DRAFT VERSION NOVEMBER 29, 2021
Typeset using LATEX default style in AASTeX62

Assessing bias in estimates of turbulence using multi-component fitting in the Orion Integral Shaped Filament

RUBY E. FULFORD,[1,2] BRIAN E. SVOBODA,[1] DAVID S. MEIER,[3,4] AND JÜRGEN OTT[1]

[1]National Radio Astronomy Observatory, PO Box O, Socorro, NM 87801 USA
[2]Steward Observatory, University of Arizona, 933 N Cherry Ave, Tucson, AZ 85719
[3]Physics Department, New Mexico Institute of Mining and Technology, 801 Leroy Place, Socorro, NM 87801, USA
[4]Adjunct Astronomer, National Radio Astronomy Observatory, Socorro, NM, 87801, USA

ABSTRACT

Understanding the physics and chemistry of molecular clouds is essential to understanding the environments in which stars form. High-mass star formation in particular remains an enigma. The Orion A molecular cloud, also known as the Orion Integral Shaped Filament (ISF) is the closest high-mass star forming region to our solar system and is therefore crucial to our understanding of star formation. Recently, the Green Bank Ammonia Survey (GAS; Friesen et al. 2017) mapped the ISF with observations of NH_3 inversion transitions. They further analyzed this data to infer the ISF's physical properties using a model that assumed a single velocity component along the line of sight. In this work we use NestFit, a Bayesian software framework, to perform multi-modal Nested Sampling on the GAS observations to fit up to two velocity components. We compare the results of the two-component model to those of the one-component model to determine whether assuming a single component biases model results. We find that the resulting distributions for many properties do not differ significantly between the two models. However, unlike the one-component model, we find that the two-component model shows a peak in velocity dispersion at the thermal speed, which implies the presence of non-turbulent, thermal gas. The excess of gas at the thermal speed over previous analyses is evidence for a greater quantity of gas unstable to gravitational collapse and efficient star formation.

Keywords: stars: formation — ISM: clouds — ISM: molecules — ISM: structure — Orion A

1. INTRODUCTION

Star formation is a fundamental astrophysical process integral to the structure and composition of the universe. While low-mass star formation is fairly well understood, the details of high-mass star formation remain an open problem (Motte et al. 2018). High-mass stars are defined as having masses $> 8\,\mathrm{M}_\odot$. The continuously increasing sensitivity and versatility of radio telescopes has recently strengthened our ability to observe molecular clouds, the regions in which star formation begins (Tan et al. 2014). Studying molecular clouds is imperative to improving our understanding of high-mass star formation. Demystifying massive star formation will supplement the current understanding of low-mass star formation to create a more complete understanding of star formation at all masses, which will inform our understanding of planetary system formation.

Molecular clouds are regions of molecular gas, dense relative to the ISM, that are largely comprised of H_2 and tiny dust grains of high visual extinctions. These grains absorb starlight and often give the regions optically dark appearances. The gas in molecular clouds undergoes thermal motion resulting from its kinetic temperature. The gas also undergoes the more random and complex turbulent motion, which can result from star formation feedback, supernovae, and other non-thermal sources. Turbulence counteracts the gravitational collapse of gas necessary for star formation, and therefore less turbulent regions form stars more efficiently (Bergin & Tafalla 2007).

The Orion Molecular Cloud, 0.45 kpc from the sun, is the nearest example a high-mass star forming region. (Genzel & Stutzki 1989; Motte et al. 2018). The Orion ISF, a filament that includes the Orion Nebula, is a smaller region of the greater Orion Molecular Cloud Complex and is the focus of this project.

Molecular clouds are frequently comprised of multiple velocity components. Velocity components refer to the sections of gas within a cloud that move at different velocities along the line of sight. When examining spectral profiles of observations towards

Corresponding author: Ruby Fulford
ref1@email.arizona.edu

molecular clouds, it is difficult to distinguish between a single component with a wide velocity dispersion and multiple overlapping narrow components, especially where the signal-to-noise ration (SNR) is low. In the past, the only way to distinguish between different numbers of velocity components has been via visual assessment of spectral profiles, which is not feasible when analyzing images with large numbers of pixels that each have unique spectral profiles. However, recent developments of automated statistical model comparison methods have made it possible to analyze a large number of spectral profiles at once and determine the number of velocity components most likely to comprise each profile.

The model used by Friesen et al. (2017) to analyze and infer the physical properties of Orion A assumes a single velocity component. However, star-forming molecular clouds are usually turbulent environments of irregular structure. ALMA observations conducted and analyzed by Hacar et al. (2018) suggest that the ISF is comprised of complex bundles of fibers. Since individual fibers are likely moving at different velocities along the line of sight, modeling the ISF as a single velocity component may not be consistent with the physical reality of the cloud. In this project, we use multi-modal Nested Sampling to fit a model of up to two components to Orion A and compare these results to those of the one-component GAS model. We aim to determine whether ignoring additional velocity components biases the results of a model, and whether using a multi-component model reveals any new information about the turbulence and star-formation efficiency of Orion A. Since Orion A is widely studied and significant to the field of massive star formation, it is important that models being used to infer its properties are as unbiased as possible.

2. OBSERVATIONS

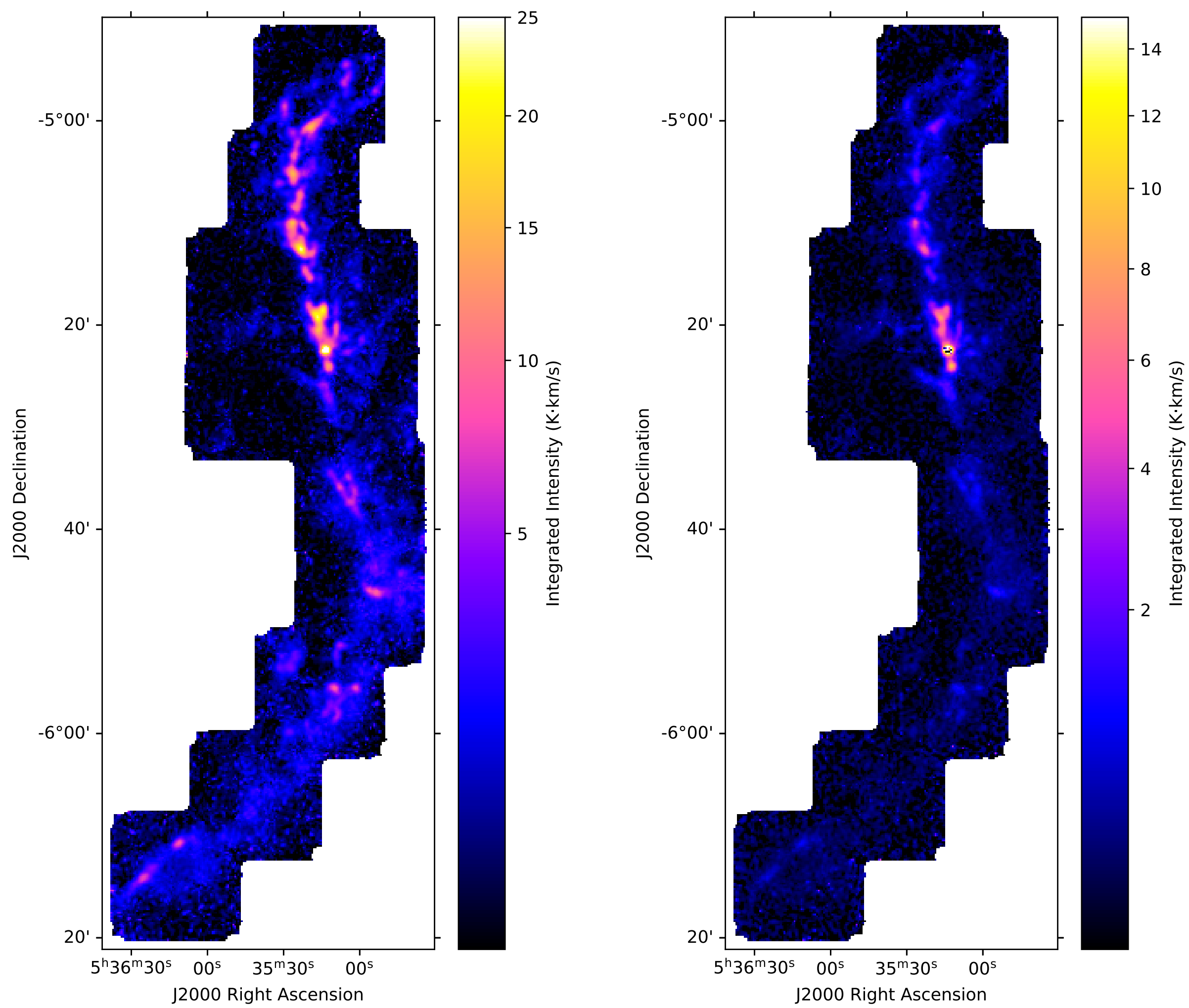

Figure 1. Moment 0 maps of NH_3 inversion transition emission toward Orion A, observed via the Green Bank Ammonia Survey. Left: (1,1) Transition. Right: (2,2) Transition.

This paper utilizes radio observations from the Green Bank Ammonia Survey (GAS), a program aiming to map all Gould Belt star-forming regions with $A_V \gtrsim 7$ mag visible from the northern hemisphere. The Gould Belt is mapped with radio emission from NH_3 and other important molecules, detected by the Green Bank Telescope. Observations from this survey have a spectral resolution of $5.7\,\mathrm{kHz}$, or approximately $0.07\,\mathrm{km\,s^{-1}}$, at $23.7\,\mathrm{GHz}$.. The first GAS data release includes observations of four regions in the Gould Belt: B18 in Taurus, NGC 1333 in Perseus, L1688 in Ophiuchus, and Orion A in North Orion. As aforementioned, Orion A is the focus of our work (GAS; Friesen et al. 2017).

GAS detects emission from the (1,1) and (2,2) NH_3 inversion transitions(Friesen et al. 2017). Inversion transitions occur when the N atom of an NH_3 molecule tunnels through the three H atoms, thus "inverting" the molecule, and splitting each of its (J,K) quantum energy levels into two levels. Inversion releases a radio photon, at $23.69\,\mathrm{GHz}$ and $23.72\,\mathrm{GHz}$ for the (1,1) and (2,2) transitions respectively. Figure 1 shows moment 0 maps of (1,1) and (2,2) transitions in the Orion ISF. The mode RMS noise of the (1,1) and (2,2) observations is approximately 0.09 K.

3. METHODS

This project utilizes Bayesian inference, a statistical inference method that builds upon prior knowledge when fitting models to data. Bayesian statistics is very useful in astronomical contexts; it allows us to compare different models in a rigorous, statically justified manner. The foundation of Bayesian statistics is Bayes' theorem:

$$p(M_i|D,I) = \frac{p(M_i|I)p(D|M_i,I)}{p(D|I)}, \tag{1}$$

where M_i is one of two or more alternate models, D is the data, and I is prior known information (Gregory 2005). In the context of our research, different models assume different numbers of velocity components for Orion A: 0 components, 1 component, and 2 components. If a pixel has 0 velocity components, it contains only noise and no signal.

$p(M_i|D,I)$, also known as the posterior distribution, denotes the probability that a model M_i is true given the data observed and any previously known information. $p(D|M_i,I)$, also known as the likelihood of the model, denotes the probability that a given model would produce the observed data. $p(M_i|I)$ is the prior distribution, which represents any previously known information about the model and its parameters. For a model with multiple parameters, the prior distribution for a parameter θ of model M_i is $p(\theta|M_i,I)$. The posterior distribution of a parameter θ is $p(\theta|M_i,D,I)$.

$p(D|M_i)$, known as the Bayesian evidence, is the probability that a set of data is observed assuming the entire model, including all parameters, is true. The evidence can be found by integrating the numerator of Equation 1 over the parameter space (Feroz & Hobson 2008). Evidence is important when comparing models because the ratio between the evidences of two models produces the Bayes' Factor, as shown below:

$$B_{ij} = \frac{Z_i}{Z_j}, \tag{2}$$

where i and j are model indices, Z is the Bayesian evidence of a model, and B is the Bayes' factor. The higher the Bayes' factor, the better of a fit model i is compared to model j (Gregory 2005).

To begin Bayesian analysis of the GAS data, we first created prior distributions for each of the five parameters that made up our model and the GAS model: centroid velocity (V_{lsr}), velocity dispersion (σ_ν), kinetic temperature (T_{kin}), excitation temperature (T_{ex}), and column density (N). We extracted the parameter values from property maps of each parameter that resulted from the model fitting described in Friesen et al. (2017). Figure 2 shows GAS model centroid velocity and velocity dispersion property maps. Property data was extracted from each map as a 2D array and flattened into a 1D array. The 1D arrays were plotted as histograms for each parameter, shown in Figure 3. We eliminated from our histograms pixels with error greater than 1 K for T_{ex} and T_{kin}, 0.1 km/s for σ_ν, 0.25 cm^{-2} for N, and 0.3 km/s for V_{lsr}. These are the same limits used to create the parameter histograms in Friesen et al. (2017). We also eliminated pixels with errors or values of 0.

We constructed β distributions to envelop each histogram, which became the prior distributions on which we based our Bayesian analysis. As evident in Figure 3, the β distributions were intentionally shaped slightly larger than the histograms to account for any true values not produced by the GAS model. For the centroid velocity histogram, two β functions were added together to account for the histogram's two modes. The x-axis endpoints along with the a and b values used to create the β distributions were changed for each parameter. The values used are displayed in Table 1.

After creating the prior distributions, we downloaded the NestFit software framework (as well as the MultiNest FORTRAN software) for use in fitting a two-component model to the GAS data. NestFit employs a Bayesian statistical method known as multi-modal Nested Sampling. This method is unique in that it provides the evidence of a model, while similar methods, such as Markov Chain Monte Carlo methods, ignore evidence and essentially treat it as a proportionality constant between the two sides of

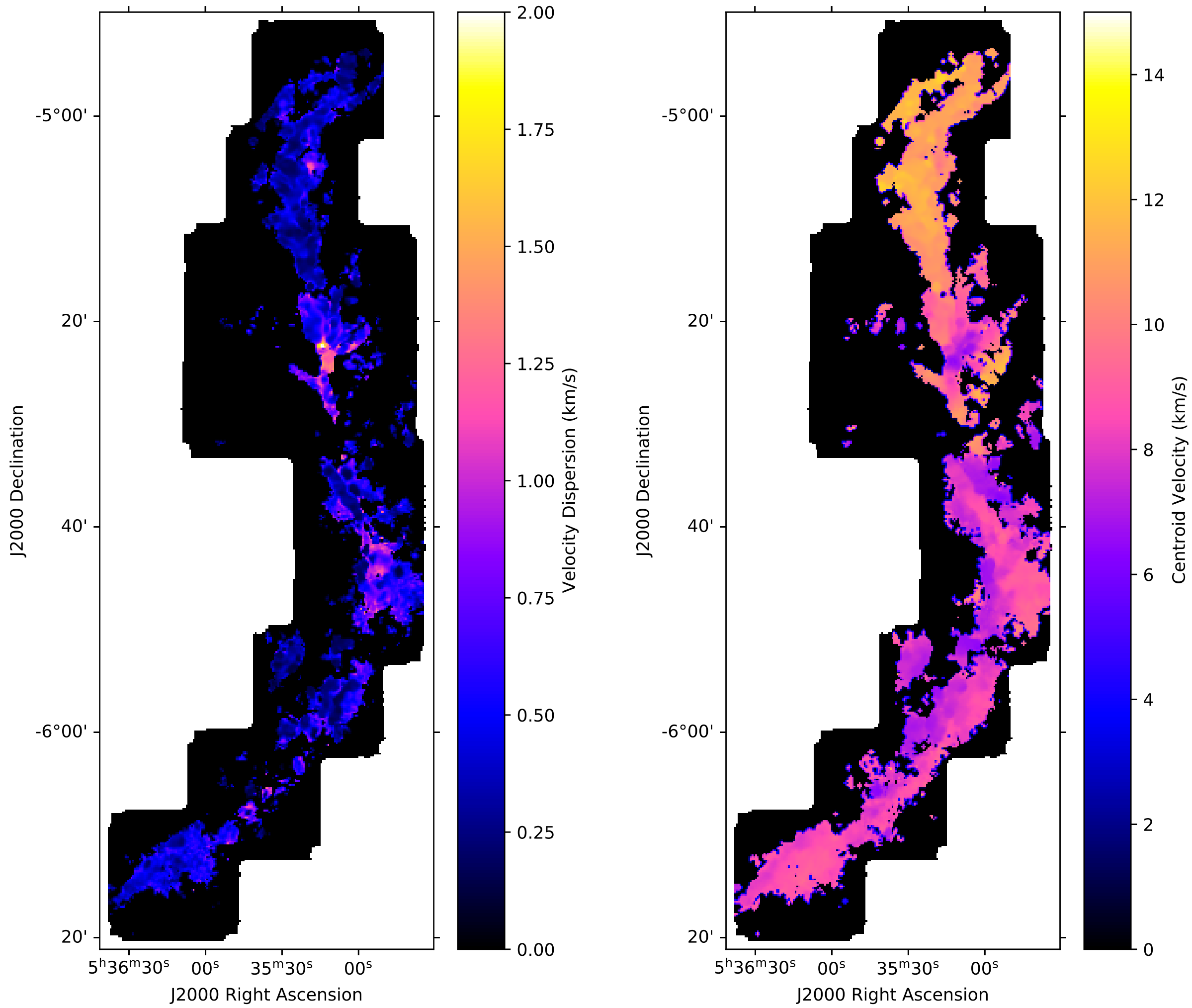

Figure 2. Left: Velocity dispersion in the ISF according to the model used in Friesen et al. (2017). Right: Same as left, but for centroid velocity

Property	a-value	b-value	x-min	x-max
V_{lsr}	6.3, 20	10, 14.7	13.3	16
T_{kin}	2	5.5	9	56
T_{ex}	1.01	3	2.8	14.5
σ_ν	1.7	6.3	0	2.2
N	3.45	6.8	13.3	16

Table 1. Parameters used to create β functions for each property. β functions were created using the ”beta” python function. Note that the velocity dispersion β function has the ”loc” parameter set to 0.032. Velocity dispersion was the only property for which the ”loc” parameter was changed from the default.

Bayes’ theorem (Feroz & Hobson 2008). Nested Sampling produces evidence values for the 2, 1, and 0 component models for each pixel. Multinest, a software integral to NestFit, enables multi-modal nested sampling and is particularly effective at handling

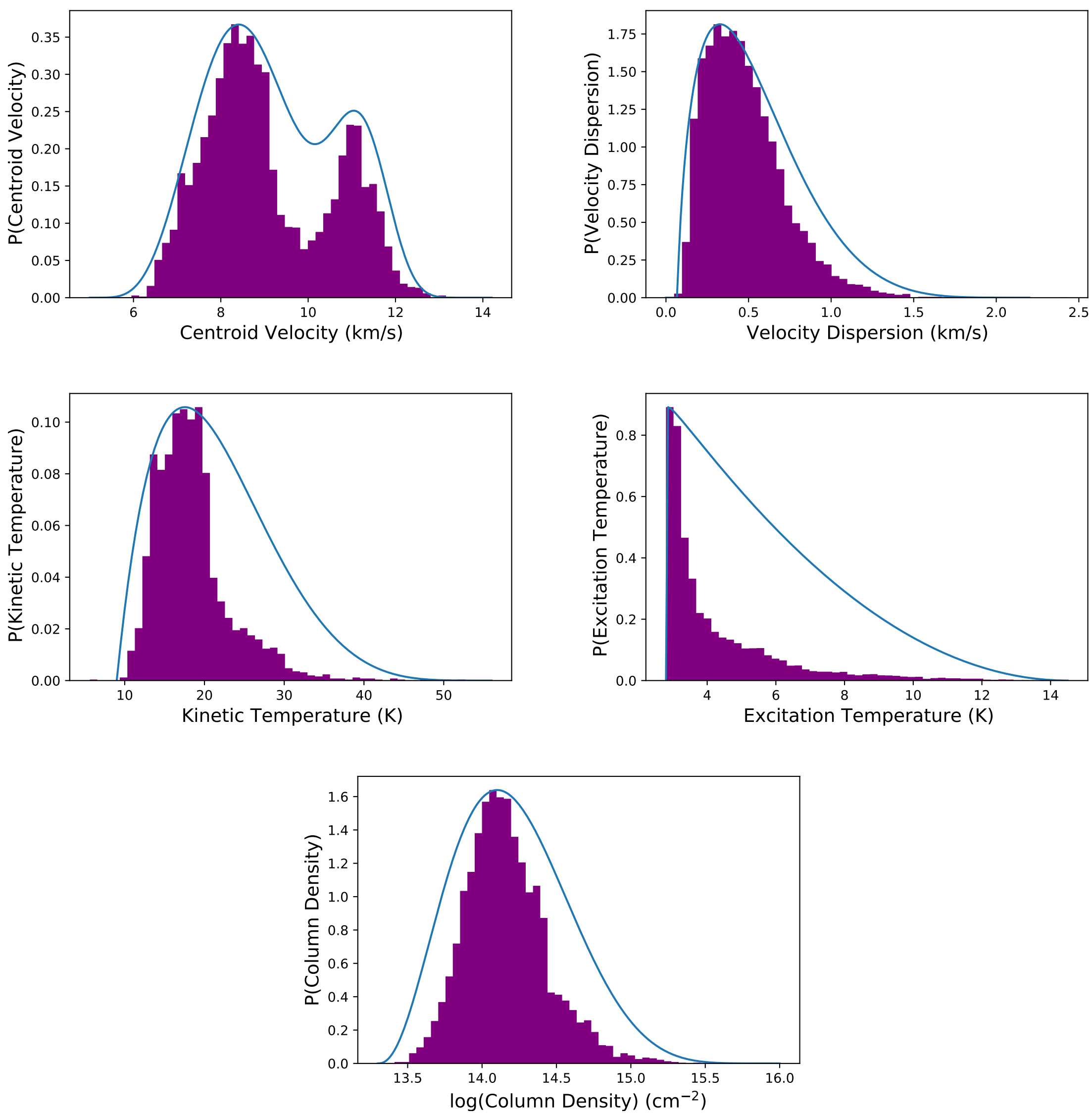

Figure 3. β distributions shaped over histograms of parameter value results from the GAS one-component model, which are used as prior distributions when fitting a NestFit model of up to two components to the GAS data. The β distributions' shapes were chosen via visual assessment; they are not statistically fit over the histograms. The shapes were chosen to eliminate any physically nonsensical values, but to still allow for any true values not detected by the GAS model.

multi-modal posteriors, so it is useful for detecting multiple velocity components (Feroz et al. 2009). Utilizing the GAS data and the priors we constructed, NestFit created posterior distributions for the each parameter and used its evidence calculations to determine the most probable number of velocity components present at each pixel. NestFit products included model data cubes for the (1,1) and (2,2) transitions as well as an HDF file containing posterior and evidence values, which we sliced using the `h5py` python package in order to analyze posteriors and velocity component results.

4. RESULTS

NestFit determined which areas of Orion A are likely to have zero, one, or two velocity components, as shown in Figure 4.

Figure 5 compares the property distributions resulting from the one-component GAS model to those resulting from our two-component NestFit model. The NestFit medians tend to be slightly lower than the GAS medians, but not by values of statistical significance. The NestFit histograms contain more values than the GAS histograms since they include values from both velocity components, which may effect the scale of the NestFit histograms.

Figures 6 and 7 map the maximum a posteri values for velocity dispersion and centroid velocity respectively at each pixel and each velocity component. The maps for the second velocity component are smaller than the first since not every pixel was found by NestFit to have two components.

Figure 8 shows the 1-0 Bayes' Factor map, which compares the probability that a pixel will contain emission of one velocity component to the probability that it will contain only noise. This Bayes' factor map is placed beside a (1,1) transition moment 0 map that has been adjusted to better display faint emission.

Figures 9-11 display posterior distributions for each of the five parameters for four example pixels. The examples chosen are a pixel with 2 velocity components, a pixel with 1 component and strong emission, a pixel with 1 component and weak emission, and a pixel containing only noise. The spectral profile for the two-component pixel is shown in Figure 12.

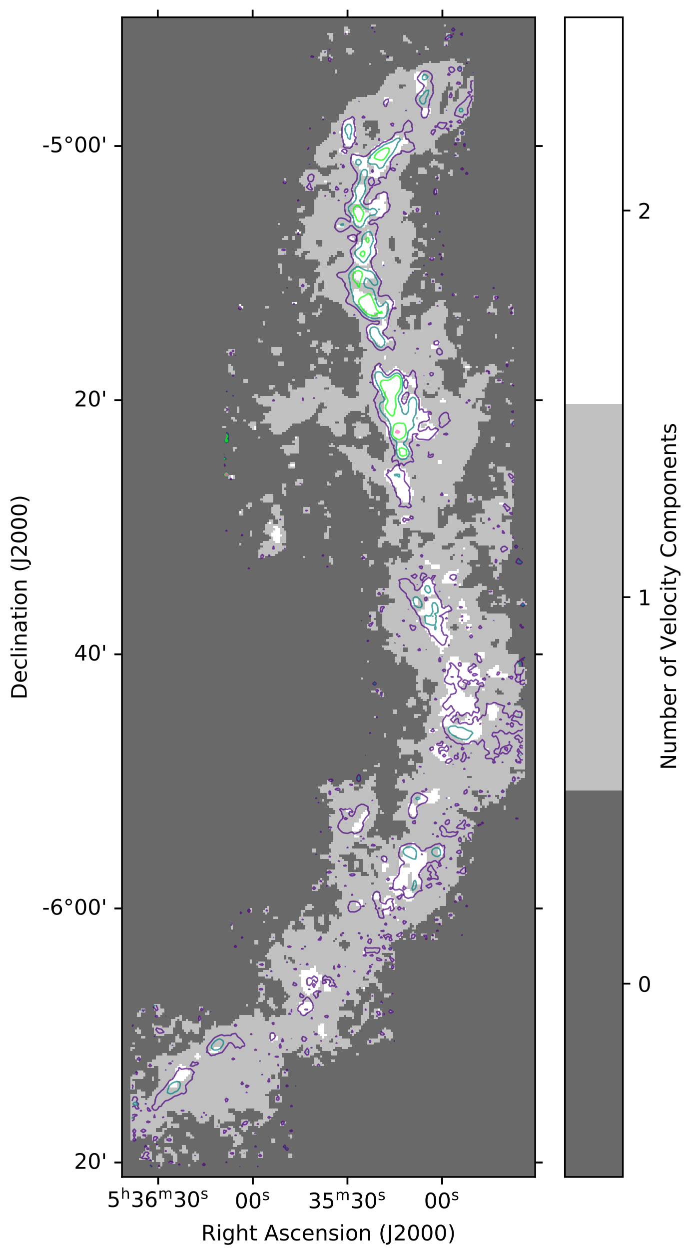

Figure 4. The number of velocity components most probable at each pixel, according to the results of two-component Nestfit modelling. The contours show the (1,1) moment 0 map, at levels of 2, 5, 10 and 40 K · km/s

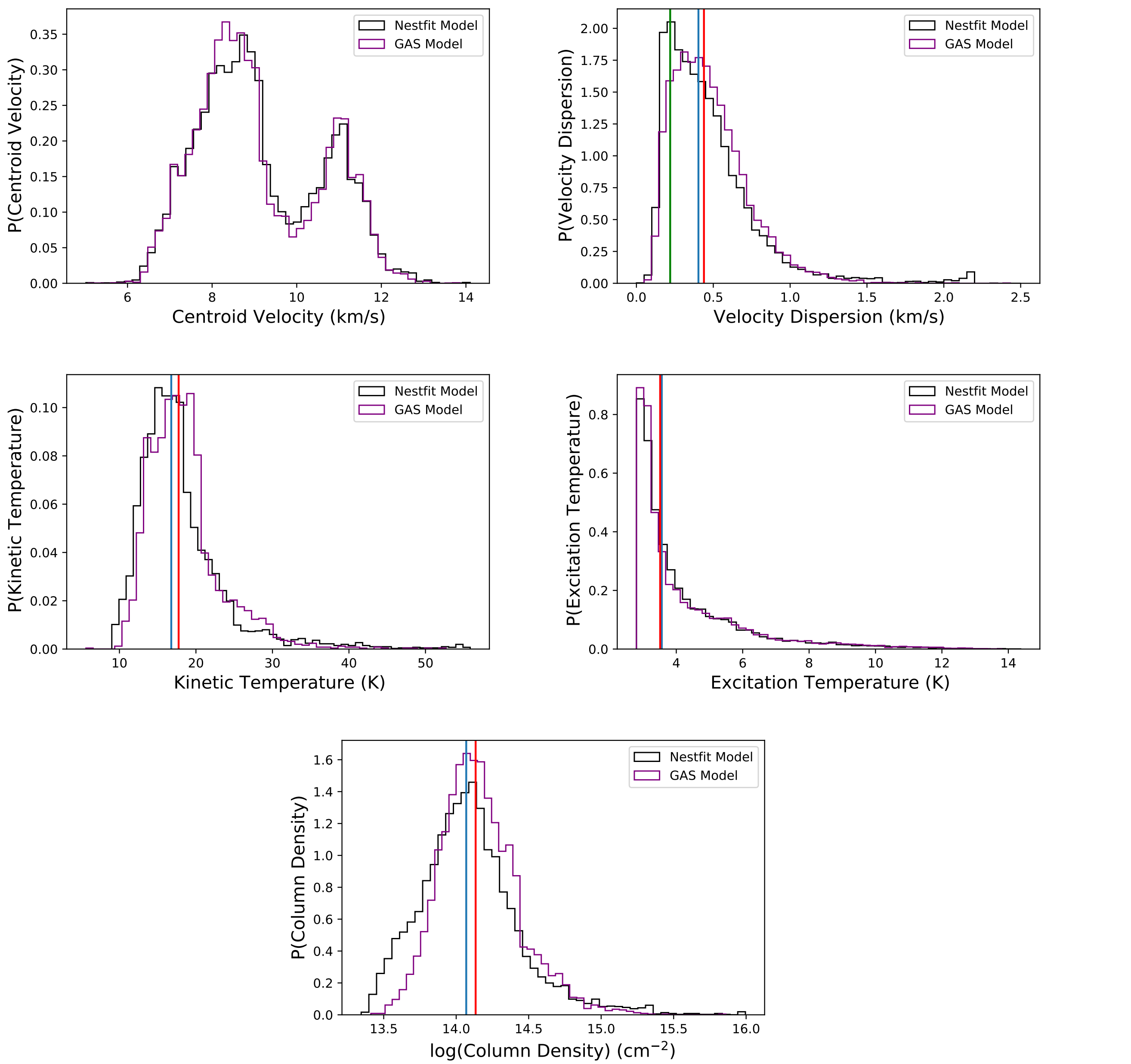

Figure 5. The results of the one-component GAS model compared to the results of the two-component Nestfit model. The red vertical lines represent the medians of the GAS distributions, and the blue vertical lines represent the medians of the Nestfit distributions. The green vertical line on the velocity dispersion plot denotes the thermal speed in Orion A, calculated using the median kinetic temperature of NH_3 in Orion A as determined by the Nestfit model.

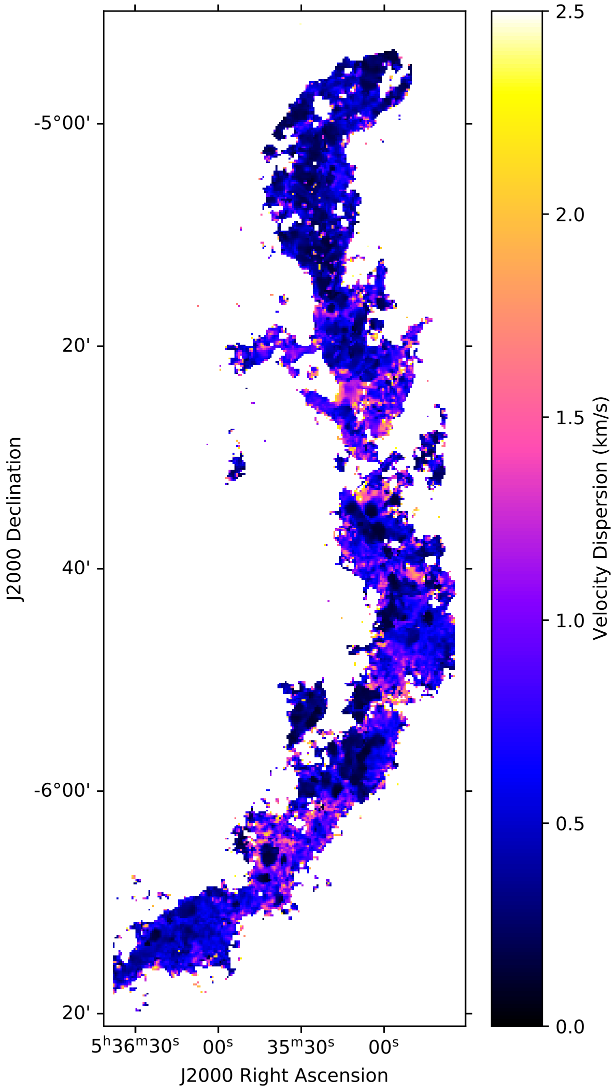

Figure 6. Maps of the maximum a posteri velocity dispersion values for both velocity components, resulting from the NestFit model fitting.

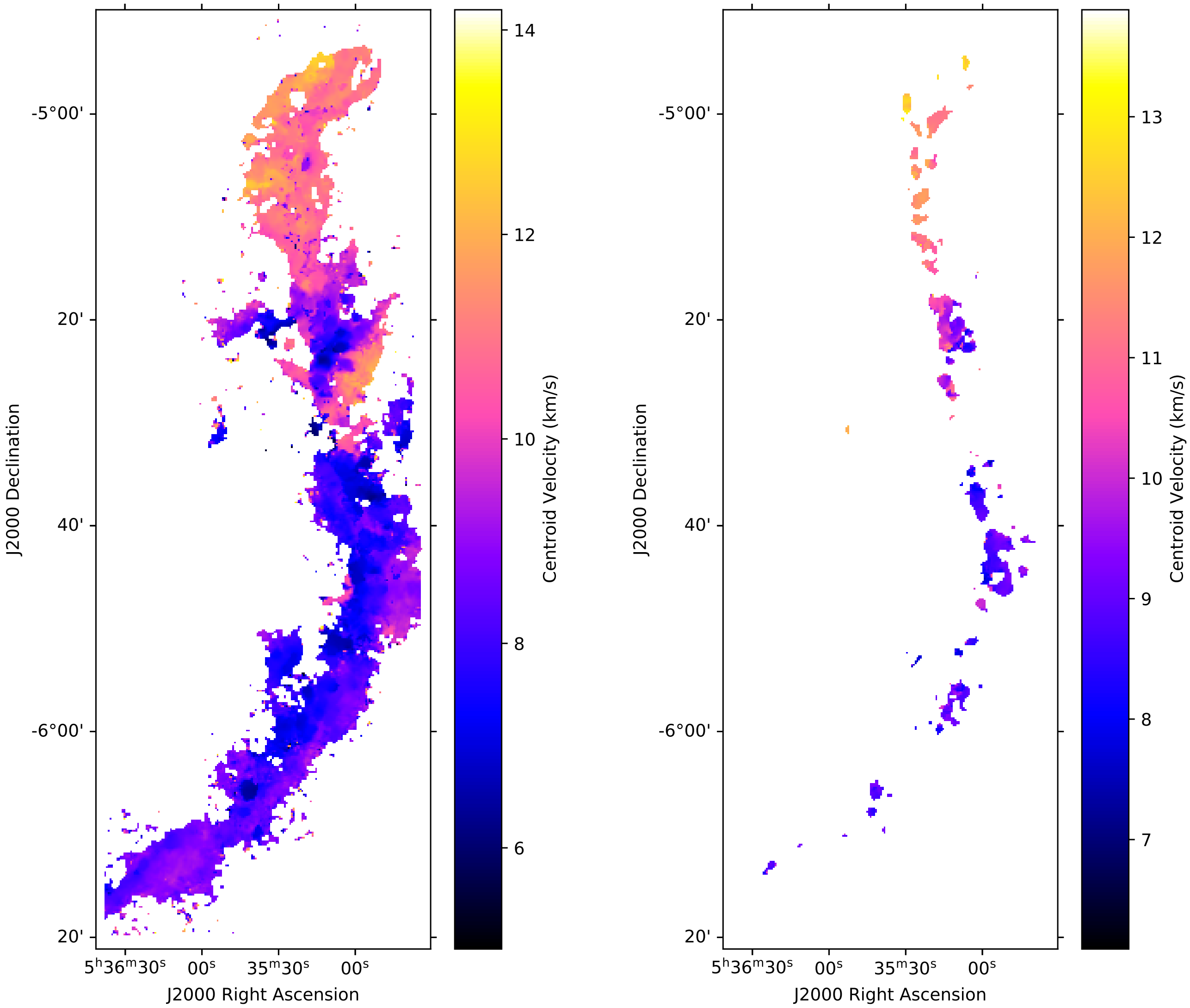

Figure 7. Same as Figure 6, but for centroid velocity

.

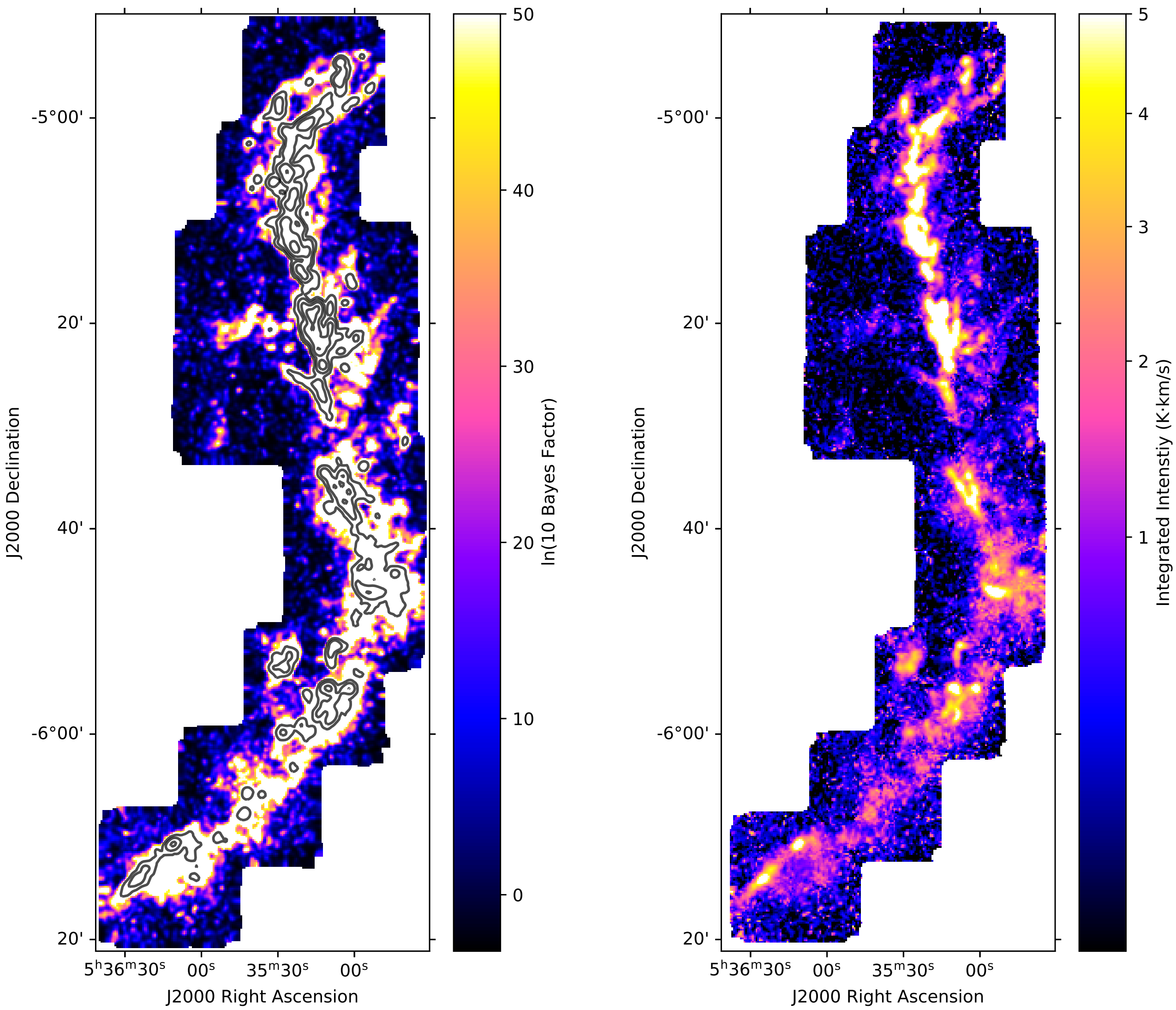

Figure 8. Left: The ratio between the Bayesian evidence for one velocity component and the Bayesian evidence for noise, also known as the 1-0 Bayes factor, for each pixel of our Orion A map. The contours show the NH_3 (1,1) integrated intensity at levels of 200, 1000, 5000, and 25000 K·kms/s. Right: A moment 0 map of the (1,1) NH_3 transition emission towards Orion A, adjusted to show faint emission.

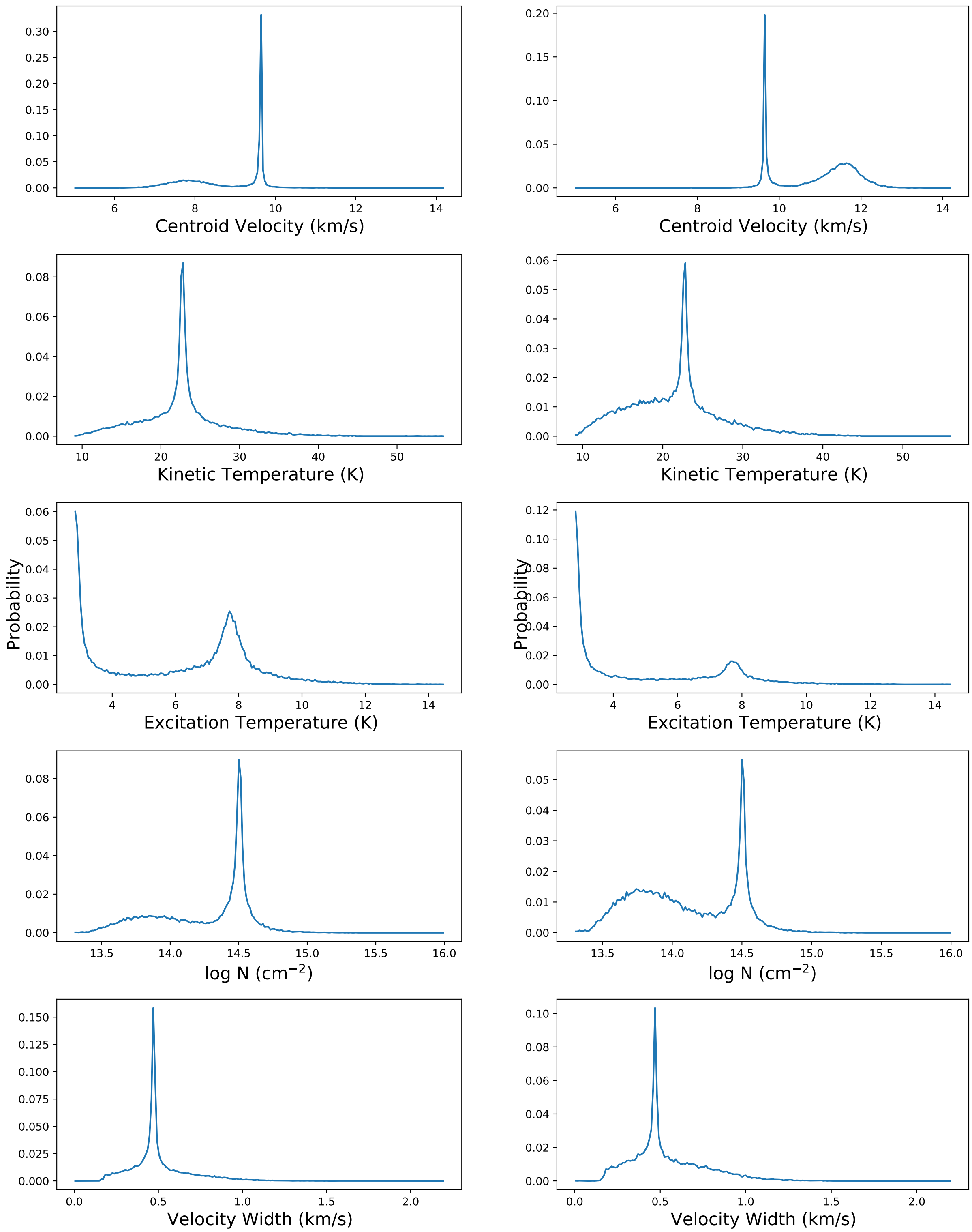

Figure 9. Posterior distributions of each parameter, resulting from Nestfit model fitting, for a pixel found to contain two velocity components. The left and right columns are posteriors for the two different velocity components.

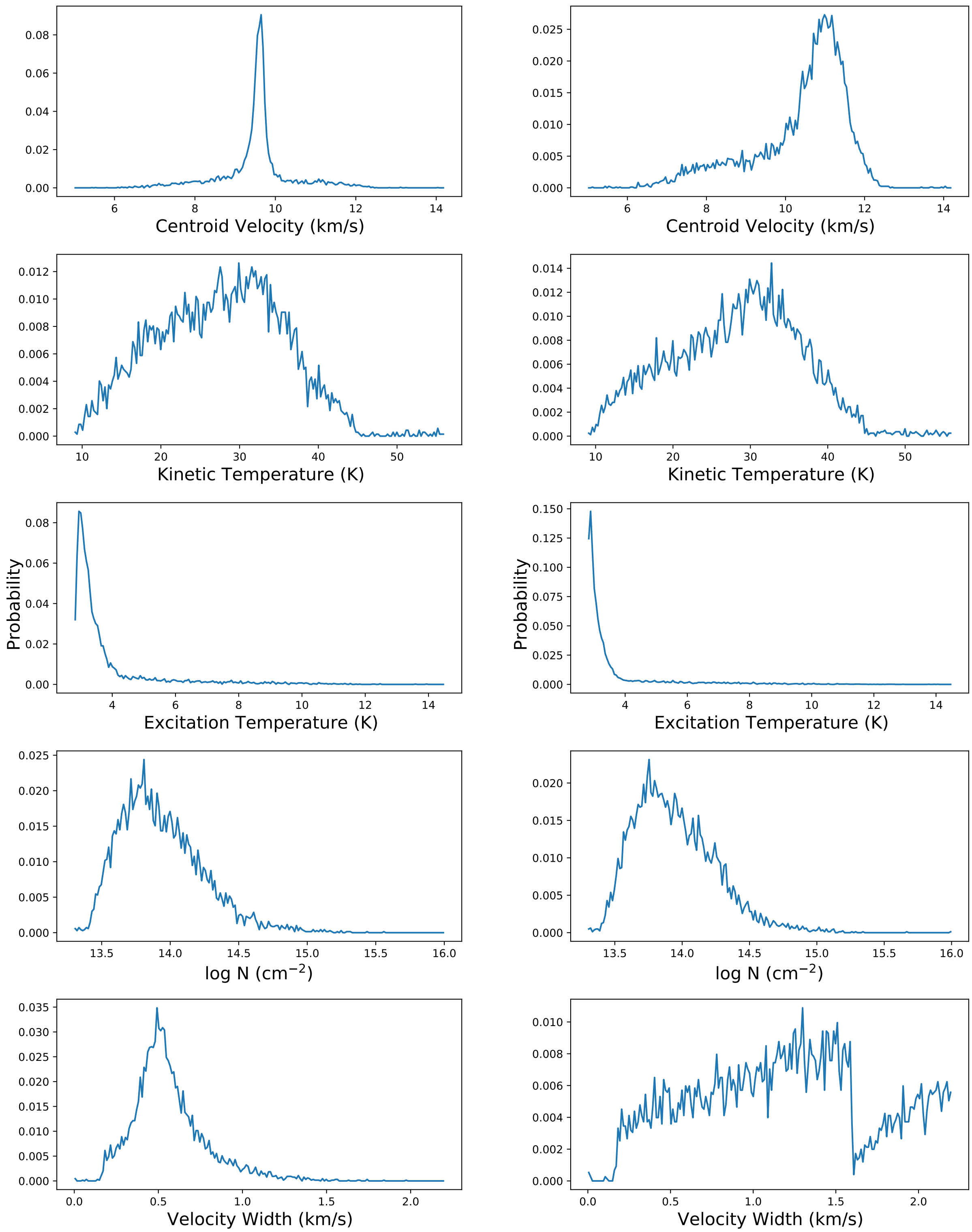

Figure 10. Posterior distributions for two pixels each containing one velocity component, one pixel containing strong emission (right) and one containing faint emission (left).

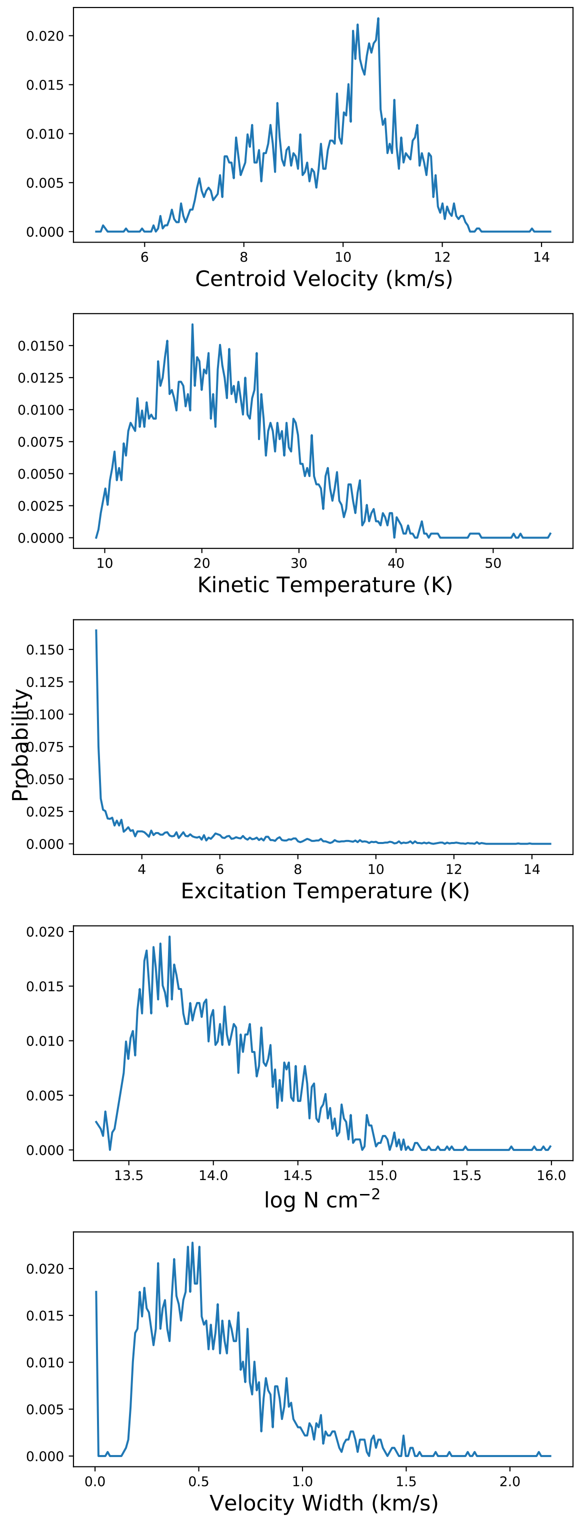

Figure 11. Posterior distributions for a pixel containing only noise.

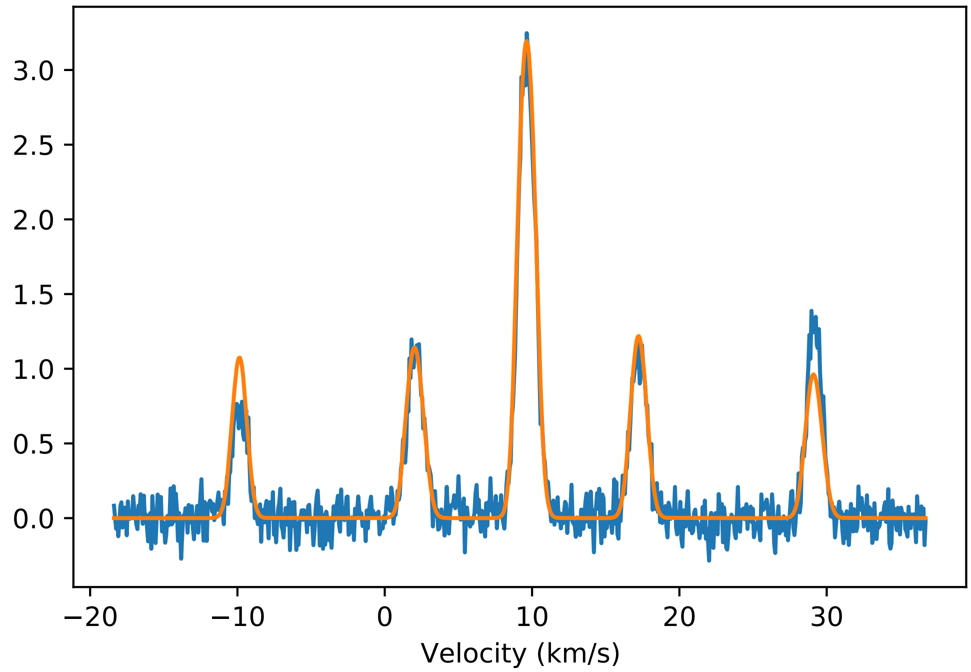

Figure 12. The observed spectral profile for a pixel found by NestFit to have two velocity components, shown in blue, compared to the model spectral profile resulting from NestFit model fitting, shown in orange.

5. DISCUSSION

The posterior plots in Figures 9-11 demonstrate that centroid velocity tends to be the best constrained parameter. Excitation temperature tends to be the worst constrained, evident from the fact that the example posterior distributions for excitation temperature appear very similar to the prior distribution.

It is interesting to compare the moment 0 map of the GAS data to the Bayes 1-0 map, as shown in Figure 8. The Bayes 1-0 map displays the faint outer structure of the filament in much better detail than the moment 0 map, even when said moment 0 map is adjusted to display faint emission. Since the moment 0 map displays integrated intensity, regions that are faint relative to the rest of the cloud are overshadowed by brighter regions. However, the Bayes 1-0 map does not focus on intensity, but rather the probability that emission is present. If a pixel clearly contains emission, it will be displayed prominently on a Bayes 1-0 map, even if said emission is not especially bright.

Figure 4 implies that areas of higher intensity are more likely to have two velocity components. The presence of multiple velocity components could be a result of ultraviolet radiation from star formation feedback eroding Orion A's gas and breaking up its uniformity, which would make dense areas with high star formation rates more likely to have multiple components. In reality, most regions of the cloud likely have multiple clumps of gas along the line of sight, not just the regions determined to have two components by our model. A complex velocity structure is consistent with the evidence presented in Hacar et al. (2018) that Orion A is made up of many filaments and fibers. However, in fainter, lower SNR regions, noisier spectral profiles can often be sufficiently fit by a single Gaussian, giving the appearance of a single velocity component even if more are present in reality. In higher SNR areas, spectral profiles are better defined, and if two or more velocity components are present it will become clear that the spectral profile cannot be fit by a single Gaussian. For this reason, it is more difficult to detect multiple velocity components when SNR is low.

Despite the probable presence of multiple velocity components, using a model that allowed for up to two components did not hugely change the parameter value distribution shapes or the median parameter values. This indicates that single-component models like the one described in Friesen et al. (2017) are reliable ways to determine the properties of Orion A, despite not taking all velocity components into account. The shape of the spectral profile for the example pixel with two components, shown in

Figure 12, suggests that the second component is more akin to a small shoulder on the highest velocity peak than a completely separate second peak. This shape may explain why disregarding the second component does not hugely alter model results.

Though the results are largely similar, a notable difference between the results of the NestFit two-component model and the results of the GAS one-component model is that the NestFit model's results show a peak in velocity dispersion at the thermal speed, about 0.22 km/s, as is evident from Figure 5. Since the NH_3 line is well resolved (the FWHM spans about 10 channels), this pile-up at the thermal speed is real, not a misleading result of low resolution. When velocity dispersion is equal to the thermal speed, only thermal motion is occurring; additional velocity dispersion indicates turbulence. Therefore, the peak at the thermal speed on the NestFit velocity dispersion histogram is representative of thermal, non-turbulent gas not detected by the GAS model. Since less turbulent gas is more efficient at forming stars, the presence of thermal gas is indicative of the star formation occurring in the ISF. Figure 6 shows that the regions of Orion A with low velocity dispersion correspond to the denser, brighter regions of the cloud as shown by Figure 1, which is where star formation is likely occurring. All the other star-forming molecular clouds studied in Friesen et al. (2017) show a velocity dispersion peak at the thermal speed, so it is logical that Orion A also displays this peak. Since Orion A hosts more active star formation than the other clouds studied, it may be more turbulent, which would explain the necessity of a multi-component model that accounts for turbulent gas when attempting to detect thermal gas in Orion A.

6. CONCLUSION

Orion A, also referred to as the Orion Integral Shaped Filament, is the closest massive star forming region to our solar system. This region was recently mapped by the Green Bank Ammonia Survey using observations of NH_3 inversion transitions. The paper in which the GAS data was released, Friesen et al. (2017), used a model to infer Orion A's properties that assumed a single velocity component. With this project we aimed to discover whether using a single-component model as opposed to a multi-component model biases results, and whether allowing for multiple components reveals new information about turbulence in the ISF. We found that significant regions of the cloud likely have two or more velocity components, but despite this, results do not differ greatly between the GAS one-component model and the NestFit two-component model. One noticeable difference between our results and those of Friesen et al. (2017) is that our two-component model finds a peak in Orion A's velocity dispersion at the thermal speed, indicating the presence of thermal, non-turbulent gas not detected by previous models. Since stars form more efficiently in non-turbulent environments, our findings are consistent with the star formation known to be occurring in Orion A. Our results could suggest that there are more dense, star-forming cores in Orion A than previously expected, since dense core gas tends not to move at supersonic speeds (Bergin & Tafalla 2007). Overall, we conclude that while a one-component model is a fairly reliable way to predict the properties of Orion A, multi-component models are best when studying the cloud's turbulence, velocity structure, and star-formation efficiency.

7. FIGURES

8. ACKNOWLEDGMENTS

We would like to thank the National Radio Astronomy Observatory, the Green Bank Observatory, and the National Science Foundation for their support on this project.

Facilities: Green Bank Telescope

Software: This research has made use of the following software projects: NestFit (Svoboda in prep) & (Svoboda 2021), Astropy (The Astropy Collaboration et al. 2018), Matplotlib (Hunter 2007), NumPy and SciPy (Oliphant 2007), IPython (Pérez & Granger 2007), h5py (Collette 2013), Cython (Behnel et al. 2011), pyspeckit (Ginsburg & Mirocha 2011), Spectral Cube (Robitaille et al. 2016), and the NASA's Astrophysics Data System.

REFERENCES

Behnel, S., Bradshaw, R., Citro, C., et al. 2011, Computing in Science & Engineering, 13, 31

Bergin, E. A., & Tafalla, M. 2007, ARA&A, 45, 339, doi: 10.1146/annurev.astro.45.071206.100404

Collette, A. 2013, Python and HDF5 (O'Reilly)

Feroz, F., & Hobson, M. P. 2008, MNRAS, 384, 449, doi: 10.1111/j.1365-2966.2007.12353.x

Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601, doi: 10.1111/j.1365-2966.2009.14548.x

Friesen, R. K., Pineda, J. E., co-PIs, et al. 2017, ApJ, 843, 63, doi: 10.3847/1538-4357/aa6d58

Genzel, R., & Stutzki, J. 1989, ARA&A, 27, 41, doi: 10.1146/annurev.aa.27.090189.000353

Ginsburg, A., & Mirocha, J. 2011, PySpecKit: Python Spectroscopic Toolkit. https://ascl.net/1109.001

Gregory, P. C. 2005, Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with 'Mathematica' Support

Hacar, A., Tafalla, M., Forbrich, J., et al. 2018, A&A, 610, A77, doi: 10.1051/0004-6361/201731894

Hunter, J. D. 2007, Computing In Science & Engineering, 9, 90, doi: 10.1109/MCSE.2007.55

Motte, F., Bontemps, S., & Louvet, F. 2018, ARA&A, 56, 41, doi: 10.1146/annurev-astro-091916-055235

Oliphant, T. E. 2007, Computing in Science & Engineering, 9

Pérez, F., & Granger, B. E. 2007, Computing in Science & Engineering, 9

Robitaille, T., Ginsburg, A., Beaumont, C., Leroy, A., & Rosolowsky, E. 2016, spectral-cube: Read and analyze astrophysical spectral data cubes. https://ascl.net/1609.017

Svoboda, B. 2021, autocorr/nestfit: Initial public release, v0.2, Zenodo, doi: 10.5281/zenodo.4470028. https://doi.org/10.5281/zenodo.4470028

Tan, J. C., Beltrán, M. T., Caselli, P., et al. 2014, in Protostars and Planets VI, ed. H. Beuther, R. S. Klessen, C. P. Dullemond, & T. Henning, 149, doi: 10.2458/azu_uapress_9780816531240-ch007

The Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123, doi: 10.3847/1538-3881/aabc4f