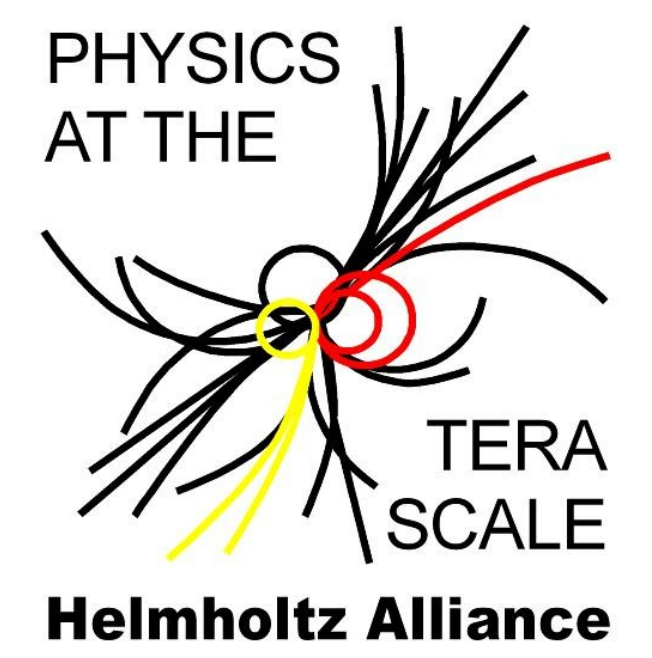


# The Analysis Centre: Statistics Tools Group.

Education, tools and support for physics analyses within the Helmholtz Alliance "Physics at the Terascale".



## Meetings and User Support.

### Informal Statistics Meetings:

- > Forum for **everyday's statistics problem**.
- > Experts **discuss** and answer questions.
- > Documented on alliance Wiki.
- > **Monthly** at DESY, trying EVO for remote people.

### Statistics Software Review:

- > **Connecting people** working on statistics projects.
- > **Exchange knowledge** and progress.
- > **Annual meetings**.

### Ideas for the future:

- > Improve collection of links/material
- > Started with links to statistics projects in Germany: [www.wiki.terascale.de/index.php/Statistics\\_Projects](http://www.wiki.terascale.de/index.php/Statistics_Projects)
- > Discussion forums, hypernews: to be decided.

## News and Announcements.

- > Subscribe to [anacentre-statistics@desy.de](mailto:anacentre-statistics@desy.de) via <https://lists.desy.de/sympa/info/anacentre-statistics>

## Mission.

For statistics tasks and problems in physics analyses, provide

- > **education**,
- > **support** and
- > development of **tools**.

### Covered topics:

- > Optimal **signal/background** separation.
- > Signal determination: advanced **fitting** techniques.
- > Determination of **limits**.
- > Data corrections (**unfolding** etc.).
- > **Systematic uncertainties**.

## People and Network.

### The core team:

- > O. Behnke, G. Flucke, C. Kleinwort, S. Schmitt (all DESY), K. Kröninger (Göttingen).
- > 10 - 50% FTE each.

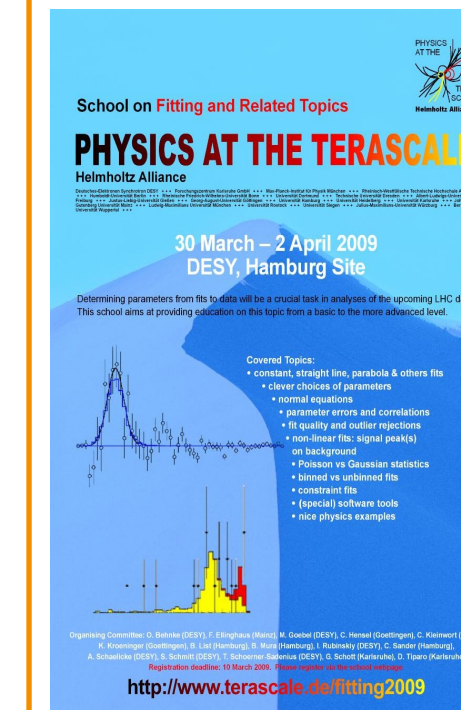
### Collaborators in the alliance and contacts:

- > People **contributing to schools** (various institutes).
- > DESY **fellows and studentships**.
- > Contact to developers of statistics related tools.
- > Established **contact to ATLAS and CMS statistics committees** via K. Kröninger/G. Schott.

## Schools/Workshops.

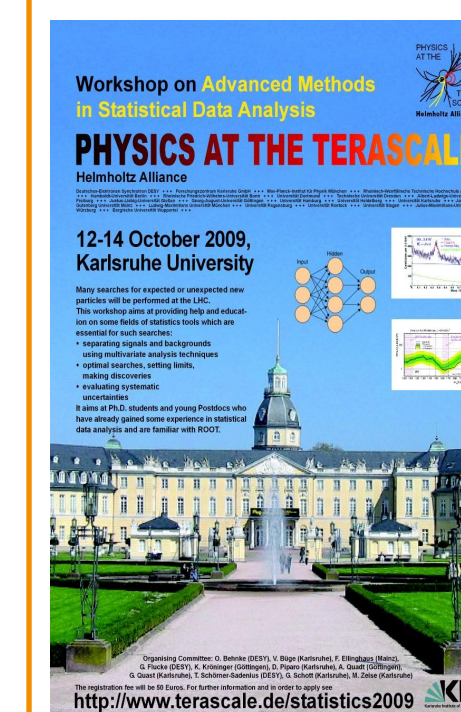
### Two educational events per year:

- > One at DESY, one at another alliance institute.
- > Combine **lectures with hands-on** exercises.
- > **Documented** for non-participants and future re-use.



### The past: 2009

- > **School on Fitting an Related Topics**
- > At DESY 30.3.-1.4.2009.
- > More than 90 registrants.
- > [www.terascale.de/fitting2009](http://www.terascale.de/fitting2009)



### Workshop Advanced Methods in Statistical Data Analysis:

- > In Karlsruhe 12.4.-14.10.2009.
- > About 50 registrants.
- > [www.terascale.de/statistics2009](http://www.terascale.de/statistics2009)

### Plans for 2010:

- > School on statistics basics: **spring 2010 at DESY Hamburg**.
- > Workshop on advanced statistics: **October 2010 in Göttingen**.

## Statistics Projects.

## Millepede II.

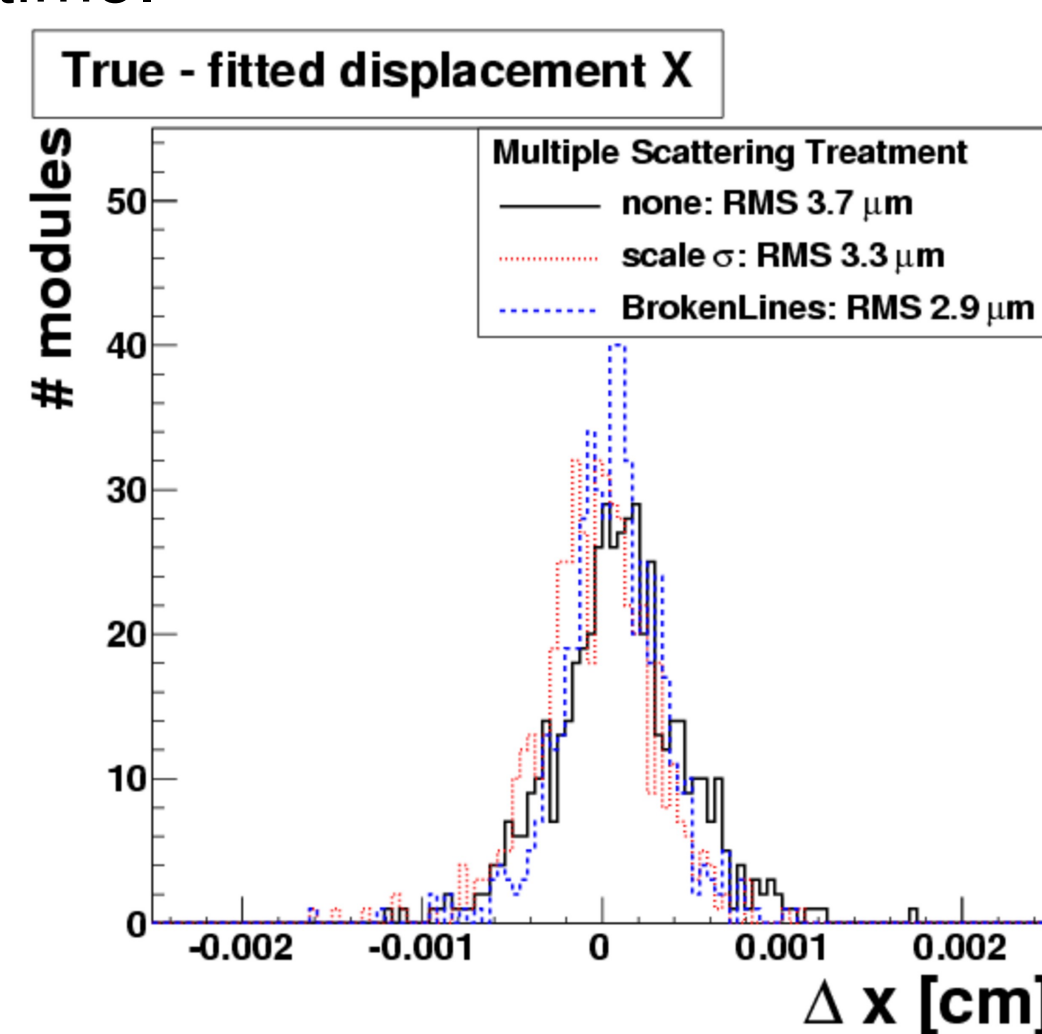
### Large linear fit problems:

- > Developed by V. Blobel (experiment independent).
- > Minimisation with two parameter classes:
  - local**: appearing in subsets of data (e.g.  $10^7$ ),
  - global**: appearing in all the data (up to  $10^6$ ).
- > Result needed for global parameters only.
- > Typical case for a **global alignment fit** with track (local) and alignment (global) parameters.

### Status in Statistics Tools Group:

- > Took over maintenance, see [www.wiki.terascale.de/index.php/Millepede\\_II](http://www.wiki.terascale.de/index.php/Millepede_II).
- > Recent improvement: Make use of bordered band matrix structure as for 'broken line' track fits.
  - ⇒ Allows alignment of silicon trackers with rigorous multiple scattering treatment in reasonable CPU time.

Toy MC alignment result ( $B = 0$  T): Improvement with rigorous multiple scattering treatment (BrokenLines)



### Core group involvement:

- > **Fitting tools made available for C++ world**:
  - > **Lvmini**: Efficient minimisation for large number of parameters.
  - > **Millepede II**: Linear fits for complex alignment or calibration problems.
  - > **Multivariate analysis: TMVA** development.
  - > **TUnfold**: unfolding program.

### Possible future projects/open topics:

- > C++ package for broken lines track fits.
- > Compare/benchmark unfolding programs.
- > ...

### Further statistics projects in Germany:

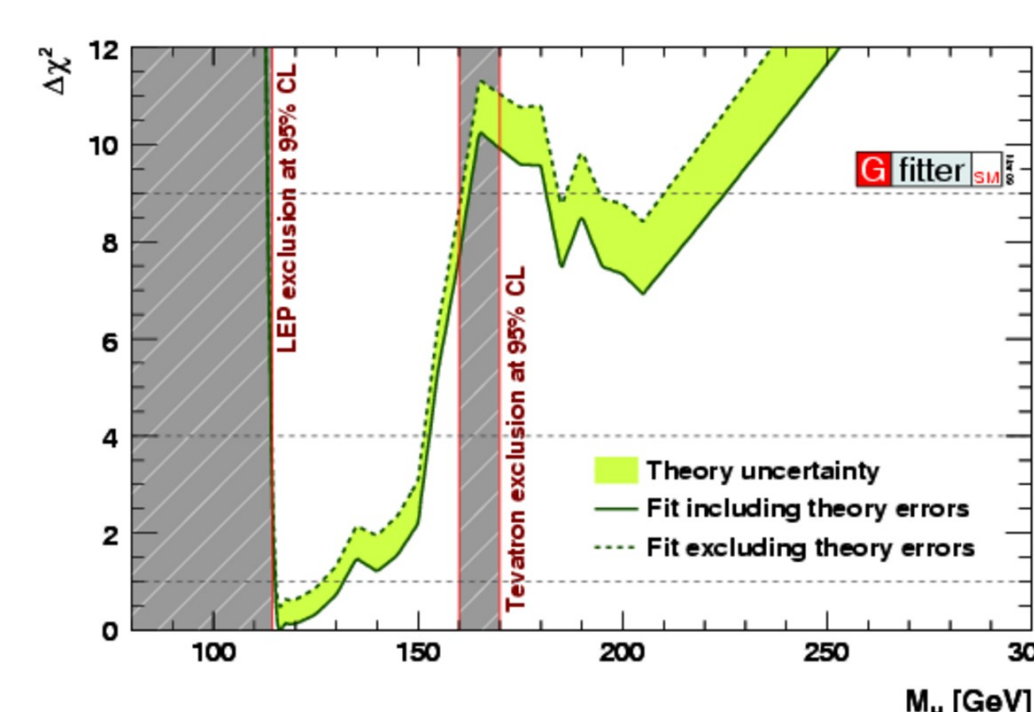
- > **BAT**: Bayesian Analysis Toolkit.
- > **Gfitter**: Generic Fitter for HEP Model Testing.
- > **Fittino**: SUSY parameters determination.
- > **RooStat[Cms]**: Modelling and combination of analysis channels etc.

## Fitting Physics Models.

### Gfitter

- > Generic fitter for HEP model testing.
- > Core packages for data handling, fitting, etc.
- > Physics plug-in packages: Standard Model and beyond.
- > [gfitter.desy.de](http://gfitter.desy.de)

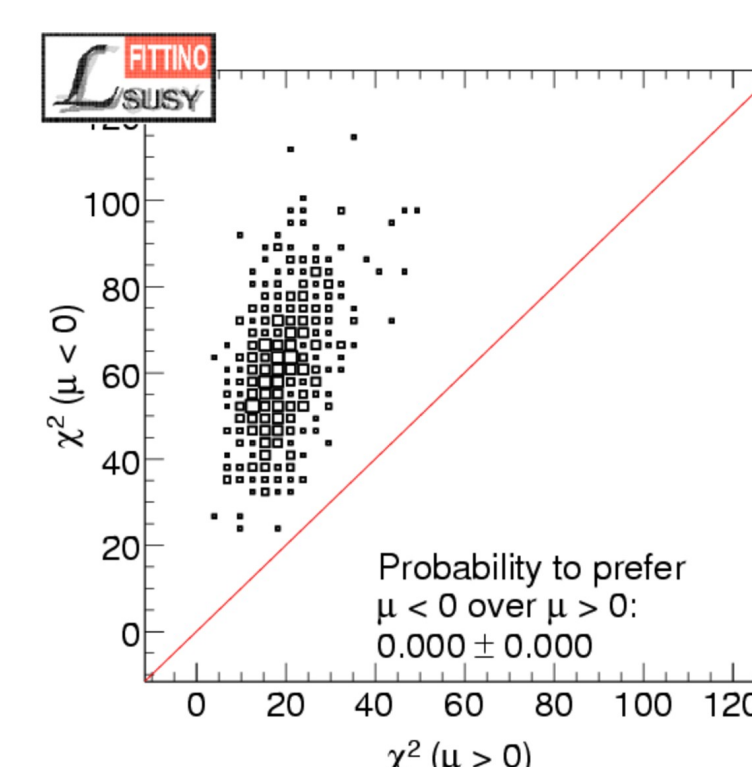
Combination of EW fit and Higgs searches



### Fittino

- > Tool for fitting SUSY parameters.
- > Collider, low energy and cosmological data.
- > [www.flc.desy.de/fittino](http://www.flc.desy.de/fittino)

Toy MC fits: Test LHC sensitivity for mSugra parameter  $\text{sign}(\mu)$



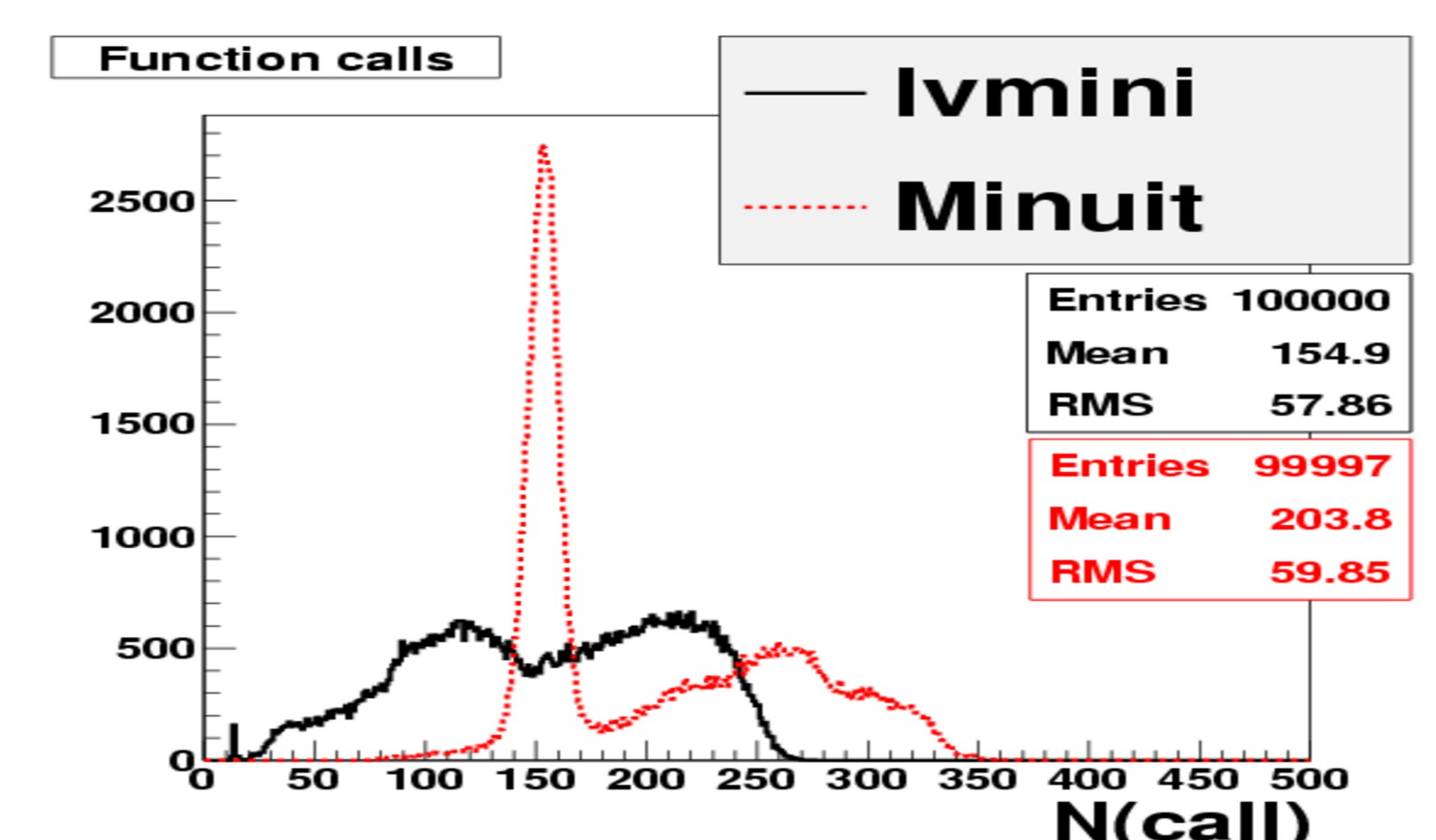
## Lvmini.

### Efficient minimisation for large number of parameters:

- > Fortran program by V. Blobel.
- > L-BFGS algorithm: matrix less quasi-Newton method.
- > Line search.
- > Few or **many** (aiming at up to  $10^5$ ) parameters.
- > User provides function **value and gradient**.

### Status in Statistics Tools Group:

- > Goal: Integration as alternative fitting mechanism in ROOT.
- > Prototype interface with ROOT setup.
- > Preliminary benchmarking with 2D-Rosenbrock function: less function calls.



## Bayesian Analysis Toolkit.

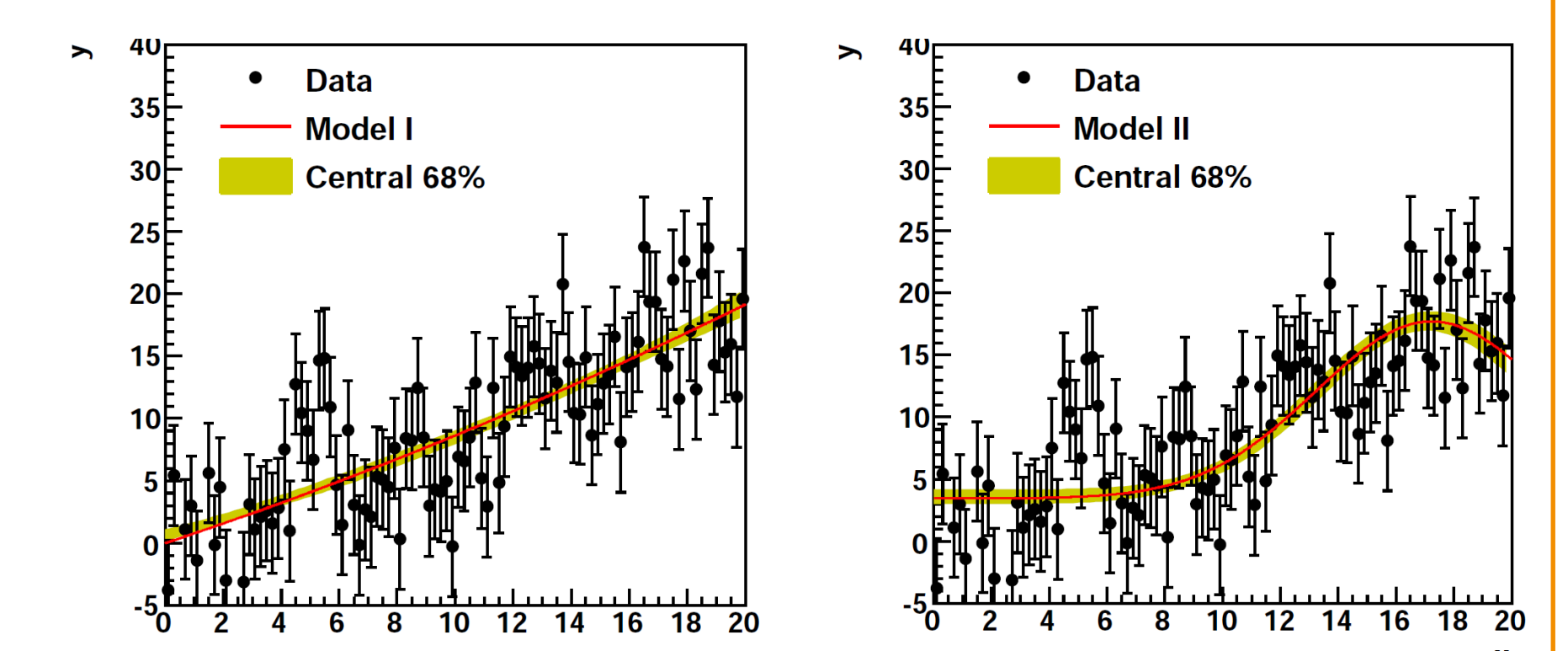
- > Toolkit for data analysis.
- > Based on Bayes' Theorem.
- > Using Markov Chain Monte Carlo.
- > Provides parameter estimation, uncertainty propagation, limit setting,...
- > [www.mppmu.mpg.de/bat](http://www.mppmu.mpg.de/bat) (arXiv:0808.2552).



New position funded by alliance!

### Example:

- > Easily fit different models to the data.
- > Uncertainty estimate from Markov Chain Monte Carlo.



## Multivariate Data Analysis.

- > Two DESY fellows started recently to contribute to core development of **TMVA**: Toolkit for **Mu**lti**V**ariate data **A**nalysis.

### Core Software

- > Implement **unit testing**.

### Support Vector Machine

- > Revise/improve implementation.
- > Implement **regression** in addition to classification.

