Data Processing and Curve Fitting of Counts Data

Liam Taylor, Gordon McIntosh, and Alaina Swanson
Division of Science and Mathematics, University of Minnesota, Morris, MN, 56267

Erick Agrimson and Kaye Smith
Department of Physics, St. Catherine University, St. Paul, MN 55105

Abstract

Various data analysis methods were explored to more accurately and consistently determine the Regener-Pfotzer (RP) maxima for high altitude cosmic ray intensity. The radiation has been measured during 15 balloon flights using Geiger-Mueller (GM) counters with five second accumulation times. Of the 15 flights, 10 of them included omnidirectional counts data, and 8 of them included vertical coincidence counts data. Count data from altitudes greater than 10 km were analyzed to determine the maxima. The data analysis methods used were moving average filtering and summation of GM counts into one minute intervals. Moving average filtering did not give reliable results, so the summation method was chosen. Once the data were summed, several different curves were fit to determine the RP maximum. The curves tested include second and third order polynomials as well as cubic spline interpolation of the data averaged over 1 km intervals. Second order polynomial fitting did not fit the data well. Third order polynomials and cubic splines gave better results. Third order polynomial fitting was chosen due to its ease of use and the similarity of the results given by the cubic spline interpolation (within 1%). The omnidirectional RP maxima occurred at an average altitude of 21.8 km ± 1.7 km, while the vertical coincidence RP maxima occurred at an average altitude of 18.5 km ± 1.1 km.

Background

Balloon payloads launched by the University of Minnesota, Morris and St. Catherine University are instrumented with GM counters, temperature, and pressure sensors to determine the RP maxima. However, they use miniature GM tubes as a part of their payload, while the University of Minnesota, Morris and St. Catherine University both use AWARE Electronics Model RM-60 and RM-80 GM counters. The RM-60 and RM-80 GM counters have much larger collection areas than a miniature GM tube. As such, a moving average filter is not an applicable method of data smoothing in this case.

The next method of data processing was reducing the counts into counts per minute in Excel using the equation

\[ \text{SUM(OFFSET(input cell,ROW(Sreference cell))\text{reference cell}1\text{m-n-a})} \]

where \( n \) is the number of cells added together, the input cell is the first cell containing counts data, and the reference cell is the cell containing the equation. All of the bolded cell are static cells. To find the corresponding altitude, we modified equation 1 to be

\[ \text{AVGARE(OFFSET(input cell,ROW(Sreference cell))\text{reference cell}1\text{m-n-a})} \]

(Eqn 2).

With this, we were able to generate altitude vs. counts per minute plots. Once the data were processed, several curves were fit. Second and third order polynomial fitting was done using Excel, and cubic spline interpolation\(^3\) fitting was done using MATLAB. Second order polynomial fitting was ineffective at determining the RP maxima, while third order polynomials and cubic spline interpolation were able to fit the data. Third order polynomial fitting was chosen over cubic spline interpolation due to its ease of use. Cubic spline interpolation required additional data wrangling to be effective, and the RP maxima determined by the curves were within a 1% difference. Once a curve is fit to the data, the RP maximum is determined by finding where the slope of the fitted curve is zero.

Data Processing and Curve Fitting

The first method of data processing used was a moving average filter. This methodology was used by Harrison et al. [2] for their high altitude coincidence counts data. However, they use miniature GM tubes as a part of their payload, while the University of Minnesota, Morris and St. Catherine University both use AWARE Electronics Model RM-60 and RM-80 GM counters. The RM-60 and RM-80 GM counters have much larger collection areas than a miniature GM tube. As such, a moving average filter is not a suitable method of data smoothing in this case.

The next method of data processing was reducing the counts into counts per minute in Excel using the equation

\[ \text{SUM(OFFSET(input cell,ROW(Sreference cell))\text{reference cell}1\text{m-n-a})} \]

where \( n \) is the number of cells added together, the input cell is the first cell containing counts data, and the reference cell is the cell containing the equation. All of the bolded cell are static cells. To find the corresponding altitude, we modified equation 1 to be

\[ \text{AVGARE(OFFSET(input cell,ROW(Sreference cell))\text{reference cell}1\text{m-n-a})} \]

(Eqn 2).

With this, we were able to generate altitude vs. counts per minute plots. Once the data were processed, several curves were fit. Second and third order polynomial fitting was done using Excel, and cubic spline interpolation\(^3\) fitting was done using MATLAB. Second order polynomial fitting was ineffective at determining the RP maxima, while third order polynomials and cubic spline interpolation were able to fit the data. Third order polynomial fitting was chosen over cubic spline interpolation due to its ease of use. Cubic spline interpolation required additional data wrangling to be effective, and the RP maxima determined by the curves were within a 1% difference. Once a curve is fit to the data, the RP maximum is determined by finding where the slope of the fitted curve is zero.

The methods of data reduction and curve fitting have reduced the uncertainty in determining the RP maximum. These methods are not without downsides, however. Moving to higher order polynomials would give better goodness of fit, but such curves make determining second order fit difficult. Future work includes,

- Expanding the number of curves available to fit the data
- Reducing noise
- Developing other methods of determining RP maxima using pressure, temperature, and time data.

As these methods continue to be developed, they will be applied to a wider range of data sets to analyze different aspects of RP maxima.

References

