BIDS Statistical Models - An implementation-independent representation of General Linear Models

Christopher Markiewicz, Katherine Bottenhorn, Gang Chen, Alejandro de la Vega, Oscar Esteban, Camille Maumet, Thomas Nichols, Russell Poldrack, Jean-Baptiste Poline, Tal Yarkoni

BIDS Statistical Models - An implementation-independent representation of General Linear Models

Authors:

Christopher Markiewicz[1], Katherine Bottenhorn[2], Gang Chen[3], Alejandro De La Vega[4], Oscar Esteban[5], Camille Maumet[6], Thomas Nichols[7], Russell Poldrack[1], Jean-Baptiste Poline[8], Tal Yarkoni[9]

Institutions:

[1]Stanford University, Stanford, CA, [2]Florida International University, Miami, FL, [3]National Institutes of Health, Bethesda, MD, [4]University of Texas Austin, Austin, TX, [5]University Hospital of Lausannne and University of Lausanne, Lausanne, Switzerland, [6]Inria, Univ Rennes, CNRS, Inserm, Rennes, France, [7]University of Oxford, Oxford, United Kingdom, [8]McGill University, Montreal, Quebec, [9]University of Texas at Austin, Austin, TX

First Author:

Christopher Markiewicz
Stanford University
Stanford, CA

Co-Author(s):

Katherine Bottenhorn
Florida International University
Miami, FL
Gang Chen
National Institutes of Health
Bethesda, MD
Alejandro De La Vega
University of Texas Austin
Austin, TX
Oscar Esteban
University Hospital of Lausannne and University of Lausanne
Lausanne, Switzerland
Camille Maumet
Inria, Univ Rennes, CNRS, Inserm
Rennes, France
Thomas Nichols
University of Oxford

Oxford, United Kingdom
Russell Poldrack
Stanford University
Stanford, CA
Jean-Baptiste Poline
McGill University
Montreal, Quebec
Tal Yarkoni
University of Texas at Austin
Austin, TX

Introduction:

The general linear model (GLM) is a mainstay of neuroimaging analysis, and particularly task-based functional MRI (fMRI) analysis. As such, there is an independent implementation in all major tool suites. This proliferation provides access to the technique to researchers from varied backgrounds, but each implementation has its own input specification. Consequently, it is non-trivial to compare methods across studies that use different tool suites or to constructing the same model across suites in order to compare the tools themselves.

The Brain Imaging Data Structure (BIDS) is a standard for organizing data from a broad range of neuroscientific experiments [2,3,5,6], including neuroimaging data, experimental events and physiological recordings. Because experiments are often carefully designed to elicit neural responses corresponding to specific proposed cognitive mechanisms, it is useful to include the intended model in the dataset as a means of documenting the experiment and providing a software-independent guide to reproduce the analysis.

Here we present BIDS Stats-Models, a specification for describing how a GLM or similar model should be fit to a BIDS dataset.

Methods:

The BIDS Stats-Models specification was inspired by the OpenFMRI [7] structure, which included descriptions of events, run-level models and group-level models, formatted for easy integration with the FSL FEAT modeling tool [8,9].

Although experimental event descriptions were originally incorporated into BIDS, this did not abstract model descriptions away from the details of any one tool. A BIDS extension proposal (BEP) was initially drafted in September 2016, using a JSON format to declare the stages of the model, and assuming the input data is organized according to BIDS. In October 2018, a workshop of 20 attendees was hosted at Stanford University to assess and refine the specification.

Reference implementations have been used to drive development of the specification. The PyBIDS library [10] began to support components of Stats-Models in June 2017, and an application to run models (FitLins [4]) was begun in October 2017. The initial version of FitLins used the nilearn[1] estimator.

Results:

The BIDS statistical model is encoded as a JSON object describing the inputs to the model, and an array of steps corresponding to levels (run, session, subject or dataset) of analysis (Fig. 1a). Each step contains a model specification and contrast specification (Fig. 1b).

Image inputs to a Stats-Model may be selected via BIDS entities, such as "task", "run", "session" or "subject", if provided. These images, along with events and time series are provided as inputs to the initial (typically run-level) step. Later steps receive as inputs the statistical maps generated as outputs in the previous steps, along with variables defined at that level in the BIDS dataset. For example, if age or experimental group are defined in participants.tsv, these variables are available at the dataset level of analysis.

Within a step, a model can be specified to indicate the dependent variables, variance structure, error distribution and software-specific parameters.

Contrasts are named, weighted sums of parameter estimates, and may be of type "t", "F" or "FEMA" (fixed-effects meta-analysis). The resulting contrast maps are the inputs to the next level; hence, to pass a raw parameter estimate to the next level, a "dummy contrast" that shares a name with its lone input may be used.

Models can be included in a BIDS dataset, in a file according to the template models/model-_[desc-]_smdl.json.

a

```
{
  "Name": "my_first_model",
  "Input":
    {
     "task": "nback"
    },
  "Description": "A simple 2-condition contrast",
  "Steps": [
    {
      "Level": "Run",
      ...
    },
    {
      "Level": "Session",
      ...
    },
    {
      "Level": "Subject",
      ...
    },
    {
      "Level": "Dataset",
      ...
    }
  ]
}
```

b

```
"Steps": [
  {
    "Level": "run",
    "Model": {"X": ["Face", "Place", "a_comp_cor_*"]},
    "Contrasts": [
      {
        "Name": "faces_vs_places",
        "ConditionList": ["Face", "Place"],
        "Weights": [1, -1],
        "Type": "t"
      }
    ]
  },
  {
    "Level": "subject",
    "DummyContrasts": {"Type": "FEMA"}
  },
  {
    "Level": "dataset",
    "DummyContrasts": {"Type": "t"}
  }
]
```

a. A schematic model demonstrating the overall structure of a BIDS Stats-Model, including an optional name and description, inputselector, and steps. The steps are arranged in a “standard” order, fitting each run as a first level model, and combining at the session, subject and finally dataset levels. **b.** An example “Steps” section for a plausible multi-run study. A design matrix with face, place and aCompCor regressors is fit to each run in the study, and a contrast between face and place parameter estimates is calculated. At the subject level, runs are combined by a fixed-effects meta-analysis (FEMA) to estimate each subject’s contrast and variance. Finally, a t-contrast is used at the dataset level to estimate the contrast with subject as a random effect.

(https://files.aievolution.com/prd/hbm2101/abstracts/abs_1174/fig1.png)

Conclusions:

Reproducibility and interoperability are core principles of open scientific practice, and common formats for representing data and algorithms are necessary to achieve these principles in practice.

The BIDS Stats-Model aims to provide an abstract, tooling-independent representation of the most common modeling approaches in neuroimaging.