

1 Cosmic Variance

In general, the Universe can only be considered to be homogeneous at the most extreme distance scales ($> 1Gpc$). At the scales which tend to be observed, matter in the Universe is not distributed uniformly and will gather together in massive clusters. Conversely, there will be empty regions of space, known as voids, where virtually no matter is present at all.

Measurements deduced from observational surveys will certainly be affected by this large-scale cosmic structure. A measurement of an arbitrary region of the sky will have an associated uncertainty far greater than any sample variance due to Poisson statistics. This uncertainty is known as *Cosmic Variance*, and will nearly always be the dominant source of error for any high-redshift cosmic survey. Given a probability distribution for the number counts, the cosmic variance is defined as:

$$\sigma_v^2 = \frac{\langle N^2 \rangle - \langle N \rangle^2}{\langle N \rangle} - \frac{1}{\langle N \rangle} \quad (1)$$

where $\langle N \rangle$ is the mean and $\langle N^2 \rangle$ is the variance[?].

Any number or density measurement derived from a galaxy population is susceptible to cosmic variance. As an extreme example, it is reported to be potentially at the 50-70% level in the HST UDF surveys[?]. Evidently, it will be particularly relevant to this study of cosmic reionization.

1.1 Factors which affect the Cosmic Variance

Fortunately, there exist numerous ways of reducing the cosmic variance of a galaxy survey to a manageable amount, mostly accepted to be $< 10\%$. As a general rule, this is achieved when a volume of $10^7 Mpc^3$ is surveyed, with a square survey area over a single contiguous region[?].

The cosmic variance of a survey depends on three factors principally:

- it decreases steadily with volume
- it is reduced for higher aspect ratio surveys
- it is also reduced when multiple sightlines are taken i.e. sparse sampling

Taking advantage of these factors, one can take steps to address the issue of cosmic variance.

1.1.1 Survey Volume

As mentioned above, the cosmic variance is heavily dependent on the volume of the survey, indicated by the steady decrease to lower percentages as larger volumes are surveyed, as in Figure 1. For volumes of $10^4 Mpc^3$, it is as much as 60% whereas for $10^7 Mpc^3$, it

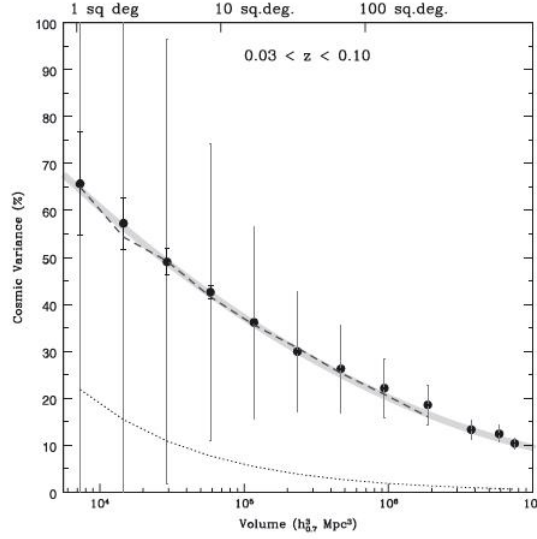


Figure 1: Graph showing the relation between the percentage cosmic variance and survey volume.

will be 10% and below.

The dashed line below the main trend on the graph shows how the percentage uncertainty varies solely due to the sample variance. This gives some idea of the gulf between the effect of cosmic variance and that of Poisson statistics.

1.1.2 Aspect Ratio

The aspect ratio of the survey is another crucial factor that can help deal with cosmic variance. For the same survey area, a long thin strip chosen over a standard square-shaped survey will decrease the cosmic variance. When taken to the extreme, this can cause a significant reduction. The same trend as for Figure 1 is shown in Figure 2, but with the grey dots corresponding to higher aspect ratio surveys. The graph illustrates that, particularly for larger volumes, increasing the aspect ratio of a survey can make a huge reduction in the cosmic variance uncertainty.

1.1.3 Sparse Sampling

One might tackle cosmic variance in a very effective manner via the method of *sparse sampling*, whereby several identical observing areas are distributed randomly across the sky. As opposed to one contiguous survey, these multiple sightlines in different directions allow a more representative portion of the sky to be surveyed.

It has been shown that cosmic variance is reduced with the number, N of such sightlines by $\frac{1}{\sqrt{N}}$. This holds irrespective of the base survey area[?].

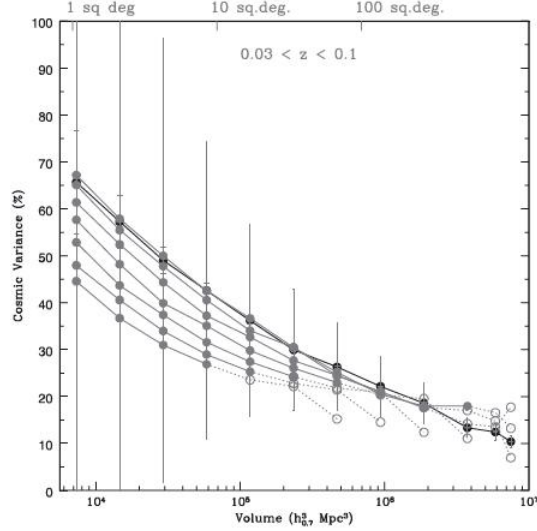


Figure 2: The same as for graph 1, this time with the additional grey points showing the dependence on aspect ratio. From the top to bottom they are 1:2, 1:4, 1:8, 1:16, 1:32, 1:64, 1:128.

1.2 General Formula

In order to address cosmic variance in the code itself, an empirical formula was employed. In a paper (Driver and Robotham, 2010), this formula was derived using data from the SDSS (Sloan Digital Sky Survey), and extrapolated out to higher redshifts[?]. Due to its extrapolated nature, it should be noted that this is an approximate formula. Nonetheless, it should output a robust estimate of the cosmic variance for given surveys. Additionally, through it one can deduce possible ways to reduce the cosmic variance to below 10%.

$$\zeta_{CV} (\text{percent}) = \frac{\left[1.00 - 0.03\sqrt{X-1}\right] \cdot \left\{219.7 - 52.4\log_{10} [AB \cdot 291.0] + 3.21\log_{10} [AB \cdot 291.0]^2\right\}}{\sqrt{NC/291.0}} \quad (2)$$

The formula employs the median redshift transverse lengths A and B , and the radial depth C , all given in Mpc , to represent the volume of each survey sample. The number of such samples i.e. the number of sightlines is N . The aspect ratio is dealt with in the amplitude term at the front of the expression, where the aspect ratio 1:128 would be submitted as $X = 128$, for example.

Although it is extrapolated from an equation that applies at $z < 0.1$, it takes advantage of the fact that cosmic variance should vary according to Poisson statistics along the radial length. This means that a survey with twice the radial depth will have $\sqrt{2}$ less cosmic variance. Hence the factor of \sqrt{C} in the denominator. Equation 2 does not, however, make any corrections for the evolution of the clustering signature of the galaxy

population, with this expected to be lower towards higher redshift. Correspondingly, any estimate output should be treated as an upper limit.

This equation has been entered into the code which, from the survey parameters input, calculates the required quantities and gives an estimate of the cosmic variance as a percentage of the number counts expected to be observed. Furthermore, it will inform the user of how many random pointings of such a survey are necessary to achieve $< 10\%$ cosmic variance.