Chapter 1

Historical Background and the State of the Art

This section of the paper is meant to give the reader a general introduction to the type of radio astronomical observations which inspired the development of the Spectral Correlation Function (SCF). The SCF is descended from the fusion of two paths for studying the Interstellar Medium (ISM). One path is the observational technique of generating spectral line maps by taking spectra from several positions in an astrophysical object and analyzing the data contained in these maps. The other path is the theoretical technique of using Magneto-Hydrodynamic (MHD) simulations to model the properties of the ISM. The SCF provides an interface between these two techniques. It is intended to be a tool for comparing the products of these two paths, as well as a method by which spectral line maps can be analyzed in their own right.

1.1 The Interstellar Medium

Outside of objects in the solar system, the most prominent astronomical objects are the stars in our galaxy. Apart from a few prominent nebulae, early astronomers thought that the galaxy was defined strictly in terms of these luminous bodies. It was only near the beginning of the twentieth century that astronomers began to suspect the presence of more material in the galaxy aside from what was visible. Only a few observers prior to this time noticed or paid much heed to the anomalous 'holes' in an otherwise uniform background of stars. In 1889, E. E. Barnard compiled an atlas of these dark patches and recognized them as clouds of obscuring material in the galaxy. The presence of such obscuring matter led other observers to posit the existence of dark, obscuring material between the stars that was not as dense or opaque as Barnard's dark clouds. Observations established that the light from distant stars was diminished by this material and corrections to distance estimates were established, drastically expanding the size-scale of the universe. Questions regarding the nature of this material remained to be addressed.

Since that time, a century of study has established a large body of knowledge about the ISM. In understanding the ISM, astronomers have come to split the ISM into its two principal components: gas and dust. The former is further subdivided into five basic categories, characterized by the names of the regions in which the gas is found (Dyson and Williams, 1980, 172 ff):

Name	Temperature (K)	Typical Form	Density (m^{-3})
Coronal Gas	5×10^5	Ionized	$< 10^4$
Diffuse Nebula (HII)	8000	Ionized and Atomic	$\geq 10^8$
Inter-Cloud	6000	Ionized and Atomic	3×10^5
Diffuse Atomic Clouds (HI)	70	Atomic	3×10^7
Cool Molecular Cloud	20	Molecular	$10^9 - 10^{10}$

Table 1.1: Classification of the Gas Components of the ISM

Astronomers study each of these regions individually, for different techniques are better suited for the study of certain regions than others. For example, the low energy environments of cool molecular clouds are particularly well suited for study using the techniques of radio astronomy because such low energy interactions produce low energy photons. Hence, the radio data sets analyzed using the Spectral Correlation Function will usually focus on these cool objects.

The dust component of the ISM exists primarily in the cooler regions and consists large atomic and molecular agglomerations called grains. These grains form from the material blown off cool stars by stellar winds. The cooler regions preserve the grains from destruction because the thermal motion of the ambient gas lacks sufficient energy to destroy the grains in collisions. These regions, characterized by high densities of molecules and dust grains, are referred to as molecular clouds. The material in this thesis deals primarily with the observations and modeling of these objects.

1.2 Spectral Lines and Spectral Line Maps

Spectroscopy has always been a tool of immense use in physics; it is even more so in astronomy. Spectroscopic observations of the ISM were first taken in the visible wavelengths, examining the prominent nebulae in the sky. Some nebulae had similar spectral features to stars near them and were categorized as reflection nebulae. Others exhibited spectral features associated with the recombination of hydrogen, indicating the presence of a large body of ionized gas. Observations continued in this fashion until 1951 when Ewen and Purcell observed interstellar neutral hydrogen in the ubiquitous 21 cm spin-coupling transition of the ground state. From this point, the door to observing the ISM in wavelengths outside the visible was opened. In 1963, the absorption caused by OH at $\lambda = 18$ cm was discovered in interstellar clouds. Soon thereafter, emission from several different molecules was discovered. The standard source of radiation at long wavelengths came from the rotational transitions of heavy molecules. The study of spectra from these molecular sources swiftly grew into a new field of astronomy.

Eventually, observational technology progressed to the point that spectra from individual portions of molecular clouds could be analyzed separately. Thus, a cloud could be

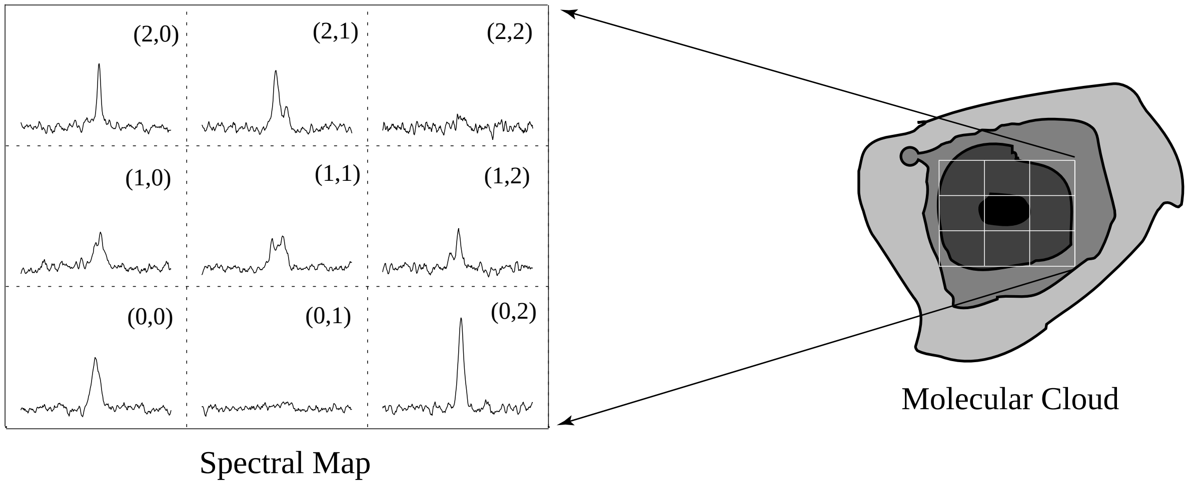

Figure 1.1: A sample spectral map.

'mapped' by taking a series of spectra over its extent, creating a grid with a spectrum at every intersection. An example of such a map appears in Figure 1.1.

Such maps have an inherently three dimensional coordinate system, with two axes representing position and the third representing the frequency of the spectrum. At every point in this three dimensional cube, there is an antenna temperature assigned. Antenna temperature is a conventional measurement in radio astronomy corresponding to the power received from incoming radiation. The use of the word 'temperature' is carried over from the first days of radio astronomy and is the same amount of power that a perfect resistor would emit in the given frequency range if the resistor were substituted in place of the antenna. In the spectra measured by radio telescopes, antenna temperature is the substitute for the intensity that one would normally find in a spectrum. When a spectral line is identified with a given absorption or emission feature, the difference between the rest frequency of the transition and the observed frequency is converted via the Doppler effect into a velocity. Thus, the cubes are often measured in position-position-velocity space.

1.3 Data Analysis Techniques

The analysis of a spectral-line data cube can be very time consuming, especially when the size of the maps grows beyond a few hundred pixels. As a consequence, several analytical tools have been developed to speed analysis. All of these tools involve condensing a large portion of the map to fewer data by assigning a characteristic number for a spectrum to represent that spectrum. Examples include assigning the peak antenna temperature or the mean velocity shift from a spectrum to that spectrum's point in the map, thereby reducing the three dimensional cube to a two dimensional map. These maps can be analyzed quickly and the salient features can be selected for more precise analysis. Such reductions can be perilous, however, for there is the risk of omitting too much data.

In addition to these rather simple methods, much effort has been invested in develop-

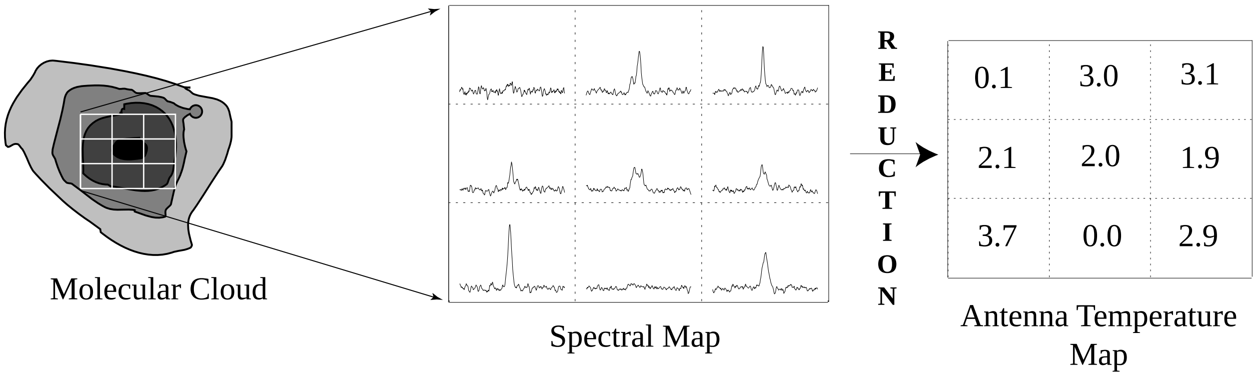

Figure 1.2: Example of creating an antenna temperature map from a spectral map by replacing each spectrum with the corresponding peak antenna temperature. The reduced maps are usually then converted to greyscale maps.

ing other analysis techniques for these immense data sets. These techniques can be divided into three basic categories:

- More rigorous analysis of the two dimensional maps
- Examinations of the three dimensional structure in the cubes
- Examining the relations between objects at different size scales in the cubes.

The following represents a brief summary of the methods used to date. It is by no means comprehensive; the only purpose is to illustrate what approaches are being tried and how the SCF differs from them. Briefly, the SCF can be summarized as a measure of how a given spectrum differs from its neighbors. These differences are further broken down into differences in shape, antenna temperature scale and velocity offset. By examining where groups of spectra are closely related and where they are not, the SCF provides new information about data cubes made from the ISM.

1.3.1 Two-Dimensional Analyses

These techniques of analysis carry the examination of two-dimensional maps to their most mathematically precise extent. As in simple analyses, the spectral maps have their constituent spectra reduced from a full spectrum to a characteristic number for that spectrum. A cartoon of this reduction appears in Figure 1.2. The most common characteristics are: peak temperature, integrated temperature, line width and line offset in velocity space. All of these characteristics are illustrated in the diagram found in Figure 1.3

Using these maps, the analyses are performed. The techniques of wavelet analysis have been brought to bear on these maps (Gill and Henriksen, 1990). These techniques center around convolving the "wavelet" function over the map at different scales, highlighting structures of a comparable shape. This analysis shares much with the concepts in Fourier analysis, except the particular analysis can select the wavelet shape.

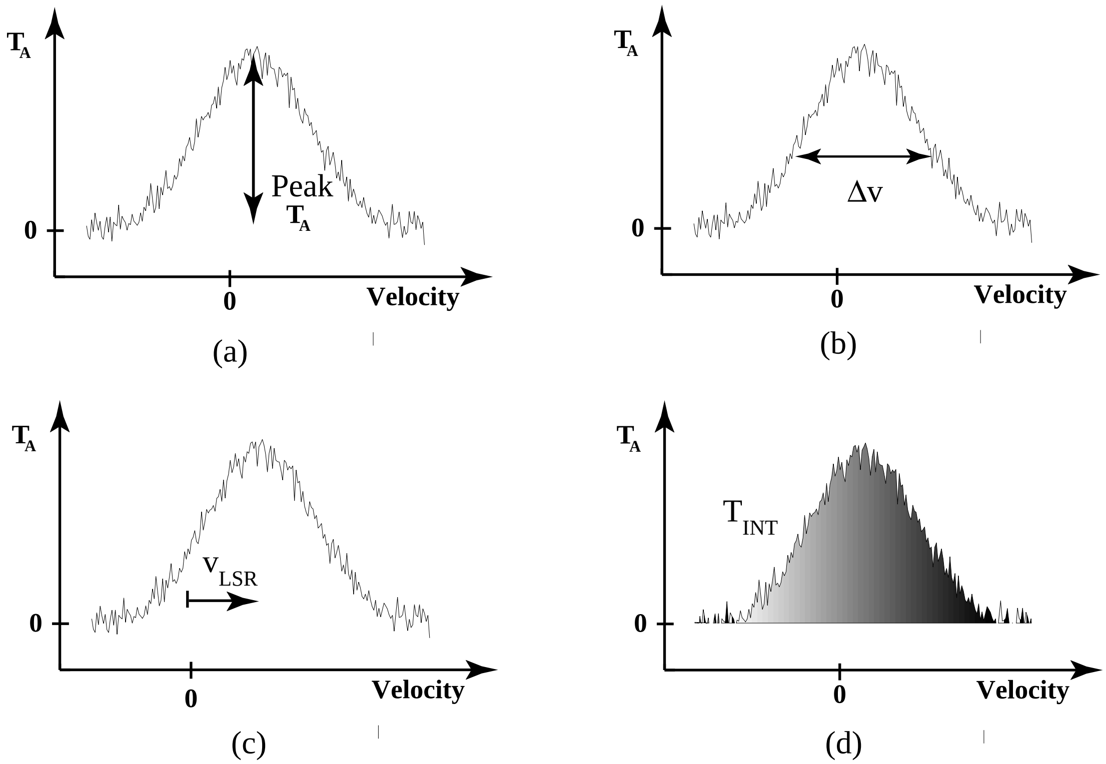

Figure 1.3: Diagram illustrating the definitions of the basic spectral line parameters. (a) Peak antenna temperature. (b) Line width (Δv). (c) Line offset (v_{LSR}). (d) Integrated temperature (T_{int}).

In contrast with the wavelet techniques, Houlahan and Scalo (1992) propose the use of structure tree statistics to characterize the properties of the ISM. This method focuses specifically on the hierarchical nature of the ISM. A map of the data with a single contour is plotted and the structure of the individual contour at a given value is examined. Specifically, the number of closed circles made by the contour is counted. As the value of the contour is raised (lowered) it splits (merges) into what are dubbed its children (parents). By examining the number of children/parents a certain contour has at a certain level, the hierarchical nature of the cloud can be characterized.

Another method of analysis is proposed by Wiseman and Adams (1994) who consider all of the clouds to be a set in a topological space. Structure is imposed upon this space by defining a pseudometric between the elements based upon their given properties. The metric functions used convert differences in the volume and mass into distances in this space. By examining the distribution of distances between these elements, a statistical analysis of the cloud is thereby performed.

All of these methods rely upon the contraction to a two-dimensional map and the subsequent sacrifice of data. The results presented by each of these methods illuminate a different aspect of each of the maps. Still, the SCF extends beyond them by considering the entirety of a spectrum and its relation to its neighbors when reducing the cubes to two dimensional maps.

1.3.2 Three Dimensional Structure Analyses

Not all techniques sacrifice the third dimension of the data. The methods discussed in the following section consider the spectral maps as a cube of data in three dimensional space. From these cubes, surfaces of constant antenna temperature can be considered and their structure examined.

The prime example of realizing the map as a three dimensional structure is found in the work of Williams *et al.* (1994). In this research, the CLUMPFIND routine is developed to examine data cubes. This routine searches for the peaks in the three dimensional array and then appropriately sorts the neighboring points into 'clumps' corresponding to the 'parent' peaks. The algorithm's primary task is to sort which neighbors belong to which peaks if the choice is not obvious (i.e. those measurements at saddlepoints). Once all the clumps are accounted for, statistical analyses on the resulting structures are performed, examining mass, peak temperature, velocity and size of the clump.

Studies of the fractal nature of the ISM offer illuminations into its structure by using these three dimensional analyses. An example of fractal analysis is presented in Elmegreen and Falgarone (1996). The fractal approach relies upon the observation that there are similarities between the two and three-dimensional structure of the ISM when it is viewed on specific size scales. The nature of these scale-similarities is characterized by the fractal dimension of the data. Different formation models for the ISM predict different fractal dimensions, thus models of the ISM can be evaluated by comparing the fractal dimensions produced by simulations with those observed in the ISM.

The SCF differs from both of these methods in that it considers the elements of a map as spectra rather than considering the whole map as a three dimensional array. By making this distinction, the SCF is able to locate similarities between structures at different rest velocities which would not appear in three dimensional analysis.

1.3.3 Scale Investigations

The recognition of similarities at different size scales motivated fractal analyses. Observations also indicate that there are relations between some observed quantities and the size scale over which these quantities are measured. The fiducial example is given by Larson (1981) who noticed that there exists a power law relationship between the width of a given line and the scale over which the width is measured (that is $\Delta v \propto \ell^a$). This observation probes the nature of turbulent motion in the ISM where the specific kind of turbulence is related to the exponent a. Such analysis is continued in other papers and the scales on which such relationships break down can be used to understand the behavior of turbulence in the ISM (Goodman et al., 1997).

In addition to testing correlations among size-scale and other simple observables, astronomers have also worked to examine more complex relationships. In these studies, the similarities between basic properties of two spectra within a map are compared in terms of the distance in pixels between those spectra. An example of this kind of analysis is the use of the autocorrelation functions (Dickman and Kleiner, 1985). In this analysis, a certain characteristic of the spectra (for example, the velocity offset) is compared between all the spectra a set distance apart. The average value of the difference between these characteristic values is averaged over the map and the result characterizes that parameter on the given size scale. The size scale is then changed and the behavior of the function is then analyzed. Other types of functions are also used but their primary purpose is the development of relations between observed spectral quantities and size scales.

A final method of interest is that of Principal Component Analysis (PCA) established by Heyer and Schloerb (1997). The key to PCA is to realize that the various line parameters in a map are related. For example, the amount of material giving rise to a line may influence both the peak antenna temperature and the width of the line. In realizing this, the data can be reduced to a few independent variables called the principal components. These components represent all significant variation in the data. The analysis of how well these principal components model the data as well as studying the maps of these components will produce more information about the clouds.

The principal difference between these methods and the SCF is that they focus upon the relationships between objects at set scales often larger than those considered in the analysis of neighbors that the SCF produces. The SCF also focuses on different qualities than do the above studies and most importantly, it preserves the three dimensional nature of the data that these studies reduce in other fashions. In some fashions, the SCF is most akin to autocorrelation functions; however, the notion of scale is much more malleable with the SCF than it is with these other analyses.

1.4 The Origins of the Spectral Correlation Function

The above methods of analysis present several manners in which the data cubes can be examined. In all cases, the methods involve the reduction of the data to a manageable and easily interpreted set. These reductions choose to include or exclude specific aspects of the data in their condensation; and naturally, much of the character of the data is lost or obscured. The SCF offers another approach to analyzing data cubes, preserving some aspects of the data and eliminating others. In particular, the SCF focuses on the similarities among constituent spectra in a data cube in shape, size, and velocity offset.

Certain reductions of data sets eliminate the uniqueness of the data set that generates the reduced data set. In other words, many different data sets will produce the same reduced data set. A particularly relevant example would be the reduced data set created by generating a histogram of the peak antenna temperatures for the constituent spectra in the cube. Naturally, a cube with the same spectra but in randomized positions will produce the same histogram. This type of uniqueness loss is the fundamental problem that the SCF was formulated to help solve. Specifically, the SCF set out to probe the uniqueness of MHD simulations of the ISM and the spectral maps that these cubes produce. The goal of any simulation of the ISM is to mimic the properties of an observed portion of the ISM as accurately as possible. Because it is impossible to establish the exact initial conditions that led to a particular object observed in a spectral map, these simulations must mimic the more general qualities of observational data. A proposed goal for the simulation projects is to develop a model that reproduces the distributions of spectral characteristics (e.g. antenna temperature, line widths, velocity offsets etc.) that are found in observations of the ISM. Falgarone *et al.* (1994) have proposed using the statistical moments of these distributions to evaluate the degree to which simulations approximate the observed qualities of the ISM. In their paper, they find that the observed maps moments matched their simulations with a high degree of accuracy. To date, this is the only published work specifically evaluating hydrodynamic simulations by comparing them with real spectral maps.

As mentioned previously, the goal of reproducing characteristic distributions suffers from a lack of uniqueness in the data set from which the distributions are generated. Hence, the placement of spectra in the data cube can be random and the same distributions will be generated. Naturally, in the observed ISM, it is impossible to formulate a cube which will have the randomized positions of a possible theoretical cube. More specifically, the spectra in an observed map of the ISM are assumed to share some characteristics with their neighbors in pixel space because the physical regions from which the spectra are formed should share some physical similarities with their neighbors in real space. The hypothesis is that the shared characteristics of the neighboring spectra can be quantified through the development of an appropriate correlation function. These observations were prompted by the Falgarone *et al.* work on comparing spectra. In particular, the spectra in their final simulations seem to have radically dissimilar spectra next to each other, in contrast with the observed data with which they compared their work. It is this work which has prompted the development of the SCF to refine the evaluation criteria for MHD models of the ISM.