Statistical Analysis of Errors

Introduction

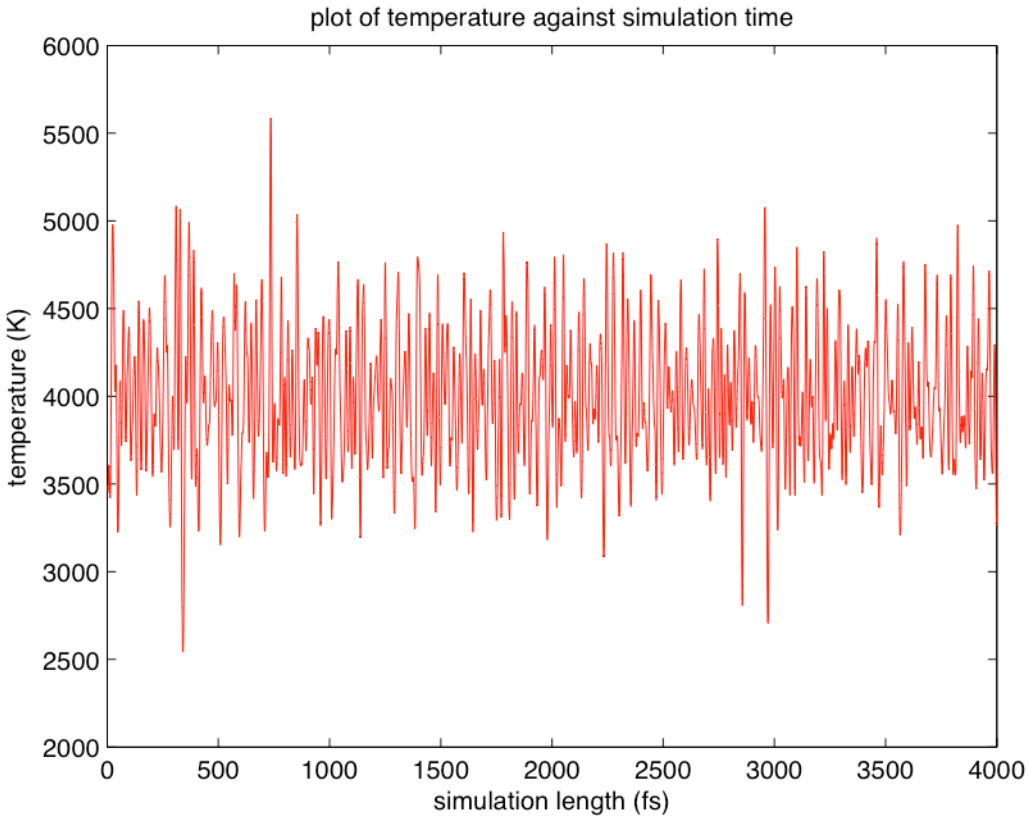
Molecular dynamics simulations model the time evolution of a system of atoms, by solving Newton's equations of motions. They are invaluable for studying the properties of minerals at the high temperature and pressures expected in the lower mantle, in particular, single-crystal elastic constants.

Data File

The *OSZICAR* file from a molecular dynamics simulation of periclase, the second most abundant phase in the mantle, can be found here:

```
/nfs/ugrad-library/SOEE2250/practical_2
```

Use `vi` or `emacs` to view the file. Of interest for this practical assignment are the lines that begin with an integer. The integers indicate the simulation time (in fs). To the right of this is the temperature (in K). The temperature was set to 4000 K, the temperature at the core-mantle boundary and you should be able to see it oscillate about this value.



PART I

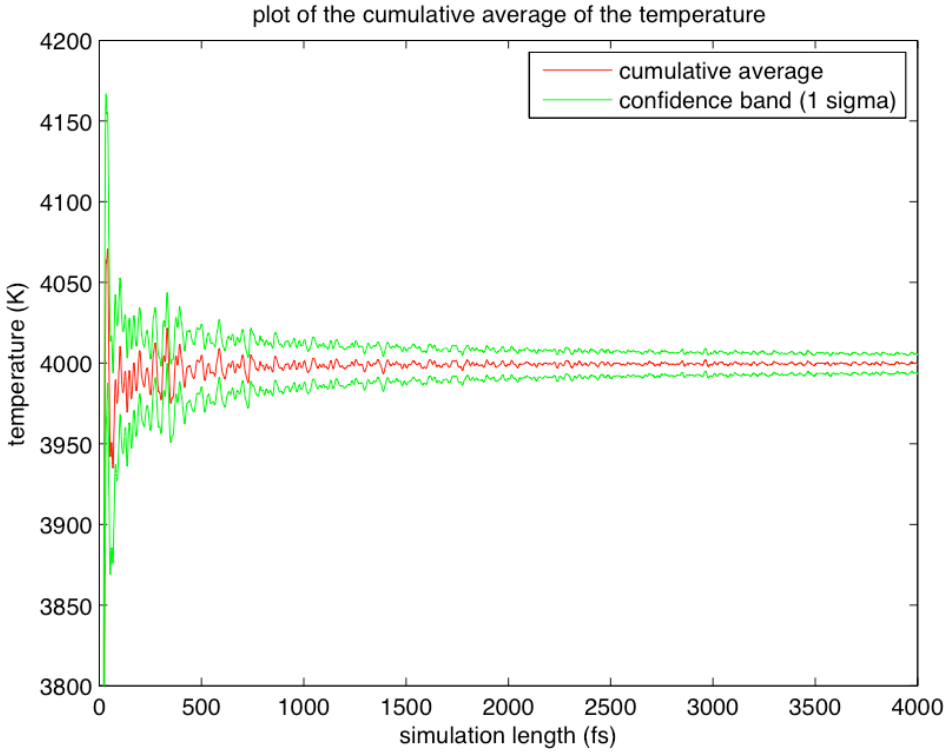
Write an `awk` script to read in the *OSZICAR* file and write the simulation time and temperature to a file called *temp.dat*. The first few lines should resemble:

```
1 3491.
2 3455.
3 3451.
4 3475.
…
```

Do not spend more than 15 minutes working on this. If after that time you are unable to do it, use the copy of the file *temp.dat*, available from the same location as the *OSZICAR* file.

PART II

Write a MATLAB script to read in the values from the file *temp.dat* and calculate the cumulative average of the temperature and its associated error (the standard deviation of the mean), as a function of simulation time. You should calculate the cumulative average at 1 fs intervals i.e. each step. Upon completion, your script should plot the cumulative average of the temperature and its associated error, as a function of simulation time. Plot the error as confidence bands. In addition, the values should be written to a file called *cumave.dat*.

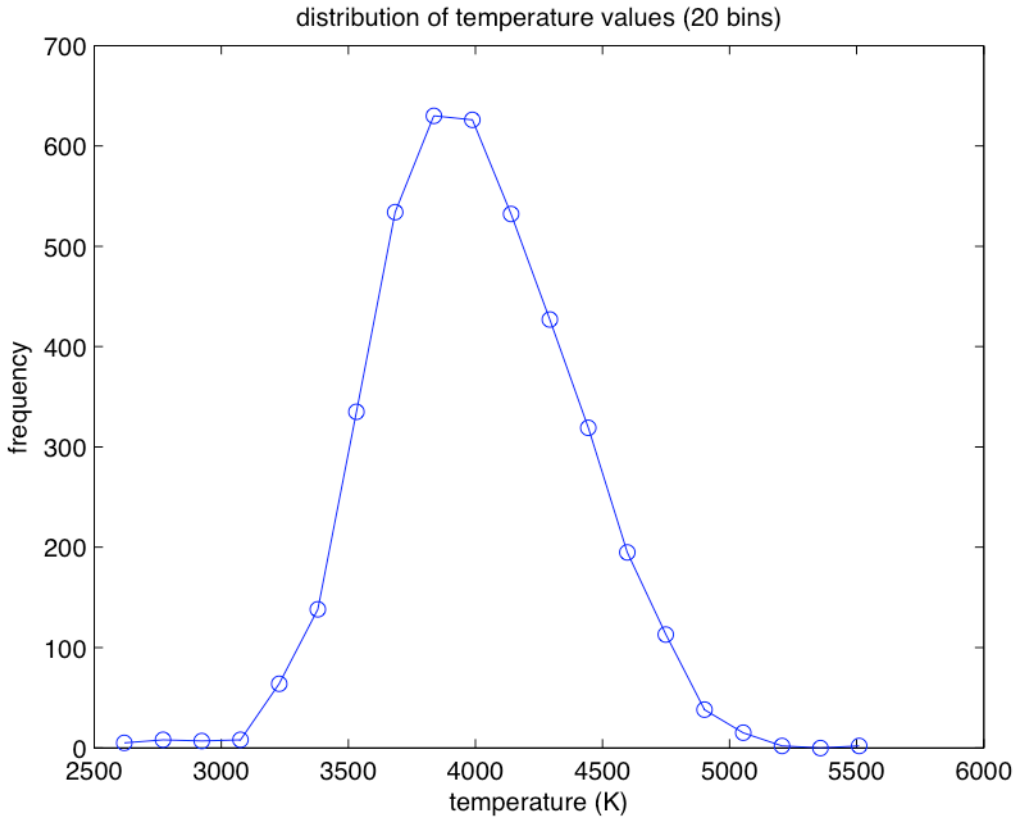


PART III

Write a MATLAB script to read in the temperature values from the file *temp.dat,* calculate and plot their frequency distribution, for a specified number of bins. Some guidelines:

1. Use `input` to ask the user for the number of bins to use.
2. Determine the temperature range sampled during the simulation i.e. the difference between the highest and lowest values. The simplest way to do this is to use the MATLAB functions: `min()` and `max()`.
3. Let the bin width = temperature range / the number of bins to use.
4. Use two nested loops (one that loops over each bin and one that loops over the temperature values) and a conditional statement, to determine the number of temperature values that fall into each bin.
5. Plot frequency against the mid-point of each bin.

Once finished, run your script using different numbers of bins, to see how the distribution develops. Your plot, when using 20 bins, should resemble:



PART IV

If you have time, look up the MATLAB function `histfit` using the `help` function. It takes a data series and plots a histogram of the values, using a given number of bins, together with a normal density function. Edit you answer to PART III, to use `histfit` to plot the temperature values.