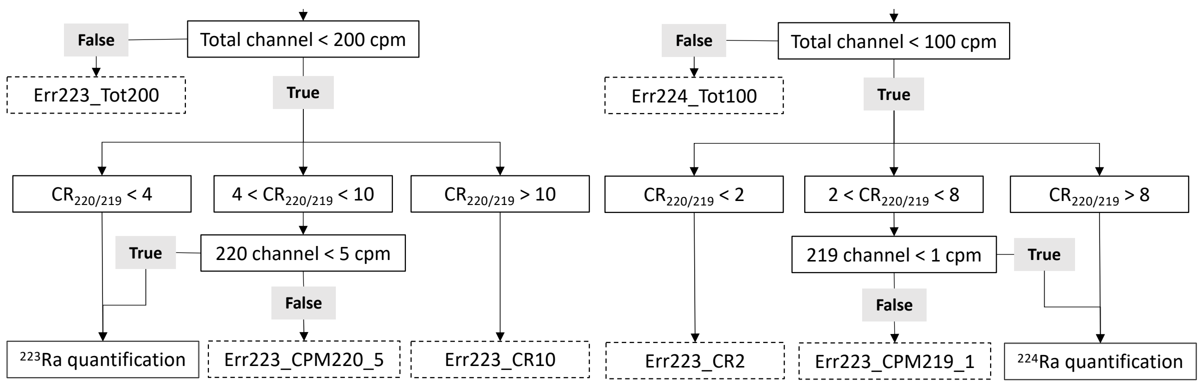


Figure 4. Flow charts of the guidelines for quantifying radium-223 (a) and radium-224 (b)

using RaDeCC apparatus (modified after Diego-Feliu et al. (2020). $CR_{220/219}$ is the count rate ratio of the 220-channel to the 219-channel.

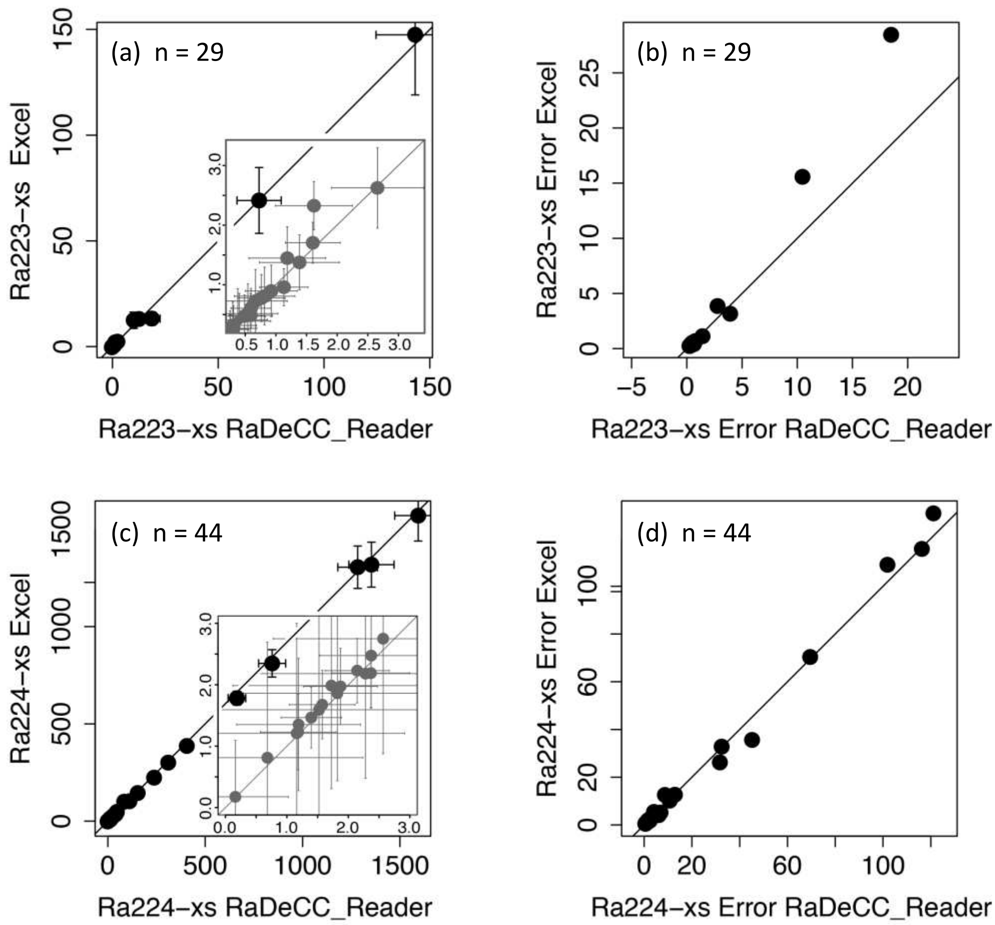


Figure 5. Validation of RaDeCC Reader outputs. Volume corrected activities and propagated uncertainties of excess radium-223 (a, b) and excess radium-223 (c, d) determined by the RaDeCC Reader program vs. an Excel implementation. Individual reads are plotted as black circles in units of dpm/m^3, relative to a 1:1 line. Inset plots (a, c) show the agreement between RaDeCC Reader program and the Excel implementation for samples in the low activity range.

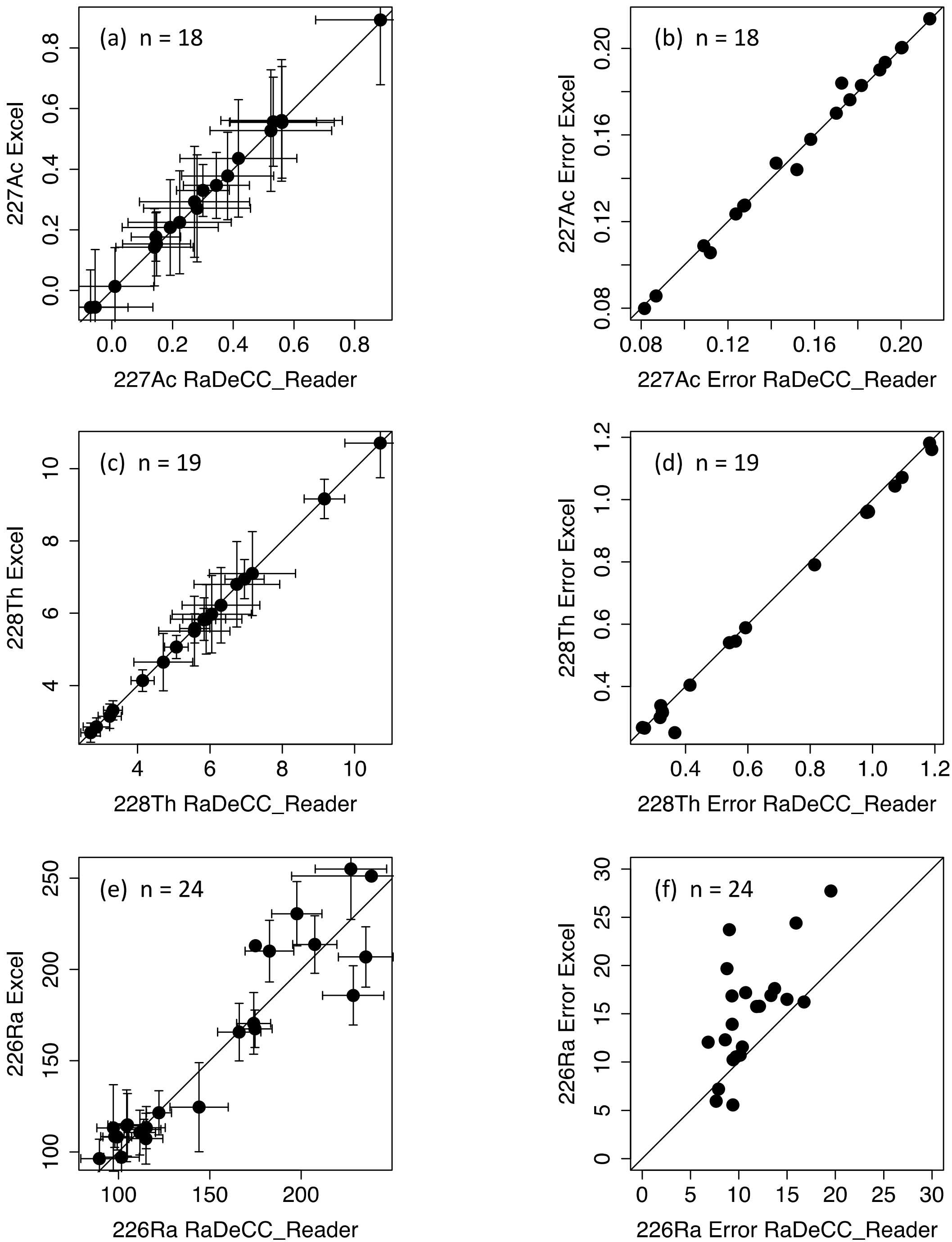


Figure 6. Volume corrected activities (dpm/m^3) of actinium-227 (a), thorium-228 (c) and radium-226 (e). Propagated uncertainties associated with the calculation of actinium-227 (b), thorium-228 (d) and radium-226 are also included. Individual samples are plotted as black circles in units of dpm/m^3, relative to a 1:1 line.

Table 1. Glossary of variable terms used, their descriptions and units.

Variable	Description	Units
Y 219 CC	Erroneously registered 219 channel chance coincidence counts	cpm
Y 220 CC	Erroneously registered 220 channel chance coincidence counts	cpm
cpm219	counts per minute (219 channel)	cpm
cpm220	counts per minute (220 channel)	cpm
cpm total	counts per minute (total channel)	cpm
corr219	cpm219 corrected for Y 219 CC	cpm
corr220	cpm219 corrected for Y 220 CC	cpm
final219	corr219 corrected for cross-channel interference	cpm
final220	corr219 corrected for cross-channel interference	cpm
E219	Detection efficiency of the 219 channel	-
E220	Detection efficiency of the 220 channel	-
$Ratio_{219/220}$	Detection efficiency ratio of the 219 and 220 channels	-
P_{219}	Probability of radon-219 decaying in the cell	-
P_{220}	Probability of radon-220 decaying in the cell	-
L_{219}	Loss resulting from the 219 channel delay and window settings	-
L_{220}	Loss resulting from the 220 channel delay and window settings	-
bkgcorr223	final219 corrected for background	cpm
bkgcorr224	final220 corrected for background	cpm
dpm223	final219 corrected for detection efficiency	dpm
dpm224	final220 corrected for detection efficiency	dpm
vdpm223	dpm223 corrected for sample volume	dpm m^{-3}
vdpm224	dpm224 corrected for sample volume	dpm m^{-3}
$^{223}Ra_i$	vdpm223 for the initial read (including supported fraction)	dpm m^{-3}
$^{223}Ra_s$	vdpm223 for the latter read (supported fraction only)	dpm m^{-3}
$^{224}Ra_i$	vdpm224 for the initial read (including supported fraction)	dpm m^{-3}
$^{224}Ra_s$	vdpm224 for the latter read (supported fraction only)	dpm m^{-3}
λ_{223}	Decay constant for radium-223	d^{-1}
λ_{224}	Decay constant for radium-224	d^{-1}
$^{223}Ra_{xs}$	Excess radium-223 activity	dpm m^{-3}
$^{224}Ra_{xs}$	Excess radium-224 activity	dpm m^{-3}
vdpm226	radium-226 activity	dpm m^{-3}
$vdpm226_{initial}$	radium-226 activity initially measured in sample	cpm m^{-3}
slope of cpm total	Gradient of counts over time in the total channel	cpm min^{-1}
m	Radium-226 conversion factor	min^{-1}
E226	Detection efficiency for radium-226	-
Volume	Volume or mass of sample water	L, kg