



Measurement Uncertainty

Evaluation according to the
ISO/IEC Guide 98-3:2008 (JCGM/WG1/100)
Guide to the expression of uncertainty in measurement (GUM:1995)
<http://www.bipm.org/en/publications/guides/gum.html>

Dario Petri
Department of Industrial Engineering
University of Trento - Trento - Italy
dario.petri@unitn.it



Outline



- ✓ Basic Terminology
- ✓ Measurement Uncertainty
- ✓ Uncertainty Evaluation Methods
- ✓ Combined Uncertainty
- ✓ Indirect Measurements



BASIC TERMINOLOGY

Basic definitions



measurement error [VIM]:
measured quantity value minus a **reference** quantity value

International Vocabulary of Metrology
a **reference** value is required

two components:

random error [VIM]:
in replicate measurements
varies in an unpredictable manner
e.g.: noise interference,
fluctuations in environmental conditions, ...

systematic error [VIM]:
in replicate measurements remains
constant or varies in a predictable manner
e.g.: lack of calibration,
time instability of instruments, ...

random errors express **variability**
not related to a reference value

bias:
average of the measured values
minus the **reference** value

Measurement data model

measurement data are modeled as **realizations of a random variable**

distribution of measurement data is represented by a **probability density function (pdf)**

concept of pdf implies continuity of values:
we neglect that measurement values are defined on a discrete scale due to finite resolution of instruments

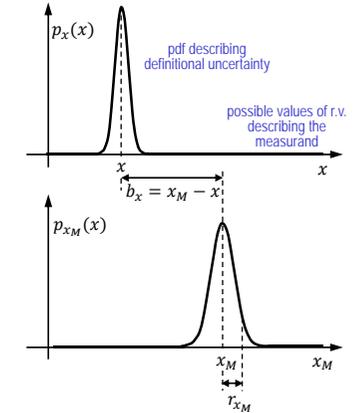
Measurement result

"true" value of the measurand (unknown) \rightarrow $x = x + d_x$ \leftarrow definitional error

random error \rightarrow $x_M = x_M + r_{x_M} = (x + b_{x_M}) + r_{x_M}$ \leftarrow systematic error

random errors has zero mean

by definition: offsets are accounted for by systematic errors



mean value x_M of repeated measurement assumed as the **best approximation** of the measurand value x
called **measured value** estimated as the (arithmetic) **average** of measurement data

Accuracy

accuracy [ISO 5725]: defined for a **single measurement**
the closeness of agreement between a **test result** and the accepted **reference value**

reference is known when calibrating an instrument (reference is provided by a measurement standard)
reference is not known when measuring: it should be the measurand "true" value, which is unknowable (VIM)

"test results" emphasizes accuracy as a feature of measuring **instruments**
an **instrument is accurate** if **each result** it produces is accurate
in specified operating conditions

a measurement is said to be **more accurate** when it offers a **smaller measurement error [VIM]**

accuracy depends on both **systematic and random errors**

Precision - Trueness

similar to measurement error which is composed of systematic and random errors \rightarrow unlike accuracy require a series of values

accuracy consists of two components [ISO 5725]:

trueness:

precision:

the closeness of agreement between:

the **average value** obtained from a large **series** of test results and an accepted **reference value**

independent test results obtained under stipulated conditions

systematic error \rightarrow related to the closeness to a reference value
concept rarely used \rightarrow depends on a reference value

random error \rightarrow related to closeness of measurement results to each other
depends only on random errors does not relate to a reference value

Precision - Trueness

precision:

feature of an instrument that indicates its capability of avoiding:

random errors

the greater the precision the less the random errors

the closer the measured values to each other

trueness:

systematic errors

the greater the trueness the less the systematic errors

the closer the average of the measured values to the reference value

accuracy:

measurement errors

the greater the accuracy the less the measurement errors

the closer the measured values to the reference value
high accuracy needs both high trueness and high precision

there are **no standardized procedures** to evaluate **accuracy** as a function of **trueness** and **precision**

Visual representations

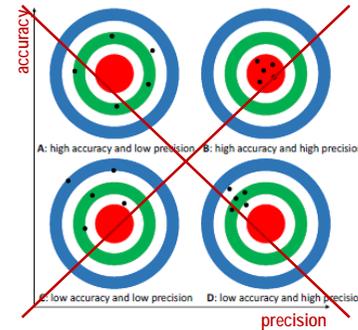
bullseye charts are often used to visually explain the concepts

reference represented by the red bullseye

black dots represent the values returned by replicated measurements

e.g., with the search phrase "accuracy and precision"

great majority of pictures found with google image search are **wrong**



(A) the concept of "high accuracy and low precision" is a nonsense: a measurement instrument **cannot be accurate and imprecise** at the same time!

(A) accuracy is related to a **single measured value**: singularly, values in (A) are located w.r.t. the bullseye as in (D), labeled "low accuracy"

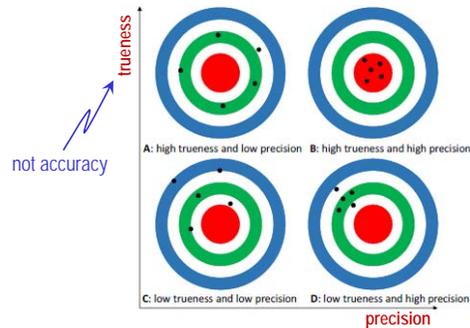
(A), (C): values are further from each other compared to (B) and (D) instrument **precision is correctly visualized** even if the bullseye is hidden (random errors are not related to a reference value) the information about the relative spread of the black dots would remain visible

(C), (D): average value is off the red bullseye: **low trueness** due to systematic errors

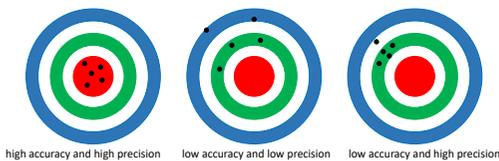
Ref.: S. Shirmohammadi, L. Mari, D. Petri, "On the Commonly-Used Incorrect Visual Representation of Accuracy and Precision," IEEE Instr. and Meas. Magazine, 2021

Visual representations

correct visualizations

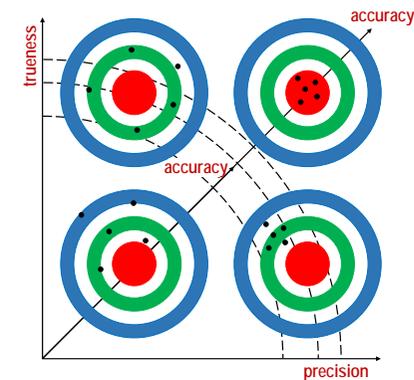


not accuracy



Visual representations

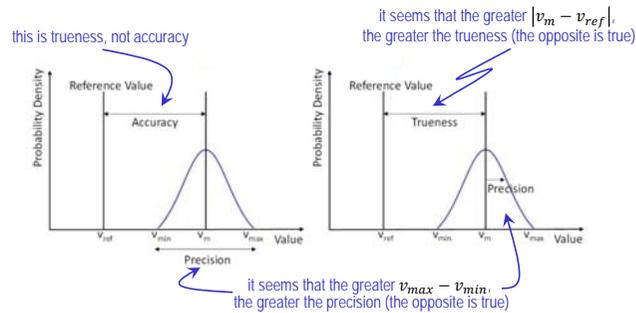
correct visualization



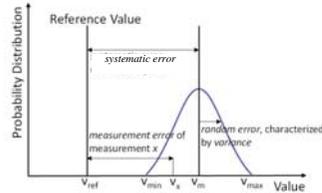
Visual representations

visual representations of measured values of **replicated measurements** of the same measurand using an instrument

these pictures are **wrong**



this picture is **correct**
assuming that a reference value is known



MEASUREMENT UNCERTAINTY

Uncertainty

accuracy can be used to characterize **instruments** since it requires the **knowledge of a reference value**, which is **not available for measurement**

not measurement results

Guide to the Expression of Uncertainty in Measurement

uncertainty of measurement (GUM):

parameter, associated with the **result** of a measurement, that characterizes the **dispersion of the values** that could reasonably be attributed to the measurand

a parameter that **summarizes the distribution** of measured values

Standard uncertainty

dispersion of measurement data about their mean value is **quantified** by the **standard deviation (std)**

an **estimate** of the std of the measured value x is called **standard uncertainty** $u(x)$

suggested notation

Expanded uncertainty

called **coverage interval**

it is often required an **interval** about the measured value that may be expected to encompass a **"large fraction"** of values distribution that could reasonably be attributed to the measurand

quantified by the **coverage probability** $p(x)$ (or **level of confidence**) of the interval

half-width of the coverage interval is called **expanded uncertainty** $U(x)$

$$(x - U(x), x + U(x)) = x \pm U(x)$$

usual notation

suggested notation

Expanded uncertainty

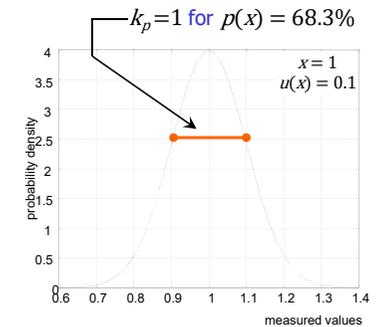
to build a **coverage interval**, measurement data **pdf** must be **known** (or assumed)

pdf **shapes** that often may be reasonably assumed:

- **normal**
 - U-shaped
 - uniform
 - Weibull
 - triangular
 - Poisson ...
- most common assumption

for normal pdf: $U(x) = k_p u(x)$

coverage factor depends on coverage probability $p(x)$

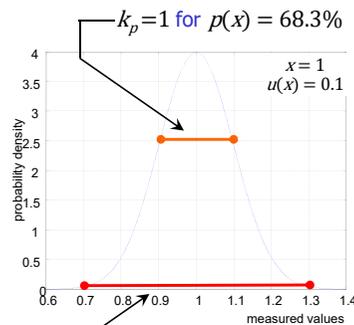


Expanded uncertainty

coverage interval $x \pm U(x)$ $U(x) = k_p u(x)$

for a normal pdf:

coverage level p [%]	coverage factor k_p
68,27	1
90	1,645
95,45	1,960
99	2,576
99,73	3



UNCERTAINTY EVALUATION METHODS

two methods of evaluation of uncertainty:

two components of uncertainty that usually (but not always) are related to **random** or **systematic** effects, respectively

Type A evaluation

by **statistical analysis** of the distribution of data from replicated measurement

Type B evaluation

by means of **non-statistical analysis** of measurement data, based on experience or other a priori information

repeated measurements enable the identification of the **effects** of random **fluctuations** of **influence factors**

e.g., on the instruments or the measurand

n **independent observations** x_k of a random variable x

under generally satisfied constraints

the **best estimate** of the **expectation** of x :

$$\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k$$

assumed as the **measured value**

dispersion of **measurement data** about the mean \bar{x} can be characterized by the **standard deviation** σ , estimated by the **experimental variance**:

$$s^2(x_k) = \frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2$$

experimental variance of the **mean** \bar{x} : $s^2(\bar{x}) = s^2(x_k)/n$

standard uncertainty of the **measured value** \bar{x} :

$$u_A(x) = s(x_k)/\sqrt{n}$$

type A method of uncertainty evaluation

measurement of a DC voltage with a multimeter:

- 5-digit instrument in high resolution mode
- 10 V range
- $n = 7$ repeated observations

readings

1	9.2585
2	9.2597
3	9.2573
4	9.2568
5	9.2586
6	9.2592
7	9.2581

measurement uncertainty evaluated using a **type A evaluation method**:

- measured (average) value: $\bar{V} = 9.2583$ V taken as the best estimate of the value of the measurand
- experimental standard deviation (of data): $s(V_k) = 0.93$ mV
- standard uncertainty: $u_A(V) = s(V_k)/\sqrt{7} = 0.35$ mV

Type B evaluation method

short-time data variability usually does **not** include **all uncertainty sources**

during measurement, values of **influence factors** are almost **constant**, but they **differ** from the instrument **calibration** ones

can't be detected by repeated measurement

↙ **systematic contribution** on measurement data that can be evaluated using a priori available information on **sensitivity to influence factors**

Type B evaluation method

a priori available information:

- previous measurement data
- experience with, or general knowledge of, the behavior and properties of relevant systems and instruments
- **manufacturer's specifications (user's manual)** ← common available information
- data provided by calibration and other certificates
- uncertainties assigned to reference data taken from handbooks

accuracy of uncertainty evaluated using **type B** methods strongly depends on **available information**:

reliability

proper use

which calls for insight based on **experience** and general **knowledge**, skills that can be learned with **practice**

Numerical example

measurement of a DC voltage with a multimeter:

- 5-digit instrument in high resolution mode
- 10 V range
- instrument reading: 9.2587 V

measurement uncertainty evaluated using a **type B evaluation method**

available information: **user manual**, which shows that the instrument characteristics (may) change with **time (aging)**

Numerical example

table extracted from user's manual

time elapsed from calibration

secondary influence properties are neglected: only 1 or (at most) 2 digits are significant

uncertainty expressed with **1 or (at most) 2 digits**

% of reading + # digits

Range	Resolution			Accuracy	
	Slow	Medium	Fast	(6 Months)	(1 Year)
300 mV	—	10 μ V	100 μ V	0.02 % + 2	0.025 % + 2
3 V	—	100 μ V	1 mV	0.02 % + 2	0.025 % + 2
30 V	—	1 mV	10 mV	0.02 % + 2	0.025 % + 2
300 V	—	10 mV	100 mV	0.02 % + 2	0.025 % + 2
1000 V	—	100 mV	1 V	0.02 % + 2	0.025 % + 2
100 mV	1 μ V	—	—	0.02 % + 6	0.025 % + 6
1000 μ V	10 μ V	—	—	0.02 % + 6	0.025 % + 6
10 V	100 μ V	—	—	0.02 % + 6	0.025 % + 6
100 V	1 mV	—	—	0.02 % + 6	0.025 % + 6
1000 V	10 mV	—	—	0.02 % + 6	0.025 % + 6

half-width Δ of the **coverage interval** due to **aging** effects:

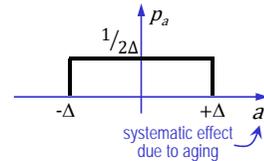
$$\Delta = 0.02\% \text{ of reading} + 6 \text{ digits}$$

$$\Delta = 2 \cdot 10^{-4} \cdot 9.2587 + 6 \cdot 10^{-4} = 2.5 \cdot 10^{-3} \text{ V} = 2.5 \text{ mV}$$

Numerical example

standard uncertainty evaluation requires to know the pdf of readings

since **no information** is available about the **distribution of readings** due to aging, the **same probability** is assigned to **all possible values**



uniform distribution

there are **no reasons** to consider some values more probable than others

assumption follows from the Maximum Entropy Principle of statistics
It is criticized

derived standard uncertainty:

$$u_B(V) = \frac{\Delta}{\sqrt{3}} = 1.4 \text{ mV}$$

$u_A(V) = 0.35 \text{ mV}$ obtained from repeated measurements is almost negligible

Type B evaluation method

systematic error b_{x_M} can be due to $N > 1$ **different uncertainty sources**:
e.g., more instruments

$$b_{x_M} = \sum_{n=1}^N b_{x_{M,n}}$$

if $N \geq 4 \div 5$, $b_{x_{M,n}}$ **uncorrelated** and of the same order of magnitude:

b_{x_M} **is almost Gaussian** with estimated variance:
Central limit theorem

$$u_B^2(x) = \sum_{n=1}^N u_{B,n}^2(x)$$

Central limit theorem

given $y = f(x_1, x_2, \dots, x_i, \dots, x_N)$,

linear combination of r.vs.

If:

- none σ_i dominates the others
- none combination coefficient dominates the others
- $N \rightarrow \infty$

then the **pdf of y is normal**,
no matter on the shapes of the pdfs associated with x_i

almost normal pdfs are often obtained for y if $N \geq 4 \div 5$

the larger N , the better the approximation

COMBINED UNCERTAINTY

Combined uncertainty

very often **influence factors**:

are affected by
random **fluctuations**

differ from the values
assumed during **calibration**

contributions to measurement results
of **both phenomena** must be considered

both type A and type B evaluation methods
must be jointly applied
and the obtained uncertainties $u_A(X)$ and $u_B(X)$ must be combined

Combined uncertainty

random fluctuations and differences w.r.t. the calibration context
are due to **different physical phenomena**

the related effects can be assumed **uncorrelated**
and **composed "in quadrature"**:

$$u_c(x) = \sqrt{u_A^2(x) + u_B^2(x)}$$

**combined
standard
uncertainty**

expanded uncertainty is then evaluated as:

$$U_c(x) = k u_c(x)$$

usual coverage factor
values: $2 \leq k \leq 3$

in the above example: $u_A(V) = 0.35 \text{ mV}$, $u_B(V) = 1.4 \text{ mV}$

$$u_c(V) \cong u_B(V) = 1.4 \text{ mV}$$

INDIRECT MEASUREMENTS

Indirect measurement

measurand value is often obtained using **mathematical computation**

e.g.: electric power in DC conditions: $P = V \cdot I$

impedance of an electric load: $Z = V / I$

mechanical power: $P = T \cdot \omega$

r.v. modeling
measurand values

deterministic
function

r.v. modeling
measured quantities

$$y = f(x_1, x_2, \dots, x_i, \dots, x_N)$$

estimate of the "best"
measurand value
(often defined as $E[y]$)

measured value
(estimate of $E[x_i]$)

$$y = f(x_1, x_2, \dots, x_i, \dots, x_N)$$

Indirect measurement

assumptions:

standard uncertainty of each input quantity $x_i, i = 1, \dots, n$ is known and its value assures **small deviations** from the measured value x_i

function $f(\cdot)$ is **fairly linear** about measured values $x_i, i = 1, \dots, n$

$f(\cdot)$ can be approximated by the first order terms of its Taylor series expansion centered on x_i

$$y - \hat{y} = \sum_{i=1}^N \frac{\partial f}{\partial x_i} (x_i - \hat{x}_i)$$

derivatives are evaluated in the measured values \hat{x}_i of the input quantities

$y = E(y)$ if the deviation of x_i from $E(x_i)$ is negligible

Limitations

derivatives can be **difficult/impossible to evaluate**

e.g.: when $f(\cdot)$ is implemented by an algorithm with if-the-else clause

if the magnitude of **all derivatives** is close to **zero**, or the **nonlinearity** of $f(\cdot)$ is significant, **higher-order terms** of the Taylor series must be considered

Mathematical model

squaring the linear approximation:

$$(y - \hat{y})^2 = \sum_{i=1}^N \left(\frac{\partial f}{\partial x_i} \right)^2 (x_i - \hat{x}_i)^2 + 2 \sum_{i=1}^N \sum_{j=i+1}^N \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} (x_i - \hat{x}_i)(x_j - \hat{x}_j)$$

taking the expectation:

$$E[(y - \hat{y})^2] = \sigma_y^2 \cong u^2(y)$$

$$E[(x_i - \hat{x}_i)^2] = \sigma_i^2 \cong u^2(x_i)$$

$$E[(x_i - \hat{x}_i)(x_j - \hat{x}_j)] = \sigma_{ij} = \sigma_{ji} \cong u^2(x_i, x_j) = u^2(x_j, x_i)$$

(estimated) covariance of x_i and x_j

$$u_c^2(y) = \sum_{i=1}^N \left(\frac{\partial f}{\partial x_i} \right)^2 u^2(x_i) + 2 \sum_{i=1}^N \sum_{j=i+1}^N \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} u(x_i, x_j)$$

Mathematical model

the degree of correlation (**correlation coefficient**) between x_i and x_j can be estimated as:

$$r(x_i, x_j) = \frac{u(x_i, x_j)}{u(x_i)u(x_j)} \quad -1 \leq r(x_i, x_j) \leq +1$$

where $r(x_i, x_j) = r(x_j, x_i)$

= 0 if measurements x_i and x_j are **uncorrelated**

$$u_c^2(y) = \sum_{i=1}^N \left(\frac{\partial f}{\partial x_i} \right)^2 u^2(x_i) + 2 \sum_{i=1}^N \sum_{j=i+1}^N \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} u(x_i) u(x_j) r(x_i, x_j)$$

Law of Uncertainty Propagation (LUP)

Special cases

if measurements of all quantities x_i and x_j are **uncorrelated**:

$$u_c^2(y) = \sum_{i=1}^N \left(\frac{\partial f}{\partial x_i} \right)^2 u^2(x_i)$$

$r(x_i, x_j) = 0,$
 $i \neq j, i, j = 1, \dots, n$

worst-case:

since $|r(x_i, x_j)| \leq 1$:

$$u_c(y) \leq \sum_{i=1}^N \left| \frac{\partial f}{\partial x_i} \right| u(x_i)$$

similar expressions hold for **relative uncertainty**: $v(y) = \frac{u(y)}{y}$

Expanded uncertainty

assumptions:

- partial derivatives of $f(\cdot)$ exist; at least one of them significantly differs from zero
- $f(\cdot)$ is **fairly linear** about y
- none of the standard uncertainties $u(x_i)$ dominates the others
- the number N of input quantities is high enough

central limit theorem (CLT)
can be applied

measurement result y is almost **normally distributed**

expanded uncertainty $U(y) = k_p u_c(y)$

coverage factor
related to a given
coverage probability

provided by the LUP

Expanded uncertainty

if the above **constraints do not hold**,
the CLT may lead to
incorrect evaluation of **expanded uncertainty**

the pdf of y can be derived using **Monte Carlo simulations**

Supplement to the GUM issued by BIPM (2008)
<http://www.bipm.org/en/publications/guides/gum.html>

Numerical example

measurement of a **DC power** by measuring
DC voltage and DC current with **two different multimeters**

- input quantities: V and I
- mathematical model: $P = V \cdot I$
- only available information: **manufacturer's specifications**

measurement uncertainties of V and I
evaluated using a **type B** evaluation method

readings:

- $V = 8.0125$ V (10 V range)
- $I = 50.105$ mA (100 mA range)

measured power: $P = 0.4015$ W

Numerical example

voltage uncertainty evaluation

DC Voltage

Range	Resolution			Accuracy	
	Slow	Medium	Fast	(6 Months)	(1 Year)
300 mV	—	10 μ V	100 μ V	0.02 % + 2	0.025 % + 2
3 V	—	100 μ V	1 mV	0.02 % + 2	0.025 % + 2
30 V	—	1 mV	10 mV	0.02 % + 2	0.025 % + 2
300 V	—	10 mV	100 mV	0.02 % + 2	0.025 % + 2
1000 V	—	100 mV	1 V	0.02 % + 2	0.025 % + 2
100 mV	1 μ V	—	—	0.02 % + 6	0.025 % + 6
1000 mV	10 μ V	—	—	0.02 % + 6	0.025 % + 6
10 V	100 μ V	—	—	0.02 % + 6	0.025 % + 6
100 V	1 mV	—	—	0.02 % + 6	0.025 % + 6
1000 V	10 mV	—	—	0.02 % + 6	0.025 % + 6

$$\Delta = 2.2 \text{ mV}$$

$$u(V) = \frac{\Delta}{\sqrt{3}} = 1.3 \text{ mV}$$

Numerical example

current uncertainty evaluation

DC Current

Range	Resolution			Accuracy	Burden Voltage
	Slow	Medium	Fast		
30 mA	—	1 μ A	10 μ A	0.05 % + 3	0.45 V
100 mA	—	10 μ A	100 μ A	0.05 % + 2	1.4 V
10 A	—	1 mA	10 mA	0.2 % + 5	0.25 V
10 mA	100 nA	—	—	0.05 % +	0.14 V
100 mA	1 μ A	—	—	50.05 % + 5	1.4 V
10 A	100 μ A	—	—	0.2 % + 7	0.25 V

* Typical at full range

$$\Delta = 45 \mu\text{A}$$

$$u(I) = \frac{\Delta}{\sqrt{3}} = 26 \mu\text{A}$$

Numerical example

since two **different instruments** are used, voltage and current measurements can be assumed **uncorrelated**

$$\frac{\partial P}{\partial V} = I = 50.1 \text{ mA} \quad \frac{\partial P}{\partial I} = V = 8.01 \text{ V}$$

$$u_{C,unc}(P) = \sqrt{\left(\frac{\partial P}{\partial V}\right)^2 \cdot u^2(V) + \left(\frac{\partial P}{\partial I}\right)^2 \cdot u^2(I)} = 0.22 \text{ mW}$$

$\left(\frac{\partial P}{\partial V}\right)^2 \cdot u^2(V) \rightarrow 0.06 \text{ mW}$
 $\left(\frac{\partial P}{\partial I}\right)^2 \cdot u^2(I) \rightarrow 0.21 \text{ mW}$

e.g., because only one instrument is used to measure both voltage and current so that $r(V, I) \cong 1$

if **worst-case** correlation is considered:

$$u_{C,wc}(P) = \left| \frac{\partial P}{\partial V} \right| u(V) + \left| \frac{\partial P}{\partial I} \right| u(I) = 0.27 \text{ mW}$$

Special cases

uncorrelated inputs:

y	$p x_1$	$x_1 \pm x_2$	$x_1 \cdot x_2$	x_1/x_2	x_1^p
$u_C^2(y)$	$p^2 u_1^2$	$u_1^2 + u_2^2$	$x_2^2 u_1^2 + x_1^2 u_2^2$	$\frac{x_2^2 u_1^2 + x_1^2 u_2^2}{x_2^4}$	$p^2 x_1^{2(p-1)} u_1^2$
$v_C^2(y)$	v_1^2	$\frac{x_1^2 v_1^2 + x_2^2 v_2^2}{(x_1 \pm x_2)^2}$	$v_1^2 + v_2^2$	$v_1^2 + v_2^2$	$p^2 v_1^2$

worst case:

y	$p x_1$	$x_1 \pm x_2$	$x_1 \cdot x_2$	x_1 / x_2	x_1^p
$u_{C,wc}(y)$	$ p u_1$	$u_1 + u_2$	$ x_2 u_1 + x_1 u_2$	$\frac{ x_2 u_1 + x_1 u_2}{x_2^2}$	$ p x_1^{p-1} u_1$
$v_{C,wc}(y)$	v_1	$\frac{ x_1 v_1 + x_2 v_2}{ x_1 \pm x_2 }$	$v_1 + v_2$	$v_1 + v_2$	$ p v_1$

APPENDIX: GENERAL REQUIREMENTS for uncertainty evaluation methods

The requirements

- ✓ The method should be **universal**: it should be applicable to all kinds of measurements and types of input data
- ✓ The **quantity used to express uncertainty** should be:
 - **internally consistent**, i.e. directly derivable from the components that contribute to it, as well as independent of how these components are grouped and of the decomposition of the components into sub-components
 - **transferable**: it should be possible to use directly the uncertainty evaluated for one result as a component in evaluating the uncertainty of another measurement in which the first result is used

Assumptions

- The **true value** of the measurand is unknown and unknowable
⇒ the error approach can't be followed
- “the **distribution of values** that could reasonably be attributed to the measurand” needs to be determined
 - we don't know the value of the measurand
 - we don't know whether it belongs to that distribution or not
- The GUM states: “It is assumed that the result of a measurement has been **corrected for all recognized systematic effects** and that every effort has been made to identify such effects”(art. 3.2.4)