

PHYSICS FOR ENGINEERING STUDENTS

Measurement uncertainty II

Systematic effects, combination and reporting of results

Lecture 5 — Book-style student notes

Learning focus: Systematic effects · Consistency tests · Weighted mean · Type A and Type B evaluation

University of Szeged · Institute of Physics

Measurement uncertainty II

Repeated observations reveal random scatter and allow the standard uncertainty of a mean to be estimated, but they do not by themselves establish that the result is unbiased. A series of closely clustered readings may all be displaced by calibration error, an imperfect physical model, environmental influence, or a consistent procedural effect. The second stage of uncertainty evaluation therefore asks what influences can shift all measurements in a similar direction and how their remaining effects can be quantified.

Learning objectives

- Distinguish random effects from systematic effects.
- Reveal bias by changing conditions and comparing independent results.
- Calculate consistency statistics, weighted means, and combined uncertainty.
- Apply Type A and Type B evaluations and report a complete result.

Central principle: Choose, evaluate, and report a probability or uncertainty model from the physical measurement process and its assumptions.

Further interpretation

A complete result is built through a chain: define and perform the measurement, identify uncertainty sources, evaluate each contribution, combine them using the measurement model, and report the estimate with its uncertainty and unit. This workflow links the statistical information from repeated measurements to non-statistical information from calibration certificates, instrument specifications, resolution, previous data, and physical judgement. The objective is not to label every effect as either good or bad, but to represent the state of knowledge consistently.

1. Random and systematic effects

Random effects vary unpredictably from observation to observation and produce dispersion. Under stable repeatability conditions they can be evaluated statistically, and the uncertainty of the mean normally decreases as the number of independent observations increases. Systematic effects act consistently over a measurement sequence and can shift the mean away from the value that would be obtained with an ideal method. Repetition under unchanged conditions does not expose a constant offset and does not reduce it.

The distinction refers to behaviour within a defined experiment. An influence that is systematic during one short sequence may vary randomly across days, instruments, laboratories, or operators. This is why changing conditions can reveal or randomize systematic effects. Randomization does not

eliminate them; it turns their between-condition variability into observable dispersion that can be evaluated statistically. Known significant biases should still be corrected whenever possible, with uncertainty assigned to the correction.

Further interpretation

Random effects vary unpredictably from observation to observation and produce dispersion. Under stable repeatability conditions they can be evaluated statistically, and the uncertainty of the mean normally decreases as the number of independent observations increases. Systematic effects act consistently over a measurement sequence and can shift the mean away from the value that would be obtained with an ideal method. A useful diagnostic is to separate within-run scatter from changes between runs, days, operators, and instruments. If the latter variation is larger, increasing the number of readings in one unchanged run will not solve the main problem; the design must deliberately vary the suspected influence or introduce an independent reference.

2. Model error: the pendulum example

For a simple pendulum at small amplitude, the period is approximated by $T=2\pi\sqrt{L/g}$, giving $g=4\pi^2L/T^2$. This expression hides assumptions: the bob is treated as a point mass, the string is massless and inextensible, the amplitude is small, air resistance and pivot friction are negligible, the support is fixed, and the gravitational field is uniform. Violation of these assumptions creates model error that can bias the inferred value of g .

$$T = 2\pi\sqrt{L/g}$$

$$g = 4\pi^2L/T^2$$

$$T \approx 2\pi\sqrt{L/g}(1 + \theta_0^2/16 + \dots)$$

The finite-amplitude correction illustrates the method. The true period is slightly longer than the small-angle expression, approximately $T \approx 2\pi\sqrt{L/g}(1 + \theta_0^2/16 + \dots)$. If the uncorrected simple formula is used, the longer measured period makes the calculated g systematically too small. The remedy is to reduce the amplitude, apply an adequate correction, or include the residual model contribution in the uncertainty budget. A precise measurement with an inadequate model can be highly repeatable but physically wrong.

Further interpretation

For a simple pendulum at small amplitude, the period is approximated by $T=2\pi\sqrt{L/g}$, giving $g=4\pi^2L/T^2$. This expression hides assumptions: the bob is treated as a point mass, the string is massless and inextensible, the amplitude is small, air resistance and pivot friction are negligible, the support is fixed, and the gravitational field is uniform. Violation of these assumptions creates model error that can bias the inferred value of g . The lesson extends beyond pendulums: every compact

formula is valid only in a defined regime. Changing amplitude, geometry, loading, or frequency and plotting the residual against that change is often more informative than collecting many nominally identical observations, because the trend tests the physical approximation directly.

3. Instrumental, environmental and procedural bias

Instrumental sources include zero error, scale-factor error, drift, hysteresis, nonlinearity, finite alignment, and imperfect calibration. A balance that reads +0.20 g at zero will return values that are consistently too high by approximately 0.20 g. If the offset is verified, it should be corrected rather than merely hidden inside a large uncertainty. Calibration uncertainty and possible drift determine how well that correction is known.

$$x_{\text{corrected}} = x_{\text{indicated}} - e_0$$

Environmental effects include temperature, pressure, humidity, electromagnetic interference, vibration, and thermal gradients. Procedural effects include parallax, reaction delay, positioning, contact force, sample preparation, and data-selection rules. The measurement model should connect each relevant effect to the output. Cause-and-effect analysis is useful because it prevents omissions and double counting: a contribution already represented by repeated scatter should not automatically be added again as a separate independent term.

Further interpretation

Instrumental sources include zero error, scale-factor error, drift, hysteresis, nonlinearity, finite alignment, and imperfect calibration. A balance that reads +0.20 g at zero will return values that are consistently too high by approximately 0.20 g. If the offset is verified, it should be corrected rather than merely hidden inside a large uncertainty. Calibration uncertainty and possible drift determine how well that correction is known. Bias sources should be linked to parameters in the measurement equation whenever possible. This turns a list of possible problems into testable predictions: a zero error gives an additive shift, a scale-factor error grows with signal, drift changes with time, and a thermal coefficient follows temperature.

4. Revealing systematic effects experimentally

A systematic effect can often be exposed by deliberately changing one influential condition while keeping the others controlled. Repeating a pendulum experiment at different amplitudes tests the small-angle assumption; changing temperature tests thermal expansion; reversing the apparatus can reveal alignment or directional offsets; interchanging components can reveal channel-dependent calibration. The result is plotted against the changed condition to search for a reproducible trend.

Comparisons between experimenters and between independent methods are especially powerful. Agreement among observers suggests that operator-dependent effects are small relative to the stated

uncertainty. Measuring g by a pendulum, free fall, and another independent technique changes the dominant assumptions and apparatus. Agreement does not prove that all systematic effects vanish, because methods can share references or corrections, but it greatly reduces the likelihood that one method-specific bias dominates.

Further interpretation

A systematic effect can often be exposed by deliberately changing one influential condition while keeping the others controlled. Repeating a pendulum experiment at different amplitudes tests the small-angle assumption; changing temperature tests thermal expansion; reversing the apparatus can reveal alignment or directional offsets; interchanging components can reveal channel-dependent calibration. The result is plotted against the changed condition to search for a reproducible trend. A controlled change is strongest when it is reversible and repeated. Alternating two amplitudes, orientations, temperatures, or channels helps distinguish a genuine condition-dependent effect from slow drift. Randomizing the order adds protection against time trends and makes the comparison compatible with statistical analysis.

5. Comparing two measurement results

Suppose two independent measurements of the same quantity are $x_1 \pm u_1$ and $x_2 \pm u_2$, where u_1 and u_2 are standard uncertainties. Their absolute difference is $D = |x_1 - x_2|$. If the uncertainties are independent, the standard uncertainty associated with the difference is $u_D = \sqrt{u_1^2 + u_2^2}$. The normalized discrepancy $E_n = D/u_D$ provides a quantitative consistency indicator.

$$D = |x_1 - x_2|$$

$$u_D = \sqrt{u_1^2 + u_2^2}$$

$$E_n = D/u_D$$

Values of E_n near or below one indicate compatibility at approximately the standard-uncertainty scale; a larger value suggests that the results or the uncertainty model require investigation. Visual overlap of intervals is a useful first check but is not a rigorous universal test. Correlation, coverage factors, distribution shape, and the meaning of the reported intervals must be considered. Inconsistency is evidence that something is missing—it is not by itself proof of which experiment is wrong.

Further interpretation

Suppose two independent measurements of the same quantity are $x_1 \pm u_1$ and $x_2 \pm u_2$, where u_1 and u_2 are standard uncertainties. Their absolute difference is $D = |x_1 - x_2|$. If the uncertainties are independent, the standard uncertainty associated with the difference is $u_D = \sqrt{u_1^2 + u_2^2}$. The normalized discrepancy $E_n = D/u_D$ provides a quantitative consistency indicator. A normalized difference is a screening tool, not a verdict. Before interpreting it, verify that both results describe the same

measurand, use uncertainties on the same statistical scale, and account for shared references. A large discrepancy motivates investigation; it does not identify which result is biased.

6. A structured search for systematic error

Detection should follow a sequence rather than rely on intuition. First repeat measurements under nominally identical conditions to estimate random variability. Next vary relevant conditions, compare observers, reverse or exchange components, and use independent methods or reference values. Examine residuals and time series for trends that simple averages conceal. Each test should be chosen because it challenges a specific assumption in the physical model.

When a discrepancy is found, determine whether a correction can be evaluated. If it can, update the result and assign uncertainty to the correction. If it cannot be removed or quantified reliably, redesigning the experiment may be necessary. Simply inflating uncertainty after an unexplained inconsistency is not always defensible, because the missing effect may not be bounded. An adequate uncertainty model should explain the observed between-condition dispersion without being tuned merely to force agreement.

Further interpretation

Detection should follow a sequence rather than rely on intuition. First repeat measurements under nominally identical conditions to estimate random variability. Next vary relevant conditions, compare observers, reverse or exchange components, and use independent methods or reference values. Examine residuals and time series for trends that simple averages conceal. Each test should be chosen because it challenges a specific assumption in the physical model. The search should be planned around assumptions rather than around a generic catalogue of errors. For each assumption, state an observable consequence and a test that would contradict it. This creates a traceable path from physical reasoning to data, correction, uncertainty contribution, or experimental redesign.

7. Combining measurements and the idea of weighting

Several independent estimates of the same quantity should not necessarily contribute equally. A measurement with smaller standard uncertainty contains more information and should have greater influence. For normally distributed independent estimates with known standard uncertainties, the natural weight is inverse variance: $w_i = 1/u_i^2$. Halving an uncertainty therefore multiplies the weight by four.

$$w_i = 1/u_i^2$$

Weighting is justified only when the estimates refer to the same measurand and their uncertainty statements are comparable. A very precise but biased result can dominate the weighted mean and make the combined value misleading. Before combination, results should be checked for consistency,

common systematic effects, and correlations. If measurements share a calibration standard, their errors are not fully independent and the simple inverse-variance formulas underestimate the combined uncertainty.

Further interpretation

Several independent estimates of the same quantity should not necessarily contribute equally. A measurement with smaller standard uncertainty contains more information and should have greater influence. For normally distributed independent estimates with known standard uncertainties, the natural weight is inverse variance: $w_i = 1/u_i^2$. Halving an uncertainty therefore multiplies the weight by four. Weighting is justified only when the estimates refer to the same measurand and their uncertainty statements are comparable. Inverse-variance weighting rewards precision only when uncertainty estimates are credible and the inputs are compatible. It cannot protect against a common unrecognized bias. A preliminary consistency check and an explicit correlation assessment therefore belong before the calculation of any combined value.

8. Weighted mean and its uncertainty

For N independent results x_i with weights $w_i = 1/u_i^2$, the weighted mean is the sum of weighted values divided by the sum of weights. The standard uncertainty of the weighted mean is the reciprocal square root of the total weight. More precise measurements pull the result toward their values, and adding an independent informative measurement increases total weight.

$$x_w = (\sum_i w_i x_i) / (\sum_i w_i)$$

$$u(x_w) = 1 / \sqrt{\sum_i w_i}$$

Example: $x_1 = 10.00$ with $u_1 = 0.20$ and $x_2 = 10.30$ with $u_2 = 0.10$ give weights 25 and 100. The weighted mean is $(25 \times 10.00 + 100 \times 10.30) / 125 = 10.24$, while its standard uncertainty is $1 / \sqrt{125} \approx 0.089$. This calculation is meaningful only after consistency is checked: the difference 0.30 compared with $\sqrt{0.20^2 + 0.10^2} = 0.224$ gives $E_n \approx 1.34$, which merits attention before presenting a final combination.

Further interpretation

For N independent results x_i with weights $w_i = 1/u_i^2$, the weighted mean is the sum of weighted values divided by the sum of weights. The standard uncertainty of the weighted mean is the reciprocal square root of the total weight. More precise measurements pull the result toward their values, and adding an independent informative measurement increases total weight. Example: $x_1 = 10.00$ with $u_1 = 0.20$ and $x_2 = 10.30$ with $u_2 = 0.10$ give weights 25 and 100. The weighted result should be accompanied by residuals or a consistency statistic. If one highly weighted value dominates, its calibration and model

assumptions deserve special scrutiny. Reporting only the compact combined uncertainty can otherwise create false confidence by hiding disagreement among the contributing measurements.

9. Randomization of systematic effects

Measurements made with different instruments, methods, laboratories, observers, orientations, or times can turn condition-specific biases into a distribution of independent results. The observed between-result standard deviation then contains information about systematic variability that remains invisible within one repeatability series. This strategy underlies reproducibility studies and interlaboratory comparisons.

$$\chi^2 = \sum_i w_i (x_i - x_w)^2$$

The expected dispersion from the reported individual uncertainties should be compared with the observed dispersion. If the observed spread is substantially larger, the uncertainty statements may omit a between-condition component. A Birge ratio or chi-square consistency statistic can formalize the comparison. Randomization allows statistical estimation of systematic variability, but it neither identifies every cause nor removes a common bias shared by all participants.

Further interpretation

Measurements made with different instruments, methods, laboratories, observers, orientations, or times can turn condition-specific biases into a distribution of independent results. The observed between-result standard deviation then contains information about systematic variability that remains invisible within one repeatability series. This strategy underlies reproducibility studies and interlaboratory comparisons. The expected dispersion from the reported individual uncertainties should be compared with the observed dispersion. Randomization is most effective when the changed conditions sample the population to which the result will be generalized. Varying only convenient factors may miss the dominant reproducibility component. The design should therefore connect laboratories, operators, instruments, or times to the intended scope of the final statement.

10. Type A and Type B evaluation

Type A evaluation obtains a standard uncertainty by statistical analysis of measured data, such as the standard deviation of repeated readings, regression residuals, or between-series variation. Type B evaluation uses other information: calibration certificates, manufacturer specifications, instrument resolution, previous measurements, physical limits, and expert judgement. Type A does not mean random error, and Type B does not mean systematic error; the labels describe the evaluation method.

$$u_{\text{rect}} = a/\sqrt{3}$$

$$u_{\text{tri}} = a/\sqrt{6}$$

$$u = U/k$$

Every contribution must ultimately be expressed as a standard uncertainty. For a rectangular distribution with half-width a , $u=a/\sqrt{3}$; for a triangular distribution, $u=a/\sqrt{6}$; for a quoted expanded uncertainty U with coverage factor k , the corresponding standard uncertainty is $u=U/k$. The assigned probability model should reflect the information available, not a habitual default.

Further interpretation

Type A evaluation obtains a standard uncertainty by statistical analysis of measured data, such as the standard deviation of repeated readings, regression residuals, or between-series variation. Type B evaluation uses other information: calibration certificates, manufacturer specifications, instrument resolution, previous measurements, physical limits, and expert judgement. Type A does not mean random error, and Type B does not mean systematic error; the labels describe the evaluation method. Type A and Type B describe routes to evaluation, not physical categories of error. Either route may quantify an influence that behaves randomly or systematically in a particular experiment. The essential requirement is to state the information source, probability model, conversion to standard uncertainty, and justification.

11. Combining uncertainty contributions

If a measurand Y depends on input quantities $X_1, \dots, X_n$ through $Y=f(X_1, \dots, X_n)$, a small input uncertainty contributes according to the sensitivity coefficient $c_i=\partial f/\partial X_i$. For independent inputs, the combined standard uncertainty is the root sum of squared contributions $c_i u(X_i)$. This general propagation law includes unit conversion and shows which inputs dominate the result.

$$c_i = \partial f / \partial X_i$$

$$u_c^2(Y) = \sum_i [c_i u(X_i)]^2$$

Correlation adds covariance terms and can either increase or decrease the uncertainty. Shared calibration, common environmental data, and common processing constants are frequent sources. An uncertainty budget should list each input estimate, source, evaluation type, probability distribution, standard uncertainty, sensitivity coefficient, and contribution. Improving a negligible term has little effect; effort should target dominant contributions that can be reduced without creating new bias.

Further interpretation

If a measurand Y depends on input quantities $X_1, \dots, X_n$ through $Y=f(X_1, \dots, X_n)$, a small input uncertainty contributes according to the sensitivity coefficient $c_i=\partial f/\partial X_i$. For independent inputs, the combined standard uncertainty is the root sum of squared contributions $c_i u(X_i)$. This general propagation law includes unit conversion and shows which inputs dominate the result. Correlation adds covariance terms and can either increase or decrease the uncertainty. An uncertainty budget is also a design tool. Ranking the squared sensitivity-weighted contributions reveals which improvement can materially reduce the final uncertainty. Correlations must be retained because a

common influence can cancel in a difference and reinforce in a sum, changing both the magnitude and interpretation.

12. Reporting a measurement result

A standard report gives the best estimate and its uncertainty with the unit, for example $L=(125.4\pm0.3)$ mm. The uncertainty is normally rounded to one or two significant digits, and the estimate is rounded to the same decimal position. Compact notation such as $L=2.357(25)$ m means 2.357 ± 0.025 m; the digits in parentheses refer to the final digits displayed.

$$X = x \pm u$$

$$U = ku_c$$

$$u_r = u(x)/|x|$$

The report must state whether the number is a standard or expanded uncertainty. For expanded uncertainty use $U=ku_c$ and give k and the intended coverage probability where appropriate. Relative standard uncertainty $u_r=u/|x|$ supports comparisons across scales. Repeatability, reproducibility, and accuracy should be distinguished: they describe agreement under same conditions, agreement under changed conditions, and closeness to a reference respectively. A complete result also identifies the measurand, method, important conditions, and traceability.

Take-home message: The mathematical expression is useful only together with the physical assumptions that justify it and the uncertainty statement that limits its interpretation.

Further interpretation

A standard report gives the best estimate and its uncertainty with the unit, for example $L=(125.4\pm0.3)$ mm. The uncertainty is normally rounded to one or two significant digits, and the estimate is rounded to the same decimal position. Compact notation such as $L=2.357(25)$ m means 2.357 ± 0.025 m; the digits in parentheses refer to the final digits displayed. The report must state whether the number is a standard or expanded uncertainty. A result is transferable only when its conditions and uncertainty convention are clear. The reported value should identify corrections, coverage factor where used, reference state, and relevant environmental or procedural conditions. This allows another reader to compare results without mistaking different measurands or confidence levels for disagreement.

13. Integrated measurement workflow and worked applications

The following compact studies connect the definitions and equations to realistic interpretation, calculation, and experimental design decisions.

Zero-offset correction

A balance shows 0.20 g with no load and 10.20 g for a nominal 10 g object. Five repetitions scatter by only 0.03 g. The small scatter demonstrates good repeatability but says nothing about the 0.20 g offset. Applying the measured zero correction gives approximately 10.00 g. If the zero check has standard uncertainty 0.02 g and the mean loaded reading has random standard uncertainty 0.01 g, their independent contributions combine to $\sqrt{0.02^2 + 0.01^2} = 0.022$ g. The corrected result is therefore reported with an uncertainty governed mainly by knowledge of the offset, not by the impressive repeatability of the loaded readings.

Pendulum amplitude test

Measure the period at amplitudes 2°, 5°, 10°, and 15° while keeping length and timing procedure unchanged. Plot the value of g calculated with the small-angle equation against θ_0^2 . A downward trend is expected because finite amplitude lengthens the period. Fitting the trend tests the theoretical correction and permits extrapolation toward zero amplitude, while the residuals reveal whether other effects remain. The exercise illustrates a general principle: vary a condition that couples to a suspected bias and compare the observed response with the model. Repeating only at 15° would reduce timing noise but would never diagnose the amplitude-dependent systematic error.

Consistency of two laboratories

Laboratory A reports 100.20 ± 0.15 and laboratory B reports 99.80 ± 0.20 in the same unit, with standard uncertainties. Their difference is 0.40 and the standard uncertainty of the difference is $\sqrt{0.15^2 + 0.20^2} = 0.25$, giving $E_n = 1.6$. This is larger than a one-standard-uncertainty compatibility threshold and deserves investigation. Possible causes include a shared measurand defined differently, calibration drift, sample instability, or underestimated uncertainty. The calculation does not establish that A is high or B is low. A reference measurement, exchange of standards, or cross-measurement of the same stable sample is needed to identify the source.

Static and dynamic spring methods

A spring constant can be found statically from the slope $F = kx$ or dynamically from $T = 2\pi\sqrt{m/k}$. The static method is sensitive to force calibration, displacement zero, and hysteresis. The dynamic method depends on timing, effective oscillating mass, damping, and amplitude. Agreement within uncertainty increases confidence because the dominant assumptions differ. Disagreement should lead to targeted tests: reverse loading direction, change attached mass, restrict amplitude, inspect residuals, and

include spring mass in the dynamic model. Combining the two estimates before resolving a significant discrepancy would hide rather than solve the physical problem.

Weighted mean with unequal precision

Three independent values are 5.10 ± 0.20 , 5.30 ± 0.10 , and 5.26 ± 0.05 . Their inverse-variance weights are 25, 100, and 400, so the most precise result has sixteen times the influence of the first. The weighted estimate is $(25 \times 5.10 + 100 \times 5.30 + 400 \times 5.26) / 525 \approx 5.25$, with standard uncertainty $1/\sqrt{525} \approx 0.044$. Before reporting this compact result, calculate the weighted residuals and chi-square. If the values scatter more than their uncertainties predict, a between-measurement component or correlation must be considered instead of trusting the small formal uncertainty.

Rectangular specification

A manufacturer states that an uncorrected effect lies within ± 0.60 unit and gives no reason to prefer any value inside the interval. Representing this information by a rectangular distribution gives $u = 0.60/\sqrt{3} = 0.346$ unit. The number 0.60 is a limit, not already a standard uncertainty. Treating it directly as u would be conservative but inconsistent with the selected probability model; dividing by $\sqrt{6}$ would instead assume a triangular distribution concentrated near zero. The uncertainty budget should record the original specification, chosen distribution, divisor, and physical justification so the conversion can be reviewed.

Correlated calibration

Two lengths are measured with the same scale and later subtracted. A common scale-factor error moves both values in the same relative direction. If the lengths are similar, much of that common effect cancels in their difference; treating the two calibration uncertainties as independent would overestimate the difference uncertainty. In a sum they reinforce. The covariance term expresses this behaviour mathematically. Correlation is therefore not a minor statistical detail: it encodes shared physical information. Uncertainty evaluation should track common references, constants, environmental corrections, and processing steps across all input quantities.

Complete reporting statement

A defensible final statement might read: “The corrected length at 20.0 °C is $L = 125.4$ mm with expanded uncertainty $U = 0.6$ mm, using coverage factor $k = 2$, corresponding to an approximate 95% coverage probability under the stated model.” This reports more than $L = 125.4 \pm 0.6$ mm. It identifies correction state, temperature condition, uncertainty type, coverage factor, and interpretation. Supporting information should describe the measurement method, traceability, dominant contributions, and any correlations. Compact parenthetical notation may be used in tables, but the surrounding text must still define what the parentheses mean.

Integrated interpretation

The chapter can be reduced to one disciplined sequence. First define the measurand and write a model that connects it to indications, corrections, and influence quantities. Next use repeated observations to describe short-term scatter, while separately challenging possible systematic effects by changing conditions, comparing methods, and using references. Convert every retained contribution to a standard uncertainty, include correlations, and examine whether independent results are statistically compatible before combining them. Finally report the corrected estimate, unit, uncertainty type, coverage information, and conditions.