

# Global Optimization and Broadband Analysis Software for Interstellar Chemistry (GOBASIC)<sup>★</sup>

Mary L. Rad<sup>★★</sup>, Luyao Zou<sup>★★</sup>, James L. Sanders III, and Susanna L. Widicus Weaver

Emory University Department of Chemistry Atlanta, GA 30322, USA  
e-mail: susanna.widicus.weaver@emory.edu

Received 2 September 2013 / Accepted 12 October 2015

## ABSTRACT

**Context.** Broadband receivers that operate at millimeter and submillimeter frequencies necessitate the development of new tools for spectral analysis and interpretation. Simultaneous, global, multimolecule, multicomponent analysis is necessary to accurately determine the physical and chemical conditions from line-rich spectra that arise from sources like hot cores.

**Aims.** We aim to provide a robust and efficient automated analysis program to meet the challenges presented with the large spectral datasets produced by radio telescopes.

**Methods.** We have written a program in the MATLAB numerical computing environment for simultaneous global analysis of broadband line surveys. The Global Optimization and Broadband Analysis Software for Interstellar Chemistry (GOBASIC) program uses the simplifying assumption of local thermodynamic equilibrium (LTE) for spectral analysis to determine molecular column density, temperature, and velocity information.

**Results.** GOBASIC achieves simultaneous, multimolecule, multicomponent fitting for broadband spectra. The number of components that can be analyzed at once is only limited by the available computational resources. Analysis of subsequent sets of molecules or components is performed iteratively while taking the previous fits into account. All features of a given molecule across the entire window are fitted at once, which is preferable to the rotation diagram approach because global analysis is less sensitive to blended features and noise features in the spectra. In addition, the fitting method used in GOBASIC is insensitive to the initial conditions chosen, the fitting is automated, and fitting can be performed in a parallel computing environment. These features make GOBASIC a valuable improvement over previously available LTE analysis methods.

**Key words.** methods: data analysis – astrochemistry – line: identification – molecular data

## 1. Introduction

Broadband spectral line surveys are useful tools in astrochemistry, because they can be used to tightly constrain physical conditions in astronomical sources by probing a range of excitation energies for each observed molecule. Over 180 molecules have been detected in the interstellar and circumstellar medium<sup>1</sup> with a large number of those being complex organic molecules (COMs), i.e., those containing six or more atoms (Herbst & van Dishoeck 2009). Many astrochemical modeling networks strive to elucidate the chemistry which leads to the formation of these COMs (see Garrod et al. 2008; Laas et al. 2011 for examples). However, these models must be benchmarked against observational datasets to determine their level of accuracy. Spectral line surveys are well-suited for this purpose.

Spectral line surveys were historically time-prohibitive because of the limited bandwidth of heterodyne receivers and associated spectrometers. However, technological advancements have led to the development of broadband instruments that enable routine line survey observations. Vast spectral datasets are now available because of the implementation of this receiver

and spectrometer technology at observational facilities such as the *Herschel* Space Observatory and the Atacama Large Millimeter/submillimeter Array (ALMA). With this new generation of observational capabilities, the spectral line survey is becoming a default part of nearly every spectral observation. The challenge now becomes how to most efficiently analyze these datasets so that the physical and chemical information can be extracted on a timescale that is comparable to observing proposal cycles. While the rotation diagram method is routinely used for this type of analysis, this approach has shortcomings in terms of dealing with blended lines and multiple physical components. Programs such as the XCLASS/MAGIX program suite (Möller et al. 2013), CASSIS<sup>2</sup>, and WEEDS (Maret et al. 2011) are also available, but often take considerable amounts of time and user input. A more global fitting approach that requires little user input would be ideal for analysis of broadband line survey information.

To meet these needs, a spectral analysis program<sup>3</sup> using global optimization methods and local thermodynamic equilibrium (LTE) assumptions was created in the MATLAB scientific computing environment. The Global Optimization and Broadband Analysis Software for Interstellar Chemistry (GOBASIC) program offers a simple and uniform data analysis platform that can be adapted to fit the specific needs of the user

<sup>★</sup> A copy of the software is available at the CDS via anonymous ftp to [cdsarc.u-strasbg.fr](http://cdsarc.u-strasbg.fr) (130.79.128.5) or via <http://cdsarc.u-strasbg.fr/viz-bin/qcat?J/A+A/585/A23>

<sup>★★</sup> The first two authors contributed equally to this manuscript.

<sup>1</sup> <http://www.astro.uni-koeln.de/cdms/molecules> (Müller et al. 2005).

<sup>2</sup> <http://cassis.irap.omp.eu>

<sup>3</sup> [http://chemistry.emory.edu/faculty/widicusweaver/gobasic/GOBASIC\\_4.4.tar.gz](http://chemistry.emory.edu/faculty/widicusweaver/gobasic/GOBASIC_4.4.tar.gz)

while being applicable to a variety of observational data analysis problems. The first generation of this program, presented herein, has been limited to the LTE approximation because that is the simplest model for lineshape analysis. While radiative transfer may be the most physically accurate model for interstellar sources, collision rates are not available for many COMs of interest (van der Tak et al. 2007). Therefore, the LTE approximation is a reasonable first approximation to use for the analysis of line surveys of chemically-complex sources. An overview of the GOBASIC program design and the results of benchmark tests are presented here.

## 2. Program design

The GOBASIC program code, written in the MATLAB scientific computing environment, has been deposited as supplementary information for this manuscript. A basic overview of the program design is given here; details of the program operation are given in a tutorial document that has been included with the program code. MATLAB is a commercially available program supported by multiple computing environments. All tests described in this paper were run in MATLAB 2014a.

The GOBASIC program accepts an observational dataset as a two column ASCII file, henceforth referred to as *xy*, formatted with frequency in MHz and intensity in antenna temperature,  $T_a$  (K). The spectrum is calibrated to main beam temperature,  $T_{MB}$  (K), using the conversion

$$T_{MB}(K) = T_a(K)/\eta, \quad (1)$$

where  $\eta$  is the efficiency of the telescope, which is defined in the program as a user-specified parameter.

Analysis of observational data is performed by comparing a spectrum to laboratory spectral information that is imported in a modified tab-delimited catalog file format, based on the standard format implemented in the JPL Spectral Line Catalog (Pickett et al. 1998) and the Cologne Database for Molecular Spectroscopy (CDMS; Müller et al. 2005). The catalog files are truncated to match the frequency range of the observational dataset. Users may also specify thresholds for line intensity, frequency uncertainty, and lower state energy when loading catalog files. Transitions that do not meet these thresholds are removed from the simulation to reduce computing time. All transitions are simulated if no thresholds are specified.

GOBASIC uses a rotational partition function of the form

$$Q(T) = \sqrt{\frac{\pi}{ABC} \left(\frac{kT}{h}\right)^3}, \quad (2)$$

where  $Q(T)$  is the partition function at temperature  $T$ ,  $A$ ,  $B$ , and  $C$  are the rotational constants of the molecule,  $k$  is Boltzmann's constant, and  $h$  is Planck's constant. The value of  $Q(T)$  is obtained through interpolation or extrapolation of the partition function information provided by the JPL and CDMS catalogs (Pickett et al. 1998; Müller et al. 2005). Vibrational partition functions are not considered. Users should be mindful of this when loading catalogs that contain transitions from multiple vibrational states.

The Einstein  $A$  coefficients of the transitions are determined using the partition function information provided at 300 K and catalog information

$$A_{ul}g_u = 2.7964 \times 10^{-22} \times \frac{I\nu_c^2 Q(300)}{e^{-E_l/300k} - e^{-E_u/kT}} \text{ (MHz)}, \quad (3)$$

where  $g_u$  is the upper state degeneracy,  $I$  is the line strength ( $I = 10^{(\text{LGINT})}$ , where LGINT is the base 10 logarithm line strength provided in the catalog),  $\nu_c$  is the transition frequency in MHz,  $Q(300)$  is the partition function value at 300 K, and  $E_l$  and  $E_u$  are the lower and upper state energies, respectively, in Joules.

The generating function for the spectrum in GOBASIC was derived from the equations used to obtain a population diagram assuming LTE conditions (Goldsmith & Langer 1999). The optical depth,  $\tau$ , is adapted from Equation 6 in Goldsmith & Langer (1999):

$$\tau(x) = \sqrt{\frac{4 \ln 2}{\pi} \frac{hN_u B_{ul} \nu_c}{c \Delta \nu}} (e^{h\nu_c/kT} - 1) \times \exp\left(-\frac{4 \ln 2 (x - \nu_c + \nu_{\text{off}})^2}{\Delta \nu^2}\right), \quad (4)$$

assuming a Gaussian line profile, where  $x$  is the spectral frequency,  $\nu_c$  is the center frequency of the transition,  $\nu_{\text{off}}$  is the Doppler shift of the transition,  $\Delta \nu$  is the full width half maximum (FWHM) of the spectral line,  $N_u$  is the column density of the upper state,  $B_{ul}$  is the Einstein  $B$  coefficient of the transition, and  $c$  is the speed of light.

Assuming LTE conditions and substituting the relationships

$$B_{ul} = A_{ul} \frac{c^3}{8\pi h \nu_c^3} \quad (5)$$

and

$$N_u = N_T \frac{g_u e^{-E_u/kT}}{Q(T)}, \quad (6)$$

where  $N_T$  is the total column density gives  $\tau$  in the form

$$\tau_c(x) = \frac{1}{4} \sqrt{\frac{\ln 2}{\pi^3}} \frac{c^2}{\nu_c^2 \Delta \nu} \frac{N_T A_{ul} g_u}{Q(T)} e^{-E_u/kT} (e^{h\nu_c/kT} - 1) g(x), \quad (7)$$

where  $g(x)$  is the Gaussian line profile

$$g(x) = \exp\left(-\frac{4 \ln 2 (x - \nu_c + \nu_{\text{off}})^2}{\Delta \nu^2}\right). \quad (8)$$

The optical depth is related to the main beam temperature by the relationship

$$T_{MB} = \frac{h\nu_c}{k} \frac{1 - e^{-\tau}}{e^{h\nu_c/kT} - 1} \eta_B, \quad (9)$$

where  $\eta_B$  is the beam filling factor  $\theta_s^2/(\theta_b^2 + \theta_s^2)$ , assuming a Gaussian power profile for both source  $\theta_s$  and beam  $\theta_b$ . Source size  $\theta_s$  can be obtained from interferometric observational studies. For the beam size,  $\theta_b$ , the half-power beamwidth of a Gaussian illuminated dish with  $-10$  db tapering at the edge is dependent on the radiation frequency, and is calculated in GOBASIC using the relationship

$$\theta_b \approx \frac{\sqrt{-8 \ln 2 \ln 0.1}}{\pi} \frac{\lambda}{D} \times \frac{180}{\pi} \times 3600 \text{ (arcsec)}, \quad (10)$$

where  $\lambda$  is the radiation wavelength corresponding to the transition frequency  $\nu_c$ , and  $D$  is the dish diameter of the telescope<sup>4</sup>. Equation (10) is appropriate for most single-dish telescopes, but the user may also specify their own beam size estimation and

<sup>4</sup> <http://www.cv.nrao.edu/course/ast534/2DApertures.html>

adapt the GOBASIC code directly for either single-dish or interferometric observations. If no source size information is available, the user can set  $\theta_s$  to 0, which in turn forces  $\eta_B$  to be unity.

The fit function is generated by summing the main beam temperature vector of each component  $j$ :

$$f(x; \mathbf{p}) = \sum_j T_{\text{MB}_j}(x) = \frac{h\nu}{k} \sum_j \frac{1 - e^{-\tau_j(x)}}{e^{h\nu/kT_j} - 1} \eta_{B_j}, \quad (11)$$

where  $\mathbf{p}$  is the fitting parameter vector  $\mathbf{p} = [p_1, p_2, \dots, p_{4j}]$ . The vector norm of the residual

$$\|\mathbf{y} - f(x; \mathbf{p}) - \mathbf{y}_{\text{previous}}\|, \quad (12)$$

is minimized, where  $\mathbf{y}$  is the observational data, while  $\mathbf{y}_{\text{previous}}$  is the previously fitted spectra. This is equivalent to a least-squares optimization. Here, a component refers to any unique spectral simulation that differs from all others in at least one characteristic, i.e., molecular identity, temperature, velocity, line width, or spatial distribution. In calculating the total simulation, GOBASIC sums the intensities of components, rather than their optical depths. In this way, each component can be treated separately for ease of fitting, and individual beam-dilution factors can be applied.

The fitting parameters for each component are the base 10 logarithm of column density  $N_T$  in  $\text{cm}^{-2}$ , the spectral line full width with half maximum  $\Delta v$  in  $\text{km s}^{-1}$ , the temperature  $T$  in K multiplied by 0.01, and the velocity shift  $v_{\text{off}}$  in  $\text{km s}^{-1}$ . The rescaling of column density and temperature ensures all fitting parameters are the same order of magnitude, offering reliable fitting results. Line width and velocity shift are converted on the fly to frequency using the Doppler relation

$$\nu - \nu_0 = \frac{v_0 v}{c}, \quad (13)$$

where  $v$  is the shift or line width in  $\text{km s}^{-1}$ , and  $c$  is the speed of light in  $\text{km s}^{-1}$ .

Multiple components can be fitted simultaneously. The number of components fitted at once is dictated entirely by the computational cost, and this number can be increased if the program is implemented in a parallel computing environment. Initial guesses for each of the parameters and upper and lower boundaries are required before optimization.

The algorithm used is the pattern search algorithm in the MATLAB global optimization toolbox<sup>5</sup>. This algorithm is extremely useful for this type of dataset, because it is a derivative free method and therefore able to solve non-smooth multivariable optimization. Unlike most conventional derivative fitting routines, the algorithm GOBASIC uses is less likely to be trapped in a particular local minimum before it traverses the whole parameter space, and as such the result is not sensitive to the initial guess of the parameter matrix. This algorithm is also well-suited for parallel processing, which allows the program to take advantage of multicore CPUs and larger computer clusters.

Analysis of many blended features is one of the main challenges of current approaches to line survey analysis. GOBASIC is well-equipped to handle this challenge because it is designed to fit an arbitrary number of molecular components simultaneously. The only limit to this approach is the available computational power, since the fitted parameter space expands on the order of  $p^{4j}$  as the number of components  $j$  increases. GOBASIC

**Table 1.** Example of the iterative fitting process.

Fit run	Molecules fitted
1	<b>CH<sub>3</sub>OH</b> CH <sub>3</sub> OH, SO <sub>2</sub> , <b>CH<sub>3</sub>CN</b> CH <sub>3</sub> OH, SO <sub>2</sub> , CH <sub>3</sub> CN, <b>HCOOCH<sub>3</sub></b> CH <sub>3</sub> OH, SO <sub>2</sub> , CH <sub>3</sub> CN, HCOOCH <sub>3</sub> , <b>CH<sub>3</sub>OCH<sub>3</sub></b>
2	<b>CH<sub>3</sub>OH</b> , SO <sub>2</sub> , CH <sub>3</sub> CN, HCOOCH <sub>3</sub> , CH <sub>3</sub> OCH <sub>3</sub> CH <sub>3</sub> OH, SO <sub>2</sub> , CH <sub>3</sub> CN, HCOOCH <sub>3</sub> , CH <sub>3</sub> OCH <sub>3</sub> CH <sub>3</sub> OH, SO <sub>2</sub> , <b>CH<sub>3</sub>CN</b> , HCOOCH <sub>3</sub> , CH <sub>3</sub> OCH <sub>3</sub> CH <sub>3</sub> OH, SO <sub>2</sub> , CH <sub>3</sub> CN, <b>HCOOCH<sub>3</sub></b> , CH <sub>3</sub> OCH <sub>3</sub> CH <sub>3</sub> OH, SO <sub>2</sub> , CH <sub>3</sub> CN, HCOOCH <sub>3</sub> , <b>CH<sub>3</sub>OCH<sub>3</sub></b>

**Notes.** Molecules in bold are allowed to vary, and all others are imported as previous fits in the program.

also allows the approach of iterative fitting, where the whole spectral analysis can be split into several fitting jobs, each one fitting several components. Caution should be exercised, because non-Gaussian lineshapes are also fitted with multiple Gaussian components. Users have the option to edit the lineshape used for fitting if they wish to adapt the program for other types of analyses. An example of the iterative fitting routine is shown in Table 1. This approach requires that molecules with the most intense spectral features be analyzed first. Additional analysis should then be performed for molecules with weaker spectra, working down in order of spectral intensity. The optimization for each new molecule can be set to take into account the results from previously analyzed molecules.

The fit results for the parameters, as well as the associated uncertainties, are displayed in Matlab terminal, and the fitted spectra are plotted versus observational data in Matlab GUI. The results and spectra can also be exported into readable tables and comma-delimited files. More information of the estimation of uncertainties of the parameters are discussed in Appendix A.

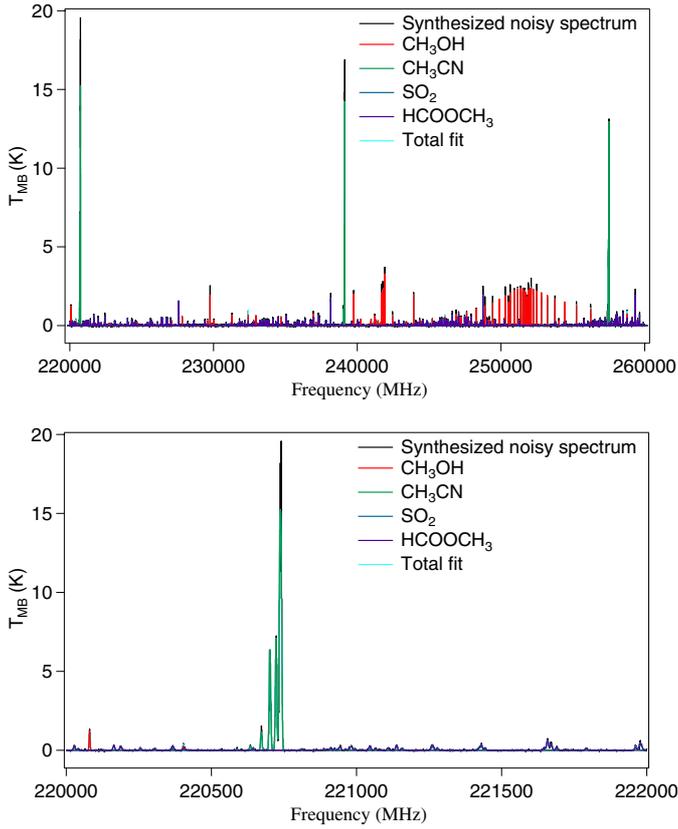
### 3. Results and discussion

#### 3.1. Performance test: synthesized spectra

The performance of the GOBASIC program was assessed by analyzing artificial spectra, which can be easily simulated in GOBASIC. Both a noiseless spectrum and a spectrum with artificial noise were simulated from four molecular components with known physical parameters. The four molecules used were methanol, methyl cyanide, sulfur dioxide, and methyl formate, which are common, highly-abundant molecules found in hot cores. Artificial noise was generated for the noisy spectrum by the Matlab random number generator. Random numbers had a Gaussian distribution, with a standard deviation of  $\sigma = 30$  mK added to the baseline, plus a standard deviation of  $\sigma = 20\%$  applied to spectral line intensity. The input parameters for spectral synthesis were chosen such that both optically thin (e.g., methanol, sulfur dioxide and methyl formate) and optically thick (e.g., methyl cyanide) conditions were included.

The fits of both the noiseless and noisy spectra were open to a large parameter space, where column density could vary across 10 orders of magnitude, temperature could vary from 0 to 1000 K, line width and shift could vary from  $-30$  to  $+30$   $\text{km s}^{-1}$ , and the initial guesses for each parameter were set at the middle point of the parameter space. The fit of the noisy spectrum is shown in Fig. 1. The fitted results and reported uncertainties

<sup>5</sup> Global Optimization Toolbox: Direct Search: patternsearch User's Guide (R2014a). Retrieved September, 2014 from <http://www.mathworks.com/help/gads/patternsearch.html>



**Fig. 1.** Fit of synthesized noisy spectrum for four molecules. *The top panel* shows the entire spectrum and *the bottom panel* shows spectral detail. The synthesized spectrum is in black, methanol in red, methyl cyanide in green, sulfur dioxide in blue, methyl formate in purple, and total fit in cyan.

are shown in Table 2. Both sets of results show excellent agreement with the input values even across such a large parameter space, and are capable of accounting for both optically thin and optically thick spectral line features. Changing initial guess values did not change the results within the uncertainty, indicating the robustness of the GOBASIC fit algorithm to find the global minimum.

### 3.2. Benchmark tests: comparison to rotational diagram method using Orion-KL survey

Performance of GOBASIC was compared with two analysis techniques: the rotation diagram method such as that used by Blake et al. (1987) in their survey of Orion-KL, and the MAGIX (Möller et al. 2013) attachment to XCLASS<sup>6</sup> (Zernickel et al. 2012; Möller et al. 2013).

The comparison of GOBASIC with the rotation diagram method used by Blake et al. (1987) was conducted using a spectral line survey of the Orion-KL star-forming region (Widicus Weaver, Sumner, & Lis, unpublished data) collected at the Caltech Submillimeter Observatory (CSO)<sup>7</sup>. In this test, analysis was conducted for methanol, sulfur dioxide, methyl cyanide, ethyl cyanide, methyl formate, dimethyl ether, thioformaldehyde, HCCCN, HNCO, CH<sub>3</sub>CCH, HDCO, and HDO.

<sup>6</sup> <https://www.astro.uni-koeln.de/projects/schilke/XCLASS>

<sup>7</sup> The CSO is operated by the California Institute of Technology, previously under contract from the National Science Foundation.

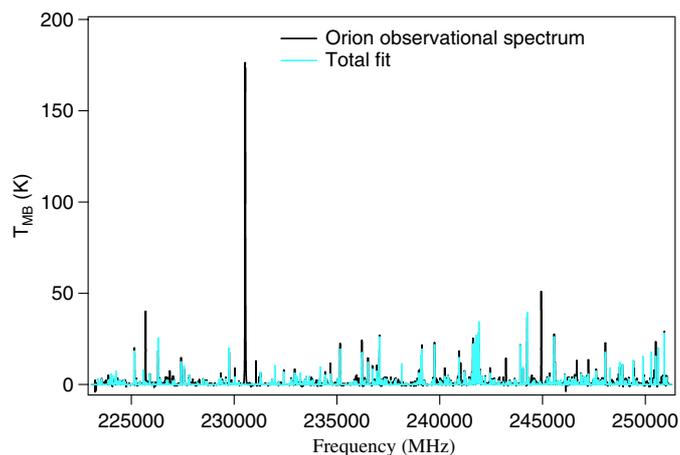
**Table 2.** Comparison of fit parameters and reported uncertainties ( $1\sigma$  in parentheses) of both noiseless and noisy simulated spectra.

	Simulated input	Noiseless spectrum <sup>a</sup>	Noisy spectrum
<b>Methanol</b>			
$N_T$ (cm <sup>-2</sup> )	$1 \times 10^{16}$	$1.0 \times 10^{16}$	$9.91(5) \times 10^{15}$
$\Delta v$ (km s <sup>-1</sup> )	5	5.0	5.04(2)
$T$ (K)	170	170.0	165.1(9)
shift (km s <sup>-1</sup> )	-2	-2.0	-1.98(1)
$\tau_{\max}$	0.0645		
<b>Methyl cyanide</b>			
$N_T$ (cm <sup>-2</sup> )	$1 \times 10^{15}$	$1.0 \times 10^{15}$	$1.035(2) \times 10^{15}$
$\Delta v$ (km s <sup>-1</sup> )	8	8.0	7.93(2)
$T$ (K)	50	50.0	49.0(2)
shift (km s <sup>-1</sup> )	10	10.0	10.11(1)
$\tau_{\max}$	1.9356		
<b>Sulfur dioxide</b>			
$N_T$ (cm <sup>-2</sup> )	$1 \times 10^{14}$	$1.0 \times 10^{14}$	$1.2(4) \times 10^{14}$
$\Delta v$ (km s <sup>-1</sup> )	15	15.0	14(4)
$T$ (K)	120	120.0	159(59)
shift (km s <sup>-1</sup> )	0	0.0	0(1)
$\tau_{\max}$	0.0083		
<b>Methyl formate</b>			
$N_T$ (cm <sup>-2</sup> )	$1 \times 10^{16}$	$1.0 \times 10^{16}$	$1.04(3) \times 10^{16}$
$\Delta v$ (km s <sup>-1</sup> )	10	10.0	10.20(8)
$T$ (K)	400	400.0	416(10)
shift (km s <sup>-1</sup> )	4	4.0	3.93(4)
$\tau_{\max}$	0.0092		

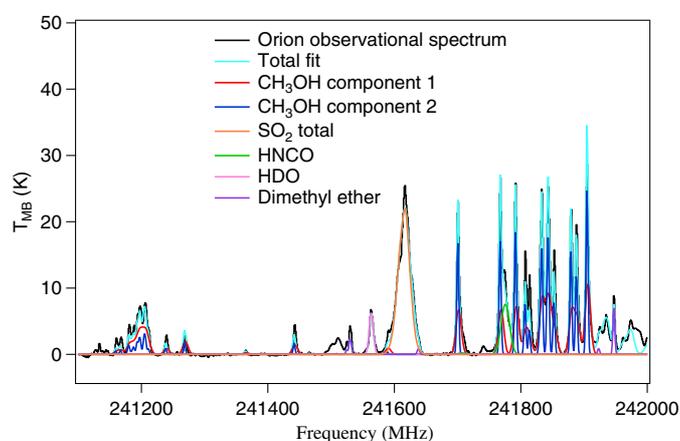
**Notes.** <sup>(a)</sup> The relative uncertainties for the results from the fit of the noiseless spectrum are on the order of  $10^{-8}$ , and are therefore not listed.

Two fit iterations were performed. The first fitting iteration included <sup>13</sup>CS, <sup>34</sup>SO, C<sup>34</sup>S, CS, formaldehyde, and OCS. These molecules have intense spectral features, but multiple transitions do not occur in the spectral range. This fit simply accounts for their contributions to the spectrum and provides no quantitative temperature or column density information. The second fitting iteration included all 12 molecules listed above, taking the first fitting iteration results into account. Lines from CO isotopologues were flagged throughout the fitting process. The fit was performed with no intensity cut-off, and was open to the large parameter space used for the synthesized spectra described above. Since the results were benchmarked against Blake et al. (1987), whose analysis did not consider beam-filling factors, the beam-filling factor was also not included in the GOBASIC analysis presented here. The beam size of Blake et al. (1987) was identical to the beam size used for the CSO observations. Therefore, the results are directly comparable. Results of the fits are shown in Figs. 2 through 6, and in Table 3.

As can be seen in Figs. 3 through 6, GOBASIC modeled the major features of the observational spectrum with few over-predictions, despite the fact that Orion-KL is an extremely complicated source with multiple temperature and velocity components for several molecules and severe line blending in some cases. The temperatures, line widths and velocity shifts of most molecules are comparable with those obtained in the previous study of Blake et al. (1987), even when the parameters were allowed to vary significantly during the optimization. Differing column densities are due to the different temperatures and line widths found between the two studies. The difference in temperatures of some molecules between the two analyses is not surprising because there is a propensity for over-prediction in



**Fig. 2.** Total fit for Orion of all reported molecules listed in Table 3. The observational spectrum is shown in black and the total fit in cyan.

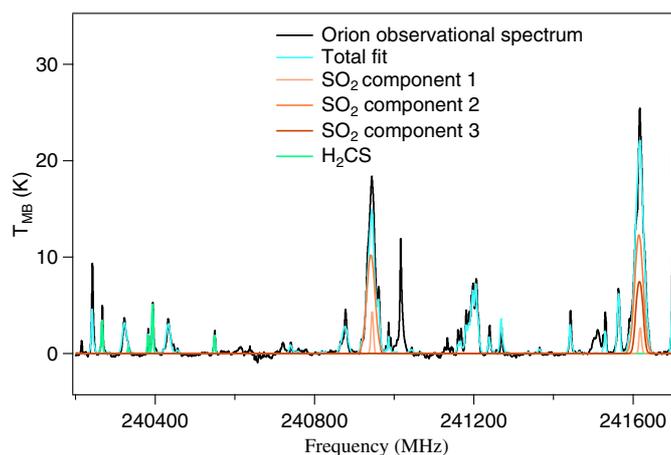


**Fig. 3.** Orion fit details for methanol, sulfur dioxide ( $\text{SO}_2$ ), HDO, HNCO, and dimethyl ether (DME). The observational spectrum is shown in black, the total fit in cyan, and molecular components are illustrated by the colors listed in the figure legend.

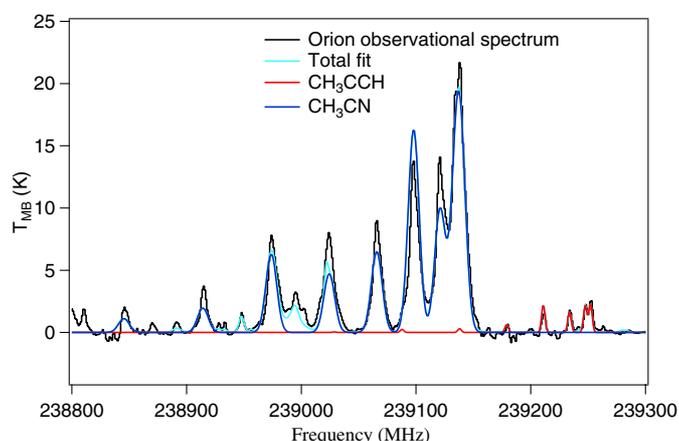
the rotation diagram analysis due to blended lines, where it is difficult to determine the amount of flux arising from a single molecule. In a rotation diagram, any over-prediction of the integrated intensities that results from line blending may change the slope of the line, and hence the predicted temperature. It can also be seen with the co-adding of spectral simulations of many molecules that the low intensity transitions also contribute significant flux that cannot be accounted for with the single-line Gaussian fitting used to produce a rotation diagram. A global analysis such as that used in GOBASIC can take these components into consideration when determining the physical conditions for a particular molecule.

Another advantage of GOBASIC is that it can be used to identify new molecules, provided that the molecular catalog is available. Rotation diagrams can be unreliable for new molecule identification because reasonable rotation diagrams may be produced from a set of unrelated weak lines, especially for complicated asymmetric rotors (Snyder et al. 2005). In GOBASIC, however, the velocity shift is uniform for all lines in a given molecular component, enabling straightforward determination of missing predicted transitions.

Also, the results of this GOBASIC test indicate that the fit was improved by both fitting multiple components simultaneously and by including previous fits of other molecules,



**Fig. 4.** Orion fit details for sulfur dioxide ( $\text{SO}_2$ ) and thioformaldehyde ( $\text{H}_2\text{CS}$ ). The observational spectrum is shown in black, the total fit in cyan, thioformaldehyde in green, and the three components of sulfur dioxide in the colors listed in the figure legend.



**Fig. 5.** Orion fit details for  $\text{CH}_3\text{CCH}$  and methyl cyanide ( $\text{CH}_3\text{CN}$ ). The observational spectrum is shown in black, the total fit in cyan, the  $\text{CH}_3\text{CCH}$  fit in red, and the methyl cyanide fit in blue.

especially in the cases when molecular lines are blended. In this benchmark, we fitted 15 molecular components of 12 molecules simultaneously on a 24-core parallel computing cluster in 60 h of machine time. But it may be impossible to fit all molecules together due to the limitation of computational power. In such cases, molecules can also be fitted individually as long as the transitions of one molecule do not overlap with another, and then included into the total fit later. If one molecule has various temperature or velocity components, or two molecules have significant overlap in transition frequencies (e.g., methyl cyanide and  $\text{CH}_3\text{CCH}$ ), simultaneous fitting is recommended.

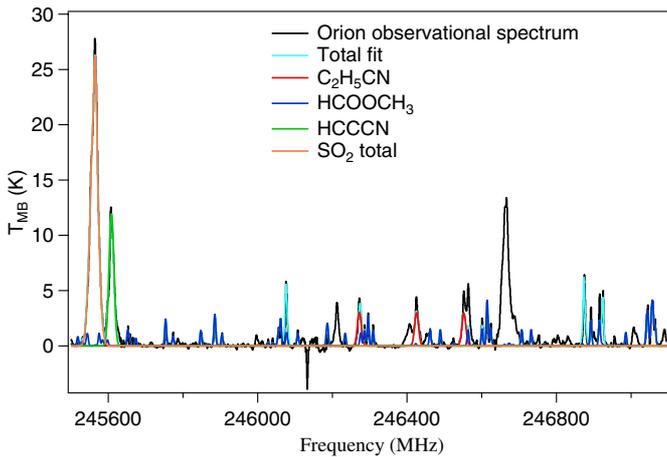
### 3.3. Benchmark tests: comparison to XCLASS/MAGIX using W3( $\text{H}_2\text{O}$ ) survey

The comparison between GOBASIC and the XCLASS/MAGIX code was conducted by analyzing a line survey of W3( $\text{H}_2\text{O}$ ) for methyl formate. The W3( $\text{H}_2\text{O}$ ) observations were collected at the CSO; these observations and the full analysis of the W3( $\text{H}_2\text{O}$ ) line survey will be presented in a forthcoming publication (Wehres et al., in prep.). In the first step of the comparison between GOBASIC and XCLASS/MAGIX, it was verified that the GOBASIC and XCLASS/MAGIX programs produced the

**Table 3.** Comparison of Orion fit parameters and previous analysis results from Blake et al. (1987).

Molecule	GOBASIC				Previous (Blake et al. 1987)			
	$T$ (K)	$N_T$ (cm $^{-2}$ )	$\Delta v$ (km s $^{-1}$ )	$v$ (km s $^{-1}$ )	$T$ (K)	$N_T$ (cm $^{-2}$ )	$\Delta v$ (km s $^{-1}$ )	$v$ (km s $^{-1}$ ) <sup>a</sup>
Methanol	134(4)	$5.5(2) \times 10^{16}$	4.41(9)	-0.26(4)	146(3) <sup>c</sup>	$3.5 \times 10^{16}$	3.3/9.6	-9/ -1.8
	200(6)	$9.3(3) \times 10^{16}$	12.1(2)	-2.1(2)				
Methyl cyanide	244(12)	$4.5(2) \times 10^{15}$	13.9(2)	-1.80(1)	274(12) <sup>d</sup>	$6.5 \times 10^{14}$	10.3	-4.0
Ethyl cyanide	148(6)	$3.45(6) \times 10^{15}$	15.1(3)	-4.4(2)	101(9) <sup>e</sup>	$9.6 \times 10^{13}$	4.0	-0.6
Methyl formate	340(22)	$3.8(2) \times 10^{16}$	6.6(1)	-0.75(8)	150(9) <sup>d</sup>	$8.2 \times 10^{14}$	11.4	-4.0
Sulfur dioxide	113(4)	$9.3(5) \times 10^{16}$	29.8(9)	1.55(3)	90(10) <sup>c</sup>	$2.6 \times 10^{15}$	3.9	-1.1
	15.1(8)	$1.3(1) \times 10^{16}$	20.9(9)	0.0(4)	106(3) <sup>f</sup>	$2.8 \times 10^{16}$	25.2	-1.3
	256(36)	$1.7(5) \times 10^{16}$	9(3)	-2.5(5)	(60) <sup>b,e</sup>	$<1.0 \times 10^{15}$	4.2	-1.3
Dimethyl ether	123(8)	$1.42(9) \times 10^{16}$	4.2(2)	-0.9(1)	(200) <sup>b,d</sup>	$<2.0 \times 10^{15}$	6.9	-6.4
	99(13)	$1.3(1) \times 10^{15}$	5.2(3)	-0.0(2)	63(5) <sup>c</sup>	$3.0 \times 10^{15}$	3.6	-1.1
Thioformaldehyde	104(23)	$6.7(1) \times 10^{14}$	18.3(4)	-2.7(2)	116(17) <sup>c</sup>	$4.8 \times 10^{14}$	4.1	-1.4
HNCO	40(4)	$3.2(2) \times 10^{15}$	19.3(5)	-1.9(5)	312(35) <sup>d</sup>	$1.3 \times 10^{14}$	10.0	-5.2
					182(45) <sup>d</sup>	$4.8 \times 10^{14}$	9.1	-2.4
CH <sub>3</sub> CCH	95(34)	$3.5(9) \times 10^{15}$	4(1)	0.4(6)	35(3) <sup>e</sup>	$\lesssim 6.5 \times 10^{14}$	~4.3	-0.6
HDCO	118(108)	$7(8) \times 10^{15}$	7(1)	0.8(6)	52(8) <sup>e</sup>	$1.0 \times 10^{15}$	3.8	0.3
HDO	114(17)	$9.9(9) \times 10^{15}$	9.6(7)	-2.4(4)	32(7) <sup>e</sup>	$5.6 \times 10^{13}$	~4.0	-0.2
					164(23) <sup>d</sup>	$4.4 \times 10^{15}$	9.1	-3.3

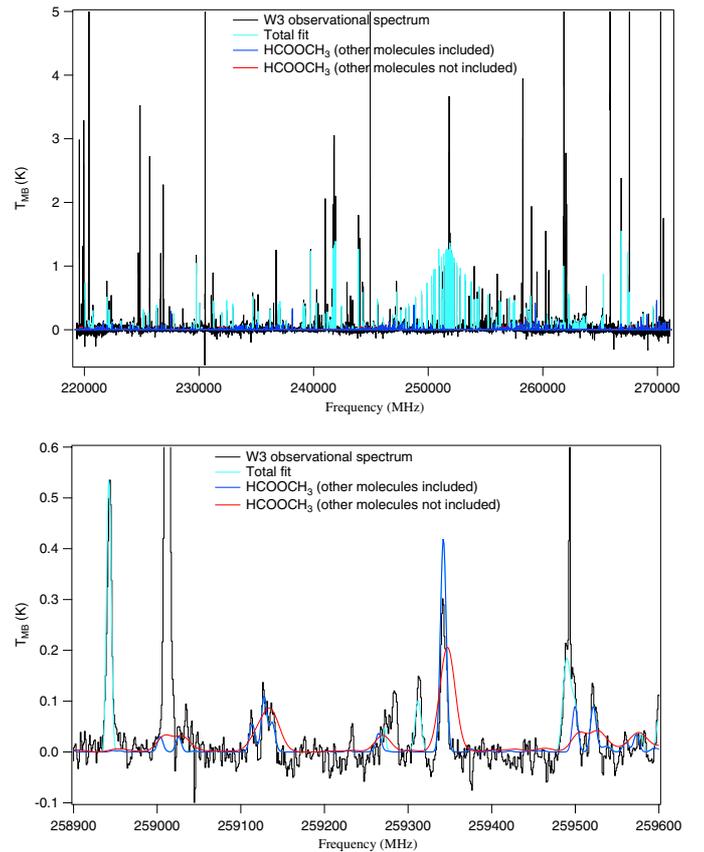
**Notes.** (a) The  $v_{\text{LSR}}$  values used for the two sets of observations are different. These values have been adjusted to be on the same  $v_{\text{LSR}}$  scale. (b) Temperature values reported in parentheses for SO<sub>2</sub> were assumed and not determined through a rotation diagram analysis. (c) Compact Ridge. (d) Hot Core. (e) Extended Emission. (f) Plateau.



**Fig. 6.** Orion fit details for ethyl cyanide (C<sub>2</sub>H<sub>5</sub>CN), methyl formate (HCOOCH<sub>3</sub>), HCCCN, and sulfur dioxide (SO<sub>2</sub>). The observational spectrum is shown in black, the total fit in cyan, the ethyl cyanide fit in red, the methyl formate fit in blue, the HCCCN fit in green, and the sulfur dioxide total fit in orange.

same simulated spectra when using the same input parameters. Subsequent tests involved analysis of identical datasets using both GOBASIC and XCLASS/MAGIX. All XCLASS/MAGIX fits were performed using the Levenberg-Marquardt algorithm, and no correction was made for source size.

In the analysis of W3(H<sub>2</sub>O), 15 small molecules such as CO and CS isotopologues were fitted in advance and included in GOBASIC as previous fits, whereas methyl formate was fitted simultaneously with five other larger molecules, including methanol, methyl cyanide, CH<sub>3</sub>CCH, dimethyl ether and sulfur dioxide. Fits both with and without consideration of beam filling factor were performed. The source size used was 3.2'' × 2.0'' (Hernández-Hernández et al. 2014). A second fit including only methyl formate, expected to give a non-physical result, was also performed for comparison. The results are presented in Table 4, and the fits are shown in Fig. 7.



**Fig. 7.** Methyl formate fit of W3(H<sub>2</sub>O). The observational spectrum shown in gray, total fit in cyan, methyl formate single-component fit without beam-filling correction in red, and methyl formate fit including other molecules and beam-filling correction in blue. The *top panel* shows the fit over the whole spectrum and *bottom panel* shows spectral detail.

The noise level of the dataset is around 0.05 K, and over half of the lines of methyl formate in the frequency range are

**Table 4.** Comparison of XCLASS/MAGIX and GOBASIC fit parameters for methyl formate on W3(H<sub>2</sub>O).

	XCLASS	GOBASIC		
		$\eta_B = 1^a$	$\eta_B \approx 0.009^{a,b}$	single component <sup>c</sup>
$T$ (K)	223.05	221(16)	202(17)	1000(140)
$N_T$ (cm <sup>-2</sup> )	$9.8(7) \times 10^{14}$	$1.06(6) \times 10^{15}$	$1.25(8) \times 10^{17}$	$8(2) \times 10^{15}$
$\Delta v$ (km s <sup>-1</sup> )	5.86	9.2(4)	8.9(5)	25(3)
$v$ (km s <sup>-1</sup> )	-0.008	0.4(2)	0.3(2)	-5.4(8)
$\tau_{\max}$ (methyl formate)		0.0019	0.27	

**Notes.** <sup>(a)</sup> Full analysis with 21 other molecules included in the fit. <sup>(b)</sup> The source size was 2.53'' based on SMA interferometry data (Hernández-Hernández et al. 2014). The beam size was calculated using Eq. (10) using a CSO dish diameter of 10.4 m. <sup>(c)</sup> Fit for methyl formate only without beam-filling correction. The temperature and width stopped at the upper bound of the fitting range.

blended or in the confusion limit. Because of the blended features, fitting methyl formate alone would produce non-physical results. It can be seen from Fig. 7 that without including other molecules, a fit of methyl formate alone tends to over-predict line widths. When including the fitted results of other molecules, GOBASIC gave a result with a reasonable set of parameters for methyl formate. Furthermore, the correction for beam-filling should give a more accurate estimate of column density, while the line width, temperature and velocity shift are not affected significantly. Although the source size information retrieved from Hernández-Hernández et al. (2014) is only the 1.3 mm continuum source size, which is expected to be different from the methyl formate source size, we can still see from Table 4 that the column density is rescaled based on the beam filling factor, while other parameters agree with the non-corrected results within the uncertainty limit. The maximum opacity of methyl formate determined in the full spectral analysis also increased because of the beam correction factor, but is still within optically thin approximation.

The results also agree with those from XCLASS/MAGIX. The linewidth differences arise because the fitting process is so different between the two programs. GOBASIC's automatic fitting process requires much less manual input than XCLASS/MAGIX, which required an initial fit of strong methyl formate lines to constrain the initial guesses, and a tight parameter space for fitting. GOBASIC, on the other hand, was able to fit, unguided, from a large range of initial guesses, and demonstrated the ability to deal with heavily blended spectral line features when a sufficient number of molecules were included in the fit.

#### 4. Conclusions

We have presented here a spectral analysis program for simultaneous global analysis of broadband line surveys, Global Optimization and Broadband Analysis Software for Interstellar Chemistry (GOBASIC). The GOBASIC program benefits from the use of a derivative-free optimization method which is appropriate for these types of non-linear, non-smooth optimization problems. The simplicity of the generating function and optimization algorithm make this program appropriate for use on most desktop and laptop computers. This program also uses parallel processing to decrease computing time, and can be adapted to run in a computing cluster environment.

The GOBASIC program has been benchmarked against both synthesized and observational datasets and the results compare favorably to those obtained using other methods. GOBASIC can be used for all molecules for which rotational spectral catalogs are available. The use of the LTE approximation allows for rapid analysis of a spectrum, enabling simultaneous spectral fitting over large spectral bandwidths for multiple molecules with multiple physical components. The program offers several advantages over other fitting routines, including robust determination of the global minimum, simple user input, simple correction for beam-filling, fast analysis time with parallel computing, and a simultaneous broadband analysis approach.

*Acknowledgements.* The authors thank both the referee and Malcolm Walmsley for very helpful comments that improved this manuscript. Funding for this work was provided by NASA *Herschel* OT1 Analysis Program RSA No. 1428755, the Emory Chemistry Early Career Research Grant, Emory SURE and SIRE programs (funding from Howard Hughes Medical Institute Grant No. 52005873), and the Emory College Undergraduate Research Matching Grant. The authors thank James G. Nagy and Yuanjun Dai for helpful discussion regarding the use of MATLAB. The authors also gratefully acknowledge Matthew Sumner for his help in collecting the Orion-KL dataset, Darek Lis for his help on spectral deconvolution and his useful input during development of the analysis program, Emerson Center for Scientific Computation at Emory University for running most of the tests, and Jacob Laas for his perl script for converting catalog files into tab-delimited text files.

#### References

- Blake, G., Sutton, E., Masson, C., & Philips, T. 1987, *ApJ*, 315, 621  
 Burrell, K. H. 1990, *Am J. Phys.*, 58, 160  
 Garrod, R., Widiuci Weaver, S., & Herbst, E. 2008, *ApJ*, 682, 283  
 Goldsmith, P., & Langer, W. 1999, *ApJ*, 517, 209  
 Herbst, E., & van Dishoeck, E. 2009, *ARA&A*, 47, 427  
 Hernández-Hernández, V., Zapata, L., Kurtz, S., & Garay, G. 2014, *ApJ*, 786, 38  
 Laas, J., Garrod, R., Herbst, E., & Widiuci Weaver, S. 2011, *ApJ*, 728, 71  
 Maret, S., Hily-Blant, P., Pety, J., Bardeau, S., & Reynier, E. 2011, *A&A*, 526, A47  
 Möller, T., Bernst, I., Panoglou, D., et al. 2013, *A&A*, 549, A11  
 Müller, H., Schlöder, F., Stutzki, J., & Winnewisser, G. 2005, *J. Mol. Struct.*, 742, 215  
 Pickett, H., Poynter, R., Cohen, E., et al. 1998, *J. Quant. Spectr. Rad. Transf.*, 60, 883  
 Snyder, L., Lovas, F., Hollis, J., et al. 2005, *ApJ*, 619, 914  
 van der Tak, F., Black, J., Schlöder, F., Jansen, D., & van Dishoeck, E. 2007, *A&A*, 468, 627  
 Zernickel, A., Schilke, P., Schmiedeke, A., et al. 2012, *A&A*, 546, A87

## Appendix A: Uncertainty estimation of fitted parameters

Denoting the optimizing parameters  $\log_{10}(N_T)$ ,  $\Delta v$ ,  $0.01T$ ,  $v_{\text{off}}$  of all components fit as a parameter vector  $\mathbf{p} = [p_1, p_2, \dots, p_j]$ , the fitting function in Eq. (11) is

$$y_i = f(x_i; \mathbf{p}) = f(x_i; p_1, p_2, \dots, p_j) \quad (\text{A.1})$$

for all  $(x_i, y_i)$  pairs in the  $xy$  dataset.

We use the equations described in Burrell (1990) to estimate the error of the parameters. The curvature matrix  $\mathbf{C}$  is defined by

$$C_{jl} = \sum_i \frac{1}{\sigma_i^2} \left( \left( \frac{\partial f(x_i; \mathbf{p})}{\partial p_j} \right) \cdot \left( \frac{\partial f(x_i; \mathbf{p})}{\partial p_l} \right) - (y_i - f(x_i; \mathbf{p})) \times \frac{\partial^2 f(x_i; \mathbf{p})}{\partial p_j \partial p_l} \right), \quad (\text{A.2})$$

summing all observational data points  $(x_i, y_i)$ . It is also assumed that all  $y_i$  have the same distribution of error, so  $\sigma_i = \sigma = \|\mathbf{y} - f(x; \mathbf{p})\| [N - \dim(\mathbf{p})]^{-0.5}$  is the rms of the residual of the spectra; this value can therefore be taken out of the summation.

Let matrix  $\boldsymbol{\epsilon}$  be the inverse of  $\mathbf{C}$ , then

$$\sigma_{p_j}^2 = \frac{1}{\sigma_{T_a}^2} \sum_i \left( \sum_l \epsilon_{jl} \frac{\partial}{\partial p_l} f(x_i; \mathbf{p}) \right)^2. \quad (\text{A.3})$$

The partial derivative of the fit function is extremely complicated. However, GOBASIC uses numerical derivatives to approximate the results. We define

$$\frac{\partial}{\partial p_j} f(x; \mathbf{p}) \approx \frac{1}{\delta} \left( f(x; p_j + \delta/2, p_l) - f(x; p_j - \delta/2, p_l) \right), \quad (l \neq j) \quad (\text{A.4})$$

$$\frac{\partial^2}{\partial p_j^2} f(x; \mathbf{p}) \approx \frac{4}{\delta^2} \left( \left( f(x; p_j + \delta/2, p_l) - f(x; p_j, p_l) \right) - \left( f(x; p_j, p_l) - f(x; p_j - \delta/2, p_l) \right) \right), \quad (l \neq j) \quad (\text{A.5})$$

and

$$\frac{\partial^2 f(x; \mathbf{p})}{\partial p_j \partial p_l} \approx \frac{1}{\delta^2} \left[ \left( f(x; p_j + \delta/2, p_l + \delta/2, p_m) - f(x; p_j + \delta/2, p_l - \delta/2, p_m) \right) - \left( f(x; p_j - \delta/2, p_l + \delta/2, p_m) - f(x; p_j - \delta/2, p_l - \delta/2, p_m) \right) \right], \quad (m \neq j, l) \quad (\text{A.6})$$

where  $\delta$  is a small increment. GOBASIC calculates the variance-covariance matrix and uncertainty of parameters from these values. In the program  $\delta$  is by default chosen as  $10^{-3}$ . Tests have shown that a value of  $\delta$  between  $10^{-6}$  to  $10^{-2}$  usually gives reasonable and consistent results. Values of  $\delta$  that are too small may cause singularity during division when calculating the numerical derivatives, due to the finite precision of float numbers in the computer, while values that are too large violate the numerical derivative approximation.

The standard uncertainty for the antenna temperature  $\sigma_{T_a}$  can be either determined based on actual observational data, or from the goodness of the fit. From observational data, the user determines the noise of the dataset and manually enters  $\sigma_{T_a}$  in GOBASIC. From the goodness of the fit,

$$\sigma_{T_a} [\text{from fit}] = \sigma = \sqrt{\frac{\sum_{i=1}^N (y_i - f(x_i; \mathbf{p}))^2}{N - \dim(\mathbf{p})}}, \quad (\text{A.7})$$

where  $\dim(\mathbf{p})$  is the dimension of  $\mathbf{p}$ , i.e., the number of parameters fitted.

The conversion of uncertainties of  $\log_{10} N_T$  and  $0.01T$  to  $N_T$  and  $T$  uses standard uncertainty propagation equations.