

# Chapter 3

## Making Maps with the SCF

In Chapter 2, the background for establishing a single function defining the correlation between two spectra was developed. The entire purpose of the SCF, however, is to analyze whole maps of spectra and thus the functions must be generalized in an appropriate way to create correlation maps of the data sets. In the following discussion, a method for doing this is established and then some examples of the correlation maps are explored in order to glean understanding about our results.

The SCF algorithm was written in the IDL software package to take advantage of the built-in functions, the multi-dimensional array processing, and graphics utilities. It would be easy to port the code to another language for speed in processing though the output would have to be analyzed using some graphics package, like PGPLOT or MONGO.

### 3.1 Creating Maps of the SCF

In order to generate the appropriate maps, a specific algorithm is developed.

1. Begin by considering a data cube with size  $X \times Y$  and spectra at each  $(x_i, y_j)$  position in the cube. For purposes of the calculation, it is assumed that the spectra are taken on an evenly spaced grid. In the discussion to follow,  $(x_i, y_j)$  represent the real right ascension and declination coordinates of the cube and the coordinates  $(i, j)$  are the pixel coordinates in the data cube.
2. For each of these spectra, a Gaussian is fitted to the data, allowing for the estimation of certain line parameters. The parameters derived from the fit are peak antenna temperature ( $T_A^*$ ), the width of the Gaussian in velocity space ( $\Delta v$ ), the center of the Gaussian ( $v_{LSR}$ ), the integrated area under the Gaussian ( $T_{int}$ ) (generated with the summation integration discussed previously), and the rms noise of the spectrum ( $\sigma_{rms}$ ). Note that most of these parameters are used only for the purpose of calculating the noise in the spectrum. The sole exception is that the width ( $\Delta v$ ) is used to determine the relevant velocity ranges around a spectrum's peak that the SCF will consider.
3. Spectra with a signal to noise ratio below a selected threshold are rejected and all correlation calculations are performed without these rejected data.

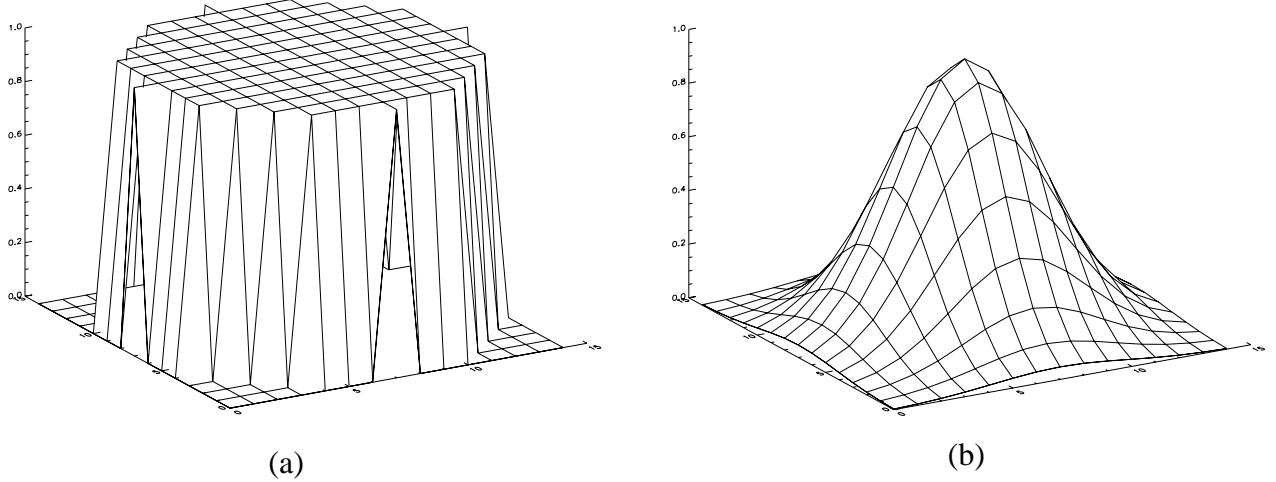


Figure 3.1: Two commonly used weight functions. (a) A flat, circular mask, used to eliminate any bias for the square directions in which the cube was sampled. (b) A Gaussian mask, also circularly symmetric, but also weights spectra closer to the base spectrum.

4. A test spectrum at  $(x_i, y_j)$  is selected to measure its correlation with nearby spectra.
5. The SCF requires two input parameters denoted by the variables  $r$  and  $q$ .  $r$  is the resolution of the SCF and represents the number of spectra for which the correlation with the test spectrum is measured. A resolution of  $r$  means to consider an  $r \times r$  box centered on the test spectrum and to compare the test spectrum with every spectrum in the box. If desired, the results of the correlation calculations can be weighted by the distance away from the test spectrum. These weight functions emphasize regions in the data cube to eliminate certain biases. Examples of weight functions appear in Figure 3.1.

The parameter  $q$  is a measure of what velocity range is compared with the other spectra in the  $r \times r$  box. All calculations discussed in the following section are to be performed only on the velocity range

$$v_{LSR}(i, j) - q\Delta v(i, j) \leq v \leq v_{LSR}(i, j) + q\Delta v(i, j) \quad (3.1)$$

The Gaussian fit parameters in equation 3.1 are taken from the fit to the test spectrum at  $(x_i, y_j)$ . All the integrations performed in the calculation of the deviation function and the SCF itself are taken over this velocity range as opposed to the full range of velocities within the spectral bandpass.

6. A comparison spectrum at  $(x_a, y_b)$  is selected with the provisions that  $a \in [i - (r - 1)/2, i + (r - 1)/2]$ ,  $b \in [j - (r - 1)/2, j + (r - 1)/2]$ ,  $a \neq i$ , and  $b \neq j$ .
7. Letting the spectrum at  $(x_a, y_b)$  equal  $T_1(v)$  and the spectrum at  $(x_i, y_j)$  equal  $T_0(v)$ , the deviation function,  $D(T_1, T_0)$ , is minimized by adjusting  $s$  and  $\ell$ . The deviation function is then normalized according to prescription in equation 2.5.

8. The above step is repeated for each spectrum in the  $r \times r$  box. The resulting values for  $S(T_{a,b}, T_{i,j})$ ,  $s_{a,b,i,j}$ , and  $\ell_{a,b,i,j}$  are averaged over all non-rejected values of  $a$  and  $b$  in the box using the selected weighting function. The resulting means are assigned to the position at  $(x_i, y_j)$  as  $S_{ij}$ ,  $s_{ij}$ ,  $\ell_{ij}$ .
9. The deviation function is then minimized between the test spectrum and all comparison spectra holding the value of  $s$  at 1. This represents the correlation between the areas of the spectra because differences in area can not be compensated by the scaling parameter  $s$ . The averaged values of the SCF and the lag,  $\ell$ , over all applicable  $a$  and  $b$  are referred to as  $S_{ij}^l$  and  $\ell_{ij}^l$  respectively.
10. The deviation function is then minimized holding  $\ell = 0$ , in an attempt to explore velocity offsets alone. The values of the SCF and  $s_{a,b,i,j}$  in this case are referred to as  $S_{ij}^s$  and  $s_{ij}^s$  respectively.
11. The last correlation calculation to be performed over the box centered at  $(x_i, y_j)$  is to calculate the SCF with  $\ell = 0$  and  $s = 1$ , the straight squared difference between each pair of spectra. The resulting averaged value is referred to as  $S_{ij}^0$ .
12. Finally, the above correlation calculations are performed for each spectrum in the cube serving as the base spectrum. This procedure yields values of the parameters  $T_A^*$ ,  $\Delta v$ ,  $v_{LSR}$ ,  $T_{int}$ ,  $\sigma_{rms}$ ,  $S$ ,  $\ell$ ,  $s$ ,  $S^l$ ,  $\ell^l$ ,  $S^s$ ,  $s^s$ , and  $S^0$  for every point  $(x_i, y_j)$ . Maps of each of these parameters can be generated and in their analysis, the results of the SCF can be analyzed.

## 3.2 Interpreting Maps of the SCF outputs

The bulk of this section is to provide some insights into how the maps generated by the SCF can be interpreted. The results are different from any other functional output generated by other analysis routines; and in applying the analysis to simple situations, more insight can be gleaned about actual data. The understanding from these maps is preliminary and is used as a semi-qualitative descriptor of the behavior of the data.

To develop well-understood situations, the algorithm was applied to small cubes of completely artificially generated spectra. It is important to note that these do not represent the data from MHD simulations in any way. Rather, they are simple constructions of data cubes which are intended to display the behavior of the SCF with only one aspect of the data cube changing in a regulated fashion.

The data cubes generated had 484 spectra arranged in a  $22 \times 22$  array. These spectra were Gaussians and had various amplitudes, widths, and/or velocity offsets. Several different configurations were analyzed including random parameters, sudden jumps, linear gradients, and constant variations in the parameters. A few general statements about the outputs from the SCF can be made.

The correlation function  $S$  represents the similarity in shape between the base spectrum and the surrounding spectral lines. In a test where the spectra were all Gaussians with

the same width, the value of  $S$  was 1 for the entire map, even for randomly distributed offsets and amplitudes. Since these all have the same shape, the correlation function accurately demonstrated the similarities of these lines.

The lag parameters  $\ell$  and  $\ell^l$  both represent the velocity by which the neighboring spectra are offset from the test spectra. For linearly changing values of  $v_{LSR}$ , this number is zero. This result is due to the fact that for any given test spectrum, the comparison spectra with higher velocities are balanced by those with lower velocities, so the mean shift is zero. Any deviations from this can be attributed to the discrete spacing of the grid points and the inability for the spectra to be shifted in order that a perfect match be obtained.

The lags are interesting because for smooth Gaussians, a velocity gradient of order  $m$  can be shown to have a corresponding lag field of order  $m - 2$ , implying a relationship similar to that of taking the second derivative. Ordinarily, recognizing the relationship would be a boon to calculations, but the relationship only holds true when the peak velocity is the only thing differing between two spectra. When their shapes and amplitudes differ, the correlation might be maximized with a lag that is not the one predicted by the second derivative of the velocity field.

The scaling factors  $s$  and  $s^s$  represent the similarities between the antenna temperatures of the spectra. When the lag is turned on, the value  $s$  represents the similarity between the integrated areas of the spectra, with a value of 1 indicating equality. When the lag is off, the value,  $s^s$ , represents the mean, for all velocities, of the ratio between the antenna temperatures from the two spectra at given velocities. A value greater than one indicates that the base spectrum is consistently higher than comparison spectrum and a value of less than one indicates the opposite case. It is important to remember that the value of  $s$  is adjusted so that the differences in antenna temperatures are minimized over the whole spectrum, and the result does not correspond to the ratio of amplitudes at a given velocity.

The remaining correlation functions contain important information in what is held constant rather than what is allowed to vary. This subtle point stems from the realization that in allowing a parameter to vary any differences due to that parameter are eliminated by the SCF analysis. For example,  $S^l$ , the correlation with adjustable lag but not scaling of the spectra is helpful for comparing the similarities of line profiles. A high value will indicate that the spectra are similarly shaped and scaled though they may not necessarily be from regions with the same bulk velocity. On the other hand, the correlation function with zero lag but adjustable scaling,  $S^s$ , measures the similarity in velocity distribution at two positions without regard to the amplitude changes.

Finally the correlation function with the lag off,  $S^0$ , indicates simply the similarity of the two spectra in size, shape and velocity distribution. Strong correlations in this function are tied to features that are large compared to the resolution of the SCF; thus changing the resolution will highlight different structures. This result is because a large value of  $S^0$  requires all the spectra in the resolution box to be very similar in all aspects.

In this section of the paper, several examples of artificially generated data are interpreted. The relevant maps are displayed and others are summarized in the text. The best approach to interpreting these data may be simply to leap in and look at some examples. In all these examples, the weight function is a uniform  $r \times r$  box.

### 3.2.1 Velocity Jump

The first example contains the results of the SCF algorithm when the spectra in two regions of a map differ from each other by a sudden jump in velocity. In this case, the jump in  $v_{LSR}$  is from  $1 \text{ km/s}$  for low values of  $y$  to  $-1 \text{ km/s}$  for high values of  $y$ . Physically, this jump corresponds to the spectra in the bottom half of the graph having a red-shift and the spectra in the top half of the graph being blue-shifted. These are the spectral maps of two bodies of gas: one coming towards the observer and one retreating from the observer. The results are shown in Figure 3.2.

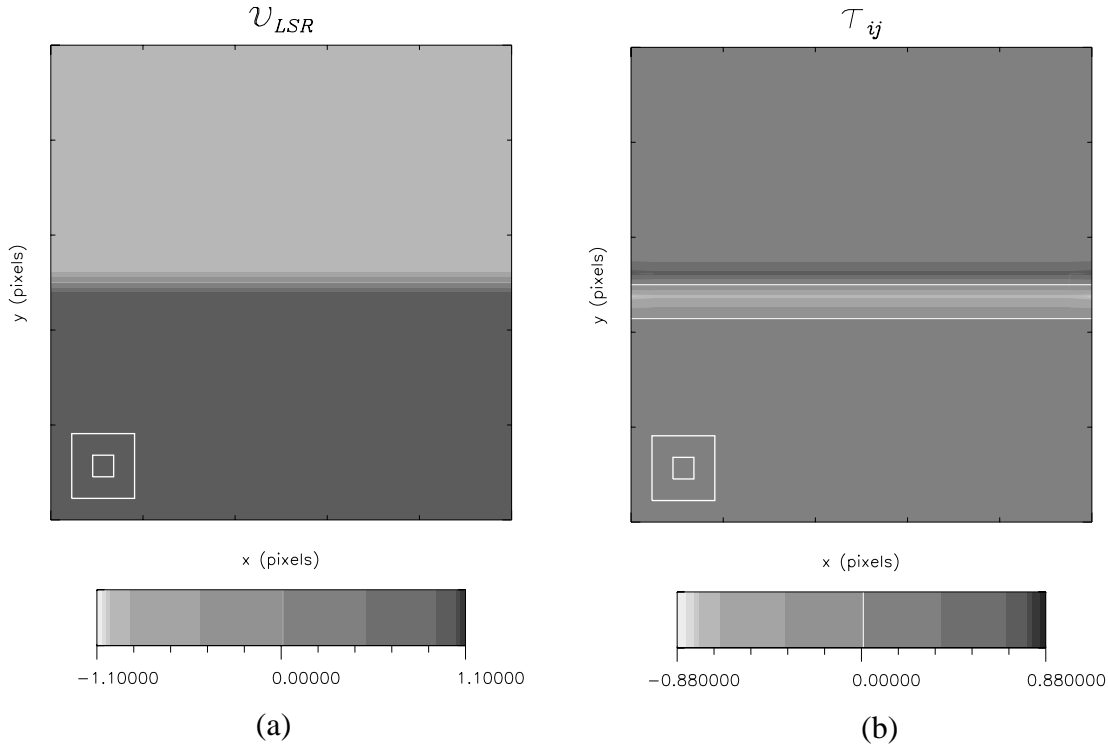


Figure 3.2: Correlation function outputs for a velocity jump. (a) The input velocity field. (b) The resulting field of lags

The figure on the left is a grey-scale representation of the fits to the  $v_{LSR}$  of the generated spectra. It depicts the sudden jump in  $v_{LSR}$  halfway up the data cube. The corresponding representation of the lag is displayed on the right. In all maps, the color bar extends for 10% beyond the actual values, which is why the extrema of the bar are at -1.1 and 1.1 respectively. The greyscale is selected to display a dynamic range of the data and color bars are provided for aid in interpreting the results.

As expected, the map displays a jump when it crosses the boundary between red-shifted spectra and the blue-shifted spectra. The drop in the value of  $\ell$  before the jump is because the spectra on the border that are red-shifted are being correlated with spectra in

the blue-shifted region that have a negative  $v_{LSR}$  relative to the spectrum with which they are being correlated. There is a corresponding jump in the blue-shifted section where the spectra are correlated with spectra that have a greater value of  $v_{LSR}$ .

In all cases, the plots have a pair of nested boxes in the lower left-hand corner. These boxes represent the size of the pixel (inner box) and the size of the SCF  $r \times r$  box where the correlations are conducted (outer box). In addition to these visual aids, an extra contour has been added at the zero level so that the examiner can distinguish between those regions with positive  $\ell$  or  $v_{LSR}$  and those with negative values.

All other plots of the map yield what is expected. The SCF function is equal to 1 over the entire plot in cases when both scaling and lag ( $S_{ij}$ ) and just lag ( $S_{ij}^l$ ) are turned on. In the case where there is no lag turned on ( $S_{ij}^s$  and  $S_{ij}^0$ ), the maps appear as shown in Figure 3.3.

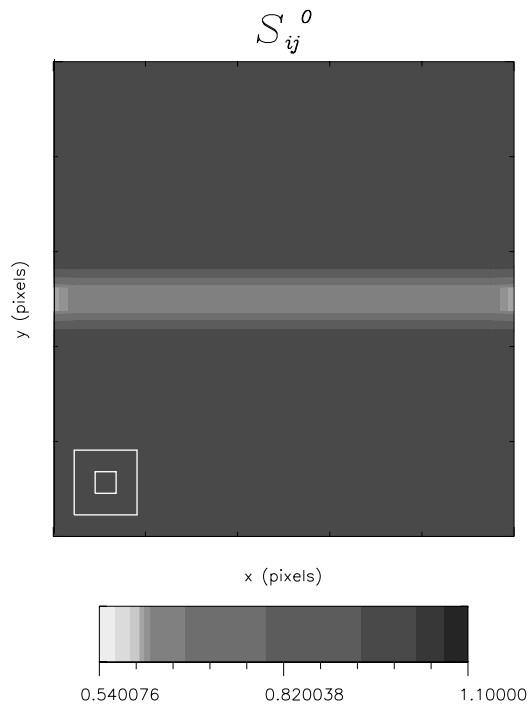


Figure 3.3: Map of  $S^0$  in the case of a jump in  $v_{LSR}$

The band in the center adequately represents the fact that the spectra in this region are not well correlated with their neighbors. The small boxes on the edge of even poorer correlation are the result of an edge effect. This effect results from the fact that pixels in the center of the map have eight spectra averaged together to yield a result while pixels on the edge only have five. It is because of effects like these that the  $r$  pixels on the border of the map are discarded in the actual analysis.

Because the artificial spectra were all of the same height, the scaling factor,  $s$  is uniformly 1. The parameters from the fitting indicate that the routine fits the Gaussians

well.

### 3.2.2 Amplitude Jump

The amplitude jump is almost exactly the same as the velocity jump, except that the jump is in the height of the Gaussian in the model spectra. The jump is from high amplitude at low  $y$  to low amplitude at high  $y$ , representing two different regions of gas, one with a higher peak antenna temperature than the other (See Figure 3.4).

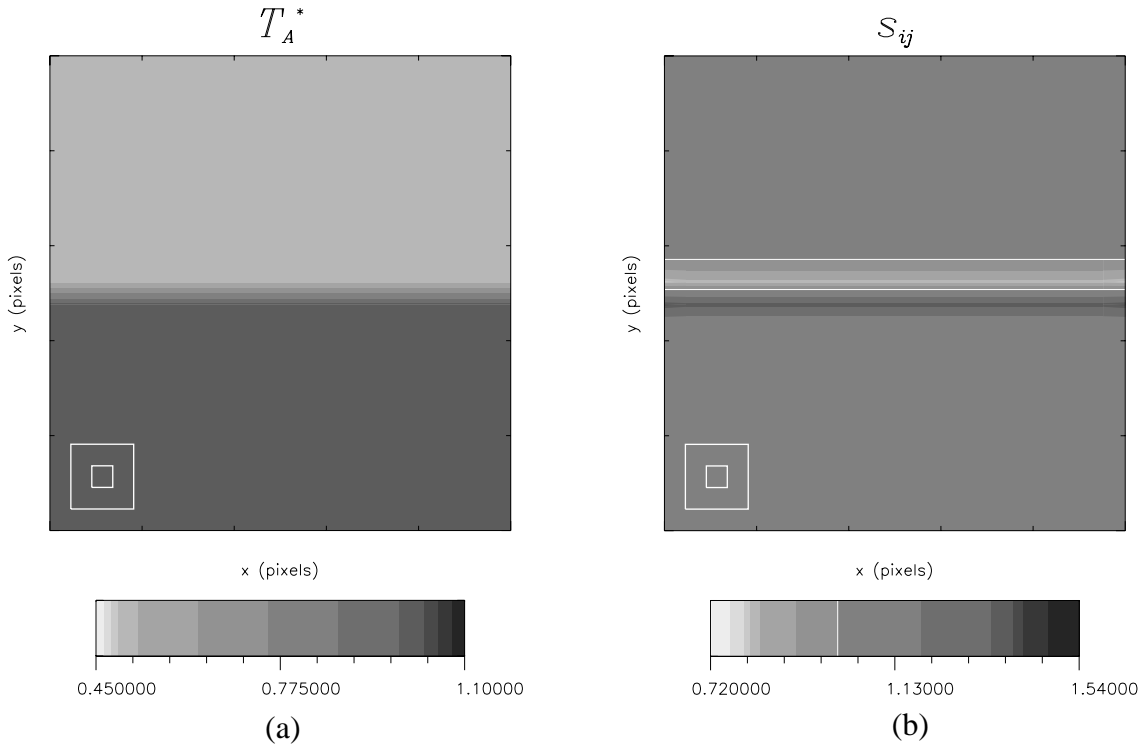


Figure 3.4: Maps of artificial data with a jump in  $T_A^*$ . (a) The amplitudes of the spectra. (b) The scaling factor  $s$ .

Again, there is an increase before and a dip after the change in the amplitude. The rise before the change may be a bit counter-intuitive; however, the way that  $s$  is defined is the amount by which the neighboring spectra are scaled up. Thus, the low spectra being scaled up to have the same form as the high spectra is represented by the jump and the high spectra being scaled down is represented by the dip. A border contour is plotted here as well, this time dividing regions with  $s < 1$  from those with  $s > 1$ . Those with  $s < 1$  indicate that the neighboring spectra had to be scaled down, so the spectrum here has, on average, a smaller integrated area than its neighbors.

The other plots generated by the SCF algorithm are similar to those generated in the case of the velocity jump. The SCF plots without scaling turned on are identical to those in the velocity jump without lag turned on (see Figure 3.3 and comments pertaining thereto).

Similarly, those plots of the SCF with scaling turned on are uniformly equal to one (with an error of 1 part in  $10^7$  due to fluctuations in the floating point calculations used on the computer).

### 3.2.3 Width Jump

While differences in lag and scale can be compensated for by the parameters  $\ell$  and  $s$  respectively, differences in shape can not be so corrected; thus the SCF can be used to measure the differences in shapes between lines. The result is that the SCF defines correlated spectra to be those with similar shape. A change in width is tantamount to a change in shape.

In this set of data, the jump is from large widths for low  $y$  to small widths for large  $y$ . The relevant plots for the case of changing widths are found in Figure 3.5.

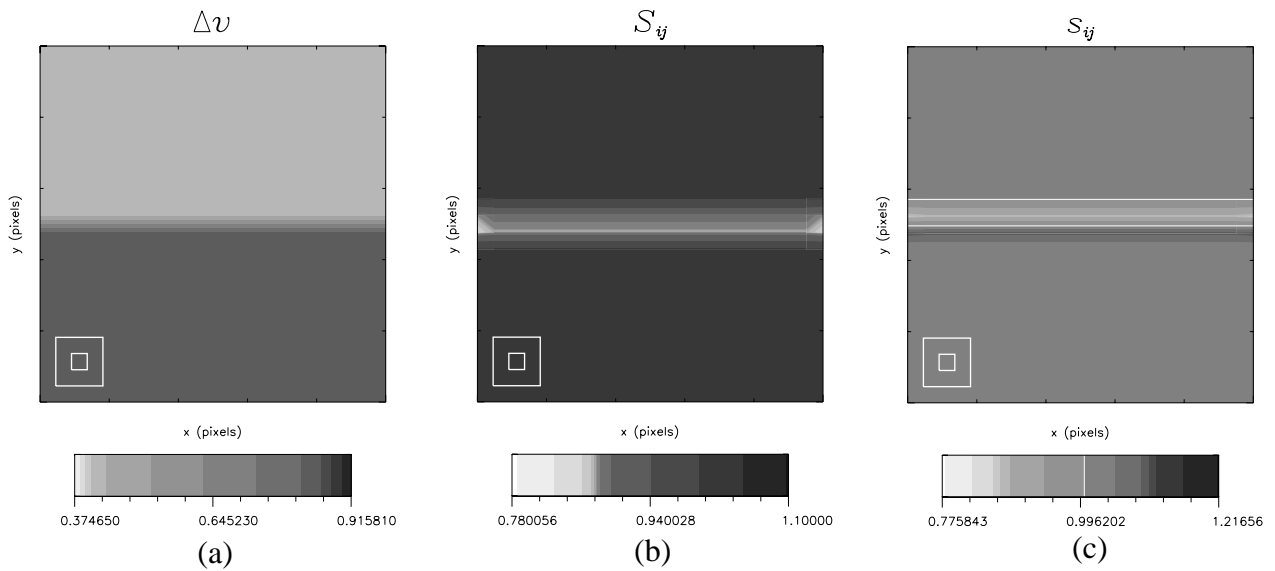


Figure 3.5: Relevant plots for a jump in Gaussian width. (a) The widths of the spectra. (b) The SCF values for the map. (c) The parameter  $s$  compensating for the differences in width.

The figure on the left represents the fit to the Gaussians, depicting the jump from large width to small width as  $y$  increases. The second plot, representing the SCF, shows how the width changing affects the data because the shapes are different along the interface between wide and narrow lines. There is less correlation between the spectra at that point. Finally, the last plot represents the value of  $s$  changing, which stands to reason because the SCF algorithm uses a changing scale factor to compensate for the differing widths.

### 3.2.4 Resolution

One of the input parameters that can be changed is the resolution,  $r$ , in order to regulate the size of the box over which the SCF parameters are evaluated. Again, the case of a changing  $v_{LSR}$  is used, but in this case there are two small clumps of gas with strong red-shifts against



background at rest relative to the observer. Like resolution in optics, the higher the value of  $r$  that is used, the more difficult it is to distinguish between two distinct features and to deduce their physical extent. The plots in Figure 3.6 illustrate this effect.

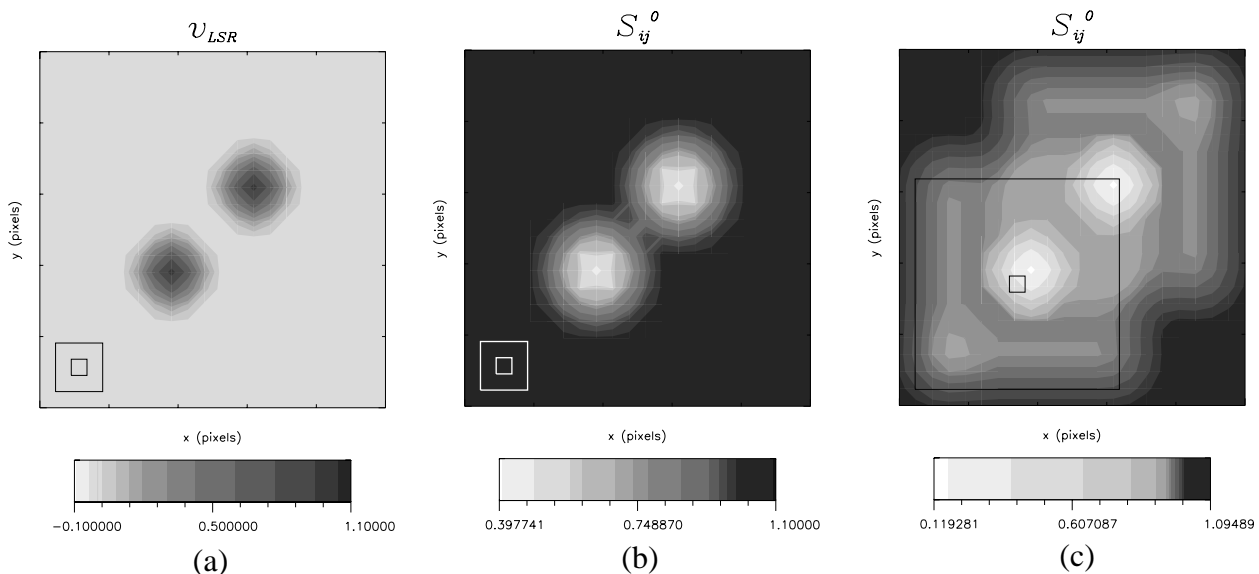


Figure 3.6: Relevant plots to display changing resolution. (a) The two-peaked velocity field. (b) The correlation function  $S_{ij}^0$  at resolution 3. (c) The correlation function at resolution 13.

For the figure with  $r=13$ , it is impossible to determine whether the gas distribution has a large region of spectra that are not perfectly correlated or a smaller region that is pulling the averages down. For this reason, the smaller resolution can be used to pick out these details.

Sometimes it can be advantageous to use a larger value of  $r$  on purpose. The correlations between gas clumps at small scales can be detected with small  $r$  values, but these features tend to mask the larger scale correlations present in a data set. To illustrate this, a data set has been generated with small-scale and large-scale structure. The relevant plots appear in Figure 3.7.

The velocity field is one with a small-scale sinusoidal modulation, on order of a pixel size, and a large jump at the center from blue-shifted velocity to red-shifted velocity. The jump is not readily distinguishable in the correlation measure at low resolution. There is a feature there; however, it cannot be discerned from the smaller scale correlations. On the other hand, if the resolution is increased, the large-scale correlation of the gas on either side of the jump becomes more apparent. The increased resolution serves to highlight this feature. The jump will become more dominant as the resolution increases, washing out the small-scale modulation in the infinite limit.

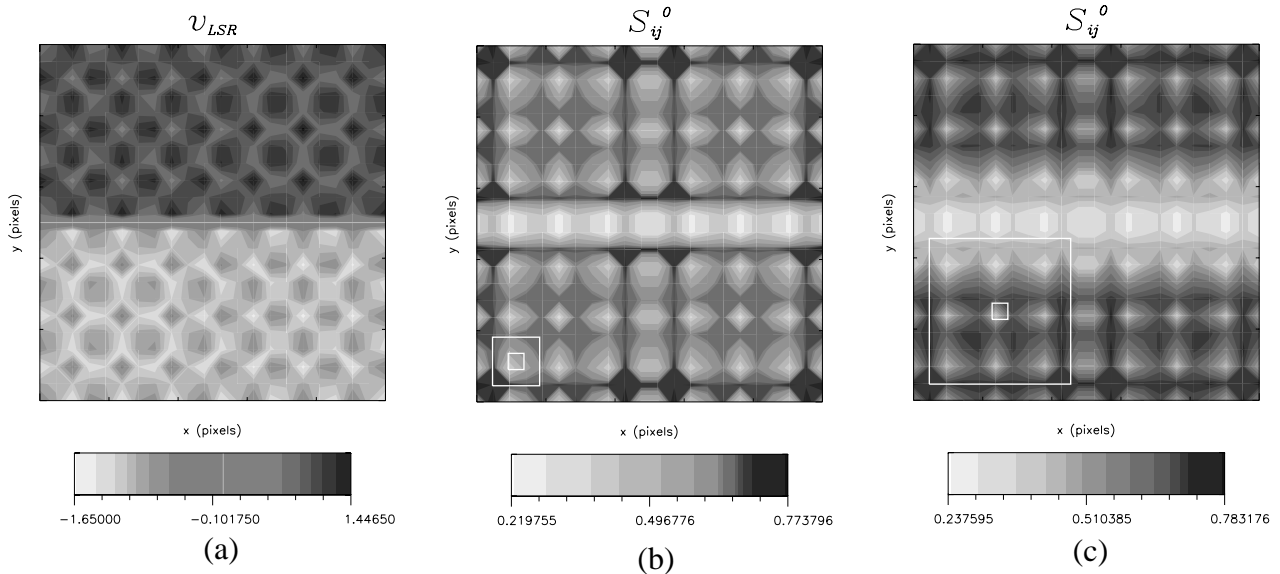


Figure 3.7: Plots displaying the ability for different resolutions to select for features of different size scales. (a) The velocity field used (b)  $S^0$  with  $r = 3$ . (c)  $S^0$  with  $r = 9$ .

### 3.2.5 Bringing it all together

The above differences are rather easy to understand individually; however, when all three basic effects are present in the same cloud, the interpretation is more difficult. In order to better understand such amalgamations of effects, the compensating factors  $s$  and  $\ell$  can be turned on and off at will.

First, the advantages of normalization immediately become apparent. We can examine a clump in a cloud and compare its values for the various flavors of the SCF. With both scaling and lag turned on, the value of  $S$  represents the best correlation one can get between two spectra. This value serves as a basis for comparison. Then, we can turn off lag ( $\ell \equiv 0$ ) and scaling ( $s \equiv 1$ ) respectively and compare the resulting  $S_{ij}^s$  and  $S_{ij}^l$ . If one of these is significantly higher than the other, we can assert that the gas has similar velocity distributions ( $S^s > S^l$ ) or that it is emitting similar amounts of radiation, but not moving coherently ( $S^l > S^s$ ). If the SCF with both lag and scaling turned off is comparable to those values with it on, we can assert that the gas within one resolution box of the central spectrum has either a roughly uniform or a completely chaotic nature, corresponding to high and low values of the SCF respectively.

The next fashion in which the results of the SCF can be compared is along the lines of the actual value of the SCF. The function, after all, is a measure of the correlation between spectra and with the normalization factor included, different parts of the map can be compared.

Additionally, maps can be compared with each other. Regions with correlation close to one are similar in shape to the spectra within the resolution box. Similarly, the calculated values of the other SCF variations,  $S^s$  and  $S^l$ , indicate similarities in intensity and velocity

distribution over the same areas.

In addition to mapping the SCF and its subsidiary variations, the parameters  $s$  and  $\ell$  that maximize the correlation for each point on the map are also calculated. These maps can be generated to give further information about the clouds.

The easiest aspect of these maps to consider is the magnitude of the lag parameter  $\ell$ . The parameter takes on large magnitude values when calculated along the border between two differently moving regions of gas. It is similar to the velocity gradient of the cloud, but by altering the resolution of the SCF, the structure of a velocity shift can be interpreted. For example, the resolution at which the jump appears the most prominent is the size scale of the physical jump.

With these realizations about the SCF and its behavior, the analysis of ‘real’ data can begin.