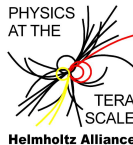


Analysis Centre: Statistics Tools Group

Gero Flucke



3rd Annual Workshop of the Helmholtz Alliance
"Physics at the Terascale"
DESY, Hamburg, November 11-13, 2009

- Who we are.
- What we are doing:
 - Education - Schools.
 - Support.
 - Tools.
- Summary and Outlook.

The Mission

Provide

- education,
- support and
- development of tools

for physics analysis within the alliance.

The People

- The **core team**: O. Behnke, G. F., C. Kleinwort, S. Schmitt (DESY), K. Kröninger (Göttingen) – only 10 – 50% FTE each.
- People contributing to schools (various institutes): Thanks!
- Since recently DESY-fellows.
- In future: studentships - you or your student?
- Network of developers of statistics related tools in Germany.

Schools/Workshops

- Combine Lectures with hands-on exercises.
- Document for non-participants and for re-use.

2009

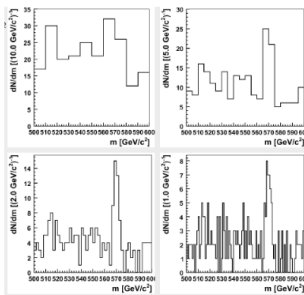
- Fitting Workshop / School (DESY, 30.3.-1.4. 2009)
<https://indico.desy.de/conferenceDisplay.py?confId=1582>
- Workshop on Advanced Methods in Statistical Data Analysis (Karlsruhe, 12.-14.10. 2009)
<https://indico.desy.de/conferenceDisplay.py?confId=1978>

Future

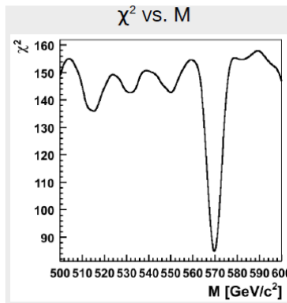
- Two schools/workshops per year.
- One about basics, one on selected topic.
- Plan 2010:
 - Basics of Statistics: DESY, Spring 2010.
 - Advanced Statistics: Göttingen, October 2010
 - Dates and topics: to be finalised.

Example: Fitting Workshop / School (DESY)

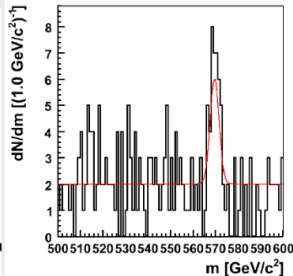
> Exercise from fitting school: Search + fit new particle with unknown mass!



Choose Binning



Signal scan



Final fit

Evaluate background
fluctuation probability: $\sim 4 \times 10^{-3}$

> Many more nice and useful fit lectures and interactive exercises:
<https://indico.desy.de/conferenceTimeTable.py?confId=1582>

Informal Statistics Meetings

- Forum for everyday's statistics problem.
- Experts discuss and answer (or follow up).
- Monthly at DESY, trying EVO for remote participation.
- Next date:
Thursday 26.11.2009, 11.00 h
- Will be announced via
anacentre-statistics@desy.de.
- Difficult format for non-Hamburg people.
- Try documentation on Analysis Centre Wiki:
www.wiki.terascale.de/index.php/Informal_meeting

⇒ User support needs to be improved!

- Statistics Software Reviews:
 - Annual meetings.
 - Connecting people working on statistics projects.
 - Exchange knowledge and progress.
- Collect links to statistics projects in Germany:
www.wiki.terascale.de/index.php/Statistics_Projects
- Collaboration in organisation of schools.
- Contact to ATLAS and CMS statistics committees:
via K. Kröninger and G. Schott, respectively.

Statistics Group involvement:

- **LVMINI**:
Efficient minimisation for large number of parameters.
- **Millepede II**:
Linear fits for huge alignment/calibration problems.
- Participation in **TMVA** development.

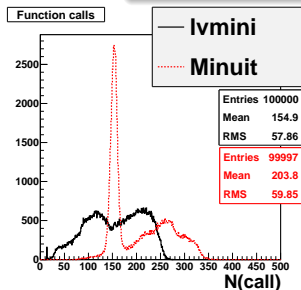
Further Statistics Projects in Germany:

- **BAT**: Bayesian Analysis Toolkit.
- **Fittino**: SUSY parameters determination.
- **Gfitter**: Generic Fitter for HEP Model Testing.
- **RooStat[Cms]**: large activity in Karlsruhe (skip here).

Overview: www.wiki.terascale.de/index.php/Statistics_Projects.

Efficient Minimisation Program (V. Blobel)

- L-BFGS algorithm:
matrix-less quasi-Newton method.
- Few or many (aiming at **up to 10^5**) parameters.
- Uncertainties: Estimate 'for free',
detailed calculation if desired (CPU increase).
- User has to provide function **value and gradient**.
- Fortran program: www.desy.de/~blobel/largesc



Status in Statistics Tools Group

- Being interfaced with ROOT.
- Preliminary benchmarking with
Rosenbrock function: less function calls.
- Goal: Integration as alternative fitting
mechanism in ROOT.

Large linear fit problems:

- Arising from minimisation with two parameter classes:

local: appearing in subsets of data (e.g. 10^7),

global: appearing in all the data ($\leq 10^6$).

- Result for global parameters only.

⇒ Mainly calibration and alignment.

Status

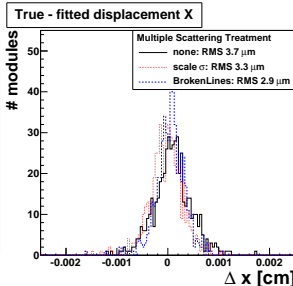
- Developed by V. Blobel.
- Maintenance by Statistics Group, see www.wiki.terascale.de/index.php/Millepede_II.
- Keep experiment independence.

Recent improvement:

Make use of band matrix structure from 'broken line' track fit.

⇒ Rigorous multiple scattering treatment in reasonable CPU time.

Toy MC ($B = 0$ T)



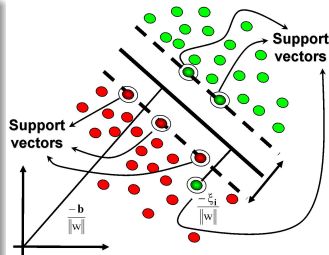
Started: Contributing to TMVA

Toolkit for Multivariate Data Analysis

- Two DESY fellows start to contribute to core development.

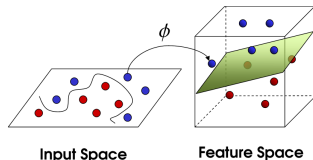
Example project: Support Vector Machine

- Build hyperplane that separates signal and background.
- Using minimal subset of all training data (support vectors).
- Non-linear SVM: map input onto a higher dimensional feature space.
- Like Neural Net, but easier to control.



Goal

- Revise/improve implementation.
- Implement regression (so far only classification).



Bayesian Analysis Toolkit

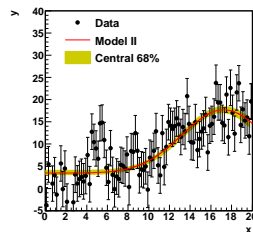
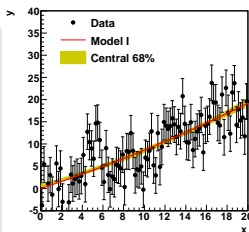
- Toolkit for data analysis.
- Based on Bayes' Theorem.
- Realised using Markov Chain Monte Carlo.
 - Parameter estimation.
- Offers
 - Uncertainty propagation.
 - Limit setting.
 - Goodness-of-fit criterion.
- www.mppmu.mpg.de/bat (arXiv:0808.2552).



New
position
funded by
alliance!

Example

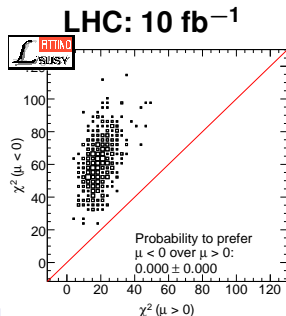
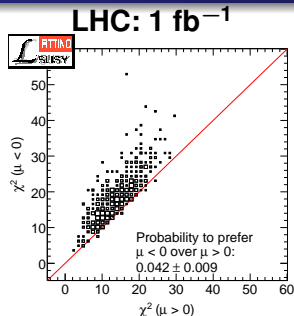
- Easily fit different models to the data.
- Uncertainty estimation from Markov Chain.



- Tool for fitting SUSY Parameters.
- Using various observables:
 - masses and widths of (s)particles,
 - edges in mass spectra,
 - branching fractions and cross-sections,
 - low energy observables (e.g. $(g-2)_\mu$),
 - cosmological data (Ω).
- Sophisticated fitting mechanisms.
- www-flc.desy.de/fittino

Physics Example at LHC

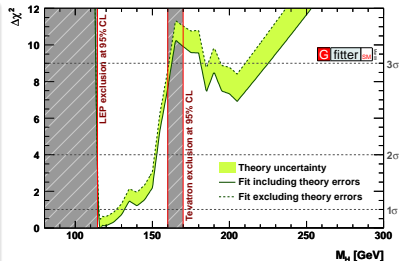
- mSUGRA: Determination of $sign(\mu)$.
- Fit toy MC:
probability to prefer wrong result.
- 4% with $L = 1 \text{ fb}^{-1}$, 0 with $L = 10 \text{ fb}^{-1}$.
- Much more in [arxiv:0907.2589](https://arxiv.org/abs/0907.2589),
Plenary talk: P. Wienemann (U. Bonn)



- Developed at Hamburg (Uni/DESY), collaborating with CERN.
- C++ code based on ROOT.
- **Core packages** for data handling, fitting, statistical analysis.
- **Physics plug-in packages**:
 - GSM: Standard model fit: EW precision data
 - G2HDM: 2HDM extension of SM
 - GOblique: BSM constraints from oblique parameters
 - GSUSY and more: to come...

Example Result:

- Combination of EW fit and direct Higgs searches.
- More results and documentation:
 - gfitter.desy.de
 - www.cern.ch/gfitter
 - Talk in QCD & EW session:
M. Goebel (U. Hamburg)



Summary

- Core at DESY and Göttingen, limited manpower in FTEs.
- Mission: education, support, tools for analysis.
- Lots of effort for schools/workshops.
- User support to be improved:
Informal statistics meetings appropriate format?
- Tools:
 - Explicit work started recently: Millepede II, LVMINI.
 - Loose network of projects in Germany.

Outlook

- Statistics Schools 2010.
- Improving user support.
- Work on tools gets full speed:
TMVA, (further) fitting tools,...