Big Maser Data: Methanol Maser Database

O.S. Bayandina[1,2] and D.A. Ladeyschikov[3]

[1]*Joint Institute for VLBI ERIC, Oude Hoogeveensedijk 4, 7991 PD Dwingeloo, The Netherlands*; `bayandina@jive.eu`

[2]*Astro Space Center, Lebedev Physical Institute, Russian Academy of Sciences, Leninskiy Prospekt 53, Moscow 119333, Russia*

[2]*Ural Federal University, 51 Lenin Str., 620051 Ekaterinburg, Russia*

Abstract. Methanol Maser Database [http://maserdb.net] is the first specialized web-interface database collecting the most up to date data on the sources of methanol molecule maser emission. Methanol masers provide us with unique information on the properties of the interstellar medium and found both in galactic and extragalactic objects. The methanol maser database is equipped with a search tool allowing an unified access to the cataloged maser data, as well as to the external data, including images and catalogs of well-known sky surveys. An ample functionality of statistical analysis is included to the web interface. An overview of the technical implementation of the database is presented.

1. Introduction

Currently in the astronomy there is a trend to use large astronomical catalogs and databases in order to study different classes of the astrophysical objects. The recent development includes databases of exoplanets (EXO.MAST, Exoplanets.org), OH maser database (Engels & Bunzel 2015), The Mikulski Archive for Space Telescopes (MAST) focused on scientific data sets in the optical, ultraviolet, and near-infrared parts of the spectrum, and others. All these databases were created with the focus on some specific phenomena, wavelength of detection or class of an astronomical object. However, the database of methanol masers is still missing to date, thus it is not possible to search across the observations of methanol masers available in literature. The creation of the unified database solve the problem of methanol masers data access thru rich opportunities of the SQL-queries.

2. Methods

For creation of the maser database we use the PostgreSQL database management system suitable for large datasets and complex SQL-queries. The database was modified with PgSphere plugin required for work with sky coordinates. Data entry to the database was done using online Vizier archive from the Strasbourg astronomical Data Center and Optical Character Recognition (OCR) system FineReader in the cases when

online data is not available. DBSCAN algorithm of Python SCIKIT-LEARN package is used for source grouping. Matching with the external datasets is done with `x-match` service of the Strasbourg astronomical Data Center. The online web-interface is written in Perl/CGI language with the support of the Javascript (including AJAX technology). Online plots in the web-interface are plotted using Plotly Python Graphing Library.

3.   Database structure

The moto "the simplest solution is almost always the best" guided us in the sense of data storage, thus it is essential to not overloadthe database with deep decompositions to simplify the SQL-queries. The actual observational data (including non-detections) is stored in the tables `met1_data` and `met2_data` for class I and class II methanol masers. The interferometric maser spot data for both class I and class II masers is stored in the table `maser_spots` to not overload the main observations table. The interferometric spot table is connected to the observation tables with a simple integer identifier `obs_id`. Result of observations grouping (done with DBSCAN algorithm) is stored in the view `all_objects` and is recalculated with each database update. The grouping table is connected to the main observation table with `grp` column, identifying a source group in format of galactic coordinates `"G55.371+0.185 (40)"` (value in brackets is a number of observations associated with this source group). External data (including fluxes) is stored locally in the following tables: `iras`, `akari`, `akari_fis`, `bgps`, `continuum`, `ego` and few others, not actually associated with methanol maser studies (`2mass`, `wise`, `gcvs`). The table `transitions` stores the known methanol maser transitions, table `papers` stores the paper information (abstract, reference, data type, etc.), and table `stat` stores the papers statistics (number of detections, etc.).

4.   Data entry

Initial data entry was done in the semi-automatic way using an information system, developed specially for the maser data. Data entry was divided into the following steps:

Step 1. Paper information entry. The entry includes paper title, authors, journal, volume, pages and other information related to the publication. Thedata was filled automatically using ADS Web API interface by entering the ADS bibcode. The observed molecules and transitions, way of targeting, and type of observations (targeted single-dish, blind single-dish or interferometry) presented in the paper also have to be specified.

Step 2. Uploading of the tables of observations in electronic format. The data may be presented in the different ways. Some papers present detection table together with non-detection table, some paper present complete target list together with detected sources characteristics. For interferometric observations, table of maser spots usually is separated from the observed targets table. All these ways of presenting the data are acceptable for entry into the database. The data is splitted into the three main table types: detection, non-detection, and coordinate list. These three types of tables cover almost all the ways of presenting data in the papers. Depending on the type of uploaded table, the information system will use the proper way of data processing in order to fit it to the database structure.

Step 3. Defining of table columns. This step is required for adding the semantic meaning to the columns of tables, uploaded in the Step 2. Developer needed to specify the columns for coordinates (in any coordinate system and units), source name, line peak velocity and intensity (in Jy, K or Jy/K units), line FWHM (optionally) and integrated line intensity (optionally). The observed methanol transition also should be specified here. Telescope beam size and distance to the source are optional fields as well. If there is a non-detection table, then this step is repeated for non-detection file, but columns are slightly different (rms instead of line peak intensity and velocity). If a table with a list of coordinates is provided, then the column for joining the detection and coordinate tables has to be specified. For example, this may be observation id or source name as presented in the original paper.

Step 4. An extra data entry. an extra data, including ASCII data for spectral lines (we currently do not use this function), images and textual descriptions associated with each observation could be added.

Step 5. Specification of the observation for each entry of extra data. The information system will find the closest observation for each entry of extra data using fuzzy match. However, this have to be double checked manually for the match between source name in observation table and source name of each extra data entry.

Step 6. Verification of data entry. In this step developer gets an overview of the data prepared for entry into the database. If everything is correct the data is added to the database.

5. Data completeness

The data entry for Class I methanol masers is currently complete and contains both single-dish and interferometric data. The number of included observations is 2725 detections and 7289 non-detections. Class I maser emission is found in 680 maser sites, not including the Central Molecular Zone. The interferometric maser spot data is available for 389 maser sites with 4740 maser spots archived in the database. The database includes the published papers with class I masers observations from the beginning of 1990. For papers older than 1990 there is sometimes not enough information about a maser observation (missing peak velocities, intensities, etc.). In that case, we do not include information from these papers to our database if the observed sources are well-known and more recent observations of class I methanol line are available for these sources. However, we include the papers where the first detection of some particular methanol maser transition was done. The total amount of papers included to the database for class I methanol masers is 105.

To date the class II maser database is in the process of data collection. The database already contains 4321 detection and 7542 non-detection and provides extensive information on the overall distribution of class II methanol masers. Class II maser emission is found in 1221 maser sites, that is close to the predicted number of class II methanol masers in the Galaxy (1200±84; van der Walt 2005). The interferometric maser spot data is available for 441 maser sites with 11858 maser spots archived in the database. The database contains most recent large surveys of class II masers, including the full 6.7 GHz methanol multibeam maser catalogue (e.g. Green et al. 2012), the 6.7 GHz methanol maser survey with the Shanghai Tianma Radio Telescope (e.g. Yang et al. 2019) and others. Total amount of currently covered papers for class II methanol papers is 38.

6.   Source grouping

The observation table `met1_data` contains observation entries for methanol masers. However, it is hard to examine the exact number of sources from this table, because there are different transitions, re-observations of sources with higher sensitivity and different velocity components. In order to get the idea of total number of detected methanol sources we have to combine observations into the source groups using the defined grouping criteria.

We have checked the different source grouping algorithms in order to fit our needs. The density-based DBSCAN algorithm was chosen because it has functionality to select a distance parameter ϵ for clustering and does not require a priori information about number of clusters. We chose $\epsilon = 60$ arcsec as the best value to find all observations within the median mean beam size (52 arcsec) of all single-dish observations in the database. To illustrate the results of source grouping we present in Figure 1 the Galactic Center region ($-0.6 < l < 0.6$, $-0.15 < b < 0.10$), the richest in methanol masers region of our Galaxy. Left panel display the spatial distribution of class I methanol maser spots, and right panel display the results of maser spots grouping. From the figure it could be seen that most dense part is combined into one large group, but several isolated groups are located in the vicinity.
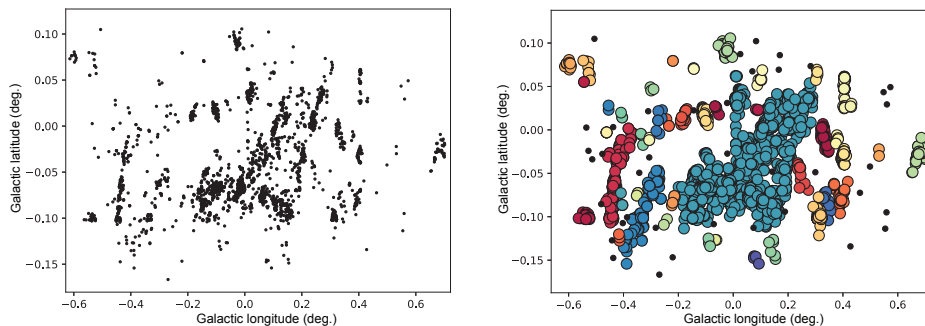

Figure 1.   *Left:* Distribution of detected class I methanol masers in the Galactic Center region. *Right:* Results of source grouping using DBSCAN algorithm with $\epsilon = 60$ arcsec. Black dots are single maser spots that have only one detections in the vicinity.

References

Engels, D., & Bunzel, F. 2015, A&A, 582, A68. `1508.06200`
Green, J. A., Caswell, J. L., Fuller, G. A., Avison, A., Breen, S. L., Ellingsen, S. P., Gray, M. D., Pestalozzi, M., Quinn, L., Thompson, M. A., & Voronkov, M. A. 2012, MNRAS, 420, 3108. `1201.0787`
van der Walt, J. 2005, MNRAS, 360, 153
Yang, K., Chen, X., Shen, Z.-Q., Li, X.-Q., Wang, J.-Z., Jiang, D.-R., Li, J., Dong, J., Wu, Y.-J., & Qiao, H.-H. 2019, ApJS, 241, 18. `1906.05801`