SAMPLING REQUIREMENTS FOR THE DETECTION OF SPATIAL TRENDS

Jean-René Roy (*Gemini Observatory* and Yvan Dutil (BOMEM)

Summary: *For surveys of physical parameters in objects where a high degree of axisymmetry is expected, at least 16 objects should be measured with adequate S/N in order to derive a robust spatial behavior such as monotonic radial variations in galaxy disks. If one is looking for more complex spatial behavior, like breaks in the rotation curves or gradients in the spatial distribution, the required sample size may need to be several times larger.*

In studies where multiplexing allows get a large sample in one observation, it may be important to establish the minimum number of objects needed in a sample to derive a robust behavior in a distribution. For example, several MCAO science cases emphasize the multiplexing gain of the larger corrected field allowed by the uniform PSF. In the text below we determine quantitatively what minimum sample sizes are needed to determine various radial behavior, e.g. a monotonic change and a gradient with one/two breaks.

Galaxies generally present a fair degree of axisymmetry in their properties. When it is appropriate to average these azimuthally – for example to determine their abundance radial distributions or rotation curves –, one can explore their radial behavior. Sampling size may become an issue. For example, how many H II regions are required to derive the radial distribution of a key physical parameter (e.g. the O/H abundance gradient) in a galaxy within a given precision? In the simplest case where a monotonic behavior can be assumed, one fits the radial distribution with a straight line (Kreysig 1988)

$$y = a + bx \tag{1}$$

If we admit a gaussian distribution for the abundance variation in y,

$$\sigma_a^2 = \left(\frac{1}{n} + \frac{\overline{x}^2}{s_{xx}}\right) s_{xy}^2, \tag{2}$$

$$\sigma_b^2 = \frac{s_{xy}^2}{s_{xx}} \tag{3}$$

where

$$s_{xx} = \sum x^2 - (\sum x)^2/n, \tag{4}$$

$$s_{yy} = \sum y^2 - (\sum y)^2/n, \tag{5}$$

$$s_{xy}^2 = \frac{s_{yy} - {s_{xy}}^2/s_{xx}}{n-2} \approx \sigma_y^2 - b^2\sigma_x^2. \tag{6}$$

For most galaxies, measurements of O/H abundances based on standard empirical calibrations give $s_{xy}^2 \sim 0.1$ dex (Dutil & Roy 1999).

We assume that the spatial distribution of H II regions is uniform as a function of galactocentric radius. This is an optimistic view since most galaxies do not display such a convenient distribution, even in the absence of a bar or of a ring which can modify the radial distribution of massive star formation (Hodge & Kennicutt 1983). However, it may be argued that a careful observer may achieve a fairly uniform sampling by choosing the H II regions appropriately. In addition, the equations above also suppose that all regions have the same weight. The different weighting schemes (based on surface brightness, signal to noise ratio or confusion limit) diminish the effective number of H II regions below the real size of the sample while reducing the value of s_{xy}^2, and behave differently against selection effects.

With a uniform spatial distribution of points between 0 and Δ_x,

$$\sum x = \overline{x}n = \Delta_x n/2 \text{ and } \sum x^2 = \overline{x^2}n = \Delta_x^2 n/3. \tag{7}$$

This leads to

$$s_{xx} = \Delta_x^2 n/4 - (\Delta_x n/2)^2/n = \Delta_x^2 n/12, \tag{8}$$

Replacing these values in equation 2, we obtain:

$$\sigma_a^2 \approx \left(\frac{1}{n} + \frac{3}{n}\right) s_{xy}^2 = \frac{4s_{xy}^2}{n} \approx \frac{4}{n}\left(\sigma_y^2 - b^2\sigma_x^2\right). \tag{9}$$

where σ_a is the uncertainty on the *extrapolated* central abundance. This uncertainty is approximately twice as large as the uncertainty on the mean abundance level $\sigma_y/\sqrt{n}$; this is caused by the contribution of the uncertainty on the slope to the extrapolated central abundance. This large contribution $(3/n)$ is due to the fact that the slope is strongly weighted by the extremum points which are few, by definition; thus more points are needed to reach a pre-defined level of precision. Therefore, four times as many points are required to infer the central abundance compared to the mean abundance level for a given error. Using this approach, we derive a similar expression for the slope of the gradient:

$$\sigma_b^2 \approx \frac{12s_{xy}^2}{n\Delta_x^2}. \tag{10}$$

One may also search for *breaks* in the radial behavior of galaxy parameters such as extinction or abundances. For radial distributions of chemical abundances, such breaks are difficult to ascertain because of the non-unique relation between abundances and the line ratio indicators (Henry & Howard 1995; Kennicutt & Garnett 1996). Breaks in the slope of abundance gradients are predicted for disk galaxies with young bars (<500 Myr), where the radial flows of the gas across the disk have not yet had the time to homogenize the abundances throughout the disk (Friedli, Benz & Kennicutt 1994; Friedli & Benz 1995); some evidence for breaks exists for a few barred galaxies (Martin & Roy 1995; Roy & Walsh 1997).

Searching for a break is equivalent of looking for a variation of the gradient between two parts of the radial distribution. In the simplest case of the break occurring in the middle of the radial distribution, half of the points are measured on each half of the radial range ($\Delta_x/2$). The mean abundance level has to be conserved either with a break or with uniform gradient slope, and this on each side of the break. For consistency, the two slopes must lead to the same abundance in the middle of the disk, and we have (see Figure 2)

$$(b_1 - b)\frac{\Delta_x}{4} = (b - b_2)\frac{\Delta_x}{4} \tag{11}$$

which reduces to

$$b_1 + b_2 = 2\,b, \tag{12}$$

where b_1 is the slope of the first part of the distribution and b_2 is the slope of the second section. In the case where $b_2 = 0$, the difference between the two gradients is equal to twice the initial slope. By adding the variance $\sigma^2_{b_1}$ and $\sigma^2_{b_2}$, we find the variance to be used for the test:

$$\sigma^2_{b_1-b_2} \approx 2\,\frac{12 s^2_{xy}}{(n/2)(\Delta_x/2)^2} = \frac{192 s^2_{xy}}{n\Delta^2_x}. \tag{13}$$

Expressed in term of signal to noise ratio, this expression translates into:

$$S/B_{b_1-b_2} = \sqrt{\frac{n}{192 s^2_{xy}}}(b_1 - b_2)\;\Delta_x = \frac{1}{4}\sqrt{\frac{n}{3s^2_{xy}}}b\;\Delta_x. \tag{14}$$

Re-doing the calculation only to extract the slope of the gradient leads to:

$$S/B_b = \frac{1}{2}\sqrt{\frac{n}{3s^2_{xy}}}b\;\Delta_x. \tag{15}$$

In order to detect a break in the slope of the gradient, one needs about four times the number of data points required to establish solely the value of the global gradient. Extending this discussion to *two breaks* in the slope, e.g. two strong gradients separated by a flattened region as predicted by some evolutionary codes of barred galaxies (Friedli, Kennicutt & Benz 1994, Friedli & Benz 1995), we derive the following equation:

$$S/B_{b_1-b_2} = S/B_{b_2-b_3} = \sqrt{\frac{n}{648 s^2_{xy}}}(b_1 - b_2)\;\Delta_x = \frac{1}{9}\sqrt{\frac{n}{2s^2_{xy}}}b\Delta_x. \tag{16}$$

About 13.5 times as many data points are needed to detect a pattern of triple slope as for measuring the mean global gradient!

A similar analysis can be done for the comparison of mean abundance levels at a given galactocentric radius. The uncertainty on the local mean abundance is estimated by:

$$\begin{aligned} \mathrm{Var}[y(x_0)] &= \mathrm{Var}[a + bx_0] \\ &= \mathrm{Var}[\overline{y} - b\overline{x} + bx_0] \\ &= \mathrm{Var}[\overline{y} - b(x_0 - \overline{x})] \\ &= \left[\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{s_{xx}}\right] s_{xy}^2. \end{aligned} \tag{17}$$

When comparing abundance levels referring to the middle of a galaxy disk ($\Delta_x/2$), this equation reduces to

$$\mathrm{Var}[y(\Delta_x/2)] = \frac{s_{xy}^2}{n} \tag{18}$$

Compared with equation 9, this shows that for a given uncertainty, it is much easier to establish the abundance in the middle of the disk than the extrapolated central abundance, since the deduced abundance level is no longer dependent on the estimation of the slope.

As example, for a typical Sb spiral galaxy with $\Delta_x = 15$ kpc, $b = -0.02$ dex/kpc and $\sqrt{s_{xy}^2} = 0.1$, one needs at least 12 H II regions in order to measure a slope at the 3σ level, and 16 regions to infer the central abundance with a precision better than ± 0.05 dex. For the same galaxy, one would need at least 48 data points to detect a break, and 162 to detect a double break in the radial distribution. However no more than four well distributed points are required to derive the mean abundance level with a precision of 0.05 dex. These results are consistent with those of Zaritsky *et al.* (1994).