Data mining, archives, virtual observatory, astrostatistics
The SAO/NASA Astrophysics Data System (ADS) is a Digital Library portal for researchers in Astronomy and Physics, operated by the Smithsonian Astrophysical Observatory (SAO) under a NASA grant.

The ADS maintains three bibliographic databases containing more than 10.9 million records covering publications in Astronomy and Astrophysics, Physics, and the arXiv e-prints. Abstracts and full-text of major astronomy and physics publications are indexed and searchable through the new ADS Labs interface as well as the traditional "Classic" search forms. A set of browsable interfaces are also available.
Title: Abundances of light elements in metal-poor stars. I. Atmospheric parameters and a new T_eff scale

Authors: Gratton, R. G.; Carretta, E.; Castelli, F.


Publication Date: 10/1996

Origin: A&A via CDS

A&A Keywords: STARS: FUNDAMENTAL PARAMETERS, STARS: ATMOSPHERES, STARS: ABUNDANCES, STARS: POPULATION II

Abstract Copyright: (c) 1996: Astronomy & Astrophysics

Bibliographic Code: 1996A&A...314..191G
Abundances of light elements in metal-poor stars

I. Atmospheric parameters and a new $T_{\text{eff}}$ scale

R.G. Gratton$^1$, E. Carretta$^{2,3}$, and F. Castelli$^4$

$^1$Osservatorio Astronomico di Padova, Vicolo dell’Osservatorio, 5, I-35122 Padova, Italy
$^2$Dipartimento di Astronomia, Università di Padova, Vicolo dell’Osservatorio 5, I-35122 Padova, Italy
$^3$Osservatorio Astronomico di Bologna, Via Zamboni 33, I-40126 Bologna, Italy
$^4$Osservatorio Astronomico di Trieste, Via G.B. Tiepolo, 11, I-34151 Trieste, Italy

Received 13 July 1995 / Accepted 12 March 1996

Abstract. We present atmospheric parameters for about 300 stars of different chemical composition, whose spectra will be used to study the galactic enrichment of Fe and light elements. These parameters were derived using an homogeneous iterative procedure, which considers new calibrations of colour-$T_{\text{eff}}$ relations for F, G and K-type stars based on Infrared Flux Method (IRFM) and interferometric diameters for population 1 stars, and the Kurucz (1992) model atmospheres. We found that these calibrations yield consistent sets of atmospheric parameters for different classes of metal-poor stars.

1. Introduction

The determination of elemental abundances in metal-poor stars is a basic constraint for models of the chemical evolution of our Galaxy, and provide wealth of data about the history of star formation (see e.g. Wheeler et al. 1989). A very important rôle is played by C, N, O, Na, Mg, and Fe, which are amongst the most abundant elements, likely produced in a variety of astronomical sites. The determination of accurate abundances for these elements...
Astronomy and Astrophysics
Self-Organized Map Index

- The Centre de Données Astronomiques de Strasbourg (CDS) (cdsweb.u-strasbg.fr) is an International Service under the Federation of Astronomical and Geophysical Data Analysis Services (FAGS) that “collects, homogenizes, distributes and preserves astronomical information.”

- The main CDS services include:
  - **SIMBAD** (simbad.u-strasbg.fr/sim-fid.pl), a reference database for the identification and bibliography of astronomical objects outside the solar system
  - **VizieR** (vizier.u-strasbg.fr/viz-bin/VizieR), a search portal to more than astronomical catalogs and published tables.
Astronomy and Astrophysics

Self-Organized Map Index

- CDS is also actively involved in the development of bibliographic information retrieval tools, notably visualized indexes to digital collections. A premier example of this tool is the ‘document map’ developed for *Astronomy and Astrophysics* ([simbad.ustrasbg.fr/A+A/map.pl](http://simbad.ustrasbg.fr/A+A/map.pl)).

- The ‘document map’ is created using the Kohonen Self Organizing Map (SOM), an “algorithm that automatically organizes documents into a two-dimensional grid so that related documents appear close to each other and general topics appear in well defined areas.”
The SIMBAD astronomical database provides basic data, cross-identifications, bibliography and measurements for astronomical objects outside the solar system.

SIMBAD can be queried by object name, coordinates and various criteria. Lists of objects and scripts can be submitted.

Links to some other on-line services are also provided.

**Statistics**

Simbad contains on 2014.07.21

- **7,543,666** objects
- **18,532,209** identifiers
- **293,904** bibliographic references
- **10,712,839** citations of objects in papers
**Example**

Other query modes:

<table>
<thead>
<tr>
<th>Identifier query</th>
<th>Coordinate query</th>
<th>Criteria query</th>
<th>Reference query</th>
<th>Basic query</th>
<th>Script submission</th>
<th>Output options</th>
<th>Help</th>
</tr>
</thead>
</table>

**basic query**: NGC 1851

*Identifier, coordinates (radius=10 arcmin), or bibcode*
Other object types: *IC (), GIC (GCI), CI* (C)

<table>
<thead>
<tr>
<th>Coordinate System</th>
<th>Coordinates</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICRS coord. (ep=J2000)</td>
<td>05 14 06.76 -40 02 47.6 (Optical) [100 100 90] C</td>
<td><a href="#">2010AJ....140.1830G</a></td>
</tr>
<tr>
<td>FK5 coord. (ep=J2000 eq=2000)</td>
<td>05 14 06.76 -40 02 47.6 (Optical) [100 100 90] C</td>
<td><a href="#">2010AJ....140.1830G</a></td>
</tr>
<tr>
<td>FK4 coord. (ep=B1950 eq=1950)</td>
<td>05 12 28.08 -40 06 10.6 (Optical) [100 100 90] C</td>
<td><a href="#">2010AJ....140.1830G</a></td>
</tr>
<tr>
<td>Gal coord. (ep=J2000)</td>
<td>244.5132 -35.0360 (Optical) [100 100 90] C</td>
<td><a href="#">2010AJ....140.1830G</a></td>
</tr>
</tbody>
</table>

Radial velocity / Redshift / cz:

V(km/s) 320.9 [1.0] / z(\~) 0.001071 [0.000003] / cz 321.07 [1.00] (\~) D 1996AJ....112.1487H

Fluxes (2):

<table>
<thead>
<tr>
<th>Band</th>
<th>Flux</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>8.80 [\sim] D ~</td>
</tr>
<tr>
<td>V</td>
<td>7.23 [\sim] D 2012AJ....144..126D</td>
</tr>
</tbody>
</table>
The VizieR Catalogue Service

http://vizier.u-strasbg.fr/

The basic VizieR search at
Strasbourg Astronomical Observatory

**Description**

- **VizieR** provides access to the most complete library of published astronomical catalogues and data tables available on line (12570 catalogues), organized in a self-documented database.

- Query tools allow the user to select relevant data tables and to extract and format records matching given criteria. Specific care has been taken for optimizing access to some very large catalogues such as UCAC2, the USNO-B1, or the 2MASS last release.
The need for a Virtual Observatory

Immense amounts of data are being produced by large telescopes using large area detectors.

Terabytes of data are now available, and Petabytes ($10^{15}$ bytes) will soon be available from frequent all sky imaging.

Vast databases are also being produced through simulations.
Astronomical Data Explosion

P. Quinn

\[ T_2 < 18 \text{ mths} \]

1990-2000

\[ \sim 100 \text{ Gb/night} \]
Data Explosion

Peter Quinn
Wavelength Coverage

The data spans the electromagnetic spectrum from the radio to the gamma-ray region.

Obtaining, analysing and interpreting the data in different wavebands involves highly specialised instruments and techniques.

The astronomer needs new tools for using this wealth of data in multiwavelength studies.
Errors of best star positions

- Hipparchus - 1000 stars
- The Landgrave of Hessen - 1000
- Tycho Brahe - 1000
- Flamsteed - 4000
- Argelander - 26000
- PPM - 400 000
- FK5 - 1500
- Tycho-1 - 1 million
- Hipparcos - 120 000
- GAIA - many million

Copenhagen University Observatory - E. Hoeg 1995
Virtual Observatories

- Provide tools for data analysis, visualization and mining.
- Develop interoperability concepts to make different databases seamless.
- Manage vast data resources and provide these on-line to astronomers and other users.
Registry: the yellow pages of the VO
PLASTIC: PLatform for AStronomy Tool InterConnection
VOTable: data stored in XML format
SIA: Simple Image Access
SSA: Simple Spectral Access
<table>
<thead>
<tr>
<th>Data Discovery</th>
<th>Spectral Analysis</th>
<th>Data visualisation and handling</th>
<th>SED building and fitting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aladin</td>
<td>VOSpec</td>
<td>TopCat</td>
<td>VOSED</td>
</tr>
<tr>
<td>Astroscope</td>
<td>SPLAT</td>
<td>VOPlot</td>
<td>Yafit</td>
</tr>
<tr>
<td>VOExplorer</td>
<td>Specview</td>
<td>VisIVO</td>
<td>easy-z</td>
</tr>
<tr>
<td>Datascope</td>
<td>Euro-3D</td>
<td>STILTS</td>
<td>GOSSIP</td>
</tr>
</tbody>
</table>

*(Workbench)*
Aladin Sky Atlas
http://aladin.u-strasbg.fr/

Description

Aladin is an interactive software sky atlas allowing the user to visualize digitized astronomical images, superimpose entries from astronomical catalogues or databases, and interactively access related data and information from the Simbad database, the VizieR service and other archives for all known sources in the field.

The Aladin sky atlas is available in three modes: a Java Standalone application, a Java applet interface and a simple previewer.
Astrogrid Workbench

http://www2.astrogrid.org/desktop

A desktop application for doing science in the Virtual Observatory.

With it, the user can discover and explore data resources, query remote catalogs, invoke remote processing tasks and construct workflows to automate tasks.

Key Features

Workbench provides the following features:

• Astroscope answers the question - what data is there available for any patch on the night sky?

• Helioscope answers the question - what Solar System data is available for a given time range?

• Task Launcher provides an easy way to query a wide range of astrophysical databases. Thus one can retrieve all stars with a certain colour from the Sloan All Sky Survey, or Brown Dwarfs from the UKIDSS WFCAM infrared survey.

• Task Launcher also enables a scientist to run a wide range of applications, such as SExtractor & Pegase, and return results in standard VO formats (such as VOTable)

Results may be saved to the user's MySpace, or to their local disk. For seamless analysis of results Workbench can pass data, via PLASTIC to viewers such as TopCat, Aladin & VOSpec.
Title: IRAS Point Source Catalog, Version 2.0
ID: ivo://nasa.heasarc/iraspsc
Description:
The IRAS Point Source Catalog, Version 2.0, is a catalog of some 250,000 well-confirmed infrared point sources observed by the Infrared Astronomical Satellite (IRAS), i.e., sources with angular extents less than approximately 0.5, 0.5, 1.0, and 2.0 arcminutes in the in-scan direction at 12, 25, 60, and 100 microns (um), respectively. Positions, flux densities, uncertainties, associations with known astronomical objects and various cautionary flags are given for each object in the catalog.
Away from confused regions of the sky, the survey is complete to about 0.4, 0.5, 0.6, and 1.0 Janskies (jy) at 12, 25, 60, and 100 microns, respectively. Typical position uncertainties are about 2 to 6 arcseconds in the in-scan direction and about 8 to 16 arcseconds in the cross-scan direction.
How to use DataScope

The NVO DataScope tool queries hundreds of astronomical services about a given location or region and organizes the information so that you can browse it, select data for download, or pass it into compatible tools for further analysis.

Starting DataScope

To start DataScope, just enter a position and size in the two fields provided and submit the query. You can enter the position as a target name, or in sexagesimal or decimal coordinates. Many formats are supported. The size is specified in decimal degrees.

Once you submit a query DataScope will start a query of registered resources that can be queried at that position. A result form will pop up and show you the kinds of responses you get.

If you wish to make sure to get fresh results you can click checkbox that skips the cache. Similarly the DataScope checks once an hour to see what resources are available to be queried. You can ask for this to be updated before your query by checking "Refresh registry" box.

A few recent queries are shown at the bottom of the page but you can leave your query off the list by clicking on the third checkbox.

Query results

The DataScope should immediately return with a page that begins to organize the results. If the results have already been cached the form will show the previously queried data. If you are starting a new query it will gather results over a period of a few minutes. A status section at the top of the query result page shows the progress of the query, the number of resources found and how long ago the query was initiated. All other data is organized into tabbed panels below the status area. You can jump to any of the panels at any time, just click on the panels' tabs at the top of the page.
VOSpec

http://esavo.esa.int/vospec/
Line flux

1.6742945316519336E-4 micron Jy
SPLAT

http://star-www.dur.ac.uk/~pdraper/splat/splat-vo/

SPLAT: Starlink SPectraL Analysis Tool
## Global list of spectra:

<table>
<thead>
<tr>
<th>Spectra</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NGC1068-NUC</td>
<td>NGC1068-NUC</td>
</tr>
<tr>
<td>NGC1068-NUC</td>
<td>NGC1068-NUC</td>
</tr>
<tr>
<td>NGC1068-NUC</td>
<td>NGC1068-NUC</td>
</tr>
<tr>
<td>NGC1068-NUC</td>
<td>NGC1068-NUC</td>
</tr>
</tbody>
</table>

## Animation controls

- **Delay:** 1
- **Loop forever:** False
- **Plot:** Create
- **Scaling option:** Auto
- **Current spectrum:**

## Capture controls

- **Start capture:** False
- **Capture to JPEG (otherwise PNG):** True
- **Basename for graphics files:** SPLAT

## Data attributes

- **Coordinate system:** Wave-length in vacuum
- **Units:** Angstrom
- **Standard of rest:** Centre of Sun
- **Date of observation:** 2000.0
- **Observatory:**
- **Longitude of observer:** E0:00:00.00
- **Latitude of observer:** N0:00:00.00
- **RA of source:** 0:00:00.0
- **Dec of source:** 0:00:00
- **Rest frequency:** 100000 GHz
- **Spectral origin:** 0
- **Source rest frame:** Centre of Sun
- **Source system:** Relativistic velocity
- **Source velocity:** 0
Specview

http://specview.stsci.edu/applet/specview_applet_run.html
Topcat

http://www.star.bris.ac.uk/~mbt/topcat/

TOPCAT: Tool for OPerations on Catalogues And Tables

TOPCAT is an interactive graphical viewer and editor for tabular data.
<table>
<thead>
<tr>
<th>Visible</th>
<th>Name</th>
<th>$ID</th>
<th>Class</th>
<th>Units</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Index</td>
<td>0</td>
<td>Long</td>
<td></td>
<td>Table row index</td>
</tr>
<tr>
<td>1</td>
<td>TARGET</td>
<td>1</td>
<td>String</td>
<td></td>
<td>Target name</td>
</tr>
<tr>
<td>2</td>
<td>RA</td>
<td>2</td>
<td>String</td>
<td>HMS</td>
<td>Right Ascension J2000</td>
</tr>
<tr>
<td>3</td>
<td>DEC</td>
<td>3</td>
<td>String</td>
<td>DMS</td>
<td>Declination J2000</td>
</tr>
<tr>
<td>4</td>
<td>RA2000</td>
<td>4</td>
<td>Double</td>
<td>degrees</td>
<td>Right Ascension J2000 (radiansToDegrees(hmsToRadians(DEC)))</td>
</tr>
<tr>
<td>5</td>
<td>DEC2000</td>
<td>5</td>
<td>Double</td>
<td>degrees</td>
<td>Declination J2000 (radiansToDegrees(dmsToRadians(DEC)))</td>
</tr>
<tr>
<td>6</td>
<td>BMAG</td>
<td>6</td>
<td>Float</td>
<td>mag</td>
<td>SuperCOS Bj magnitude</td>
</tr>
<tr>
<td>7</td>
<td>BMAG_ERR</td>
<td>7</td>
<td>Float</td>
<td>mag</td>
<td>BMAG error (fake value for demo data)</td>
</tr>
<tr>
<td>8</td>
<td>RMAG</td>
<td>8</td>
<td>Float</td>
<td>mag</td>
<td>SuperCOS R magnitude</td>
</tr>
<tr>
<td>9</td>
<td>RMAG_ERR</td>
<td>9</td>
<td>Float</td>
<td>mag</td>
<td>RMAG error (fake value for demo data)</td>
</tr>
<tr>
<td>10</td>
<td>SGFLAG</td>
<td>10</td>
<td>Short</td>
<td></td>
<td>SuperCOS Star/Galaxy flag: 1=galaxy, 2=star, 3=unclass, 4=none</td>
</tr>
<tr>
<td>11</td>
<td>galaxy</td>
<td>11</td>
<td>Boolean</td>
<td></td>
<td>Flag indicating a galaxy (sgflag == 1)</td>
</tr>
<tr>
<td>12</td>
<td>star</td>
<td>12</td>
<td>Boolean</td>
<td></td>
<td>Flag indicating a star (sgflag == 2)</td>
</tr>
<tr>
<td>13</td>
<td>VEL</td>
<td>13</td>
<td>Integer</td>
<td>km/s</td>
<td>Velocity/redshift - some from literature ZCAT</td>
</tr>
<tr>
<td>14</td>
<td>VEL_ERR</td>
<td>14</td>
<td>Integer</td>
<td>km/s</td>
<td>Nominal velocity error (fake value for demo data)</td>
</tr>
</tbody>
</table>
449 pairs found
New table created by match: 9: match(1,7) (449 rows)
VOPlot

http://vo.iucaa.ernet.in/~voi/voplot.htm
### Column Metadata

Click on a row to choose a Column Id.

<table>
<thead>
<tr>
<th>Column Id</th>
<th>Column Name</th>
<th>UCD</th>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1</td>
<td>_RAJ2000</td>
<td>POS_EQ_RA_MAIN</td>
<td>Original</td>
</tr>
<tr>
<td>$2</td>
<td>_DEJ2000</td>
<td>POS_EQ_DEC_M...</td>
<td>Original</td>
</tr>
<tr>
<td>$3</td>
<td>RAJ2000</td>
<td>POS_EQ_RA_MAIN</td>
<td>Original</td>
</tr>
<tr>
<td>$4</td>
<td>DEJ2000</td>
<td>POS_EQ_DEC_M...</td>
<td>Original</td>
</tr>
<tr>
<td>$5</td>
<td>Diam</td>
<td>EXTENSION_DIAM</td>
<td>Original</td>
</tr>
<tr>
<td>$6</td>
<td>Dist</td>
<td>POS_GAL_HC</td>
<td>Original</td>
</tr>
<tr>
<td>$7</td>
<td>Age</td>
<td>TIME_AGE</td>
<td>Original</td>
</tr>
<tr>
<td>$8</td>
<td>Nc</td>
<td>NUMBER</td>
<td>Original</td>
</tr>
<tr>
<td>$9</td>
<td>RV</td>
<td>VELOC_HC</td>
<td>Original</td>
</tr>
</tbody>
</table>

### Operator Calculator

```
+  -  *  /  log  ln
sqrt  pow  dexp  exp  cos  acos
sin  asin  tan  atan  torad  todeg
```

Enter column name: 

Enter expression: 

Enter unit: 

[Help] [Close] [Add]
STILTS

http://www.star.bris.ac.uk/~mbt/stilts/

STILTS: Starlink Tables Infrastructure Library Tool Set

Commands:

• **tcopy** - Table format converter
• **tpipe** - Generic table pipeline processing utility
• **tmatch2** - Two-table crossmatcher
• **tjoin** - Trivial side-by-side multiple-table joiner
• **tcube** - N-dimensional histogram calculator
• **tcat, tcatn** - Multiple-table concatenaters

Two VOTable-specific commands:

• **votcopy** - VOTable encoding translator
• **votlint** - VOTable validity checker

Facilities:

• format conversion
• data and metadata manipulation and display
• sorting
• row selections
• column calculation and rearrangement
• crossmatching
• statistical calculations
• histogram calculation
• data validation
• VO service access (experimental)

At version 1.2 (July 2006) a couple of Virtual Observatory service access commands have also been introduced on an experimental basis:

• **regquery** - Registry Query
• **multicone** - Multiple Cone Search
Yafit: Yet Another Fitting Tool

Usage: fit [-help] [-debug]
model=<model-file>
[modelfmt=ymodel|galaxev|starburst99|svotar|sideways-vot]
obs=<obs-file>
[smoother=square|point]
[scale=true|false]
[fitcalc=chi2|poisson|unscaled]
[gui=true|false]
[summary=<out-file>]
[bestfits=<out-table>]
[bestfitsfmt=<out-format>]

Usage: plotmodel [-help] [-debug]
in=<model-file>
[ifmt=ymodel|galaxev|starburst99|svotar|sideways-vot]

Usage: plotobs [-help] [-debug]
in=<obs-file>
• Written in Python/C
• Interactive/Batch mode
• Uses a set of user defined templates and a combination of correlation and fitting algorithms
• Flexible and modular architecture
Astrostatistics

- Are these stars/galaxies/sources an unbiased sample of the vast underlying population? Sampling
- When should these objects be divided into 2/3/... classes? Multivariate classification
- What is the intrinsic relationship between two properties of a class (especially with confounding variables)? Multivariate regression
- Can we answer such questions in the presence of observations with measurement errors & flux limits? Censoring, truncation & measurement errors
• When is a blip in a spectrum, image or datastream a real signal?  
Statistical inference  
• How do we model the vast range of variable objects (extrasolar planets, BH accretion, GRBs, ...)?  
Time series analysis  
• How do we model the 2-6-dimensional points representing galaxies in the Universe or photons in a detector?  
Spatial point processes & image processing  
• How do we model continuous structures (CMB fluctuations, interstellar/intergalactic media)?  
Density estimation, regression
How often do astronomers need statistics?  
(a bibliometric measure)

Of ~15,000 refereed papers annually:

1% have `statistics’ in title or keywords
5% have `statistics’ in abstract
10% treat variable objects
5-10% (est) analyze data tables
5-10% (est) fit parametric models
The state of astrostatistics today

The **typical** astronomical study uses:
- Fourier transform for temporal analysis (Fourier 1807)
- Least squares regression (Legendre 1805, Pearson 1901)
- Kolmogorov–Smirnov goodness–of–fit test (Kolmogorov, 1933)
- Principal components analysis for tables (Hotelling 1936)

Even traditional methods are often misused:
- Six unweighted bivariate least squares fits are used interchangeably in $H_0$ studies with wrong confidence intervals
  \[ \text{Feigelson & Babu ApJ 1992} \]
- Likelihood ratio test (F test) usage typically inconsistent with asymptotic statistical theory
- K–S g.o.f. probabilities are inapplicable when the model is derived from the data
  \[ \text{Babu & Feigelson ADASS 2006} \]
Under-utilized methodology

- modeling (MLE, EM Algorithm, BIC, bootstrap)
- multivariate classification (LDA, SVM, CART, RFs)
- time series (autoregressive models, state space models)
- spatial point processes (Ripley’s K, kriging)
- Non-detections (survival analysis)
- image analysis (computer vision methods, False Detection Rate)
- statistical computing (R)
A new imperative: Virtual Observatory

Huge, uniform, multivariate databases are emerging from specialized survey projects & telescopes:

- $10^9$-object catalogs from USNO, 2MASS & SDSS opt/IR surveys
- $10^6$- galaxy redshift catalogs from 2dF & SDSS
- $10^5$-source radio/infrared/X-ray catalogs
- $10^3$-$4$-samples of well-characterized stars & galaxies with dozens of measured properties
- Many on-line collections of $10^2$-$10^6$ images & spectra
- Planned Large-aperture Synoptic Survey Telescope will generate $\sim 10$ Pby

The Virtual Observatory is an international effort underway to federate these distributed on-line astronomical databases.

Powerful statistical tools are needed to derive scientific insights from extracted VO datasets
(NSF FRG involving PSU/CMU/Caltech)
Some methodological challenges for astrostatistics in the 2000s

• Simultaneous treatment of measurement errors and censoring (esp. multivariate)

• Statistical inference and visualization with very-large-N datasets too large for computer memories

• A user-friendly cookbook for construction of likelihoods & Bayesian computation of astronomical problems

• Links between astrophysical theory and wavelet coefficients (spatial & temporal)

• Rich families of time series models to treat accretion and explosive phenomena

https://asaip.psu.edu/resources/on-line-training