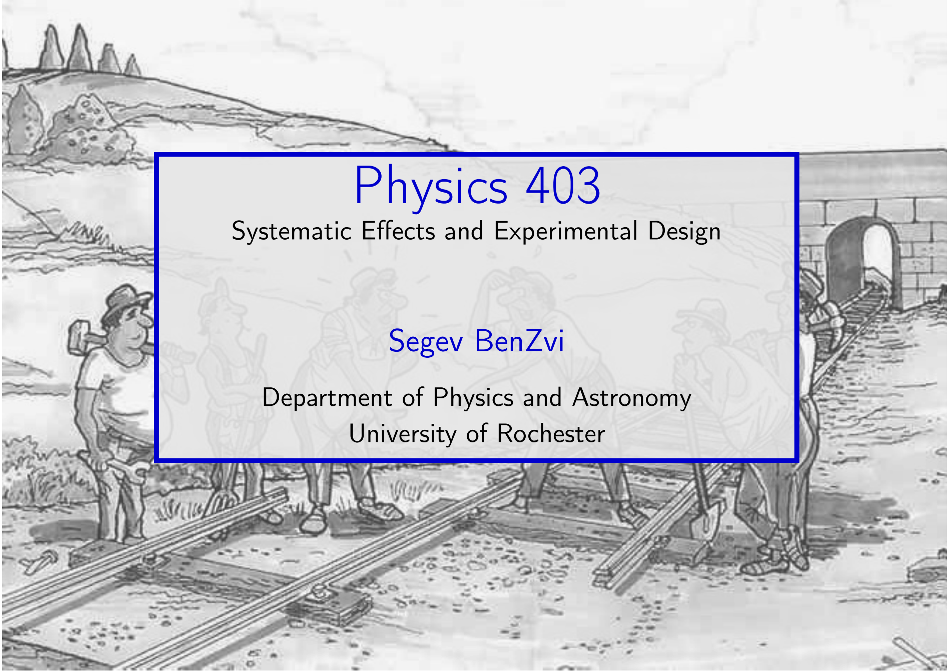


Physics 403

Systematic Effects and Experimental Design

Segev BenZvi

Department of Physics and Astronomy
University of Rochester

Statistical Uncertainties

Frequentist: how much would a result fluctuate upon repetition of the measurement? Implies some knowledge/assumptions about PDFs...

Example

We count photons → Poisson distribution. For $N = 150$, $N \pm \sqrt{N} \approx 150 \pm 12$. Note: we assumed $\hat{N} = N$.

Example

Time from a digital clock → Uniform distribution. E.g., $t = 23$ s with 1 s resolution, $\text{var}(t) = (b-a)^2/12 = 1/12$, so $t = 23 \pm 0.3$ s

Example

Efficiency of a detector → Binomial distribution. Record 45 out of 60 particles: $\hat{\epsilon} = 45/60 = 0.75$, $\text{var}(\hat{\epsilon}) = \epsilon(1-\epsilon)/N$, so $\epsilon = 0.75 \pm 0.06$

Statistical vs. Systematic Uncertainties

So what is a systematic uncertainty?

- "Systematic error: reproducible inaccuracy introduced by faulty equipment, calibration, or technique." [1]
- Who agrees with this?

Statistical vs. Systematic Uncertainties

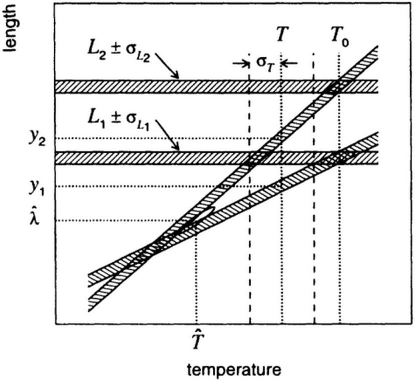
So what is a systematic uncertainty?

- "Systematic error: reproducible inaccuracy introduced by faulty equipment, calibration, or technique." [1]
- Who agrees with this?
- A neglected effect is a mistake, not an uncertainty [2]
- Some confusion is caused by the term systematic *error*, because an "error" in common language means a mistake (implying fault or incompetence), while we mean an uncertainty

  Systematic effects is a general category which includes effects such as background, selection bias, scanning efficiency, energy resolution, angle resolution, variation of counter efficiency with beam position and energy, dead time, etc. [3]

Usually expressed in form like this $A = 10.2 \pm 1.2\ \text{(stat)} \pm 2.3\ \text{(sys)}$

Estimating Systematics using Data



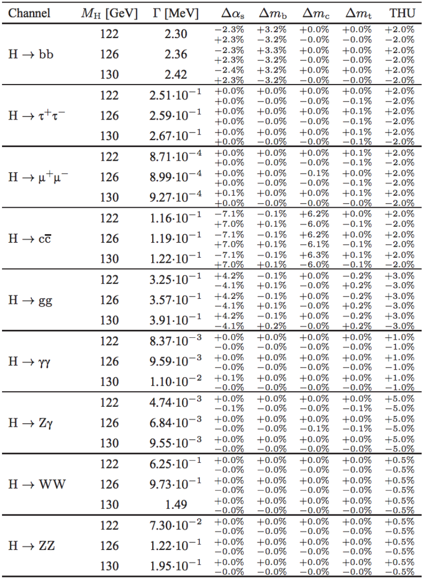
Lengths measured by two metal rulers at different temperatures [4]

- It's common to think of a systematic as something that affects all of your data equally, but this need not be the case
- Recall the example of two thermally expanding rulers from Cowan [4]:

$$y_i = L_i + c_i(T - T_0)$$

- The intersection of two lines from the data indicate a systematic offset in temperature ΔT
- This offset will affect different parts of the data in different ways

Parameter and "Theory" Uncertainties



Channel	M_H [GeV]	Γ [MeV]	$\Delta\alpha_s$	Δm_b	Δm_c	Δm_t	THU
$H \to bb$	122	2.30	−2.3% +2.3%	+3.2% −3.2%	+0.0% −0.0%	+0.0% −0.0%	+2.0% −2.0%
	126	2.36	−2.3% +2.3%	+3.3% −3.2%	+0.0% −0.0%	+0.0% −0.0%	+2.0% −2.0%
	130	2.42	−2.4% +2.3%	+3.2% −3.2%	+0.0% −0.0%	+0.0% −0.0%	+2.0% −2.0%
$H \to \tau^+\tau^-$	122	$2.51 \cdot 10^{-1}$	+0.0% +0.0%	+0.0% −0.0%	+0.0% −0.0%	+0.0% −0.1%	+2.0% −2.0%
	126	$2.59 \cdot 10^{-1}$	+0.0% +0.0%	+0.0% −0.0%	+0.0% −0.0%	+0.1% −0.1%	+2.0% −2.0%
	130	$2.67 \cdot 10^{-1}$	+0.0% +0.0%	+0.0% −0.0%	+0.0% −0.0%	+0.1% −0.1%	+2.0% −2.0%
$H \to \mu^+\mu^-$	122	$8.71 \cdot 10^{-4}$	+0.0% +0.0%	+0.0% −0.0%	+0.0% −0.0%	+0.1% −0.1%	+2.0% −2.0%
	126	$8.99 \cdot 10^{-4}$	+0.0% +0.0%	+0.0% −0.0%	−0.1% −0.0%	+0.0% −0.1%	+2.0% −2.0%
	130	$9.27 \cdot 10^{-4}$	+0.1% +0.0%	+0.0% −0.0%	+0.0% −0.0%	+0.1% −0.0%	+2.0% −2.0%
$H \to c\bar{c}$	122	$1.16 \cdot 10^{-1}$	−7.1% +7.0%	−0.1% +0.1%	+6.2% −6.0%	+0.0% −0.1%	+2.0% −2.0%
	126	$1.19 \cdot 10^{-1}$	−7.1% +7.0%	−0.1% +0.1%	+6.2% −6.1%	+0.0% −0.1%	+2.0% −2.0%
	130	$1.22 \cdot 10^{-1}$	−7.1% +7.0%	−0.1% +0.1%	+6.3% −6.0%	+0.1% −0.1%	+2.0% −2.0%
$H \to gg$	122	$3.25 \cdot 10^{-1}$	+4.2% −4.1%	−0.1% +0.1%	+0.0% −0.0%	−0.2% +0.2%	+3.0% −3.0%
	126	$3.57 \cdot 10^{-1}$	+4.2% −4.1%	−0.1% +0.1%	+0.0% −0.0%	−0.2% +0.2%	+3.0% −3.0%
	130	$3.91 \cdot 10^{-1}$	+4.2% −4.1%	−0.1% +0.2%	+0.0% −0.0%	−0.2% +0.2%	+3.0% −3.0%
$H \to \gamma\gamma$	122	$8.37 \cdot 10^{-3}$	+0.0% −0.0%	+0.0% −0.0%	+0.0% −0.0%	+0.0% −0.0%	+1.0% −1.0%
	126	$9.59 \cdot 10^{-3}$	+0.0% −0.0%	+0.0% −0.0%	+0.0% −0.0%	+0.0% −0.0%	+1.0% −1.0%
	130	$1.10 \cdot 10^{-2}$	+0.1% −0.0%	+0.0% −0.0%	+0.0% −0.0%	+0.0% −0.0%	+1.0% −1.0%
$H \to Z\gamma$	122	$4.74 \cdot 10^{-3}$	+0.0% −0.1%	+0.0% −0.0%	+0.0% −0.0%	+0.0% −0.1%	+5.0% −5.0%
	126	$6.84 \cdot 10^{-3}$	+0.0% −0.0%	+0.0% −0.0%	+0.0% −0.1%	+0.0% −0.1%	+5.0% −5.0%
	130	$9.55 \cdot 10^{-3}$	+0.0% −0.0%	+0.0% −0.0%	+0.0% −0.0%	+0.0% −0.0%	+5.0% −5.0%
$H \to WW$	122	$6.25 \cdot 10^{-1}$	+0.0% −0.0%	+0.0% −0.0%	+0.0% −0.0%	+0.0% −0.0%	+0.5% −0.5%
	126	$9.73 \cdot 10^{-1}$	+0.0% −0.0%	+0.0% −0.0%	+0.0% −0.0%	+0.0% −0.0%	+0.5% −0.5%
	130	1.49	+0.0% −0.0%	+0.0% −0.0%	+0.0% −0.0%	+0.0% −0.0%	+0.5% −0.5%
$H \to ZZ$	122	$7.30 \cdot 10^{-2}$	+0.0% −0.0%	+0.0% −0.0%	+0.0% −0.0%	+0.0% −0.0%	+0.5% −0.5%
	126	$1.22 \cdot 10^{-1}$	+0.0% −0.0%	+0.0% −0.0%	+0.0% −0.0%	+0.0% −0.0%	+0.5% −0.5%
	130	$1.95 \cdot 10^{-1}$	+0.0% −0.0%	+0.0% −0.0%	+0.0% −0.0%	+0.0% −0.0%	+0.5% −0.5%

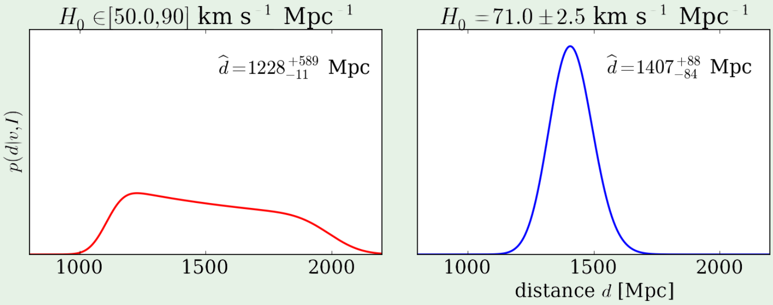
- When we repeat a measurement with a systematic uncertainty, we expect to get the same result each time
- If the uncertainty is due to measured parameters or theory, then try to improve those measurements and calculations
- Example: Higgs partial widths and % errors due to α_s, m_b, m_c, m_t, theoretical uncertainties [5]
- Improve by going to higher order in perturbative QCD, reduce spacing in lattice QCD simulations, etc.

The Bayesian Viewpoint

In the Bayesian framework, since uncertainties reflect degree of belief rather than just the spread of repeat measurements, it's straightforward to incorporate a "parameter uncertainty":

Example

Recall: distance to a galaxy given recession velocity v with uncertainties in H_0. Must select a prior on H_0 (uniform, Gaussian, ...):



$H_0 \in [50.0, 90]$ km s^{-1} Mpc^{-1}

$H_0 = 71.0 \pm 2.5$ km s^{-1} Mpc^{-1}

Systematic Uncertainties and Bias

We defined bias earlier in the course and equated it with systematic uncertainty. This is true in the case where we know there is a bias but its exact size is unknown. But there are other possibilities:

1. Bias is **known**, with **known** size; so we correct for it. Not a systematic
2. Bias is **known**, but exact size is **unknown**. This is a systematic uncertainty
3. Bias is **unknown** and **unsuspected**. Nothing to be done.

"Known knowns, known unknowns, and unknown unknowns..."

Example

Have measured lengths from an expanding steel ruler but don't know T when data were taken.

Handling the Unknown Unknowns

What makes systematic uncertainties scary:

- If you are unaware of a systematic effect in your data, you can get internally consistent results with an impressive χ^2 goodness-of-fit and still be completely wrong
- Barlow on systematics: "You never know whether you have got them and can never be sure that you have not – like an insidious disease." [6]
- The best scientists seem to have a sixth sense for picking out systematic effects and estimating their importance. There is an element of black magic to it...

Identifying Systematic Effects

If you're not a wizard, don't despair! Here are some things people do to try to catch systematic effects [2]:

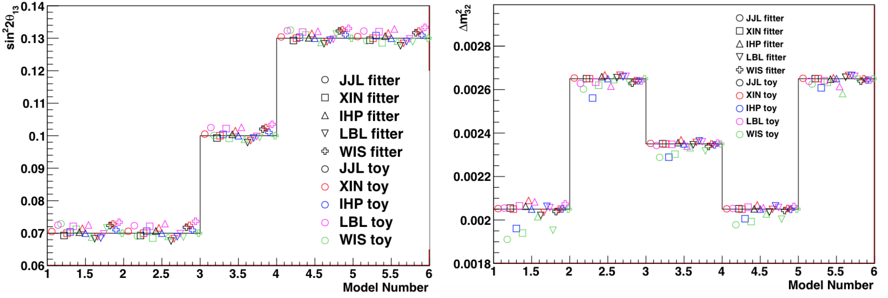
- Split the data into subsets and analyze them separately
- Vary cuts, bin sizes, etc. and explore the effect on the results
- Change parameterizations or fit techniques
- Perform independent analyses and check differences in outcomes

With experience you can learn how to identify and reduce systematic effects using these techniques

Remember, this is not a recipe. We must account for systematics if we want to publish believable deviations from expectations... but your time and resources are finite, so it's also important to understand when to cut your losses

Example: Data Challenge using Known Inputs

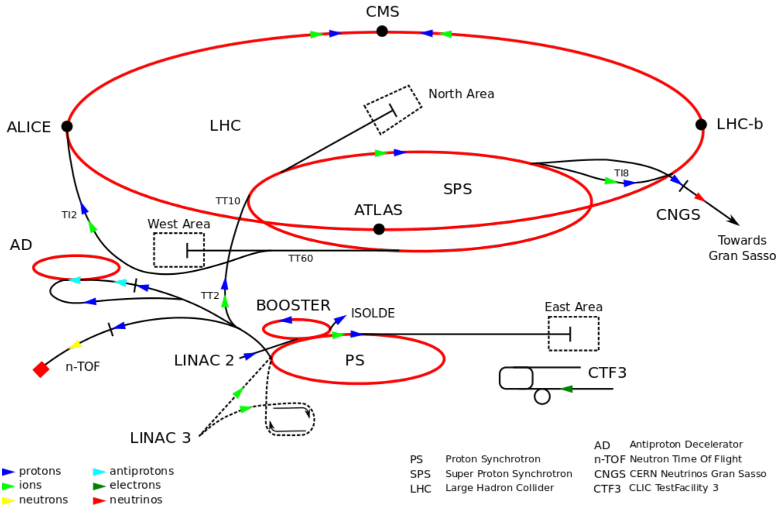
A straightforward way to pick out biases in an analysis is to perform a data challenge. Create a simulated dataset with known inputs and see if the inputs are recovered with or without the introduction of a bias



Above: recovered values of $\sin^2 2\theta_{13}$ and Δm^2_{32} from several different toy simulations and fitters in Daya Bay data challenge. From [7]

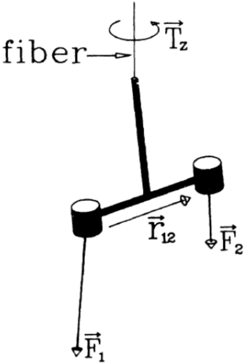
Redundancy in Analysis

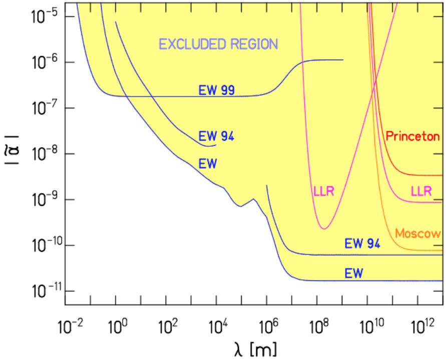
If you can afford it, build two independent experiments



Systematics-Dominated Measurements

Test of the **weak equivalence principle** using a torsion balance [8]



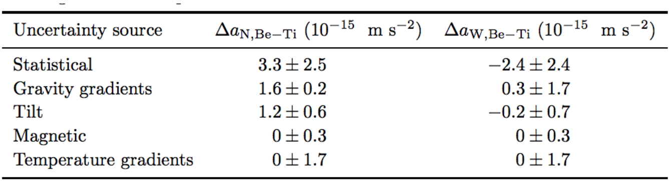


Also able to test for scaler or vector charge Yukawa couplings of the form

$$V(r) \propto \tilde{\alpha} \frac{\tilde{q}_i \tilde{q}_A}{r} \exp -r/\lambda$$

Consider a Variety of Systematic Effects

- The error budget of a torsion balance experiment [8]:

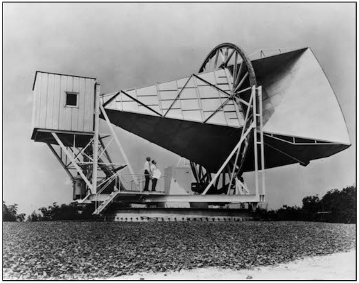


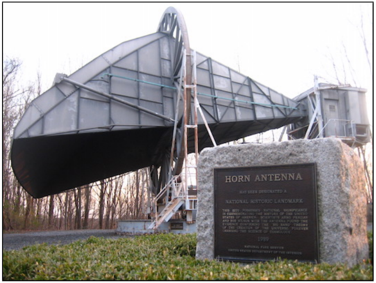
Uncertainty source	$\Delta a_{\mathrm{N,Be-Ti}}$ (10^{-15} m s^{-2})	$\Delta a_{\mathrm{W,Be-Ti}}$ (10^{-15} m s^{-2})
Statistical	3.3 ± 2.5	-2.4 ± 2.4
Gravity gradients	1.6 ± 0.2	0.3 ± 1.7
Tilt	1.2 ± 0.6	-0.2 ± 0.7
Magnetic	0 ± 0.3	0 ± 0.3
Temperature gradients	0 ± 1.7	0 ± 1.7

- Careful planning to eliminate sidereal (daily) modulations due to Earth's rotation, temperature and pressure cycles, etc.
- Measured effect of magnetic gradients on the device by attaching a permanent magnet onto the outside of the balance vacuum vessel
- Measured effect of 15 K m^{-1} temperature gradients by placing heated and cooled copper plates next to the balance
- Estimated effect of gravity gradients due to the sun, Galactic Center, Galactic dark matter halo, etc.

Finding an Anomaly

The Holmdel Horn Antenna was built by Bell Labs for Project Echo, a passive radio communications project in the early 1960s





A. Penzias and R. Wilson decided in 1963 to use the horn as an astronomical receiver and they began to observe Cas A, a radio-bright Galactic supernova remnant

Source of Excess Antenna Temperature

Problem: antenna temperature (noise) was 6.7 K, or 3.5 K higher than expected. After waiting months for the problem to go away [9], Wilson and Penzias begin investigating systematic effects:

- **Pigeon Poop**: Nesting pigeons removed from the horn
- **Seams in telescope assembly** could change its characteristics; sealed with aluminum tape and conducting glue
- **Ambient Noise**. Pointed telescope at New York City, no major change in radio background
- **Atmosphere**: 2.3 ± 0.3 K from atmospheric absorption (expected)
- **Instrumental**: 0.9 ± 0.4 K due to ohmic losses and backlobe response
- **Nuclear Fallout**. Radiation from a 1962 high-altitude nuclear test was considered, but noise should have decreased over time

Ultimate conclusion: unpolarized, isotropic, steady-state 3.5 ± 1.0 K astrophysical noise floor [10] connected to the Big Bang [11]

Techniques for Dealing with Systematic Effects

These cases illustrate some best practices in the design and conduct of experiments:

- Try to anticipate systematic effects on a measurement *before* taking data and design the experiment to minimize them

Example

If you know that there is a temperature-dependent component to your measurement (e.g., measuring length with a steel ruler), calibrate the ruler and record T for each measurement of length

- When taking data, evaluate the effects of cuts, binning, and other choices in the analysis. Vary your conditions and see what happens

Example

Was radio background from New York City interfering with the Holmdel antenna? Checked by taking data "on source" and "off source"

Techniques for Dealing with Systematic Effects

Suppose you design a check and it fails, whatever that means. What do you do? From Barlow [2] (verbatim):

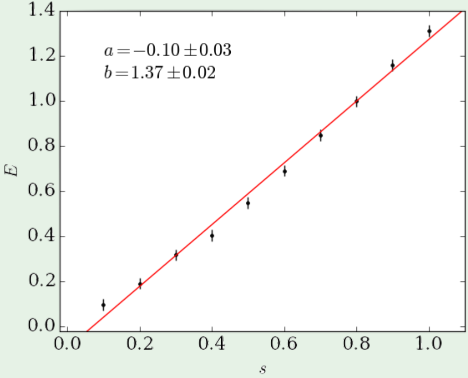
1. Check the test. A mistake may lie there. Find it and fix it.
2. If that doesn't work, check the analysis for mistakes.
3. Worry. Maybe with hindsight an effect is reasonable. (Why are the results of my ruler measurements different after lunch? Oh right, it's warmer in the afternoon.) This check now becomes an evaluation.
4. Worry. This discrepancy is only the tip of the iceberg. Ask colleagues, look at what other experiments did.
5. As a last resort, incorporate the discrepancy into a systematic uncertainty.

Note how the quantification of an uncertainty is the last resort. What you don't want to do is just slap a huge error bar on the result in the name of being "conservative." **Why not?**

Case Study: Fitting an Inappropriate Function

Example

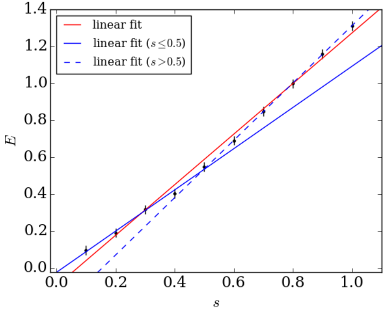
Suppose you have a calorimeter that gives you a signal s, which is related to energy by $E = s + 0.3s^2$ [2].



You take data and fit a straight line $E = a + b \cdot s$, and use the values $\hat{a}$ and $\hat{b}$ in your analysis.

Case Study: Fitting an Inappropriate Function

- You find that $\chi^2 = 16.94$ with 8 degrees of freedom, which is large but not unreasonable. (What is the approximate p-value?)
- So you stick with the linear fit, but as a check you calibrate (i.e., fit) the subranges $0 \leq s \leq 0.5$ and $0.5 < s \leq 1$ separately:



- Result: the slopes are 1.17 ± 0.03 and 1.57 ± 0.06, definitely not agreeing within statistical uncertainties.

Case Study: Fitting an Inappropriate Function

- You follow the procedure for dealing with systematic effects (check, re-check, worry) but fail to spot that the linear calibration is itself inadequate.
- Result: you incorporate a systematic uncertainty of $1.57 - 1.37 = 1.37 - 1.17 = 0.2$ into the slope b, reporting

$$b = 1.37 \pm 0.02 \pm 0.20$$

Is this reasonable?

Case Study: Fitting an Inappropriate Function

- You follow the procedure for dealing with systematic effects (check, re-check, worry) but fail to spot that the linear calibration is itself inadequate.
- Result: you incorporate a systematic uncertainty of $1.57 - 1.37 = 1.37 - 1.17 = 0.2$ into the slope b, reporting

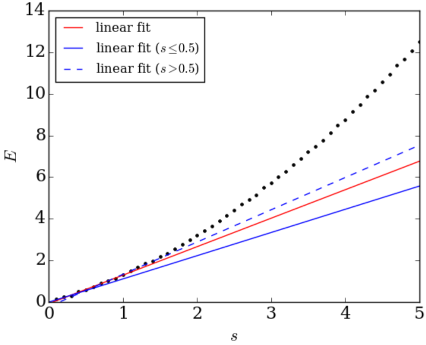
$$b = 1.37 \pm 0.02 \pm 0.20$$

  Is this reasonable?

- In the region $0 \leq s \leq 1$ this systematic uncertainty seriously overstates the error.
- Look again at the fit. The slope 1.37 is a pretty reasonable description of the data

Case Study: Fitting an Inappropriate Function

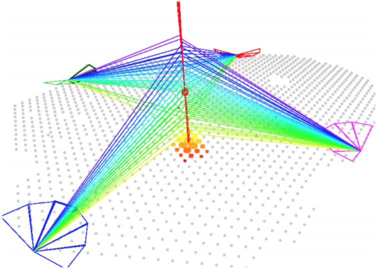
- What happens if the "calibration" of $E(s)$ is extrapolated to $s = 5$?



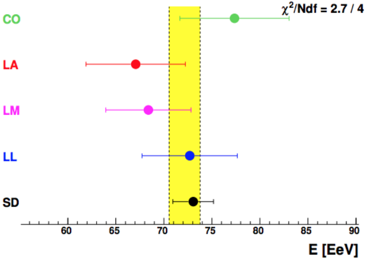
- The linear extrapolation is clearly no good. Not only that, but the systematic uncertainty is worthless for describing the calibration offset
- Lesson: there is no correct procedure for incorporating a check that fails, but folding it into the systematics is probably wrong and should be avoided unless there is no alternative

An Example of "No Alternative"

The error budget for energy scale in an atmospheric calorimeter [12]:



Source	Uncertainty
Fluorescence Yield Y	14%
p, T, e Effects on Y	7%
Calibration	9.5%
Atmosphere	4%
Reconstruction	10%
Invisible Energy	4%
Total	22%



Reconstruction: differences between two reconstruction methods that couldn't be reconciled at the time of publication

Note quadrature sum of uncertainties.

Summary

- Systematic uncertainties are a frequentist concept; for a Bayesian, there is no distinction and all such uncertainties can be dealt with using marginalization
- Still, it's useful to break out uncertainties into statistical and systematic components, as this (usually) makes clear which part of the error bar depends on how much data we took
- When conducting an experiment, one tries to identify systematic effects before, during, and after data-taking.
- There is no recipe for doing this right but there are some "best practices" that good researchers try to follow
- After all efforts have been made to eliminate systematic effects, the remaining uncertainties become systematic uncertainties.
- It is important not to inflate systematics, but in the real world, sometimes you do have to cut your losses and go with a reasonable uncertainty

References I

[1] R Bevington. *Data Reduction and Analysis for the Physical Sciences*. New York: McGraw-Hill, 1969.

[2] Roger Barlow. "Systematic Errors: Facts and Fictions". In: *Conf. on Adv. Stat. Techniques in Particle Physics*. Durham, England, 2002, pp. 134–144. arXiv: hep-ex/0207026 [hep-ex].

[3] *Notes on Statistics for Physicists, Revised*. 2001. URL: https://ned.ipac.caltech.edu/level5/Sept01/Orear/frames.html.

[4] Glen Cowan. *Statistical Data Analysis*. New York: Oxford University Press, 1998.

[5] S Heinemeyer et al. "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties". In: (2013). Ed. by S Heinemeyer. arXiv: 1307.1347 [hep-ph].

[6] R.J. Barlow. *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences*. New York: Wiley, 1989.

References II

[7] M.C. McFarlane. “Observation of Antineutrino Oscillations from Spectral Distortions at the Daya Bay Reactor Neutrino Experiment”. PhD thesis. University of Wisconsin-Madison, 2014.

[8] T.A. Wagner et al. “Torsion-balance tests of the weak equivalence principle”. In: *Class.Quant.Grav.* 29 (2012), p. 184002. arXiv: 1207.2442 [gr-qc].

[9] J. Bernstein. *Three Degrees above Zero: Bell Laboratories in the Information Age*. New York: Cambridge, 1987.

[10] A. A. Penzias and R. W. Wilson. “A Measurement of Excess Antenna Temperature at 4080-Mc/s”. In: *Astrophys.J.* 142 (1965), pp. 419–421.

[11] R.H. Dicke et al. “Cosmic Black-Body Radiation”. In: *Astrophys.J.* 142 (1965), pp. 414–419.

References III

[12] B.R. Dawson et al. "Hybrid Performance of the Pierre Auger Observatory". In: *Proc. 30th ICRC*. Merida, Mexico, 2007. arXiv: 0706.1105 [astro-ph.HE].