ANSTO/ENV/TN1-2005

Australian Government

ansto

A work instruction for the preparation and processing of raw radon data from field sites

by

S. Chambers, W. Zahorowski, L. Mokhber-Shahin,
O. Sisoutham and S. Werczynski

Prepared within the Institute for Environmental Research
Australian Nuclear Science and Technology Organisation

Last updated: 05/09/2006

Table of contents

TABLE OF CONTENTS 2

OVERVIEW 3

DATA PREPARATION 3

DATA TIME CHECK 4

DESPIKING 6

CONCATENATION 8

DETERMINING CALIBRATION COEFFICIENTS 8

DETERMINATION OF INSTRUMENTAL BACKGROUND 9

TROUBLE SHOOTING 9

RESPONSE TIME 10

AGGREGATION TO HOURLY DATA 11

CALIBRATION TO CONCENTRATIONS 12

- PREPARING THE HOURLY BACKGROUND 12
- PREPARING HOURLY CALIBRATION COEFFICIENTS 13
- DERIVING CONCENTRATIONS 13

APPENDIX 1 15

APPENDIX 2 16

Overview

This document is aimed at introducing an entry level technician to radon data processing and assumes that Microsoft Excel is the primary processing tool at hand. As such, descriptions have been made as straightforward and comprehensive as possible.

Each month, raw data from the radon detector (and corresponding meteorological data, where available) should be downloaded, checked and archived in preparation for integration to a processed radon database.

The naming convention for radon detector data files is SSMMMYY.*, where SS is a two character site identifier, MMM is the month and YY an integer year index. The data file extension of the radon detector output is *.csv to facilitate the importing of data to an Excel spreadsheet.

Particularly in cases where the controlling computer is not on the network (which would enable e-mail advisories of problems) it is important to check each monthly file as it arrives in case there has been a malfunction of a detector component or the data logger. If an inconsistency in the data file is observed, or data values are outside of predefined threshold values, an attempt should be made to diagnose the cause of the problem by noting the combination of data series in which problems were observed. Once the likely cause has been identified, the local staff where the detector is installed should be notified, and advice provided as to how to rectify the suspected problem.

Apart from the rudimentary monthly quality checks noted above, the post processing of radon data is usually only completed in 12 month blocks. This is so because, with the information about the detector's calibration and background count not continuously available, time series of these quantities need first to be generated for the time period over which the processing will take place. This is usually most convenient to do over one complete annual cycle.

The remainder of this document outlines the necessary steps to prepare a 12 month set of raw radon data for submission to a permanent radon data archive.

Data Preparation

Import a complete year of data (12 files) for a given site into separate sheets of an Excel workbook.

Each data file should contain column headers appropriate to the logger data storage format. Most ANSTO data loggers store data in the following series order:

Year, DOY, DOM, HOD, Spare, Flow, LLD, ULD, GasM, Pres, HV, ExmA, InmA, Temp, Bat

Year - Year of observations
DOY - Day of year (Julian day)
DOM - Day of month

HOD - Hour of day (30 minute intervals)

Spare - *This column is left for future use (but may contain an approximately calibrated radon concentration depending on software version).*

Flow - Flow rate through the detector (L min^{-1})

LLD - Counts per 30 minutes above the "Lower level discriminator" voltage. This is the uncalibrated radon signal.

ULD - Counts per 30 minutes above the "Upper level discriminator" voltage. A diagnostic signal only, not routinely used in post processing.

GasM - Revolutions of the gas meter impeller, the uncalibrated flow rate.

Pres - Pressure (Pa). The differential pressure between the detector delay volume and ambient. At some sites this signal is not calibrated and a signal is in milli-volts only. If it is a calibrated value, it should be ~100Pa.

HV - High voltage setting of the PMT power supply (Volts),

ExmA - External blower current draw (milli-amps). This blower is used to draw sample air through the detector, and usually operates at a lower current than the internal blower.

InmA - Internal blower current draw (milli-amps). This blower is used to circulate air within the main detector delay chamber.

Temp - The temperature (C) as measured by the data logger inside its enclosure.

Bat - Logger battery voltage (V). Should not drop below 10V (~9.6V is critical).

Data time check

Occasionally there are problems with the data logger (either during logging, or during download to the PC) that result in missing, duplicate or corrupt data. Consequently, it is necessary to look through each page of the workbook (ie. each month of data) to ensure that the data time stamp increases in a regular manner (ie. day numbers are sequential and that each day contains 48 samples).

Start by checking the number of data records for the month at the bottom of the page. Depending on the month there should be the header line plus (28*48=1344), (29*48=1392), (30*48=1440) or (31*48=1488) lines of data in the month. Next, generate a quick time-series plot of the DOY and DOM columns and confirm that they both increment regularly from beginning to end.

If either of these checks identifies a problem it will be necessary to check the whole month of data line by line. One way to do this is to auto-generate a second "Time" series in the first available blank column to the right (e.g. *Figure 1*).

Year	DOY	DOM	HOD	Spare	Flow	LLD	ULD	GasM	Pres	HV	ExmA	InmA	Temp	BatV	Time
4	1	1	0	5643	32.4	1885	1394	648	120	604.1	1157	1630	11	12.13	0:00
4	1	1	30	22926	32.7	7646	5786	655	120	604.1	1157	1630	10.9	12.16	0:30
4	1	1	100	45045	32.9	15019	11608	658	120	604.1	1157	1630	10.9	12.13	1:00
4	1	1	130	57999	33	19337	15069	661	120	604.1	1157	1631	10.9	12.15	1:30
4	1	1	200	66648	33.2	22220	17622	665	121	604.2	1157	1630	10.9	12.13	2:00
4	1	1	230	72225	33.3	24079	19037	666	121	604.2	1158	1631	11	12.15	2:30
4	1	1	300	76098	33.1	25370	20194	662	120	604.2	1158	1631	11	12.14	3:00
4	1	1	330	77046	32.9	25686	20476	658	120	604.1	1159	1632	11	12.13	3:30
4	1	1	400	78828	32.6	26280	20964	653	119	604.1	1159	1633	10.9	12.16	4:00
4	1	1	430	71793	32.5	23935	19178	650	119	604.1	1159	1633	10.7	12.17	4:30
4	1	1	500	48756	32.5	16256	13222	650	119	604.1	1159	1633	10.6	12.12	5:00
4	1	1	530	32466	32.2	10826	8827	644	119	604.1	1160	1633	10.4	12.17	5:30
4	1	1	600	21513	32	7175	5797	641	118	604.1	1159	1633	10.3	12.17	6:00
4	1	1	630	14400	32.1	4804	3851	643	118	604.1	1159	1633	10.2	12.14	6:30
4	1	1	700	10377	32.4	3463	2703	648	119	604.1	1159	1633	10.3	12.13	7:00
4	1	1	730	8247	32.7	2753	2132	655	120	604.1	1159	1632	10.3	12.18	7:30
4	1	1	800	7128	33.1	2380	1815	663	121	604.1	1158	1632	10.4	12.14	8:00
4	1	1	830	6618	34	2210	1670	680	123	604.2	1157	1631	10.5	12.16	8:30
4	1	1	900	6069	35.5	2027	1492	710	127	604.2	1151	1623	10.9	12.17	9:00
4	1	1	930	5460	36.1	1824	1294	722	126	604.2	1152	1618	12.3	12.13	9:30
4	1	1	1000	5451	37.3	1821	1292	747	125	604.2	1155	1619	14.5	12.15	10:00

***Figure 1**: Comparison of logger time stamp (HOD) and newly generated "Time" series.*

It is then easier to page down the data file and compare individual values of the original "HOD" series to the newly generated "Time" series. If the times diverge, there is a problem with the data that needs to be addressed. [Of course, it is also possible to automate the above process in a Visual Basic (or similar) program, if Excel is not being used].

Problems in the data files are usually manifest in one of four ways:

1. A data record has been duplicated.
   Solution: delete the duplicate record.
2. There is one (or many) missing data records.
   Solution: insert the necessary number of rows in the data file, and manually fill in the "Year DOY DOM HOD" columns with the correct time stamp information. Fill the missing data block with values of -9999 (e.g. *Figure 2* the highlighting in this figure is just to make the change more obvious). It should be noted that when there has been a problem with logger communication, the "DOM" value is often wrong for several days. However, so far the DOY record has always been correct. That is why it is valuable to make a simple line plot of these two columns as a check.
3. Data has been corrupted during recording to the logger or download to the PC. Sometimes, during the download (or logging) of one 30 minute sample, the logger will report a block of nonsensical data (e.g. *Figure 3*). This can be a ***very*** large amount, and is usually first noticed by a file size larger than 100-120k of the ASCII file (the bad portion of data is sometimes longer than the whole month of good data).
   Solution: Carefully identify and remove the block of corrupt data. Usually this only results in 1 or 2 missing 30 minute data records. Treat these missing records as in point 2 above.
4. Sometimes the data from the beginning of a month appears to be missing, but it has actually been appended to the end of the previous month's data file.
   Solution: cut and paste the data from the previous file.

Year	DOY	DOM	HOD	Spare	Flow	LLD	ULD	GasM	Pres	HV	ExmA	InmA	Temp	BatV
4	92	1	0	5700	33.9	1904	1392	679	97	604.3	1158	1631	11.1	12.12
4	92	1	30	5598	34.3	1870	1353	687	100	604.3	1159	1631	11.1	12.1
4	92	1	100	5916	34.3	1976	1443	686	100	604.3	1159	1632	11.1	12.1
4	92	1	130	5868	34.2	1960	1390	684	99	604.3	1159	1632	11.1	12.11
4	92	1	200	5679	34.3	1897	1401	687	100	604.3	1159	1632	11.1	12.1
4	92	1	230	5811	34.6	1941	1402	693	101	604.3	1159	1632	11.1	12.13
4	92	1	300	5901	34.6	1971	1444	692	101	604.3	1159	1632	11.2	12.12
4	92	1	330	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999
4	92	1	400	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999
4	92	1	430	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999
4	92	1	500	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999
4	92	1	530	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999
4	92	1	600	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999
4	92	1	630	6126	36.1	2046	1502	723	106	604.3	1157	1631	10.8	12.13
4	92	1	700	5946	37.9	1986	1476	759	111	604.3	1153	1629	11.1	12.12
4	92	1	730	6138	39.5	2050	1471	790	114	604.3	1151	1626	11.9	12.14
4	92	1	800	6177	39.7	2063	1504	795	113	604.3	1153	1624	13.3	12.11
4	92	1	830	6219	40	2077	1458	801	113	604.3	1155	1624	15.1	12.09
4	92	1	900	5937	40	1983	1366	801	112	604.2	1157	1625	17	12.07
4	92	1	930	5925	40.1	1979	1369	803	111	604.3	1158	1626	18.8	12.06
4	92	1	1000	5694	40.1	1902	1285	803	111	604.3	1159	1628	20.4	12.08

Figure 2: Example of missing data record replacement. Six new lines were inserted, the time tag series were filled in manually (green shading) and the missing data records filled with -9999.

Year	DOY	DOM	HOD	Spare	Flow	LLD	ULD	GasM	Pres	HV	ExmA	InmA	Temp	BatV
4	57	26	1000	6936	33.8	2316	1723	676	112	604.2	1152	1621	11.5	12.13
4	57	26	1030	6957	33.9	2323	1736	678	112	604.2	1151	1621	11.2	12.15
4	57	26	1100	7512	33.8	2508	1883	677	114	604.2	1151	1621	11	12.15
4	57	26	1130	7797	34.1	2603	1962	682	116	604.2	1150	1620	10.9	12.12
4	57	26	1200	7995	35.4	2669	2022	708	118	604.2	1149	1619	10.9	12.11
4	57	26	1230	8532	34.8	2848	2088	697	115	604.2	1151	1620	11.1	12.09
4	57	26	1300	8172	34.9	2728	2039	699	115	604.2	1151	1620	11.1	12.11
4	57	26	1330	7968	34.7	2660	1980	695	114	604.2	1151	1621	11	12.14
0	68	26	3.9	1460	58	78	22	4.1	1164	642	0.57	2.18	1+011	2003
0	374	26	34.3	1462	1070	687	122	604	1164	1641	10.2	12.1	01+01	2+200
30	863	26	34.9	1625	1198	699	122	604	1163	1641	10	12.1	01+01	2+200
0	707	26	34.1	1573	1138	682	119	604	1163	1642	9.9	12.1	01+01	2+200
30	641	26	34.2	1551	1154	684	119	604	1162	1641	9.7	12.1	01+01	2+200
0	683	26	34.7	1565	1123	694	120	604	1162	1641	9.5	12.1	01+01	2+200
30	530	26	35.1	1514	1099	703	122	604	1161	1639	9.5	12.2	01+01	2+200
0	845	26	35.3	1619	1213	706	122	604	1162	1639	9.6	12.1	01+01	2+200
30	479	26	35.1	1497	1087	702	121	604	1162	1639	9.6	12.2	01+01	2+200

Figure 3: Example of the start of a corrupt data block.

Despiking

Once the data have been regularised in time they should be checked for spikes. Generate a plot (line plot with markers) of the raw radon data (LLD channel) for each month. Scale the plot so that the typical variability in radon is full scale on the y-axis (ignoring the calibration peaks and -9999 values, e.g. *Figure 4*).

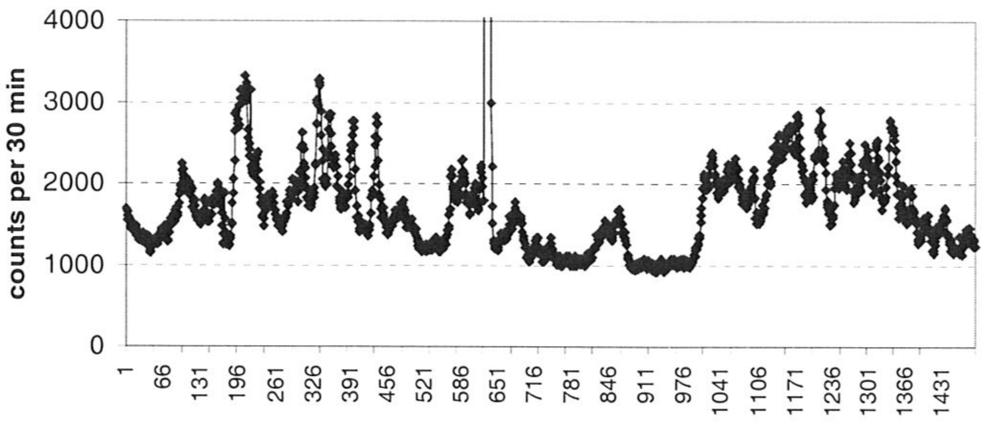

Figure 4: Line plot of 1 months data scaled to ignore the calibration peak near element 650. Note that there are so many points that individual samples are difficult to identify.

It is necessary to perform a visual scan of every sample in the record. To do this, stretch the plot out far enough so that the individual points can be clearly identified. The final plot should be approximately the width of 110 standard Excel columns (e.g. from column **A** to column **DF**). The final data spacing should be similar to that in *Figure 5*. Visually scan the time series checking for obvious spikes each of which could include 1 or several bad data points.

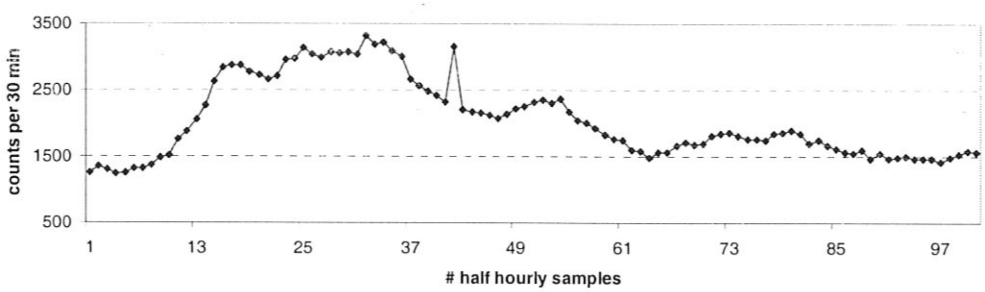

Figure 5: Portion of a 1-month line plot of radon data that has been stretched to 110 columns wide in Excel. Notice that the individual points can now be identified. Abrupt spikes in the data, as shown between points 37 and 49 need to be identified.

The current processing convention is not to modify the raw data. Consequently, when spikes are found, their location should be identified so that the corresponding data points can be removed from the processed data. A convenient way to mark the location of spikes is using an indicator series. An indicator series is a column of zeros and ones, where values of 1 and 0 correspond to good and bad data, respectively.

Data may be "bad" for a variety of reasons, so indicator series are generated to identify bad portions of data for each type of problem. In this manner, it is possible to selectively remove or replace processed data points that are suspected to have been influenced by each process. Each indicator series should be named IndXX, where XX is an incremental integer counter. An Excel comment should be entered in the cell containing the indicator series label to describe the purpose of the particular indicator series.

To generate an indicator series for spikes as observed above: (i) go to the first available blank column in the spreadsheet, (ii) enter the label "Ind01", (iii) right-click to "Insert Comment" and type "Indicator series to identify spikes in raw data", (iv) fill the entire column with the value "1", and (v) each time that a spike is observed (as in *Figure 5*), change the value from "1" to "0" (e.g. *Figure 6*) and highlight the row (as an easy visual cue that this data point has been modified).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	DOY	DOM	HOD	Spare	Flow	LLD	ULD	GasM	Pres	HV	ExmA	InmA	Temp	BatV	Ind01
2	32	1	0	3699	33.4	1237	845	669	114	603.8	1164	1639	8.2	12.17	1
3	32	1	30	3648	33.5	1220	858	671	114	603.8	1164	1639	8.4	12.18	1
4	32	1	100	3750	33.8	1254	859	676	115	603.9	1164	1639	[illegible]	[illegible]	1
5	32	1	130	3723	33.5	1245	895	670	114	603.9	1164	1640	[illegible]	[illegible]	1
6	32	1	200	3627	33.6	1213	839	672	115	603.9	1165	1639	[illegible]	[illegible]	1
7	32	1	230	3564	33.8	1192	833	676	115	603.9	1165	1639	[illegible]	[illegible]	1
8	32	1	300	3366	33.8	1126	791	676	115	603.9	1165	1640	[illegible]	[illegible]	1
9	32	1	330	3336	33.8	1116	764	677	115	603.9	1165	1640	8.9	12.17	1
10	32	1	400	3618	34.2	1210	853	684	115	603.9	1164	1640	8.9	12.15	1
11	32	1	430	4598	34.3	1930	990	686	116	603.9	1165	1639	8.9	12.16	0
12	32	1	500	3780	34.3	1264	882	687	116	603.9	1165	1639	9	12.15	1
13	32	1	530	4014	34.2	1342	956	684	116	603.9	1165	1640	9	12.2	1
14	32	1	600	4347	34.3	1453	1045	686	116	603.9	1165	1639	9.1	12.17	1
15	32	1	630	4557	34.3	1523	1076	687	115	603.9	1165	1640	9.1	12.17	1
16	32	1	700	4650	34.3	1554	1098	687	116	603.9	1165	1640	9.2	12.18	1
17	32	1	730	4611	34.5	1541	1125	690	116	604	1165	1640	9.2	12.18	1
18	32	1	800	4725	34.9	1579	1161	698	118	603.9	1164	1639	9.3	12.16	1
19	32	1	830	4797	35.1	1603	1164	703	118	604	1164	1638	9.5	12.17	1

Indicator series to identify spikes in raw data

Figure 6: Example of an indicator series, comment box and zero value identifying a spike in the raw data.

Concatenation

When all 12 months have been regularised and spikes identified, make a new page in the spreadsheet (call it "30-min Raw"), and concatenate all 12 files to make a complete year of data on the one page. Insert a row at the top of the sheet and copy across the header from one of the other monthly sheets.

Determining calibration coefficients

A calibration is performed by injecting radon from a well characterised source into the detector for 4-6 hours[1]. After a typical injection period, the concentration of radon within the detector takes ~5 hours to return to the ambient radon concentration. Thus the detector is effectively offline for at least 9 hours, or 18 samples (depending on the exact timing of the injection period and ambient radon concentration, this time may vary by several samples).

Each calibration peak needs to be identified, copied and processed, after which that part of the data record is excluded from further analysis, again by using an indicator series.

Generate a new indicator series for the identification of calibration events, and place an appropriate comment label in the heading text. Generate a separate temporary series (called LLD2) that is the product of the original LLD series and the calibration indicator series (which is initially composed entirely of the value "1"). Make a line plot of LLD2. Calibration events will be obvious on this plot. By selecting the graph,

[1] The radon detector data logger schedules calibration events that occur at midnight every 28 days.

and pointing the mouse to a calibration peak, the data record number corresponding to a calibration event will be shown. It is convenient to place the plot of LLD2 in the top half of the screen, and use the Excel "Freeze Panes" option so that the lower half of the screen can move independently to the location of calibration events. In the lower portion of the screen, scroll to the calibration event (or use the Excel GOTO function). Make a quick line plot of about 30-40 samples centred on the calibration peak value. Starting with the initial sharp rise in LLD, insert a block of 18 zeros in the corresponding indicator series (if there was some problem with the timing of the calibration peak, this length may need to be changed).

Insert a new page in the Excel workbook to hold the calibration event data. Starting 3 samples before the calibration event, copy a block of 25 samples (or sufficient samples to include the entire calibration peak and 3 points either side). Paste this calibration peak data into the new calibration page worksheet. Repeat this step for each of the calibration pulses for the year.

The next step is to determine the monthly calibration coefficient. Go to the calibration page in the worksheet. Delete all but the following three columns of data: DOY, LLD and Flow. Derive each calibration coefficient as shown in Appendix 1.

Determination of instrumental background

The next step is to determine the instrumental background. Check in the electronic log file for the dates of background checks. Alternatively, plot the detector flow rate for the year. Adjust the scale to be from -10 to just above the maximum flow rate. There should be several instances over the year when the flow rate is zero (not -9999), but the rest of the detector components are functioning normally (blower current draws should also be close to zero). It is likely that such periods are background checks, and should be approximately 1 day in length.

Make a new page in the spreadsheet and label it "Background". Copy all data from these periods of zero flow rate to this page and delete all but the following columns: **DOY DOM HOD Flow LLD.**

From the time that the detector blowers are initially turned off, it takes at least 5 hours (10 half lives of Pb^{214}, $t_{1/2}$=26.8min) before counts on the head are representative of instrumental background values. To safely account for this effect, delete the first 10 hours of zero flow rate data. Calculate the average count rate of the remaining period of zero flow (last 14 of the 24 hours, total of 28 samples). This is the 30-minute instrumental background. Multiply this value by 2 to get the hourly background. In the 30-minute raw data page, generate an indicator series with values of zero from one sample before, to one sample after each period of zero flow rate for background observations (this is because the instrument is usually turned on/off partway through a 30-minute data period).

Trouble shooting

Apart from problems with the logger, calibration or background events, there are other occasions that may lead to bad or suspicious radon data (e.g. a leak developing in the tank, or a failure in one of the blowers). There are a number of other parameters recorded by the logger that need to be within prescribed bounds (which are detector specific) otherwise the data may be affected. These additional parameters include: flow rate, high voltage (HV) setting, tank pressure, blower current draw and battery voltage.

With the assistance of the person responsible for the detector, determine reasonable bounds for each of these parameters. Then create an indicator series corresponding to each of these parameters. Use a combination of line plots and the Excel data filter function to determine when values from each of these categories are beyond reasonable limits, and adjust their respective indicator series accordingly.

There will now be a number of individual indicator series. Generate one overall (i.e. "master") indicator series as the product of each of the individual series (such that if any one value of the individual series is zero then the master series value will also be zero).

Start a new column in which to apply the master indicator series to the LLD column using a logical command that leaves the values unchanged if the indicator series is "1" and inserts the value of "-9999" if the indicator series has a value of "0". (An example of this command to be entered into the destination cell is:

= IF(XX_i=1, YY_i , -9999)

where XX is the column label of the master indicator series and YY is the LLD column label. This statement can be interpreted as:

"if the i^{th} element of the indicator series is "1" then set the i^{th} element of the current column to the value of the i^{th} element of the LLD column, otherwise set the i^{th} element of the current column to -9999"

Response time

Next it is necessary to account (in part) for the ~45 minute response time of the detector. Make a new page in the spreadsheet and call it "30-min Lagged". Copy the contents of the entire "30-min Raw" sheet into the 30-min lagged sheet. Perform a Paste Special, with VALUES ONLY, so that none of the formulas remain from the original sheet. In this copy of the data, delete all but the following columns: **DOY**, **HOD**, **Temp**, **Flow**, **Pres**, **LLD2** (the new version of LLD after multiplying by the master indicator series), and **Ind** (the master indicator series).

Next, lag the **Flow**, **Pres**, **LLD** and **Ind** columns by one sample (30 minutes). To do this, select the first data value in these 4 columns, then choose EDIT → DELETE. The four columns should all move up by one sample. Go to the bottom of the page. There should be one record missing from the bottom of those 4 columns. Pad this line with -9999 values, and set the value of the indicator series for that sample to zero. **Do not lag the DOY, HOD or Temp columns.**

Aggregation to hourly data

Next the data needs to be aggregated from 30 minute to 60 minute resolution. This can be done in a number of ways: e.g. do the aggregation manually in Excel, use an external time series analysis package, or write a short program. (If you are using Excel, copy the whole page into a new, temporary working sheet until the conversion is complete). The important steps are as follows: (i) consecutive pairs of the **Temp, Flow, Pres and Ind** series need to be **averaged**, and (ii) consecutive pairs of the **LLD** series need to be **summed**. So the number of elements in the file will reduce from 17520 to 8760.

The indicator series then needs to be filtered. All values of less than 1 should be converted to "0". That is, any hourly blocks where one of the two 30 minute readings was bad will be rejected.

Make a new sheet in the workbook and call it "Hourly Radon". Copy the newly generated 8760 element series (Temp, Flow, Pres, LLD and Ind) to this new sheet. Ensure that the data is preceded by the following columns (that need to be generated manually): **Station, Year, Season, Month, Week, Day of year, Day of Month, Hour of day** (e.g. *Figure 7*). This allows for better flexibility in data selection once the data have been imported to a data-base, or Statistica.

Site	Year	Season	Month	WOY	DOY	DOM	HOD	Temp	Flow	Pres	LLD	Ind
HT	2003	Win	Jan	1	1	1	0	22.95	32.45	127	-9999	0
HT	2003	Win	Jan	1	1	1	1	23.3	32.1	126	-9999	0
HT	2003	Win	Jan	1	1	1	2	23.5	32	125.5	-9999	0
HT	2003	Win	Jan	1	1	1	3	23.4	31.8	125	-9999	0
HT	2003	Win	Jan	1	1	1	4	23.4	31.6	124	-9999	0
HT	2003	Win	Jan	1	1	1	5	23.4	31.4	124	-9999	0
HT	2003	Win	Jan	1	1	1	6	23.4	31.4	123.5	-9999	0
HT	2003	Win	Jan	1	1	1	7	23.3	31.35	124	-9999	0
HT	2003	Win	Jan	1	1	1	8	23.25	31.2	123.5	-9999	0
HT	2003	Win	Jan	1	1	1	9	23.1	31.45	124	15218	1
HT	2003	Win	Jan	1	1	1	10	23.15	31.3	124	15363	1
HT	2003	Win	Jan	1	1	1	11	23.25	31.1	123	15518	1
HT	2003	Win	Jan	1	1	1	12	23.5	30.95	122	13788	1
HT	2003	Win	Jan	1	1	1	13	24.4	31	122	11347	1
HT	2003	Win	Jan	1	1	1	14	25.1	31.4	122.5	9123	1
HT	2003	Win	Jan	1	1	1	15	25.65	31.4	123	8178	1
HT	2003	Win	Jan	1	1	1	16	25.55	31.4	123	7920	1
HT	2003	Win	Jan	1	1	1	17	25.55	31.45	123	7928	1
HT	2003	Win	Jan	1	1	1	18	25.35	31.45	123.5	8064	1

Figure 7: Example of data in the new "Hourly Radon" page.

It is likely that summing the pairs of 30-minute LLD data has changed the value of the bad/missing data indicator in some cases. Filter the hourly LLD series for all values less than zero but not equal to -9999, change them to -9999.

Calibration to concentrations

Preparing the hourly background

The next step is to prepare hourly values of the instrumental background for the year. There are three objectives to this background estimation process.

1. Correct the values of background between the last measured value and midnight, 1-Jan of the current year. Until now, these values were extrapolated from the last two known readings;
2. Generate the background values for the present year up to the last recorded measurement;
3. Extrapolate from the time of the last measurement to midnight, 1-Jan of the following year (values to be used until the next background observation is available).

Firstly, determine the most recent background value from the previous year's data (eg. September in *Figure 8*). Then convert the 30 minute background values determined from the current year to hourly values.

As shown in *Figure 8*, make a list of months from the last measurement of the previous year to the end of the current year. Enter the background estimates in the appropriate months. Use the Excel interpolation function (Edit : Fill : Series) to linearly interpolate between the first two and then the last two points separately. Using the data interval (slope) determined from the second interpolation, extrapolate the monthly background estimates forward to January of the following year. Take the average of the December-January pairs at the start and end of the year to get the starting and ending background values for the year.

Date	60min BG
Jul-03	
Aug-03	
Sep-03	131.2
Oct-03	
Nov-03	
Dec-03	
Jan-04	
Feb-04	
Mar-04	152.1
Apr-04	
May-04	
Jun-04	
Jul-04	
Aug-04	
Sep-04	167.5
Oct-04	
Nov-04	
Dec-04	

Date	60min BG	
Jul-03		
Aug-03		
Sep-03	**131.2**	
Oct-03	134.7	
Nov-03	138.2	
Dec-03	141.7	
Jan-04	145.1	***143.4***
Feb-04	148.6	
Mar-04	**152.1**	
Apr-04	154.7	
May-04	157.2	
Jun-04	159.8	
Jul-04	162.4	
Aug-04	164.9	
Sep-04	**167.5**	
Oct-04	170.1	
Nov-04	172.6	
Dec-04	175.2	
Jan-05	177.8	***176.5***

Figure 8: Example of the steps involved in determining the instrumental background at a site for the year 2004.

Go to the "Hourly Radon" page in the spreadsheet. Make a new column called background. Enter the first Dec-Jan average value as the first element in this column (Midnight, 1-January), and enter the second Dec-Jan average value as the last element in this column (23:30, 31 December). Enter the values for the other 1 or 2 estimates that were made during the year at times/dates that correspond to the middle of the respective background measurement periods. Use the Excel interpolation function to interpolate between each pair of points to generate an hourly background series for the year. To correct last years values, interpolate between the last real measurement, and the first of the Jan-Dec average figures calculated for the present year and replace those numbers in the previous years file.

When several years of radon data are available, an alternative method to estimate the background is to make a plot of background count vs number of days since commissioning. Then fit a line to this plot, and use the slope (together with the background at the time that the detector was commissioned) to estimate all future background values.

Preparing hourly calibration coefficients

Make a new column and copy the monthly calibration coefficients to their appropriate data locations in the series. Linearly interpolate between pairs of calibration coefficients.

To obtain values prior to the first calibration of the month, interpolate between this value and the last coefficient of the previous year (if it is the first year of data, assume a constant calibration coefficient from the time of commissioning until the time of the first calibration pulse).

For the time being, assume a constant calibration coefficient between the last value and the end of the year. Correct this portion of the previous year's data using the values obtained from the interpolation between the first calibration of the current year, and last calibration of the previous year.

Deriving concentrations

The next step is to convert the raw counts (LLD) into hourly radon concentrations (^{222}Rn). The background (BG) is removed, and the calibration coefficient (C_{Rn}) applied as indicated below:

$$^{222}Rn = 1000.\left(\frac{LLD - BG}{3600}\right).\left(\frac{1}{C_{Rn}}\right)$$

It is useful to perform this as a two step process (i.e. remove the background, then apply the calibration coefficient). Make a new column and calculate LLD-BG. Then make a second column to calculate the radon concentration.

Next it is necessary to estimate the accuracy of the hourly measurements. For this purpose, the standard deviation of hourly counts is used. Since radioactive decay counting is a Poisson process, the square root of the hourly count yields the standard deviation. The standard deviation of the hourly counts is expressed in two ways: (1) as a concentration (mBq/m^3), and (2) as a percent of the net count.

$$sd(mBq/m^3) = 1000\left(\frac{\sqrt{LLD}}{3600}\right)\cdot\left(\frac{1}{C_{Rn}}\right)$$

$$sd(\%) = 100\left(\frac{\sqrt{LLD}}{LLD - BG}\right)$$

Once the calibration has been performed it is necessary to check for calibrated radon values below zero (but not equal to -9999). These occur when radon concentrations are very low. Set these values to 1 mBq m^{-3}. Removing the estimated background count sometimes results in slightly negative net counts. This is because we estimate the background as an average over approximately 1 or more days of half-hourly values. Statistical variability, as well as possible diurnal change in the background value, will occasionally result in the estimated background being larger than the observed signal at very low concentrations.

Lastly it is necessary to generate a "readme" title page for the spreadsheet to make the file self contained. Should the file need to be referred to at a later date by a different person, they should be able to get all the information they require about the dataset from that page. An example of such a page is presented in Appendix 2.

The final hourly radon data page can then be exported Statistica (or similar package) for further analysis.

Appendix 1

Determining the monthly calibration coefficient.

Arrange the remaining data in the calibration page of the worksheet such that there are 12 adjacent blocks of data like that shown to the right, one for each event.

There are four important portions of the calibration data block, which have been coloured in the example opposite for clarity.

1. Counts at ambient radon concentrations prior to the calibration pulse (first 3 points in blue).
2. Counts during the injection phase (4-6 hours when flow from the source is directed into the detector - coloured in pink). There is usually little response to the injection for the first 30 minutes.
3. Counts during the flushing phase (coloured in green), usually lasts 4-5 hours unless ambient radon concentrations are near baseline (very low), in which case it may take ~6 hours.
4. Counts at ambient radon concentration after the calibration event (last 3 points in blue).

DOY	LLD	Flow
55	1850	32.2
55	2056	32.4
55	2181	32.4
56	2298	31.9
56	5435	32.3
56	15593	32.2
56	21773	32.0
56	26813	32.1
56	29523	32.0
56	30943	32.1
56	31938	31.8
56	**32243**	32.0
56	29387	31.5
56	20410	31.2
56	13845	31.1
56	9377	31.2
56	6305	31.4
56	4617	31.6
56	3711	31.2
56	3189	31.0
56	2894	30.9
56	2690	31.0
56	2719	31.1
56	2717	30.9
56	2589	30.8

There are a number of steps that need to be performed on each calibration pulse data set. An example of the results of these steps for the data above is given here to the right.

Pre-calibration ambient	2029.0	cp30min
Post-calibration ambient	2675.0	cp30min
Change in ambient	646.0	cp30min
Time period of cal	19.0	# samples
Change in ambient per sample	34.0	
Calibration peak counts	32243	cp30min
# samples to calibration peak	9.0	
Ambient at cal peak	2335.0	cp30min
Net calibration peak	29908.0	cp30min
Peak multiplier	1.07	
Adjusted Peak	32001.6	cp30min
Peak counts per second	17.8	cps
Source Strength	21.2	kBq
Radon yield	2.67	Bq/min
Flow @ max	32.0	L/min
Conc. Rn-222 @ max	83.5	Bq/m3
Calibration Factor	0.213	cps/Bq/m3

1. Calculate the average 30-minute count for ambient radon concentration prior to the calibration.
2. Calculate the average 30-minute count for ambient radon concentration after the calibration.
3. Calculate the change in ambient count rate over the calibration period.
4. Determine the length (in number of 30 minute samples) of the calibration event (injection + flushing periods).
5. The ratio of the change in count rate to length of the calibration period yields the change in count rate per 30-minute sample during the calibration period.
6. Find the maximum count rate over the calibration period.
7. Count the number of samples from the start of the injection cycle to the peak (maximum) count rate.

8. Use the initial ambient count rate and the change in ambient per 30-minute block to determine the ambient concentration at the time that the peak count rate was observed (ie. Assuming a linear change in ambient concentration with time).
9. Subtract the ambient count rate from the maximum count rate to yield the net peak counts.
10. Multiply the net peak counts by the peak multiplier to account for the fact that the radon concentration within the detector tank would not quite be in equilibrium with the source after 4 hours.
11. Calculate the number of counts per second for the adjusted net peak 30-minute count rate.
12. Note the radon source strength and yield.
13. Read off the flow rate through the detector at the time of maximum counts.
14. The ratio of the radon yield (Bq/min) to the flow rate (L/min) gives the concentration (Bq/L) inside the detector. Convert this to Bq/m^3 by multiplying by 1000.
15. The ratio of peak counts per second to the radon concentration in the detector gives the calibration coefficient (cps/Bq m^{-3}).
16. Compare the 12 coefficients. They should all be fairly similar (a fluctuation within about 5-10% of their overall magnitude is normal). Over a long period (1 or several years), a gradual reduction in the sensitivity of the head to radon will be observed. Investigate calibration periods that yield particularly strange coefficients, and omit them if necessary.

This method was devised for the Cape Grim detector (where there is not a large diurnal change in ambient radon concentration – thus, the assumption of linear change in ambient concentration from start to end of the calibration period is relatively accurate). For continental sites, particularly where strong nocturnal inversions are encountered, this assumption may not be as valid. Furthermore, under these conditions, the three points before and after the calibration peak may not be similar (ie. The ambient concentration leading up to, or from the calibration pulse could be changing rapidly). Under these conditions, the pre-calibration and post-calibration ambient values may be better approximated by using the single closest non-calibration value to the peak.

Appendix 2

Example contents of a "ReadMe" Title page for processed radon spreadsheet.

General Notes

This file contains hourly radon concentrations (mBq m^{-3}) and related data recorded at {Station Name} for the year 200?.

Time stamp

All times reported are in local time. The convention adopted for the data time stamp is such that the time stamp is recorded at the end of the observation period. That is, the 12:00 observation represents the calibrated net counts summed between 11:01 - 12:00.

Invalid data

All periods of invalid or missing data are flagged with -9999.

Experimental error

For concentrations below XXX mBq m^{-3}, the statistical counting error is at least three times larger than the error resulting from uncertainty of the strength of the calibration source, and is therefore the main source of experimental error in this concentration range. The error is shown for all hourly concentrations as standard deviation of hourly total counts (expressed in mBq m^{-3}). The uncertainty associated with the calibration source, which is traceable to a NIST standard, is estimated to be +- 4%. For concentrations above XXXX mBq m^{-3} the statistical error becomes equal to, and then progressively smaller than, the error associated with the calibration source. The lower limit of detection (radon concentration at which there is a counting error of 30%) is XX mBq m^{-3}.

Concentration Corrections
Data have not been STP corrected. Due to the continual flushing of the tank, tank air temperature is approximately equivalent to outdoor ambient temperature. Tank pressure is approximately 100 Pa above ambient pressure.

Principle of operation
Details can be found in :
Whittlestone S. and Zahorowski, W. (1998) Baseline radon detectors for shipboard use: Development and deployment in the First Aerosol Characterisation experiment (ACE 1). J. Geophys. Res. 103, 16,743-16,751.

Contact persons

Dr Wlodek ZAHOROWSKI
Address: Australian Nuclear Science and Technology Organisation (ANSTO), PMB1, Menai, NSW, 2234, Australia
E-mail: wza@ansto.gov.au
Phone: +61 2 9717 3804
Fax: +61 2 9717 9260

or

Dr Scott CHAMBERS
Address: Australian Nuclear Science and Technology Organisation (ANSTO), PMB1, Menai, NSW, 2234, Australia
E-mail: szc@ansto.gov.au
Phone: +61 2 9717 3058
Fax: +61 2 9717 9260

Acknowledgements
If the dataset included in the file is to be used in any form in a report or publication, both {Local collaborator} and ANSTO as well as the contact persons shown above should be acknowledged.

Users of the dataset are encouraged to discuss any aspects of the dataset with the contact persons.

Other notes
(mention any peculiarities of the dataset, or comments made by station staff)