Data re-interpretation: an user perspective

Sabine Kraml
LPSC Grenoble

Marseille Workshop on Scientific Data Preservation
19-21 November 2012 *Centre de Physique des Particules de Marseille*

Motivation

- LHC is the high energy frontier machine to explore the TeV scale and provide answers to many key questions in particle physics.

  ➡ Search for the Higgs boson

  ➡ Search for New Physics beyond the Standard Model

- Need to interpret LHC results in the contexts of all kinds of models of new physics; crucial if we are to unravel the correct theory and determine its parameters.

- The complexity of a) the experimental analyses and b) the possible new physics models requires active collaboration of experimentalists and theorists —the whole HEP community— to fully exploit the LHC potential.

- Makes persistence and long-term use(ability) of LHC results extremely important

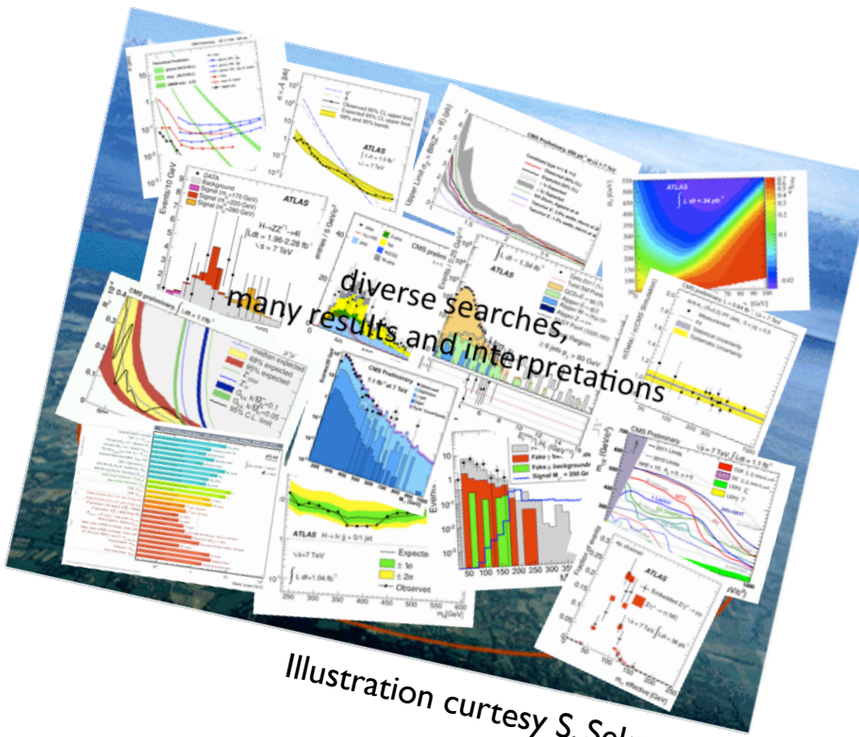


Illustration curtesy S. Sekmen

Example: Higgs search

- The discovery of the Higgs boson is a tremendous first success for the LHC experimental program

- Next: need to determine whether it is a SM Higgs (and only the SM Higgs)

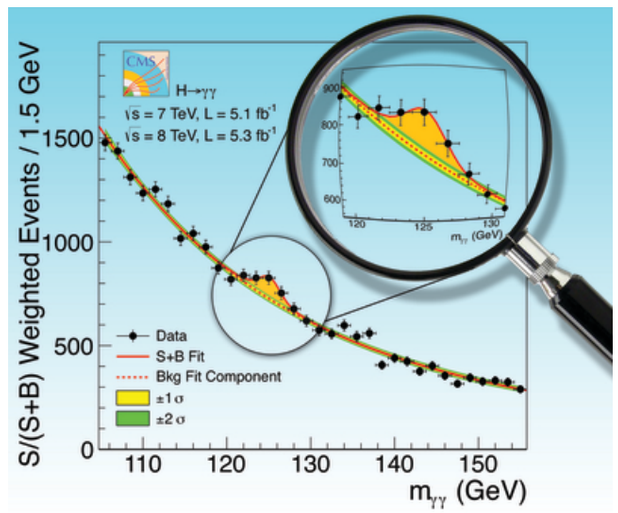
  - ➡ is it the SM Higgs?

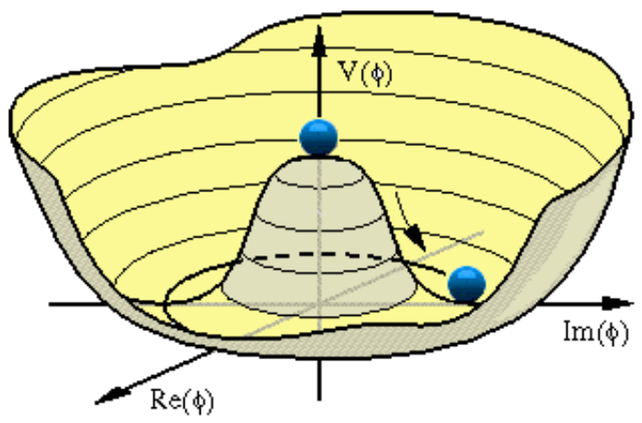
  - ➡ is it fully responsible for electroweak symmetry breaking?

  - ➡ is there more than one Higgs?
    (contributing to the 125 GeV signal / to EWSB?)

- Precise measurements in a variety of production and decay modes.

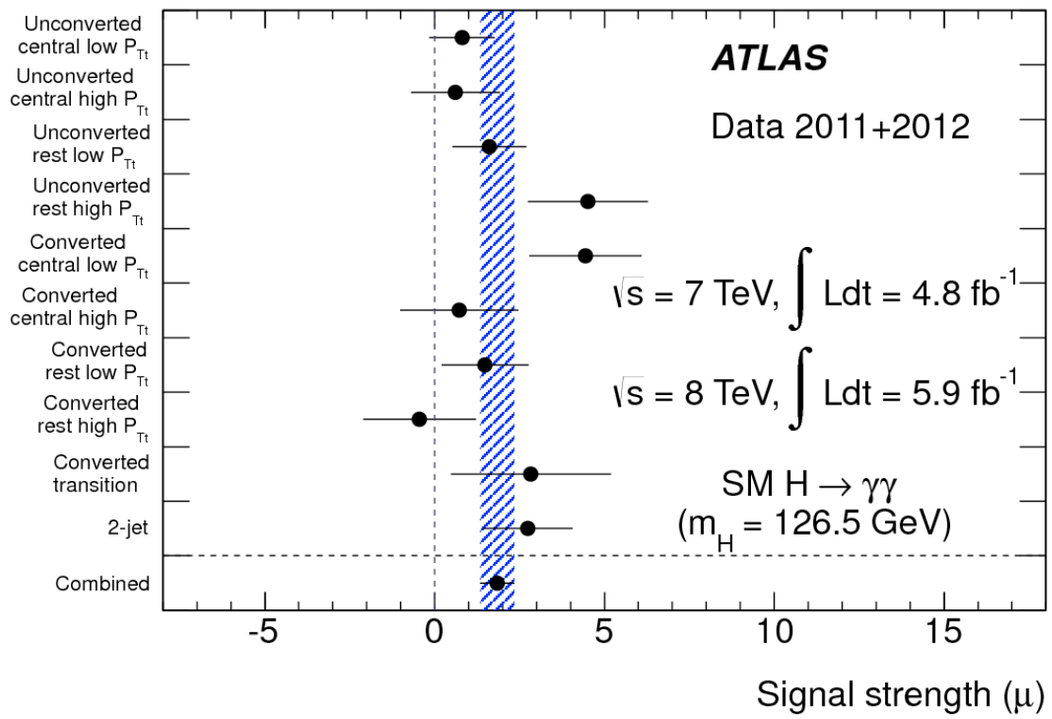
- Fits and tests of various models; need to be able to put all information together.

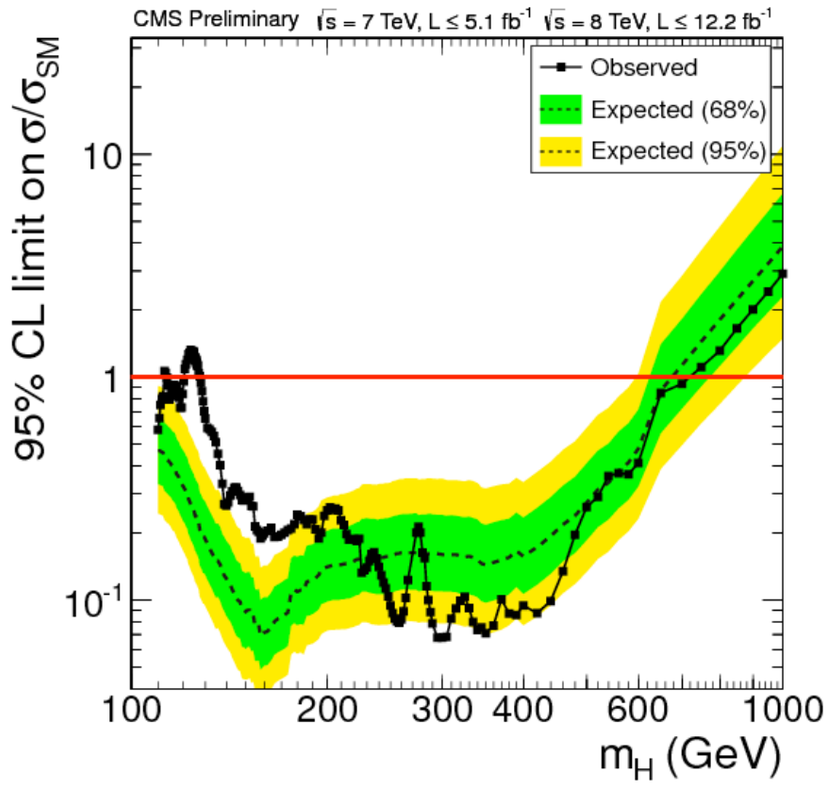




Is the Higgs mechanism as simple as envisaged in the SM?

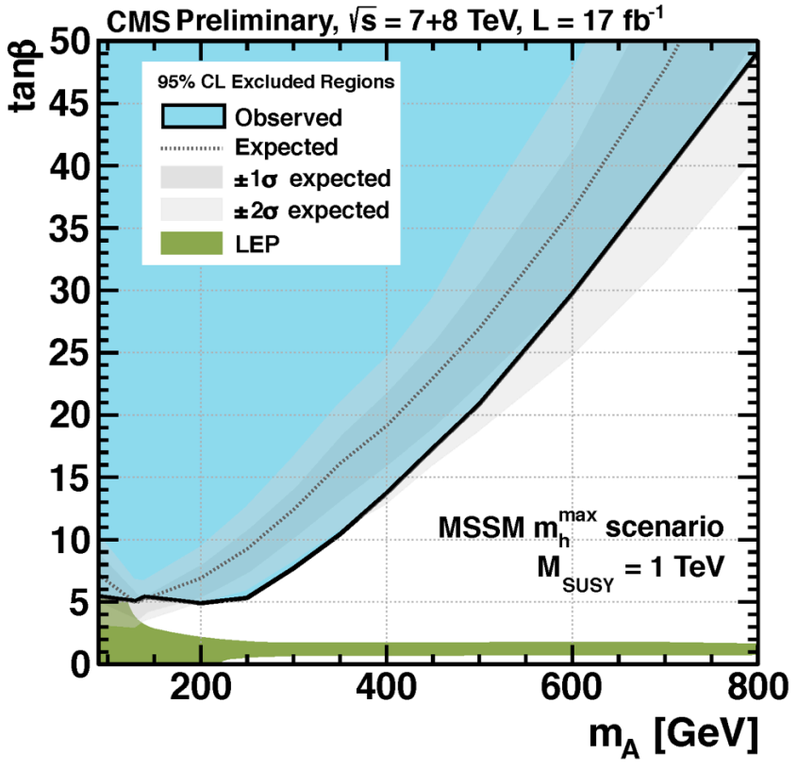
Need to (be able to) put all information together





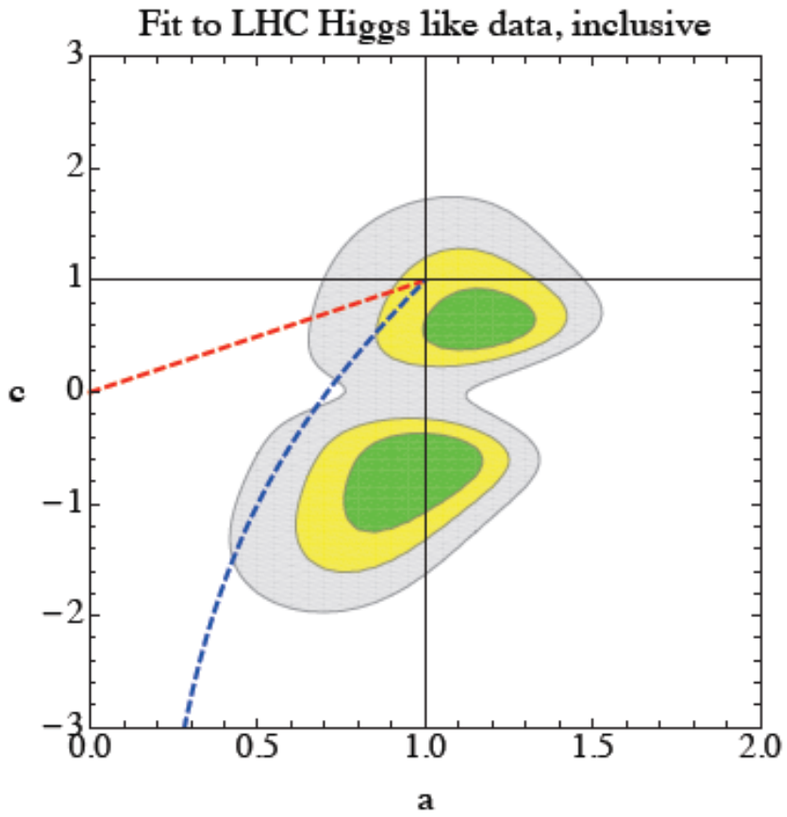
Detailed channel-by-channel information (separated into production and decays) is necessary in order to test non-standard Higgs scenarios.

NB experiments assume SM composition of production modes; this may easily differ in BSM models → need also event rate information etc to recast analyses.

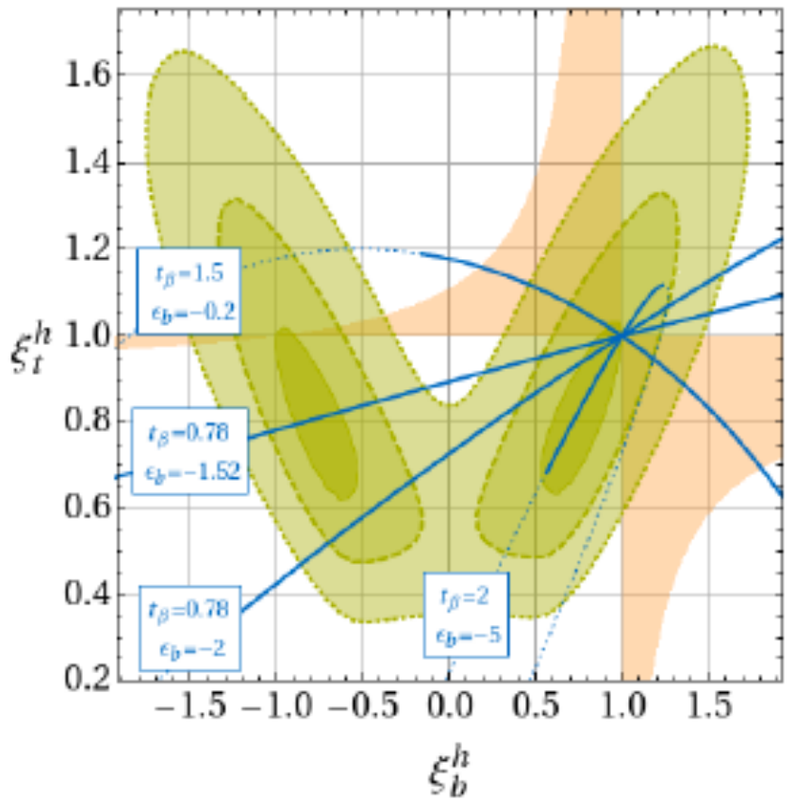


Theorists perform fits with different parameterizations of deviations from SM couplings and/or non-SM contributions

Effective Lagrangian, arXiv:1202.3697



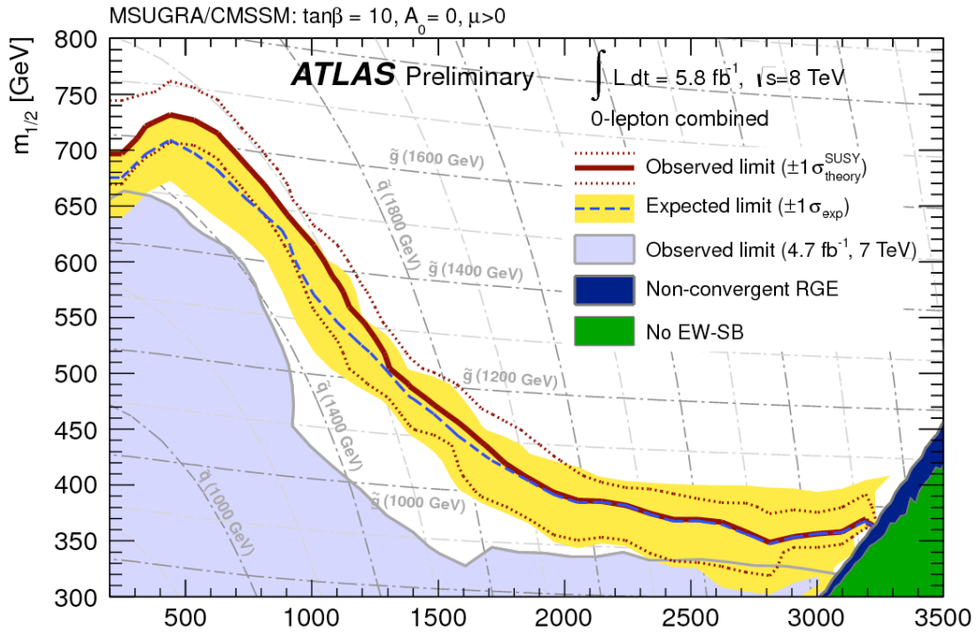
2HDM, arXiv:1210.2465

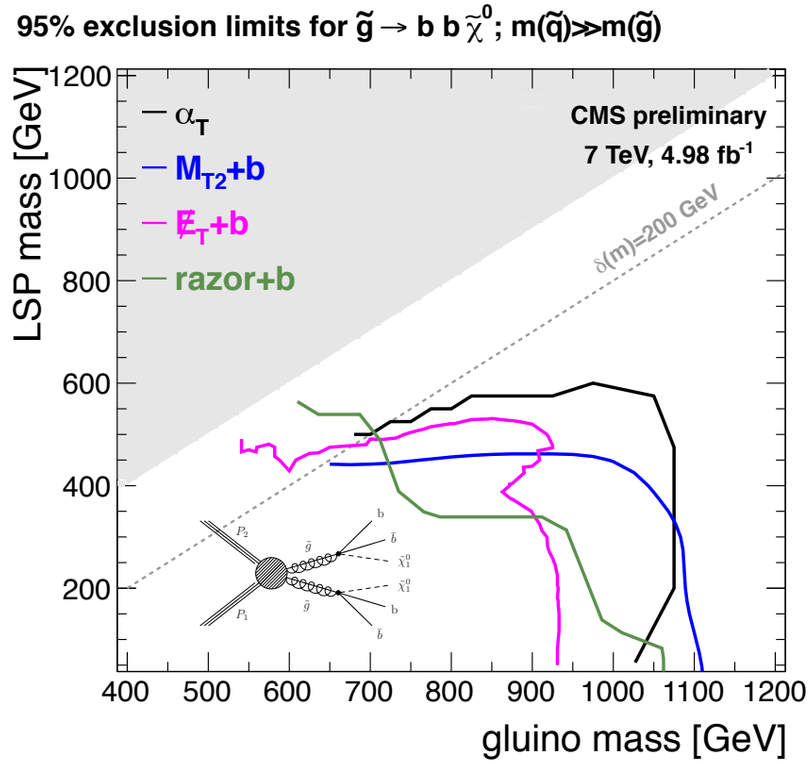


just 2 examples, lots of papers appearing on this topic .....

New Physics (BSM) searches

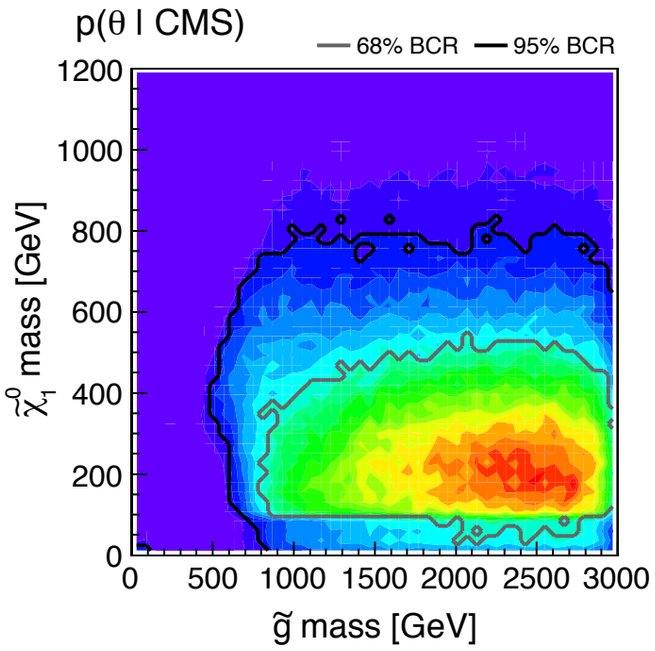
- ATLAS and CMS perform searches for new physics in many different channels.
- The collaborations typically interpret their results within constrained models, e.g. the CMSSM, or within topology-based "Simplified Models" (SMSs).
- However, constrained models and SMSs always have specific assumptions built in (mass ratios, branching fractions, etc).
- SUSY (and BSM in general) has much larger variety of signatures.
- Need to interpret LHC results in the contexts of all kinds of models of new physics; crucial if we are to unravel the correct theory and determine its parameters ⇨ community-wide effort !

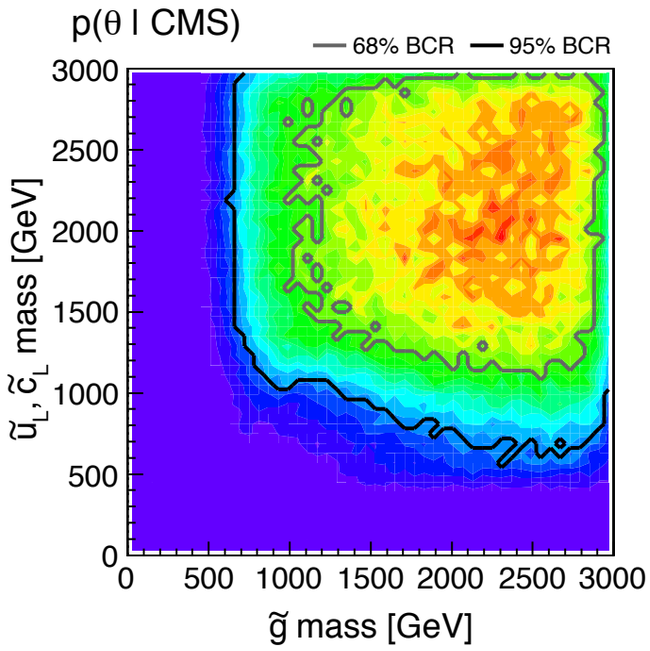


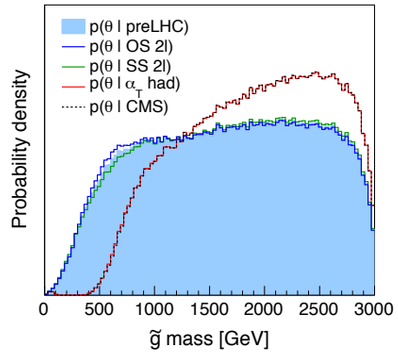


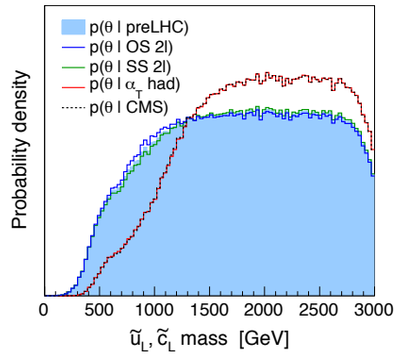
Example: phenomenological MSSM

- In arXiv:1109.5119, we interpreted the results of SUSY searches published by the CMS collaboration based on the first 1 fb-1 of data taken during the 2011 LHC run at 7 TeV within the phenomenological MSSM (pMSSM).
- The pMSSM is a 19-dimensional parametrization of the MSSM that captures most of its phenomenological features. It encompasses and goes beyond, a broad range of more constrained SUSY models.
- This allowed us to obtain more generic conclusions on how the current data constrain the MSSM.









What we would like to do (e.g.)

- Many pMSSM scenarios that escape current searches have large production cross section but suffer from low signal significance because of small mass splittings, low missing energy, etc
  - "EW-inos": superpartners of electroweak gauge and Higgs bosons
  - "Natural SUSY" : light stops, rest of SUSY heavy
- Analyses at 8 TeV not necessarily more performant than those at 7 TeV because of harder cuts.

➡ Want to

- combine standard SUSY searches (e.g. trileptons) with non-MET searches (e.g. RPV SUSY) and with Higgs searches (in particular leptonic channels)
- devise new analyses to improve sensitivity and re-assess limits

.... as well as increase discovery potential

Interpretation of LHC results

- The complexity of a) the experimental analyses and b) the possible new physics models requires active collaboration of experimentalists and theorists —the whole HEP community— to fully exploit the LHC potential.

- A common standard for the information to provide would immensely help this task. (it would actually help not only the interpretation of results but also comparisons within/ across experiments, data preservation efforts, etc, etc)

- Besides our own (physics) interest in making the most out of the LHC data, we may soon be seriously mandated by the funding agencies to work much more openly towards this aim ...

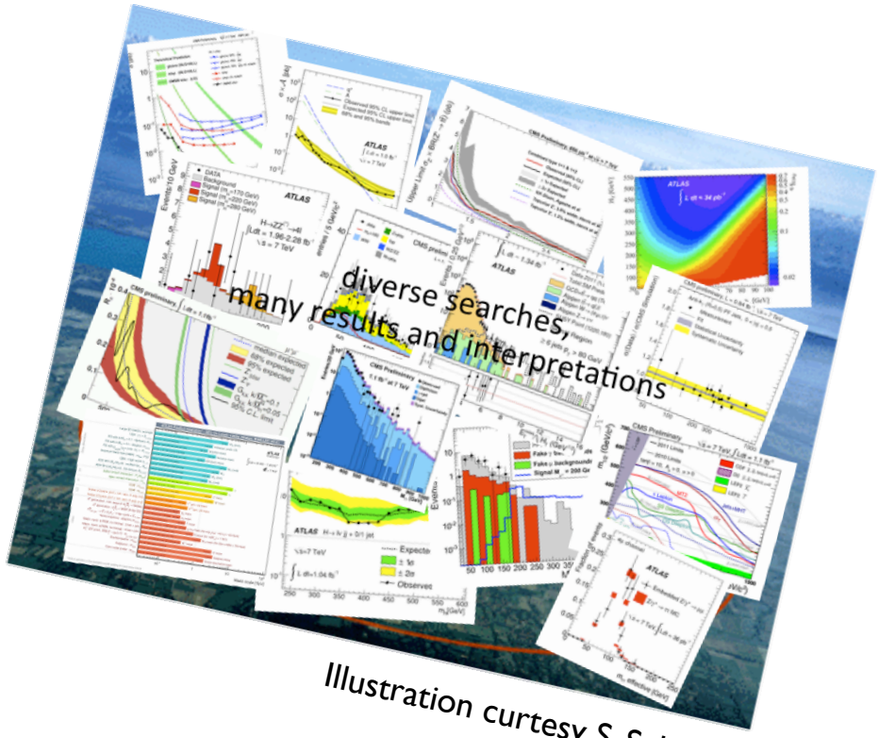


Illustration curtesy S. Sekmen

THE EUROPEAN PHYSICAL JOURNAL C - PARTICLES AND FIELDS
Volume 72, Number 4 (2012), 1976

SPECIAL ARTICLE - TOOLS FOR EXPERIMENT AND THEORY

arXiv:1203:2489

Searches for New Physics: Les Houches Recommendations for the Presentation of LHC Results

S. Kraml[1], B.C. Allanach[2], M. Mangano[3], H.B. Prosper[4], S. Sekmen[3,4] (editors),
C. Balazs[5], A. Barr[6], P. Bechtle[7], G. Belanger[8], A. Belyaev[9,10], K. Benslama[11],
M. Campanelli[12], K. Cranmer[13], A. De Roeck[3], M.J. Dolan[14], T. Eifert[15], J.R. Ellis[16,3],
M. Felcini[17], B. Fuks[18], D. Guadagnoli[8,19], J.F. Gunion[20], S. Heinemeyer[17],
J. Hewett[15], A. Ismail[15], M. Kadastik[21], M. Krämer[22], J. Lykken[23] F. Mahmoudi[3,24],
S.P. Martin[25,26,27], T. Rizzo[15], T. Robens[28], M. Tytgat[29], A. Weiler[30]

Abstract

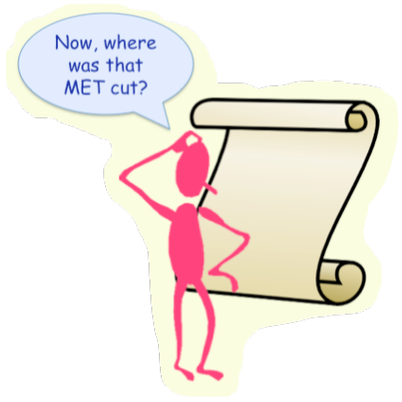
We present a set of recommendations for the presentation of LHC results on searches for new physics, which are aimed at providing a more efficient flow of scientific information between the experimental collaborations and the rest of the high energy physics community, and at facilitating the interpretation of the results in a wide class of models. Implementing these recommendations would aid the full exploitation of the physics potential of the LHC.

Nature and categories

- Analysis description
  - a. Clear, explicit & complete description of the analysis
  - b. Common analysis database (analysis codes)

a. "mandatory"
b. "desirable"

- Detector modeling
  - a. Efficiency maps
  - b. Public fast detector simulator



- Analysis dissemination
  - a. Crucial numbers of results
  - b. Full likelihood function (analytic and/or numerical form)

- Interpretation of BSM search results confidence levels, etc ....
- Higgs searches channel-by-channel information
- Analysis design disjoint sets of events

LH recommendations - some remarks

- In the Recommendations, we think it useful to clearly distinguish between
  - ★ experimental result — whatever is actually observed, i.e. the outcome of an analysis, such as event count or the measurement of a physical observable,
  - ★ and interpretation — the comparison of the experimental results to particular theoretical models
- Many of the experimental publications already implement several of the basic recommendations ⇨ work towards an agreement on a **common standard**.
- The sum of our recommendations goes substantially beyond current practice.
- Useful not only for non-collaboration groups or individuals performing (re-)interpretation studies; a common standard will also greatly facilitate the comparison and combination of analyses within and across the LHC collaborations, and help long-term data preservation efforts.
- Recommendations focus on what information should be provided, not how this should be done.

1. Analysis description

Recommendation 1a: Provide a clear, explicit description of the analysis in publications. In particular, the most crucial information such as basic object definitions and event selection should be clearly displayed in the publications, preferably in tabular form, and kinematic variables utilised should be unambiguously defined. *Further information necessary to reproduce the analysis should be provided, as soon as it becomes available for release, on a suitable common platform.*

Rivet, HEPdata, ...

Recommendation 1b: The community should identify, develop and adopt a common platform to store analysis databases, collecting object definitions, cuts, and all other information, including well-encapsulated functions, necessary to reproduce or use the results of the analyses, and as required by other recommendations.

1. Analysis description

Recommendation 1b: The community should identify, develop and adopt a common platform to store analysis databases, collecting object definitions, cuts, and all other information, including well-encapsulated functions, necessary to reproduce or use the results of the analyses, and as required by other recommendations.

Comments:

- The analysis database should also be capable of storing any analysis-related software that may be provided alongside the analysis.
- Rivet and HEPdata provide examples of such a platform, possibly supported by the inSPIRE indexing and searching infrastructure. Their functionality could be adapted to accommodate further needs.
- Phenomenologists' approach towards 1b: common platform of analysis codes by users of experimental results.

2. Detector modeling

Recommendation 2a: Provide histograms or functional forms of efficiency maps wherever possible in the auxiliary information, along with precise definitions of the efficiencies, and preferably provide them in standard electronic forms that can easily be interfaced with simulation or analysis software.

Recommendation 2b: The community should take responsibility for providing, validating and maintaing a simplified simulation code for public use, reproducing the basic response of the LHC detectors. The validation and tuning of this tool should be based on comparisons with actual performance plots, and/or other inputs, made available by the experiments along the lines of Recommendation 2a. Limits of validity should be investigated and clearly documented.

3. Analysis dissemination

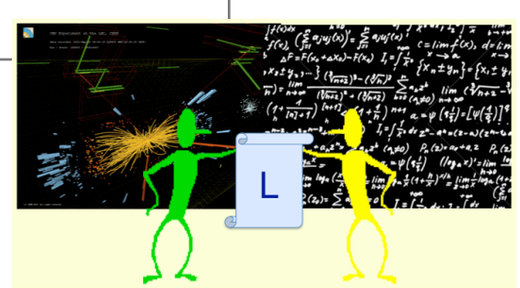
Recommendation 3a: Provide all crucial numbers regarding the results of the analysis, preferably in tabulated form in the publication itself. Further relevant information, like fit functions or distributions, should be provided as auxiliary material.

Addendum: [...] Results should be quoted without inclusion of systematic/theoretical uncertainties external to the experiment.

Towards publishing likelihoods

Recommendation 3b: When feasible, provide a mathematical description of the final likelihood function in which experimental data and parameters are clearly distinguished, either in the publication or the auxiliary information. Limits of validity should always be clearly specified.

Recommendation 3c: Additionally provide a digitized implementation of the likelihood that is consistent with the mathematical description.



4. Interpretation of results

So far our recommendations concern generally the presentation of experimental results, irrespective of whether they report a signal or are used to set exclusion limits.

Let us now turn to the interpretation of these results, the presentation of confidence intervals, parameter inference and limit setting in particular models:

Recommendation 4: *In the interpretation of experimental results, preferably provide the final likelihood function (following Recommendations 3b/3c).* When this is not possible or desirable, provide a grid of confidence levels over the parameter space. The expected constraints should be given in addition to the observed ones, and whatever sensitivity measure is applied must be precisely defined. Modeling of the acceptance needs to be precisely described.

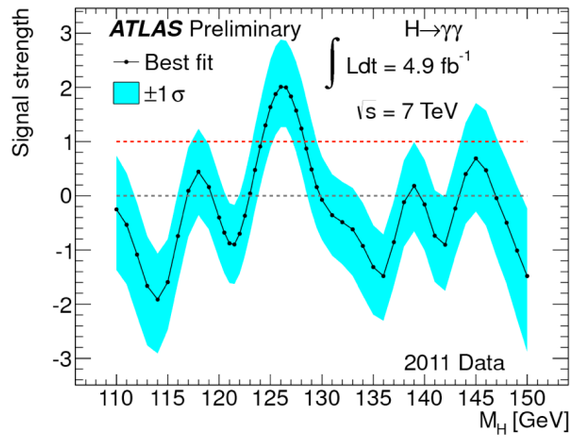
NB this applies equally to phenomenologists' interpretation studies as to interpretations of results in experimental papers.

5. Higgs searches

Recommendation 5: For Higgs searches, provide all relevant information on a channel-by-channel basis for both production and decay processes.

NB this is crucial in the context of multiple or composite Higgs boson models! Indeed, different Higgs models weight various possible production mechanism and decay distributions differently.

It is moreover very instructive to give the best-fit signal strengths as function of the SM Higgs boson mass for all available channels, along with error bands, as this facilitates testing deviations from SM couplings.



6. Analysis design

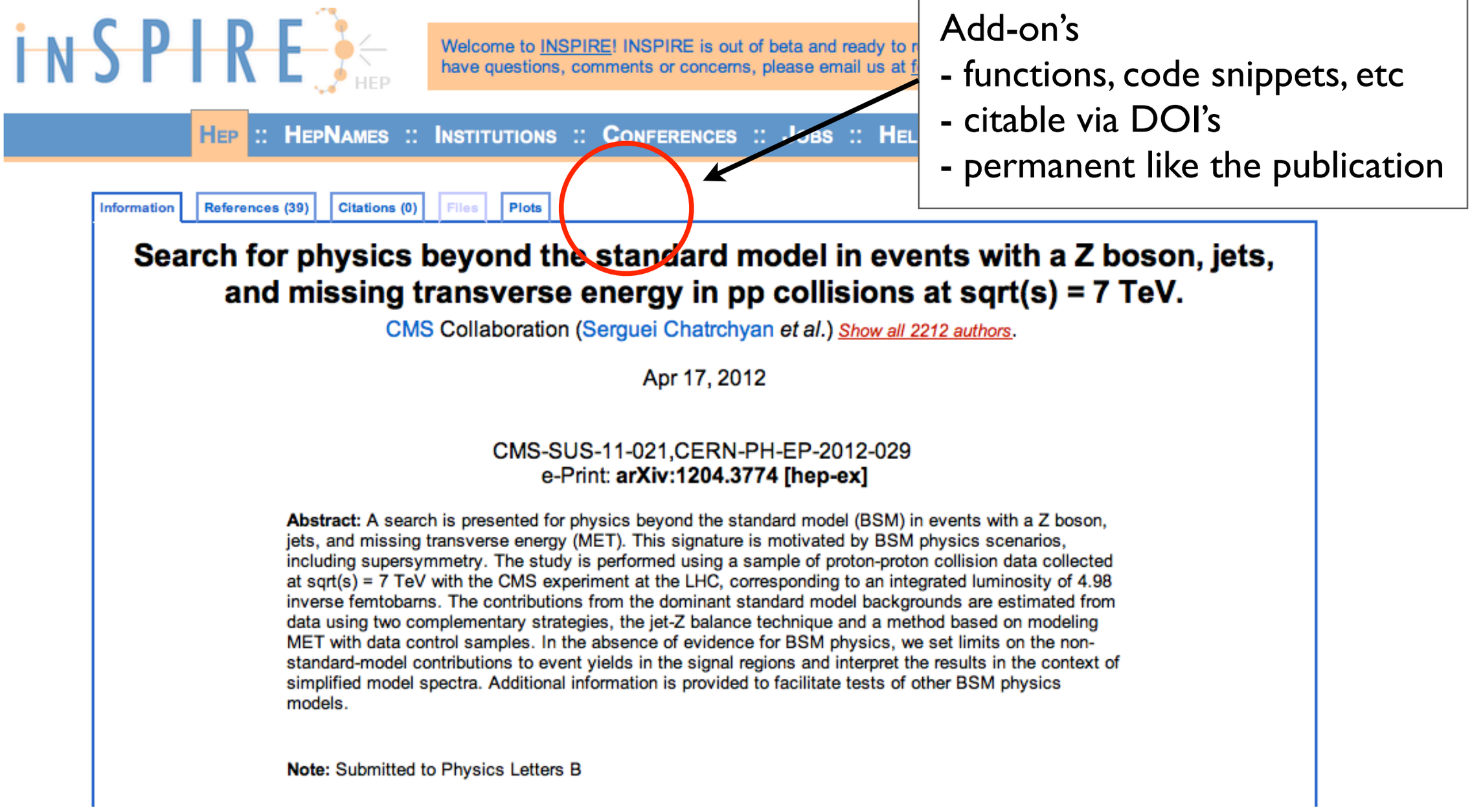
Recommendation 6: When relevant, design analyses and signal regions that are based on disjoint sets of events.

Conclusions

- In order to fully exploit the LHC physics potential, we need
  - ✓ to be able to (re-)interpret LHC data in the contexts of the broadest possible range of theoretical scenarios (cf. Les Houches recommendations)
  - ✓ a comprehensive approach to the storage, persistence and future use of LHC results.
- Work towards a common standard for presentation/preservation of results.
- Added value for the experiments, and the community as a whole:
  - ✓ faster and more precise feedback on the implications of the LHC results.
  - ✓ greatly facilitate the comparison and combination of analyses within and across the LHC collaborations, as well as the assessment of the physics potential of future facilities.
  - ✓ possibility to re-assess results in view of new discoveries.
- The tools needed to provide extended experimental information will require some dedicated efforts in terms of resources and manpower, to be supported by both the experimental and the theory communities.

Next step:
practical solutions

Discussion of extended inSPIRE services



Add-on's
- functions, code snippets, etc
- citable via DOI's
- permanent like the publication

Contacts: Till Eifert (ATLAS), Sezen Sekmen (CMS), Salvatore Mele (inSPIRE), M. Mangano and I

Fast Simulators for the LHC

11-12 June 2012 *CERN*
Europe/Zurich timezone

Search

Overview
Timetable
Registration
Registration Form
List of registrants
LH Recommendations

Dates: from 11 June 2012 09:00 to 12 June 2012 18:00
Timezone: Europe/Zurich
Location: *CERN*
Room: TH Conference Room
Chairs: Mangano, Michelangelo
Kraml, Sabine
Sekmen, Sezen

Additional info: This workshop has been motivated by the recently published "Les Houches Recommendations for the presentation of LHC results", arXiv:1203.2489, which emphasize the important role of public fast detector simulators in maximizing the use of LHC results, and suggest the HEP community to take responsibility for providing, validating and maintaining tools for fast simulation.

The workshop aims to bring together the developers of the existing and upcoming tools, the experts from experiments, and the current and potential users in order to thoroughly discuss fast simulators, and address topics such as:

- current status and shortcomings

- object implementation, difficult topologies

- validation

- input/output formats, common analysis tools

Workshop on publishing likelihoods
(for LHC results)

21-23 Jan 2013 at CERN