Gravity Modelling–Forward Modelling Of Synthetic Data

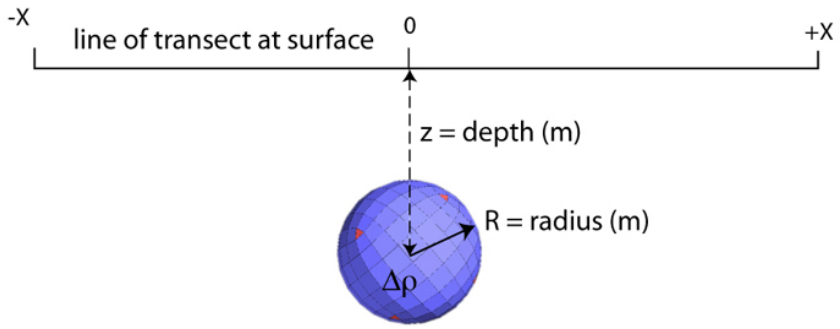
After completing this practical you should be able to:

The aim of this practical is to become familiar with the concept of **forward modelling** as a tool for interpreting geophysical anomalies. We will do this by implementing a forward model (using Excel) which predicts the Bouguer gravity anomaly for a spherical object of specified density (relative to surrounding rock), depth and size (radius). We will use **standard statistical tests** to quantify how well, or not, a particular model fits a set of observed data.

After completing this practical you should be able to:

- Use Excel to perform complex calculations multiple times using formulae.
- Use the spherical forward model to predict the gravity anomaly measured at the surface generated by a buried spherical object of any size, density or depth.
- Be able to use a simple forward model to estimate the physical and geological parameters of an object consistent with the measured gravity anomaly across that object.
- Use standard statistics such as the mean, standard deviation and root mean square deviation to assess the "goodness of fit" between model predictions and observations.

Task 1.
Implement a model which predicts the Bouguer gravity anomaly measured at the surface along a transect across the centre of a buried spherical object (*please review pages 245-247 in Chapter 8 of Lillie for additional details*).



$\Delta\rho$ = density difference (g.cm^{-3}), i.e. density of object - density of country rock.

X = distance (m) measured along transect with origin (i.e. zero) centered over centre of mass of object.

The relevant equation we need to use is;

$$gz(x) = 0.02794\,\Delta\rho\,R^3\left(z/(x^2 + z^2)^{3/2}\right)$$

where gz(x) is the predicted model Bouguer gravity anomaly at position x. The x position is the distance along the traverse measured from the zero point which is located above the centre of our model sphere (see cartoon above). To calculate this

anomaly at any position x we also need to define a value for $\Delta\rho$, the density contrast, R, the radius of the sphere and z, the depth to the centre of the sphere. So as we move from one position to the next along the transect the x position changes (and so the x variable is incremented/changed) but the variables for density contrast, radius and depth do not change (because we want to calculate gz at all positions for the same spherical object).

To begin with let us set up a model so that it calculates the gravity value at a set of stations along a 100 km transect beginning at -50km and ending at +50km with measurements made at stations at 2km intervals (i.e. -50, -48, -46…0…46, 48, 50 km). So to begin with we need to set up a column in Excel to store our x-position values, and to make sure we convert the km values to metres (see below).

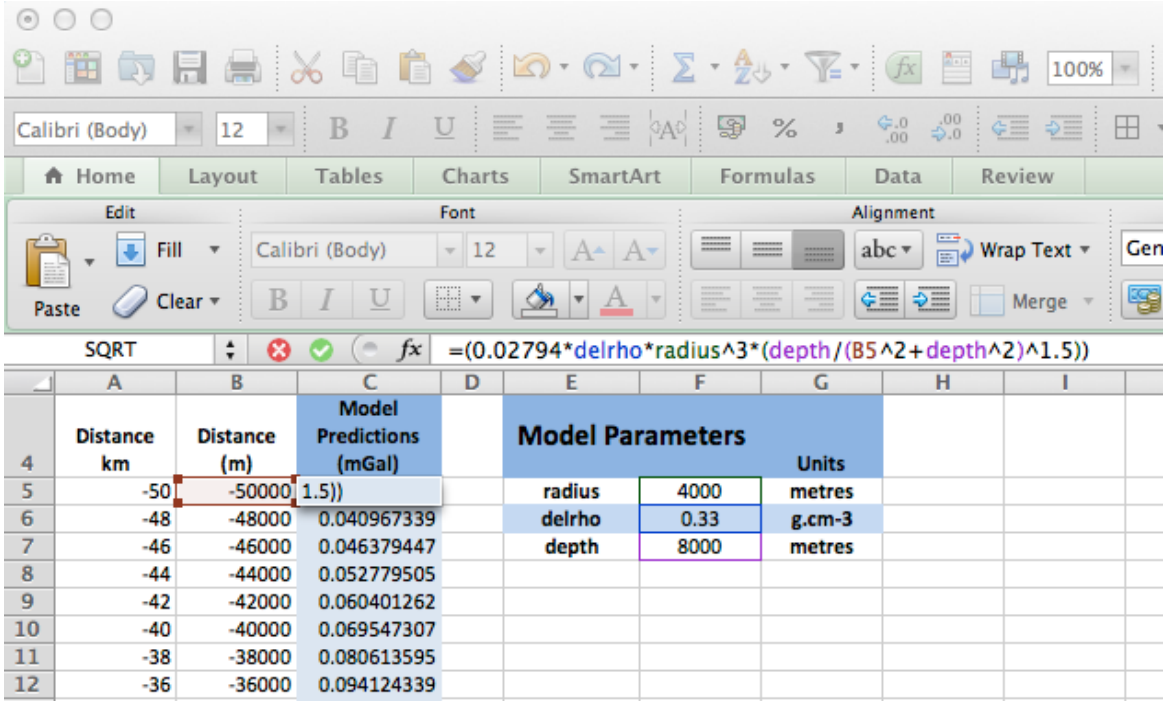
x position
(use cell reference for cell in Excel column of x-positions)

$$g_z(x) = 0.02794\ \Delta\rho\ R^3\ (z/(x^2 + z^2)^{3/2})$$

delrho (density contrast) **radius** **depth** — these 3 parameters we will declare as named variables in the workbook

Note: This equation is scaled for use with $\Delta\rho$ expressed in units of **g.cm^{-3}** and distances **R, x** and **z** in **metres**.

Once we have a column of x position values we then need to calculate the Bouguer gravity anomaly value, gz, for each of those positions using the equation above. The x-position value in the equations is taken directly from our x-position column using the appropriate cell reference. In the example below, for the first x-position this is cell B5.



	A	B	C	D	E	F	G	H	I
4	Distance km	Distance (m)	Model Predictions (mGal)		Model Parameters		Units		
5	-50	-50000	1.5))		radius	4000	metres		
6	-48	-48000	0.040967339		delrho	0.33	g.cm-3		
7	-46	-46000	0.046379447		depth	8000	metres		
8	-44	-44000	0.052779505						
9	-42	-42000	0.060401262						
10	-40	-40000	0.069547307						
11	-38	-38000	0.080613595						
12	-36	-36000	0.094124339						

A useful and elegant way to deal with the other three model parameters, i.e. the **density contrast**, the **radius** and the **depth** of the sphere is to declare these as named variables. This means that instead of a cell reference, or actual number, we will tell Excel that we are going to use a name to refer to a variable value. And, whenever this name is used in an equation Excel should use the value stored in a cell that is explicitly linked to that name. This means we can use variable names directly in writing our equations (as shown in the example above).

So, in our case, let's use the words *delrho*, *radius* and *depth* as our variable names. As shown in the example above, type these names into three cells within a blank column. Next to each name, in the adjacent column, enter an initial value for this variable (can be any value, so just use the ones in the example for now). Including the units for each variable is also good practice (in a separate third column as above). Now all we need to do is tell Excel that each of our named variables is linked to each of the adjacent cells. To do this select the cell to the right of the variable name, right click with the mouse and select **Define name …** from the pop-up menu (so F7 for depth, F6 for delrho and F5 for radius in this example). Check that the correct name is being linked to the correct cell reference, and if correct, click OK. Do the same for each of the other named variables. Note there are notes on how to do this at the end of the notes for the first laboratory session, and there is also a screen capture video (via YouTube) with commentary on how to do this accessible via links on the Moodle page for this lab.

Once you have defined each of the three named variables we need you can now enter the equation for the first x-position as in the example above. So note that we refer to the position x in the usual way, i.e. by using the cell reference that points to the x-values in our x-column (here column B), BUT we refer to the other three model variables using their names (that we defined earlier) as shown in the example above. Once you've entered the equation for the first position then simply select the cell containing the result (here C5) and do a "fill down" by selecting the right hand corner of the cell and dragging down to the end of your x positions. This will propagate the equation, incrementing the x position value as required but importantly NOT incrementing the values for delrho, radius and depth…which is exactly what we want.

Once you have your model working and predicting a model gz value for each of the x-positions, plot a simple x-y graph showing the predicted gravity value on the y-axis and the x-position on the x-axis. This should plot as a smooth bell shaped curve (see example in Task 3).

Task 2.

To evaluate how the predicted gravity anomaly changes as a function of each of the three model variables it is useful to examine a model/s where we change only one model variable at a time. Use your model to plot graphs to illustrate the predicted Bouguer gravity anomaly profile along the transect for the following conditions and make notes about the relative amplitude and shape of the anomaly for and between each example. ***Excel Tip:*** *Use the* ***Format axis*** *option on your chart (select and right click on y-axis in chart frame) to fix the maximum and minimum values scale on the y-axis. This will prevent the chart from rescaling each time, which obscures the changes you wish to view.*

Depth	Radius	Density Contrast	
10 km	4km	0.33	
20 km	4km	0.33	**Variable depth**
30km	4km	0.33	
10 km	4km	0.33	
10 km	2km	0.33	**Variable size (radius)**
10 km	1km	0.33	
20 km	4km	0.66	
20 km	4km	0.33	**Variable rock (density contrast)**
20 km	4km	0.16	

Specific behaviour to record might include;

1. With radius (R) and density contrast ($\Delta\rho$) held constant does the **amplitude** of the predicted gravity anomaly increase or decrease with increasing depth (z)? **Why** does it do what it does?

2. With radius (R) and density contrast ($\Delta\rho$) held constant does the **width** of the predicted gravity anomaly increase or decrease with increasing depth (z)? **Why** does it do what it does?

3. With depth (z) and density contrast ($\Delta\rho$) held constant does the **amplitude** of the predicted gravity anomaly increase or decrease with increasing radius (R)? **Why** does it do what it does?

4. With depth (z) and density contrast ($\Delta\rho$) held constant does the **width** of the predicted gravity anomaly increase or decrease with increasing radius (R)? **Why** does it do what it does?

5. With depth (z) and Radius (R) held constant does the **amplitude** of the predicted gravity anomaly increase or decrease with decreasing density constrast ($\Delta\rho$)? **Why** does it do what it does?

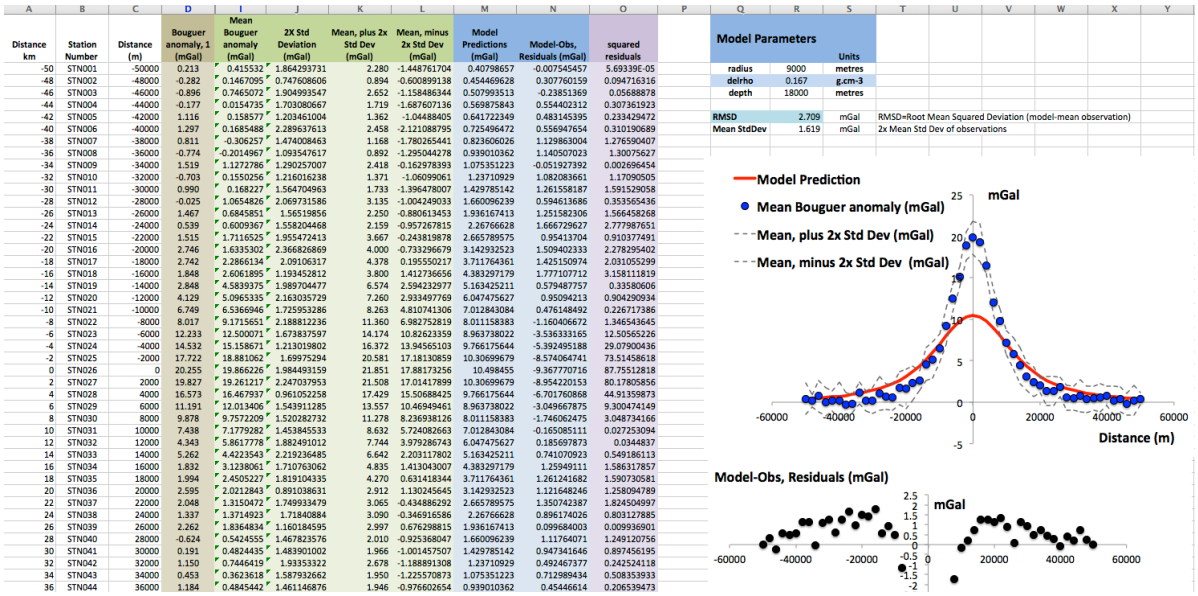
6. With depth (z) and Radius (R) held constant does the **width** of the predicted gravity anomaly increase or decrease with decreasing density constrast ($\Delta\rho$)? **Why** does it do what it does?

7. Is the response to increasing or decreasing radius (R) and density contrast ($\Delta\rho$) similar, if so why?

Task 3.

Down load the spherical model synthetic data from the Moodle site. The data are provided as a simple ascii text file (no formatting) and as an Excel file. The simplest way to open the ascii text file is to click the link, then select and copy the data (make sure you select ALL the columns and headings) and then paste this into an empty Excel workbook page (use *Paste Special->Ascii text*). Or simply download the Excel format file and open this.

Use your forward model you developed for Task 2 above to estimate the depth and radius of a spherical object that is consistent with the observed gravity data (initially, assume a density contrast, $\Delta\rho$ of 0.33). An example layout is shown below.



Distance km	Station Number	Distance (m)	Bouguer anomaly, 1 (mGal)	Mean Bouguer anomaly (mGal)	2X Std Deviation (mGal)	Mean, plus 2x Std Dev (mGal)	Mean, minus 2x Std Dev (mGal)	Model Predictions (mGal)	Model-Obs, Residuals (mGal)	squared residuals
-50	STN001	-50000	0.213	0.415532	1.864293731	2.280	-1.448761704	0.40798657	-0.007545457	5.69339E-05
-48	STN002	-48000	-0.282	0.1467095	0.747608606	0.894	-0.600899138	0.454469628	0.307760159	0.094716316
-46	STN003	-46000	-0.896	0.7465072	1.904993547	2.652	-1.158486344	0.507993513	-0.23851369	0.05688878
-44	STN004	-44000	-0.177	0.0154735	1.703080667	1.719	-1.687607136	0.569875843	0.554402312	0.307361923
-42	STN005	-42000	1.116	0.158577	1.203461004	1.362	-1.04488405	0.641722349	0.483145395	0.233429472
-40	STN006	-40000	1.297	0.1685488	2.289637613	2.458	-2.121088795	0.725496472	0.556947654	0.310190689
-38	STN007	-38000	0.811	-0.306257	1.474008463	1.168	-1.780265441	0.823606026	1.129863004	1.276590407
-36	STN008	-36000	-0.774	-0.2014967	1.093547617	0.892	-1.295044278	0.939010362	1.140507023	1.30075627
-34	STN009	-34000	1.519	1.1272786	1.290257007	2.418	-0.162978393	1.075351223	-0.051927392	0.002696454
-32	STN010	-32000	-0.703	0.1550256	1.216016238	1.371	-1.06099061	1.23710929	1.082083661	1.17090505
-30	STN011	-30000	0.990	0.168227	1.564704963	1.733	-1.396478007	1.429785142	1.261558187	1.591529058
-28	STN012	-28000	-0.025	1.0654826	2.069731586	3.135	-1.004249033	1.660096239	0.594613686	0.353565436
-26	STN013	-26000	1.467	0.6845851	1.56519856	2.250	-0.880613453	1.936167413	1.251582306	1.566458268
-24	STN014	-24000	0.539	0.6009367	1.558204468	2.159	-0.957267815	2.26766628	1.666729627	2.777987651
-22	STN015	-22000	1.515	1.7116525	1.955472413	3.667	-0.243819878	2.665789575	0.95413704	0.910377491
-20	STN016	-20000	2.746	1.6335302	2.366826869	4.000	-0.733296679	3.142932523	1.509402333	2.278295402
-18	STN017	-18000	2.742	2.2866134	2.09106317	4.378	0.195550217	3.711764361	1.425150974	2.031055299
-16	STN018	-16000	1.848	2.6061895	1.193452812	3.800	1.412736656	4.383297179	1.777107712	3.158111819
-14	STN019	-14000	2.848	4.5839375	1.989704477	6.574	2.594232977	5.163425211	0.579487757	0.33580606
-12	STN020	-12000	4.129	5.0965335	2.163035729	7.260	2.933497769	6.047475627	0.95094213	0.904290934
-10	STN021	-10000	6.749	6.5366946	1.725953286	8.263	4.810741306	7.012843084	0.476148492	0.226717386
-8	STN022	-8000	8.017	9.1715651	2.188812236	11.360	6.982752819	8.011158383	-1.160406672	1.346543645
-6	STN023	-6000	12.233	12.500071	1.673837597	14.174	10.82623359	8.963738022	-3.536333165	12.50565226
-4	STN024	-4000	14.532	15.158671	1.213019802	16.372	13.94565103	9.766175644	-5.392495188	29.07900436
-2	STN025	-2000	17.722	18.881062	1.69975294	20.581	17.18130859	10.30699679	-8.574064741	73.51458618
0	STN026	0	20.255	19.866226	1.984493159	21.851	17.88173256	10.498455	-9.367770716	87.75512818
2	STN027	2000	19.827	19.261217	2.247037953	21.508	17.01417899	10.30699679	-8.954220153	80.17805856
4	STN028	4000	16.573	16.467937	0.961052258	17.429	15.50688425	9.766175644	-6.701760868	44.91359873
6	STN029	6000	11.191	12.013406	1.543911285	13.557	10.46949461	8.963738022	-3.049667875	9.300474149
8	STN030	8000	9.878	9.7572209	1.520282732	11.278	8.236938126	8.011158383	-1.746062475	3.048734166
10	STN031	10000	7.438	7.1779282	1.453845533	8.632	5.724082663	7.012843084	-0.165085111	0.027253094
12	STN032	12000	4.343	5.8617778	1.882491012	7.744	3.979286743	6.047475627	0.185697873	0.0344837
14	STN033	14000	5.262	4.4223543	2.219236485	6.642	2.203117802	5.163425211	0.741070923	0.549186113
16	STN034	16000	1.832	3.1238061	1.710763062	4.835	1.413043007	4.383297179	1.25949111	1.586317857
18	STN035	18000	1.994	2.4505227	1.819104335	4.270	0.631418344	3.711764361	1.261241682	1.590730581
20	STN036	20000	2.595	2.0212843	0.891038631	2.912	1.130245645	3.142932523	1.121648246	1.258094789
22	STN037	22000	2.048	1.3150472	1.749933479	3.065	-0.434886292	2.665789575	1.350742387	1.824504997
24	STN038	24000	1.337	1.3714923	1.71840884	3.090	-0.346981586	2.26766628	0.896174026	0.803127885
26	STN039	26000	2.262	1.8364834	1.160184595	2.997	0.676298815	1.936167413	0.099684003	0.009936901
28	STN040	28000	-0.624	0.5424555	1.467823576	2.010	-0.925368047	1.660096239	1.11764071	1.249120756
30	STN041	30000	0.191	0.4824435	1.483901002	1.966	-1.001457507	1.429785142	0.947341646	0.897456195
32	STN042	32000	1.150	0.7446419	1.93353322	2.678	-1.188891308	1.23710929	0.492467377	0.242524118
34	STN043	34000	0.453	0.3623618	1.587932662	1.950	-1.225570873	1.075351223	0.712989434	0.508353933
36	STN044	36000	1.184	0.4845442	1.461146876	1.946	-0.976602654	0.939010362	0.45446614	0.206539473

Model Parameters		
		Units
radius	9000	metres
delrho	0.167	g.cm-3
depth	18000	metres

RMSD	2.709	mGal	RMSD=Root Mean Squared Deviation (model-mean observation)
Mean StdDev	1.619	mGal	2x Mean Std Dev of observations

This can be achieved by plotting the output of your model on the same graph as the observed gravity data (use the Series option when inserting your chart/graph to add more than one set of data to the x-y graph) and varying the model input parameters until the predicted model profile is similar to the observed profile (visual assessment).

Make sure that you use ALL the data (five values for each station) and a **valid statistical assessment** of how well your model fits the observations (see notes at end of this manual). Five measurements are available at each gravity station along the transect so should calculate the mean value of the anomaly at each station along with its standard deviation, and compare your predicted model values with the mean observed value.

Summary of statistics, residuals and RMSD as a model misfit estimate

The **mean**, μ, of a sample of N values of x drawn from a population (e.g. a number of measurements of gz made at a single gravity station) is;

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

The mean value is an *estimate* of the true value (which we wish to know, but do not know) based on the measurements we have (this assumes that the measured data are normally distributed about a single, discrete true value). In Excel you can use the AVERAGE function to calculate the arithmetic mean of a range of values in a column or row, e.g. =AVERAGE (C2:C52) will return the average value of all numbers stored in rows 2 to 52 of column C, or =AVERAGE(B2:G2) would return the mean of all the numbers in row 2 across columns B, C, D, E and G.

The **variance** of the sample is similar to the mean squared difference from the mean value for an unbiased data set (note use of N-1 not N for a sample);

$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu)^2$$

The **standard deviation** of the sample is just the square root of the variance, i.e. σ. This is a good estimate of the 'spread' of the data around the mean value, and represents the standard (or average) variation between the individual observations and the mean value. This is a good statistic for quantifying the uncertainty, or likely range, for the true value we are trying to estimate using the mean. We can say that there is a 67% probability of the true value falling within the range of the estimated mean ± 1σ, and a 95% probability of the true answer falling within the range of the mean value ± 2σ.

The Excel function STDDEV.S can be used to calculate the standard deviation of a range of values, e.g. =STDEV.S(C2:C52) will return the sample standard deviation for the numbers in column C from row 2 to row 52.

The **standard error of the mean** is an estimate of the reliability of the mean estimate made from the sample (it is sort of like the standard deviation of multiple different estimates of the true mean, based on estimates made from different samples drawn from the true population of measurements) and is given by;

$$se = \frac{\sigma}{\sqrt{N}}$$

You could calculate the standard error of the mean of a set of values stored in column C, rows 2 to 52 using Excel using; =STDDEV.S(C2:C52)/(SQRT(COUNT(C2:C52))). The function COUNT(cell1:cell2) returns the number of values between cell1 and cell2, inclusive, so here it returns the number of values, N.

An estimate of how well a model fits a set of observed data can be obtained using a similar measure to the standard deviation of a sample. In this case though we calculate the **root mean squared deviation** value, i.e. the square root of the mean squared difference between the model predictions m_i and the observed data, x_i.

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(m_i - x_i)^2}$$

The so called "**Root Mean Squared Deviation**", **RMSD**, is an estimate of the average deviation (i.e. difference or 'error') between our model predictions, m_i, at the data locations and the observations, x_i, made at those same locations. The difference between the model predictions m_i and the observations x_i at each location are called the **residuals**.

A plot of the residuals, on the y-axis, versus the data locations (i.e. the x positions) on the x-axis provides a useful tool for assessing the quality of your model. The ideal model will have low residuals (i.e. small difference between the predicted model values and the actual observed values) AND the residuals should be randomly distributed around the zero value along the whole transect. Poor fitting models will likely show structure, or a pattern, to the distribution of the residual values. The RMSD value is thus calculated as the square root of the average of all the squared residuals determined at every point in the model (i.e. model prediction minus observation, squared). NOTE: There is just a single RMSD value (one number) for the whole model, not one for each data location/point. This value is effectively a measure of the average deviation (i.e. error, or misfit), between our model and the observations, averaged over all data points, N.

The objective of modelling is to minimise this value, that is, find a model with the smallest overall error or misfit. It provides a useful way of quantifying how well a particular model fits the data, and for ranking different but perhaps quite similar looking models…so the model with the smaller RMSD is "better" than the one with a larger RMSD. Remember we need to also examine the distribution, or pattern, of the residuals in assessing how "good" a model might be. The RMSD could be calculated using Excel functions. First calculate the **squared residuals** for each location and store these in a new column, say column O, e.g. =(M2-N2)^2, (fill down for column O) where your model predictions are in column M and your mean observations are in column N.

Then the RMSD for the whole model is given by (store value in a new cell);

=SQRT(AVERAGE(O2:O52))

How do you decide if a good model is good enough? Once you find a model that looks good (i.e. has a small RMSD and randomly distributed residuals) how do you decide if it is acceptable or not? A simple way to do this in a quantitative manner is to simply compare the RMSD value (i.e. the average error between your model and the observations) and the average error (or uncertainty) on the actual observations (i.e. the average standard deviation on the observations). If the RMSD value is less than or similar to the average uncertainty (i.e. average std dev) on your observations then the model is acceptable.