A database of charged cosmic rays

RICHARD TAILLET¹, DAVID MAURIN², FRÉDÉRIC MELOT²,

¹ LAPTh, CNRS, université de Savoie
² LPSC, CNRS
taillet@lapth.cnrs.fr

Abstract: This poster presents a new on-line database and associated on-line tools (data selection, data export, plots, etc.) for charged cosmic-ray measurements, available at http://lpsc.in2p3.fr/cosmic-rays-db. The experimental setups (type, flight dates, techniques) from which the data originate are included in the database, along with the references to all relevant publications. The database relies on the MySQL5 engine. The web pages and queries are based on PHP, AJAX and the jquery,jquery.cluetip,jquery-ui, and table-sorter third-party libraries. In this first release, we restrict ourselves to Galactic cosmic rays with $Z \leq 30$ and a kinetic energy per nucleon up to a few tens of TeV/n. This corresponds to more than 200 different sub-experiments (i.e., different experiments, or data from the same experiment flying at different times) in as many publications. We provide tools to sort and visualise the data. New data can be submitted, providing the community with a collaborative tool to archive past and future cosmic-ray measurements. Any help/ideas to further expand and/or complement the database is welcome (please contact crdatabase@lpsc.in2p3.fr).

Keywords: icrc2013, database, cosmic rays.

1 Introduction

Since the discovery of cosmic rays (CR) a century ago, instrumental capabilities have steadily improved. A large variety of types of experiments (balloon- or satellite-borne, flown on a shuttle, installed on the international space station, or ground-based experiments) and techniques have been used (nuclear emulsions, drift chambers, Cerenkov counters, spectrometers...) to refine our knowledge of the CR composition and spectrum.

Interestingly, most of the even oldest CR measurements are not outdated yet. Indeed, many instruments are designed to focus on specific CR species: neither all instruments have the isotopic resolution capabilities, nor all species have been measured repeatedly. In the last twenty years, efforts have also been more devoted in measuring the CR composition at higher energy than refining the low energy data; this situation is currently changing with the AMS-02 experiment installed on the International Space Station since May 2011. Some old experiments are also useful when one wishes to inspect a possible charge-sign dependence (22 year cycle) of the Solar modulation effect as a function of the Sun polarity, as first suggested by [1] and further studied in [2]. For all these reasons, we believe it is worth providing an archival database of CR measurements to the community.

CR data are the backbone of Galactic CR propagation studies. In the last twenty years, anti-protons and positron fraction measurements have also become a strong probe for dark matter indirect searches [3, 4]. A database will therefore be useful to any researcher in these fields, but also to CR experimentalists who wish to compare their data to previously published ones. A previous effort in this direction was presented in [5]. We present here an independent, contextualised and more complete data compilation, along with many user-friendly interfaces and tools to use them.

1.1 Content of the database

In this section, we first describe the information gathered in the database, and the data themselves. We then present how this information is organised in a MySQL framework.

1.2 Definitions

CR data are connected to experiments, analyses, and publications. The need to define “sub-experiments” arises because i) an experiment may consist of several detectors, or ii) an instrument may have flown several times, or over distinct periods. Data from a sub-experiment often involve several CR species, the analyses of which are published in one or several papers. For the sake of clarity, the following keywords/definitions are used in the database:

Experiment Name associated with the instrument (CREAM, AMS). To identify unnamed balloons, we use the syntax Balloon (YYYY), and a further distinction is made if a balloon was flown several times: a comma-separated list of years Balloon (1966,1967) is used if the data were analysed and published for each flight; a plus-separated list Balloon (1967+1968) is used if the data resulted from the combined analysis of the flights.

Sub-experiment Sub-detector name or experiment name concatenated with the flight number and data taking period (YYYY/MM), with start and stop dates separated by a hyphen for durations over a month. [Balloon (1972/07), CREAM-I (2004/12-2005/01), CREAM-II (2005/12-2006/01), Ulysses-HET (1990/10-1997/12)].

Cosmic-ray quantity Combination (sum, ratio, etc.) of...
measured CR specific\cite{1}. It can be an elemental (e.g., C), isotopic (e.g., $^1\text{H}$), or leptonic (e.g., $e^+$) flux, or any ratio of these quantities such as B/C, $^{10}$Be/Be, $e^+/e^+ + e^-$, etc. The keyword SubFe is used for the group $Z = 21 - 23$, but no other charge group is defined for now.

**Energy axis** Published data may use as the energy axis: CR total energy $E_{\text{tot}}$, rigidity $\mathcal{R} = pc/Ze$ ($p$ is the momentum, $c$ the speed of light, and $e$ the electron charge), kinetic energy $E_k = E_{\text{tot}} - m$ ($m$ is the CR mass), kinetic energy per nucleon $E_{k/n} = E_k/A$ ($A$ is the atomic number). In the database, we allow four representations of the energy unit and axis: $[\text{GeV}]/[\text{GeV}/\text{MeV}]$ for $E_{\text{tot}}$, $[\text{GV}]/[\text{GV}/\text{MV}]$ for $E_k$, and $[\text{GeV}/\text{MeV}]/[\text{GeV}/\text{MeV}]$ for $E_{k/n}$.

**Publication** Referred or non-referred reference (journal or conference proceedings) providing CR quantity data from (sub-)experiments. A publication is usually attached to a single (sub-)experiment and it contains different CR measurements, but there are a few exceptions. Over time, some of these publications may be superseded by newer analyses: a specific entry of the database allows to keep track of deprecated analyses and references.

**Data** CR quantity measurement and uncertainties at one or several energy bins. Sect. 1.3 gives a complete description of a data entry.

The fact that combinations of CR quantities are themselves measurable CR quantities introduce a subtleties in the choice of how to handle the database. One could be tempted to fill the database with all useful combinations of data (e.g., the often used B/C ratio) from published quantities (e.g., B and C fluxes). However, the number of combinations that can be formed is large (for $Z < 30$, as many elements and about a hundred isotopes can be combined), and the procedure to combine the errors on the measurements is not always sound.

For these reasons, we decided to fill the database with the published quantities only. We leave the task of extracting the most complete dataset (for a given CR quantity) to the Data Extraction interface (Sect. 2.2), which combines the data found directly in the database, and those obtained by looking for all combinations of data leading to this quantity (see Sect. 2.2).

### 1.3 Data description and units

The structure of a CR data entry (energy, energy range, measurement and uncertainties) for any measured quantity is as follows:

- **(E)** ‘Central’ energy given in the publication (unit is $[\text{GeV}]/[\text{GeV}/\text{MeV}]$ if the energy axis is $E_{\text{tot}}$ or $E_k$, $[\text{GV}]/[\text{GV}/\text{MV}]$ for $E_k$, and $[\text{GeV}/\text{MeV}]/[\text{GeV}/\text{MeV}]$ for $E_{k/n}$). If only the bin range (see below) is given in the publication, the geometric mean $\langle E \rangle = \sqrt{E_{\text{min}}E_{\text{max}}}$. is used.

- **Bin range** Energy range (same unit as $\langle E \rangle$). If only $\langle E \rangle$ is given in the publication, $E_{\text{min}} = E_{\text{max}} = \langle E \rangle$.

- **Value** Measured CR quantity in unit of $[\langle E \rangle^2 \text{m}^2 \text{s}^{-1}]$ if this is a flux, or unit-less if this is a ratio. The data correspond to top-of-atmosphere (TOA) quantities, i.e. modulated by the Sun’s activity.

- **Stat Err** Statistical error (same unit as $\text{Value}$).

- **Syst Err** Systematic error (same unit as $\text{Value}$); set to 0 if not given in publication.

### 1.4 Database structure description

The database engine is MySQL5, hosted at the Laboratoire de Physique et Cosmologie (LPSC, Grenoble, France) on a backed up server. The tables are organized as follows:

- **exp** Name, type, web site (if available), and flight date.

- **subexp** Link to the experiment it belongs to, name, description of the apparatus, flight details (launch location and the number of flights for balloons), flight dates, distance to the Sun [AU], and Solar modulation level [MV].

- **publi** Bibliographic reference, web link, publication year, $\text{BibTex}$ entry (taken from the Astrophysics Data System ADS\cite{2}), and link to other publications (if more recent analyses exist).

- **subexp.publi** Bridge table linking entries from publi to one or several entries of subexp.

- **element** Name, mass, atomic number, charge, etc. for CR quantities (isotopes, elements, $\bar{p}$, $e^-$, and $e^+$).

- **data** Type (flux or ratio of element), energy axis, energy, bin range, value, statistic and systematic errors.

- **user** Contact details of administrators (persons authorised to change and validate submitted data).

**validation** Contact details of persons submitting new data (see Sect. 2.3 for the New Data interface), validation date, and identity of the person (user) who validated the data.

Each entry in a table is associated with a unique identifier. These identifiers are used to link elements from one table to another (for example, several sub-experiments can be linked to a single experiment).

---

2. A CR must be a stable species with respect to the confinement in the Galaxy, i.e. with an effective lifetime $\gtrsim$ kyr.

3. The Solar modulation parameter is not a direct product of an experiment analysis. It is generally estimated from the measured TOA fluxes and an assumption about the interstellar flux.
2 Website, interfaces, and example plots

The CR database website \texttt{http://lpsc.in2p3.fr/cosmic-rays-db} is hosted by the LPSC laboratory website, and is based on a LAMP solution.\footnote{7} Authentication uses the \texttt{https} protocol to ensure a good level of confidentiality (only administrators own credentials to access protected areas). All web pages are written using the PHP (Hypertext PreProcessor) language, with a global structure made in AJAX (Asynchronous JavaScript and XML). The third-party libraries jQuery, jQuery-UI, jQuery-Cluetip, and tablesorter are also used.

The website is based on tabs, in which the user is guided by HELP boxes (identified by question mark icons). We give below a brief description of the implemented tabs:

**WELCOME** Quick description and organisation of the database, log of the latest changes, and link to download the database content formatted for the USINE propagation code.

**EXPERIMENTS/DATA** List of available data sorted by experiment names or dates. A list of experiment acronyms is given.

**DATA EXTRACTION** Main interface to retrieve data in ASCII files, ROOT macros and plots, and Bib\TeX references for the selection.

**ADMIN** Shown for authenticated users only: internal checking of the database content, validation of submitted data.

**LINKS** Standard useful (here GCR-related) web links.

**NEW DATA** Interface to submit new data which will appear in the database after validation by authorised users.

As underlined previously, native data (i.e. data directly from publications) are listed and accessed from **EXPERIMENTS/DATA** (Sect. 2.1). In the **DATA EXTRACTION** tab (Sect. 2.2), native data and matching combinations of native data are combined to provide the most complete list of data found for user-selected quantities/criteria. Adding new data is possible from the **NEW DATA** tab (Sect. 2.3).

2.1 Data access from **EXPERIMENTS/DATA** tab

The list of published data can be ordered by experiment name or date. The publication references related to this sub-experiment are then listed along with the quantities measured (older analyses/publications of the same data are indicated). The most useful actions/pop-up informations available for the user are:

- experiment description (name, type, official web page);
- sub-experiment description (name, data periods, instrument description [mouse-over name], experimental setup picture [magnifying glass icon]);
- data for each publication (from a sub-experiment). A click on [data] pops-up a window that (a) summarises all the information on the sub-experiment, along with the ADS link of the reference and (b) shows the data (see Sect. 2.3 for their format).

By default, CR data for nuclei and anti-nuclei are given as a function of kinetic energy per nucleon, whereas leptons are given as a function of the kinetic energy. Whenever the energy axis is rigidity, the flag [Rigidity] is added after the data.

2.2 Selection and tools from **DATA EXTRACTION** tab

The user can select the quantity to display through the selection interface within the tab. A mandatory step is the quantity selection (Flux or ratio selection), for which a few predefined choices are proposed. For a ratio, both the numerator and denominator selection boxes must be filled (auto completion is enabled). The other optional selection criteria (Refine search criteria) are:

- Energy axis: to be selected among EKN, EK, R or Etot;
- Flux rescaling: multiplies the flux values and errors by \(\langle E \rangle^a\) (useful for presentation purpose);
- Energy range: restricts the energy range allowed;
- (Sub-)Experiment names: list of comma-separated names (partial names allowed, e.g., CREAM, BESS);
- Time interval: selects only experiments falling into the selected period (format is YYYY/MM).

Fig. 1: Snapshot of the result of the data extraction operation. This pop-up window appears after the selection step is completed. Buttons, links, and tables give access to raw data and plots, see text for details.

7. The acronym LAMP refers to a stack of free open source softwares: Linux operating system, Apache HTTP server, MySQL database software, and PHP.
8. \texttt{http://lpsc.in2p3.fr/usine}
9. \texttt{http://root.cern.ch}
• Show also data from combinations: tick box to add in the search the data points obtained from combinations of ‘native’ data.

Hitting the Extract Selection button pops-up a new window with the data extracted from the user selection. This is shown in Fig. 1 organised in three panels (click on hide/show to collapse/expand each panel):

1. Plots and exports for the selection: Get ROOT Macro, Data Files, Plot, and USINE File buttons return i) a ROOT executable C++ file database_plot.C to re-generate and/or modify the plot; ii) a tar-ball database_plot.tar.gz of ASCII files containing the data (one file per sub-experiment); iii) a high-resolution image database_plot.png of the plot; iv) a USINE-compliant file database_plot.USINE (i.e., that can be used as an input of the USINE propagation code).

2. List of experiments found for the selection: summary of the data sorted by (sub-)experiment (name, publication, number of data, etc.). The Get BibTex and Latex cite buttons provide respectively a BibTeX file (bibtex.bib) to be included in the references, and the text to cite this selection in the LaTeX document. As for the Export/Files tab (see Sect. 2.1), links to the experiment website and the ADS publication are provided.

3. Data for the selection: data in a table (see Sect. 1.5 for the content description) sorted by experiment name or energy. An asterisk denotes the data obtained by combinations of native data of the database.

We remind that, with the default search criteria (i.e., none), all analyses of a CR quantity by the same instrument show up in the result as long as they correspond to different data taking (or analysed) periods (i.e., different sub-experiments). Most of the time, these data are independent, but in a few cases, the analysed periods overlap. It happens, for instance, for the Voyager data (launched in 1977 and still taking data). In that case, it is up to the user to decide which data sets are relevant for her/his analysis, and exclude it using the (Sub-)Experiment names or Time interval selection box.

2.3 New data from the Add Data tab

This tab allows anyone to interactively enter new data. This is an essential part of the database as it provides the community with the possibility to contribute to the completion of the database (either by adding data from new instruments, or adding missing data from older experiments).

Submitting new data consists in two parts: submitter identification (contact details) and data submission. For the first three steps, the submitter has to fill in info about the experiment, sub-experiment and publication, with the choice of adding info to an existing entry in the database, or to add a new one: the latter action pops-up a new window in which the submitter is guided—HELP boxes are provided for each item—to fill the necessary informations, which match the keys of the database structure described in Sect. 1.4. Each time a new entry is submitted, the submitted element becomes available for further submissions, though it does not appear yet in the database (i.e., in Experiments/Data and Extract data tabs).

Once the three previous steps are completed, the last action is the submission of the CR data. A template file for the required format (and units) of the data is provided (see Sect. 1.4). Only one CR quantity at a time (with as many energy bins as desired) can be submitted. Before the final submission of the data, the Data graphical check button pops-up a summary of the uploaded file, along with a plot of the submitted data for a last visual inspection. At this stage, the submission process can still be cancelled if any mistake is spotted.

For each submitted entry (experiment, sub-experiment, publication, data), an email is sent to the administrators of the database content Validation tools from the Admin tab are then used (format check, completeness, etc.) to authorise the addition in the database (pending validation, the data are inserted in the database with a ‘not validated’ flag, and do not appear on the web site).

3 Conclusions

We strongly encourage submissions of new data, as well as user comments and suggestions of improvements and new features.

Acknowledgment: We acknowledge the support of the french PNHE. The ICRC 2013 is funded by FAPERJ, CNPq, FAESP, CAPES and IUPAP.

References


10Based on the ROOT library http://root.cern.ch To execute type root database_plot.C (the data are hard-coded). The errors displayed correspond to the quadratic sum of statistical and systematic uncertainties.
11These files are useful to quickly prepare scientific manuscripts based on LATEX and BibTEX.
12For now, only the authorised persons are the developers of the database. Any person wishing to get involved in further developments of the database is welcome to contact us.