



# Tutorial review of error evaluation in experimental water research at the example of membrane filtration

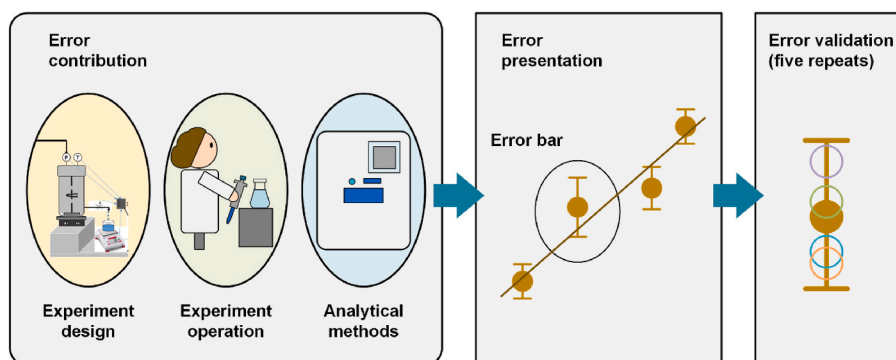
Alessandra Imbrogno, Minh N. Nguyen, Andrea I. Schäfer\*

*Institute for Advanced Membrane Technology (IAMT), Karlsruhe Institute of Technology (KIT), Hermann-von-Helmholtz-Platz 1, 76344, Eggenstein-Leopoldshafen, Germany*

## HIGHLIGHTS

- Error estimation methods are lacking in experimental water research.
- A practical method for experimental error estimation is proposed.
- Identification and quantification of error sources in membrane filtration is demonstrated.
- Error propagation is used to estimate experimental errors in membrane filtration.

## GRAPHICAL ABSTRACT



## ARTICLE INFO

Handling Editor: Dr Y Yeomin Yoon

### Keywords:

Error propagation  
Standard deviation  
Analytical error  
Calibration & validation  
Reproducibility & repeatability

## ABSTRACT

Experimental water research lacks clear methodology to estimate experimental error. Especially when natural waters are involved, the characterization tools bear method-specific artifacts while the varying environmental conditions prevent regular repeats. This tutorial review identifies common mistakes, and proposes a practical procedure to determine experimental errors at the example of membrane filtration. Statistical analysis is often applied to an insufficient number of repeated measurements, while not all error sources and contributions are considered. This results in an underestimation of the experimental error. Variations in relevant experimental parameters need to be investigated systematically, and the related errors are quantified as a half of the variation between the max and min values when standard deviation is not applicable. Error of calculated parameters (e.g. flux, pollutant removal and mass loss) is estimated by applying error propagation, where weighing contributions of the experimental parameters are considered. Appropriate judgment and five-time repetition of a selected experiment under identical conditions are proposed to validate the propagated experimental error. For validation, the five repeated data points should lie within the estimated error range of the error bar. The proposed error evaluation procedure is adaptable in experimental water research and intended for researchers to identify the contributing factors of an experimental error and carry out appropriate error quantification and validation. The most important aim is to raise awareness of the necessity to question error methodology and reproducibility of experimental data, to produce and publish high quality research.

\* Corresponding author.

E-mail address: [andrea.iris.schaefer@kit.edu](mailto:andrea.iris.schaefer@kit.edu) (A.I. Schäfer).

## 1. Introduction

Experimental error is defined as the difference between the measured value and the ‘best estimated value’ that is calculated as the average of multiple experimental measurements under the same conditions (Taylor, 1997; Gauch Jr, 2006; Vaux et al., 2012). Statistical methods (e.g. standard deviation, Grubbs’ tests for detecting outliers, and analysis of variance (ANOVA), see summary in Table S1) are used to verify differences between independent samples and determine error. However, such methods are applicable only when a large number of data points (30–40) is available (Liu et al., 2016; Böcking et al., 2019) and not for a few repetitions of an experiment (Haddaway and Verhoeven, 2015; Baker, 2016; Sekharan et al., 2022). Researchers are often under pressure to publish quickly in a culture of ‘publish or perish’ and even hype research data to emphasize novelty (Millar et al., 2020). This culture invites shortcuts in data handling (Baker, 2016; Beck et al., 2017; Edwards and Ashkanasy, 2018; Prager et al., 2019), specifically the lack of appropriate error analysis, data validation and transparent data reporting (Crawford et al., 2012; Stark, 2018; Winchester, 2018; Sholl, 2019; Le Phuong et al., 2020). In fact, several concerning reports of irreproducible data have been published (Baker, 2016; Winchester, 2018).

**Reproducibility** of data is defined as the possibility to obtain results for the same experiment or procedure by different operators, with different instruments in different facilities or laboratories (Reilly, 2007; Goodman et al., 2016). This concept differs from **repeatability**, which refers to the possibility of obtaining similar results by the same operator (in the same laboratory), with the same instrument and at the same operative conditions (Reilly, 2007). Repeatability evaluation allows the operator to know that the result is not random, while reproducibility evaluation enables results to be transferable between research groups, which benefits the research community. Besides providing information to the research community about the materials, instruments, experimental conditions, raw data and computational tools or software for analysis (this practice is termed **pre-reproducibility**) (Stark, 2018), a more transparent error analysis method helps the operators in achieving reproducible and repeatable data.

Water research encounters some challenges that render the

validation of reproducibility and error analysis difficult. These include: i) design of experiments and choice of experimental conditions may result in observed trends to be within the experimental error, ii) variability in human operation during the experiment, water sample analysis, data analysis, and reporting, iii) intrinsic variability of natural water composition and membrane material properties (porosity and pore size distribution). Many factors contribute to the fact that results from the experiments at the same conditions are not repeatable, requiring careful calibration and verification of interferences (from water components and potential artifacts) (Bisutti et al., 2004; Peng and Escobar, 2005; Lalia et al., 2013; Xu et al., 2013; Liang et al., 2022, 2023).

This review aims to: i) aid ‘sound judgment’ understanding of the main contributing factors to an experimental error in water research and particularly membrane filtration; ii) provide a practical procedure to calculate the error of unmeasured parameters (such as solute mass adsorbed, retention, and rate of disappearance) using error propagation (via ‘max-min’ or ‘statistical’ methods, depending on the quantity of data points taken at the same conditions); and iii) demonstrate the importance of validating the calculated error through five repeats of a selected experiment, which is either the experiment with an outlying result or the most representative experiment (of which results appear in multiple data sets). The error analysis procedure discussed in this tutorial review is summarized in Fig. 1. This study highlights the need of critical examination of data quality and data reporting in water research processes with membrane filtration as an example, while providing a practical and meaningful procedure to estimate experimental errors in the treatment processes. The experimental errors allow an evaluation of reproducibility and repeatability in the context of overwhelming literature on new membrane materials, complex water matrices and diverse process conditions.

## 2. Origin of experimental errors

Error sources that compromise the quality of results are categorized into: i) experimental design issues that cause under- or overestimated errors (such as inappropriate sensors used for experimental filtration control); ii) mistakes made by the operator due to inexperience or

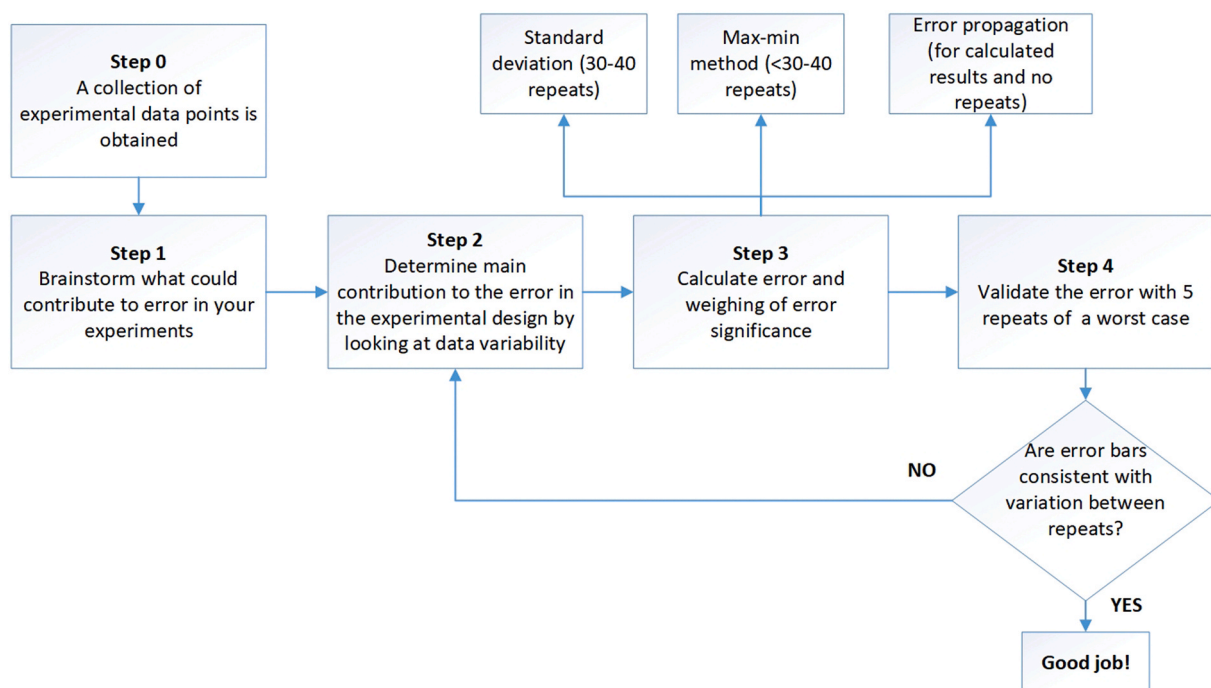


Fig. 1. Schematic of the proposed error analysis procedure.

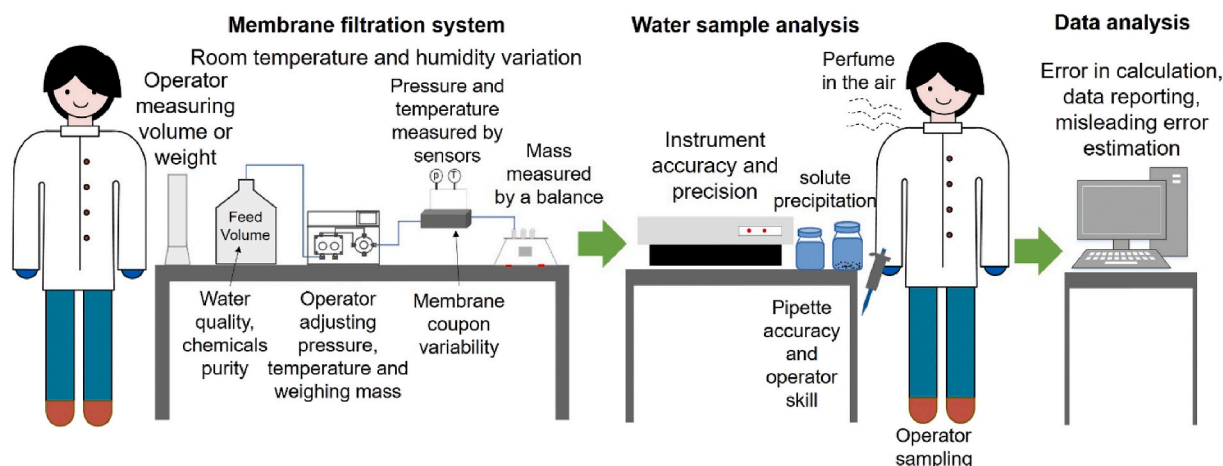


Fig. 2. Schematic of the error sources in a membrane filtration system, water sample analysis and data analysis.

inattentiveness; and iii) mistakes in calculations, simplified assumptions, and erroneous or ambiguous application of statistical analysis (for instance, the use of standard deviation, mean, and median in data reporting) (Vecchi et al., 2009; Ort et al., 2010; Beck et al., 2022). Fig. 2 schematically illustrates some sources of error associated with membrane filtration and water analysis for organic contaminants.

### 2.1. Error from experimental design

The experimental design involves the selection of experimental conditions (such as contaminant concentration, pressure, temperature and filtration volume) to observe significant variations of data points when the experimental conditions are varied. Identifying the trends in experiments is only possible if the differences between the data points throughout the data set (the data points reported in the exact figure) are larger than the associated errors. Under- or overestimating errors may effectively mask the relevant trends or result in misinterpretation of data.

A graphical representation of a hypothetical data set and different examples of error terms relevant in experimental research (accuracy, precision, repeatability, reproducibility, systematic error and random error), which causes data deviation from the hypothetical values, is

shown in Fig. 3. The 'true' value (Fig. 3A) of a measurement is a theoretical value that cannot be determined due to inevitable error in measurements. Hence only the best estimated value can be quantified from a repeated set of measurements. When a measurement is repeated at identical conditions, there is a chance that the points in the data sets vary from the 'true' value (accuracy in Fig. 3B) and each other (precision in Fig. 3C). Data accuracy indicates that measured values are close to the 'true' value (Gauch Jr, 2006; Miller and Miller, 2018). When the measurements are repeated multiple times under identical conditions, it is possible to determine the closeness of the repeated measurements to each other, which is defined as precision (Gauch Jr, 2006; Miller and Miller, 2018). The points in the data sets may not overlap due to unpredictable variations between data points from duplicate experiments. These variations are defined as random errors (JCGM, 2008; Miller and Miller, 2018). By repeating a chosen experiment or measurement multiple times, it is possible to quantify and reduce random errors. Some examples of random errors and minimization strategies are presented in Table S2. Outliers, which are data points that differ significantly from the data trend, may result from experimental error and can be identified by experiment repetition (Taylor, 1997). When the repeated measurements in the data sets vary marginally from each other (Fig. 1C), the data obtained have high repeatability (Reilly, 2007). Repeatability

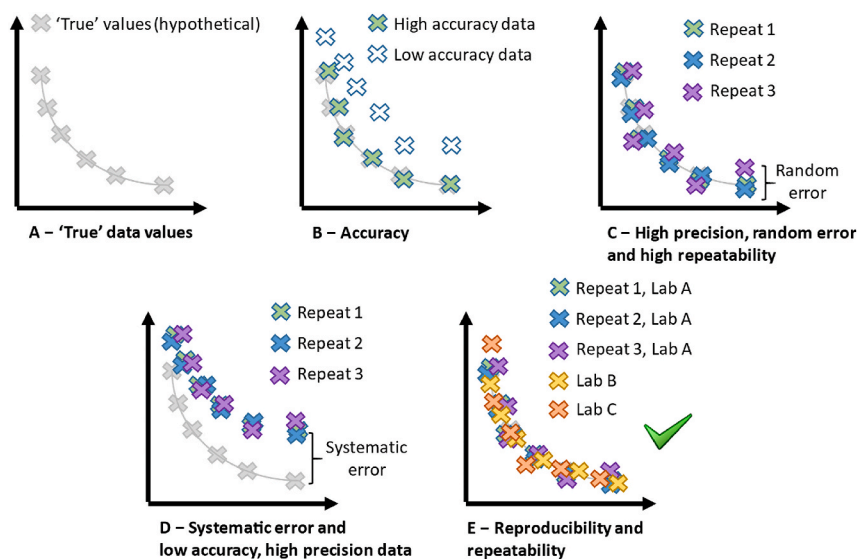


Fig. 3. Graphical illustration of A – true (hypothetical) values, B – high and low accuracy data values, C – data precision, random error, and repeatability, D – systematic error resulting in low quality data, and E – high reproducibility and high repeatability confirming data quality.

**Table 1**  
Error source identification in experimental design and examples.

Experimental step	Example of error source	How to identify and/or prevent the error
Sample preparation	Equipment tolerance (balance, pipette, glassware), dilution, sample contamination/matrix interference, glassware storage and cleaning	Check sources that introduce artifacts (pipetting, dilution, analyte loss, matrix compound, contamination). Monthly verification of equipment function (balance, pipette, pH meter and conductivity meter) with standards and recording of calibrations in lab/instrument books.
	Membrane coupon variability	Measure pure water flux before each experiment and discharge coupons with unusual permeability or evident damage.
Experimental system operation	Sensor tolerance (temperature, humidity and pressure sensors)	Monitoring of experimental parameters (such as fluctuation of sensor recording) and environmental conditions (lab humidity and temperature). Stepwise permeate mass increase is observed if the balance is not sensitive enough.
	Inadequate balance sensitivity	Compare the expected and measured value of an operative parameter (e.g. pressure, pump flow rate and temperature) with other functional sensors.
	Variation in the operative condition setting	
Analytical instrument operation	Analytical instrument error	Perform instrument calibration with standards before sample analysis and periodic monitoring of equipment performance with quality standards. Determine the limit of detection and quantification systematically.
	Interference of sample matrix or sample preparation - use of filters to remove organic particulates in sample preparation can cause loss of trace contaminants (McMahon, 1973; Horowitz et al., 1992; Nguyen et al., 2020) Contamination from perfume, insect repellent, organic solvents, detergents, building materials	Systematically identify interference of water matrix (salinity, organic matter, ions, etc.) in calibration (see for example, Table S7).  Avoid the use of volatile compounds in water analytical lab that analyses organic micropollutants. Regular analysis of blank.

refers to the possibility of obtaining similar results by the same operator (in the same laboratory) with the same instrument and at the same operative conditions (Reilly, 2007). Repeatability evaluation allows the operator to know that the result is not random.

However, repeatable data can deviate from the (hypothetical) ‘true’ value due to a systematic error in the measurement (Fig. 3D). **Systematic errors** occur when multiple measured values deviate in the same pattern from the true value and they are not minimized by repeating the experiment (JCGM, 2008; Miller and Miller, 2018). Such errors may arise from problematic sensor recording or inaccurate analytical equipment settings and calibration issues; examples are reported in Table S2. The experimental design or analysis needs to be revised, although it may be challenging to identify some sources of systematic errors.

Performing experiments with many uncontrolled operational parameters leads to a large data variability and, ultimately, irreproducibility between different laboratories (Cox, 1961; Moffat, 1985; Stark, 2018). Examples of errors encountered in experimental design are shown in Table 1. Such errors can be categorized into sample preparation, experimental and analytical operation errors. Some examples of error sources and how to minimize these errors in membrane filtration research are shown in Table S2. Measures to reduce experimental errors are summarized in Table S3.

Error sources do not contribute equally to the experimental error and a ‘weight’ can be assigned to each error. The weight is related to the significance of a contributing error source and it should be calculated with partial derivatives (where the relationship between a variable parameter and the target result is determined while all the other parameters are kept constant) (Ku, 1966; Dietrich, 2017). However, it is not practical to determine the weight with a partial derivative for each operative parameter by performing intensive experimental investigation or complicated calculations. Instead, the weighing needs to be rooted in sound judgment.

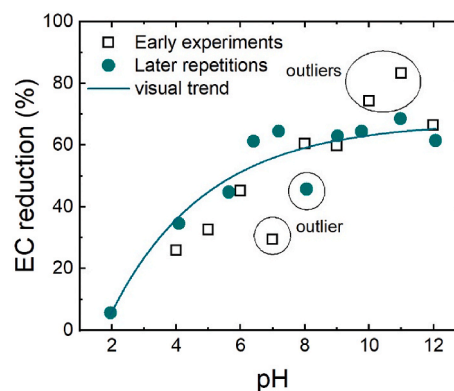
A more practical procedure based on ‘sound judgment’ is hence proposed: i) identify all the parameters that may contribute to the experimental error by looking at the error sources in the experimental design; ii) calculate the variation of the parameters between experiments and report as absolute and relative error from subtracting the maximum to the minimum values (namely the max-min method); iii) apply a common-sense practice where parameters with larger variations may likely contribute more to the experimental error; iv) calculate the

final experimental error with the error propagation where the variable parameters with the larger variation are accounted.

## 2.2. Experimental error caused by the operator

Error coming from the operator (typically omitted in error calculations) can occur in various stages, such as mistakes in feed preparation, inconsistency in manual setting of operative conditions, and sampling time deviated from the experiment plan (Kotek and Mukhametzianova, 2012; Kuselman et al., 2013; Di Pasquale et al., 2015; Kim et al., 2015; Liang et al., 2022). Inexperienced (or careless) operators are prone to making more mistakes (Rao, 1993). To minimize human operator error, it is necessary to: i) train and supervise the operators, ii) quickly identify the error or mistake by documenting the experimental process, iii) obtain a sufficient number of data points in a figure such that both the trend and scattering in data are visible (7–8 data points are suggested), and iv) perform repeats of data points to verify potential outliers.

An example of an ‘inexperienced’ operator is shown in Fig. 4. Dead-end filtration experiments (with a stirred cell) performed to remove salt ions and reduce electrical conductivity (EC) at the beginning of doctoral research were repeated later. The data reported as ‘later repetitions’



**Fig. 4.** Comparison of experiments performed by an inexperienced researcher and repeated once experienced (NF270 membrane, 9.8 bar, fluoride 50 mg/L and humic acid 80 mgC/L, NaHCO<sub>3</sub> 1 mM, NaCl 10 mM). Possible outliers are circled. Adapted from Owusu-Agyeman et al. (Owusu-Agyeman et al., 2017).

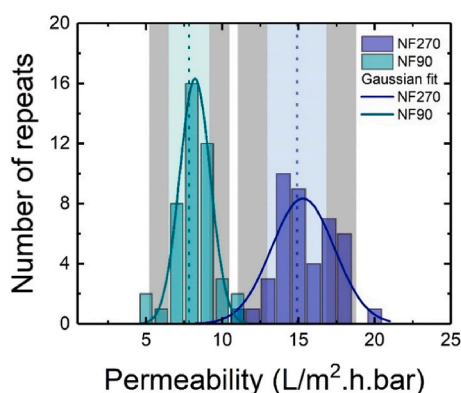
were adapted from Owusu-Agyeman et al. (2017), while the ‘early’ data set was not published. Outliers should not normally be removed from a data set in a publication; however, in this case, the whole data set was omitted by the operator due to inexperience.

Data scattering and variability are more significant for the ‘early’ experiments. When the experiments were repeated later in the project, the data trend improved as the student was trained and more experienced.

### 2.3. Error caused by variability between experiments

An example of repeated measurements in membrane research is the pure water permeability, which displays a normal (or Gaussian) distribution due to the experimental and coupon variability. Fig. 5 shows the permeability variation for commercial nanofiltration coupons (NF90 and NF270). A **normal distribution** is obtained where the repeated measurements vary equally above and below the mean value. If each filtration experiment procedure includes a pure water permeability measurement, sufficient permeability data to carry out statistical analysis are available throughout 30–40 filtration experiments. Permeability varies for each membrane due to random errors from the membrane coupon variability and the filtration operation (such as pressure adjustment, temperature, pure water quality and surrounding conditions).

When the permeability is measured with different coupons and under the same experimental conditions, many similar data points exist. This results in a Gaussian (or normal) distribution of permeability where the error can be quantified using standard deviation. The **standard deviation** indicates how much the data are spread from the mean value in a Gaussian distribution (Box et al., 1978; Taylor, 1997; International Organization for Standardization, 2006). The water permeability error, in this case, is reported as the standard deviation (with the same units as the permeability) of  $\pm 2.0$  and  $\pm 1.4$  L/m<sup>2</sup>.h.bar for NF270 and NF90, respectively. The average permeability (arithmetic mean) values of NF270 and NF90 are  $14.9 \pm 2.0$  and  $7.8 \pm 1.4$  L/m<sup>2</sup>.h.bar, respectively, which can be accepted as ‘best estimated value’. Considering the extreme values of the normal distribution curve in Fig. 5, it can be concluded that NF270 coupons with values below 11 and above 19 L/m<sup>2</sup>.h.bar and NF90 membrane coupons with permeability below 5 and above 12 L/m<sup>2</sup>.h.bar (more than 25–30% variation from the mean value) should be discarded to minimize the random errors.



**Fig. 5.** Gaussian distribution of the pure water permeability obtained from multiple measurements for NF270 (42 repetitions) and NF90 membranes (45 repetitions). Green and blue shades indicate the standard deviation of the permeability values around the mean. Grey shades indicate the respective permeability range in which 95% of values were contained. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

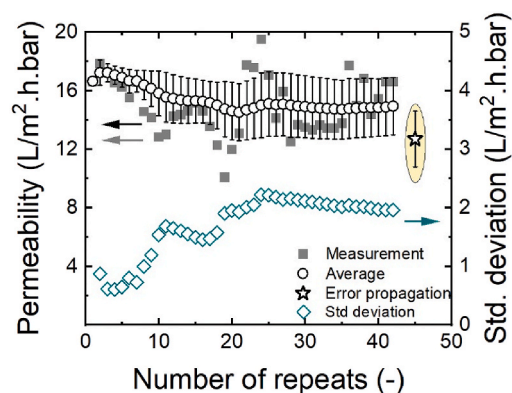
### 2.4. Incorrect use of error analysis methods in data reporting

Issues that fall in this category include: i) reporting significant digits of experimental data with exaggerated decimal numbers; ii) applying improperly statistical analysis (see Table S1 for examples and conditions); and iii) applying incorrect error propagation for non-statistical errors (Jochum et al., 1981; Tellinghuisen, 2001; Lee and Chen, 2008; Vaux et al., 2012).

Specific guidelines are reported to define the number of significant digits in a measurement and report significant digits, such that the reported number of significant digits is meaningful and the error value has the same decimal place as the measured value (Taylor, 1997; Dym, 2004). Examples and general rules are shown in Table S4. Fig. 6 provides an example of when statistical analysis is applied to insufficient experimental data. Permeability from 40 repetitions with different membrane coupons of the same type (NF270) are reported as grey squares. The mean/average values with error bars as standard deviation from varying number of repeats are reported as open circles. Standard deviation values are shown as open diamonds. The star data point corresponds to the permeability of a membrane coupon with an error bar calculated using the error propagation method as described later.

When standard deviation is applied to a number of repeats below 20, an underestimation of error can occur, as shown by the small error bars that do not cover the data variability range (grey data points). This is significant especially for numbers of repeats below 10, where the standard deviation reported as an error bar results in significant error underestimation by a factor of 4. By increasing the number of repeats, the error in permeability increases, and for repetitions above 30, the error in permeability becomes constant as the error bars with 30–40 repetitions do not change significantly. This is consistent with literature, where it has been reported that a quantification of experimental error as standard deviation is valid when: i) a minimum number of repeats at the same operative conditions is at least 30 to 40 as a ‘rule of thumb’ (Greenwood and Sandomire, 1950; Ku, 1966; Van Belle, 2011), and ii) a normal or Gaussian distribution is obtained when the frequency of results are plotted (Taylor, 1997; Gauch Jr, 2006). However, it is impossible in water research to repeat long and complex experiments 30–40 times. Hence, this method is restricted to a limited number of parameters, while an alternative approach is required for more typical experiments.

Such an alternative is the error propagation method, which can be applied to estimate the error bar for a single measurement (the star data point). It is important to note that the error bars calculated with the error propagation represent the ‘best estimated value’ (or average) and indicate where this value is likely to be located. The error propagation method will be described in the next section. A good error analysis results in an error bar similar to the error bar obtained from multiple repeats, and the expected value of the membrane permeability lies within the error bar of the star data point calculated by error propagation. Error



**Fig. 6.** Variation of NF270 water permeability and SD as a function of the number of repeated measurements. Error bars represent the standard deviation.

propagation can be applied in experimental research for error estimation when repetition of an experiment is not possible due to time- and resources constraints. In summary, factors that introduce error and how to reduce it must be identified before an experiment.

### 3. Error quantification and validation

In this section, an overview is given on: i) error quantification as absolute and relative errors for experimental parameters, ii) standard deviation, median and average for repeated measurements in analytical calibration, and iii) error propagation method for calculated parameters.

#### 3.1. Absolute and relative errors

Errors from measured parameters resulting from sample preparation (such as mass weight with analytical balance, measuring volume with cylinders or volumetric flasks), filtration operation (for example, pressure, temperature, pH, conductivity variation from respective sensors), and sample analysis (for example, concentration) can be reported as absolute or relative errors. The **absolute error** ( $\Delta y_{abs}$ ) is the difference between the measured value and the best estimate of the 'true' value (Taylor, 1997; JCGM, 2008). The absolute error has the same unit of the measured value and it is calculated with Eq. (1) (Taylor, 1997).

$$\Delta y_{abs} = |y_{exp} - \bar{y}| \quad (1)$$

where  $y_{exp}$  is the measured value and  $\bar{y}$  is the absolute arithmetic average of multiple repeated measurements at the same conditions. Absolute errors can be used as individual experimental errors of measured repeated parameters (such as pressure, temperature and permeate mass). Repetition of measurements or experiments may increase the accuracy as the mean or average value is closer to the 'true' value (Gauch Jr, 2006; Miller and Miller, 2018).

The **relative error** ( $\Delta y_{rel}$ , can be in %) is the ratio between the absolute error and the best estimated value (Taylor, 1997; JCGM, 2008). It is used in this work in error propagation because absolute errors from different quantities and units cannot be summed together. The relative error is calculated with Eq. (2) (Taylor, 1997).

$$\Delta y_{rel}(\%) = \frac{\Delta y_{abs}}{\bar{y}} \cdot 100 \quad (2)$$

For measurable parameters, the tolerance of the measuring device used for a measurement (such as a volumetric flask to prepare a solution or a pipette used to dose a small volume of solution) should be considered as this contributes as an error source to the measurement quality. **Tolerance** of a measuring device is defined as half the difference between the highest and lowest value that can be measured with the device at the same condition, and it is often reported in the manufacturer specifications. While the tolerance reported by manufacturers does not specify errors dependent on the specific application (for example, the tolerance reported on a measuring cylinder may be valid only for water at a certain temperature), this can help operators to appropriately select the measuring device. Choosing devices with the lowest tolerance is a way to reduce measurement errors and improve accuracy. For instance, the volume measurement accuracy of  $\pm 1$  to  $\pm 5\%$  for glassware, such as graduated cylinders and beakers, affects the final measured concentration by the same range (1–5%) (Liang et al., 2022). The chosen tolerance must be appropriate to the purpose, as unnecessary low tolerance will result in excessive cost.

#### 3.2. Standard deviation, mean and middle value of measurable parameters

The standard deviation ( $\sigma$ ) of multiple measurements (plotted in a Gaussian distribution as shown in Figs. S1 and S2) can be reported as absolute error and is calculated with Eq. (3) (Box et al., 1978; Taylor,

1997; International Organization for Standardization, 2006);

$$\Delta y_{abs} = \sigma = \sqrt{\frac{\sum (y_i - \bar{y})^2}{N}} \quad (3)$$

where  $y_i$  are the measured values,  $\bar{y}$  is the arithmetic mean value and  $N$  is the number of repeated measurements (Taylor, 1997). Errors determined with statistical analysis (e.g. standard deviation) are classified as 'Type A' by the 'Guide to the Expression of Uncertainty in Measurement' (Woolliams, 2013). The **arithmetic mean** is the sum of all measurements divided by the number of measurements. If the number of repeats is below 30, half the difference between the maximum ( $y_{max}$ ) and minimum ( $y_{min}$ ) measured values (namely the **max-min method**) can be reported instead of standard deviation and it is considered as an absolute error ( $\Delta y_{abs}$ ) calculated with Eq. (4) (Krzywinski and Altman, 2013);

$$\Delta y_{abs} = \frac{y_{max} - y_{min}}{2} \quad (4)$$

The main difference between the max-min method and standard deviation is that the middle value ( $y_{mid}$ ) is used as representative experimental data instead of the arithmetic mean value ( $\bar{y}$ ) used in statistical analysis. The value of  $y_{mid}$  is calculated with Eq. (5) (Miller and Miller, 2018);

$$y_{mid} = \frac{y_{max} + y_{min}}{2} (\neq \bar{y}) \quad (5)$$

#### 3.3. Error quantification for analytical tools

In environmental/water research, the data quality depends often on the quality of the analytical measurements. Performing instrument calibration using known standards is relevant to determine accuracy and validate analytical results. In this section, different instrument calibrations are used to show how experimental error of standard measurements is influenced by the instrument accuracy, precision and the range of concentration applied.

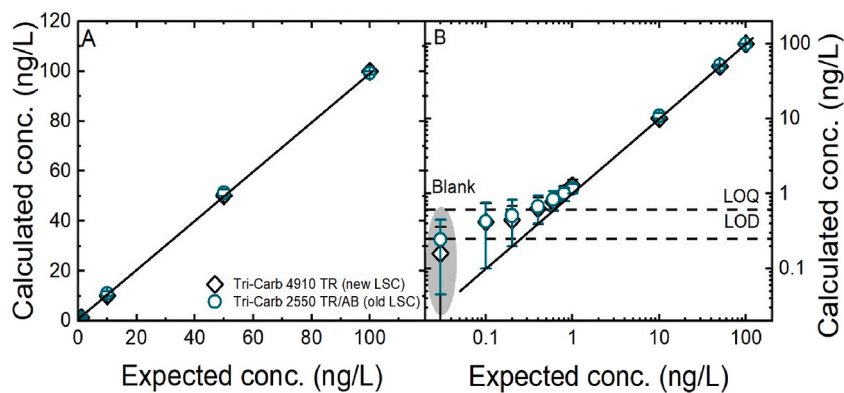
To calibrate an analytical instrument used in water analysis, the instrument signal response (in y-axis) of a particular analyte used as standard is plotted against the known concentration of the analyte solution (in x-axis). The linear relationship (regression) is valid for a limited range of concentrations, and the **limit of detection** (LOD) and **limit of quantification** (LOQ) determine the values where the measured values deviate from linear regression (Committee, 1987; Shrivastava and Gupta, 2011). LOD is defined as the lowest concentration at which the analyte can be distinguished from the blank, while LOQ is the lowest concentration where the analyte can be quantified in linear regression (Shrivastava and Gupta, 2011; Miller and Miller, 2018). The analytical error is determined as standard deviation by plotting the repetitions of quality standard measurements in a normal distribution (for examples, see Figs. S1 and S2).

Two liquid scintillation counters (LSC), two total organic carbon (TOC) analyzers, and a liquid chromatography-organic carbon detector (LC-OCD) were selected as examples of water analytical tools commonly used in water research applied to membrane filtration. Details of the standard type, range of concentration and instrument model are reported in Table S5.

##### 3.3.1. Liquid scintillation counter (LSC)

The standard deviation for each radiolabeled estradiol concentration was calculated using Eq. (5) and it was applied as an error bar in the calibrations for LSC reported in Fig. 7A. The calibration data are reported in the log-scale in Fig. 7B for two LSC models (instrument details are reported in Table S5) to visualize the deviation from the linear regression when the measured concentration approaches LOD and make the error bars more visible compared to Fig. 7A.

The most precise results are obtained in the concentration range from 1 to 100 ng/L with standard deviation values in the range of  $\pm 0.13$  ng/L



**Fig. 7.** A) Calculated radiolabeled estradiol concentration from measured radioactivity as a function of expected concentration in the 0 (blank) to 100 ng/L range. B) Log-scale representation of calibration reported to show LOD. Error bars are the standard deviation from the normal distribution in Fig. S1.

(13%) at 1.00 ng/L and 2 ng/L (2%) at 100 ng/L. Below 1 ng/L, the mean concentration deviates from the linear regression, which means the instrument has lower accuracy at such a low concentration range. The relative error increases with decreasing concentration, reaching 60% at 0.4 ng/L. The largest error bars are observed when the measurement approaches 0.2 ng/L with errors of  $\pm 0.3$  (150%) at concentrations below 0.2 ng/L. This indicates that the experimental measurements in the sub-ng/L concentration range give the largest error. If the relative error is over 100%, the analytical result is not significant and should be reported as below LOD. From visual inspection, the LOD should be above 0.2 ng/L. Similarly, the LOQ is visible from the deviations of measured data from the linear regression for both LSC instruments at concentrations approaching 0.4 ng/L. Both the LOD and LOQ estimated from visual inspection should be similar to the calculated LOD (0.2–0.3 ng/L) and LOQ (0.4–0.6 ng/L) as reported in Table S6.

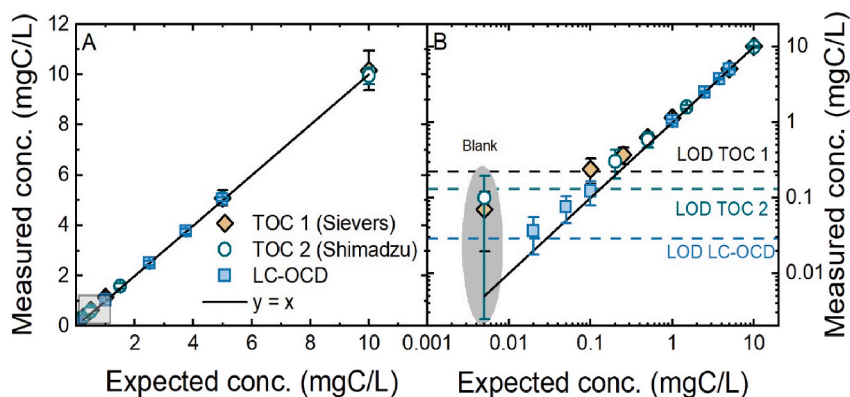
### 3.3.2. Organic carbon analyzers

Total organic carbon (TOC) analyzers are used to measure the total organic and inorganic (such as carbon acid salts and dissolved carbon dioxide) dissolved carbon concentration in water samples without distinguishing between different carbon sources. Liquid chromatography–organic carbon detector (LC-OCD) combines organic carbon detection with liquid chromatography for the separation and quantification of organic matter fractions with different molecular weights. The need to determine natural organic matter fractions at low concentrations requires LC-OCD to have a lower detection limit (25  $\mu\text{gC/L}$ ) compared to TOC (0.2 mgC/L) (Huber et al., 2011). Interferences from the natural water matrix, such as salts, microbial contamination and organic compounds (see Table S7), can introduce errors in the analysis and this results in erratic concentrations. For this reason, it is necessary to calibrate

with standards of known concentrations to evaluate instrument performance, and then standards with different water matrix compositions to identify possible interference. Calibrations and standard deviations for TOC (two instruments) and LC-OCD instrument are reported in Fig. 8.

The standard deviation (reported as an error bar) decreases with increasing TOC concentration when approaching LOD, especially for TOC 1 and 2 (model type in Table S5). LOD can be identified for both TOC analyzers at a concentration of  $0.20 \pm 0.05$  mgC/L because the TOC concentration lies within the blank measurement and is not in the linear regression range.

LOD is visible at a lower value between 0.01 and 0.02 mgC/L (10–20  $\mu\text{gC/L}$ ) for LC-OCD as the data point deviates from the linear regression (Fig. 8) and the signal cannot be distinguished from the background noise at a concentration below 0.01 mgC/L in Fig. S3. LOQ is obtained at 0.05 mgC/L (50  $\mu\text{gC/L}$ ) as reported in Table S6 and calculated with Eq. S1. This result indicates that LC-OCD is more sensitive than TOC in detecting organic carbon at concentrations below 0.2 mgC/L. Huber et al. (Huber and Frimmel, 1991) obtained a lower LOD (1.5  $\mu\text{gC/L}$ ) and LOQ (5  $\mu\text{gC/L}$ ) that were calculated as 3 and 10 times the standard deviation of blank measurements, respectively, by using the same standard (potassium hydrogen phthalate) as this work. The different LOD and LOQ may be attributed to a different system model and set-up, which results in different instrument sensitivity and higher expertise in operation. LC-OCD is used for organic carbon fractionations and real water quality analysis, where in some cases, organic fractions (e.g. phenol-rich compounds) can be adsorbed by the LC column, leading to variation in the range of concentrations that can be measured with the OCD. This mass loss can be identified by cross-checking the total concentration from LC-OCD with other instruments (TOC). LC-OCD is sensitive to environmental contamination, such as fumes from laboratory or



**Fig. 8.** A) Measured TOC concentration as a function of potassium hydrogen phthalate (KHP) standard concentration in the 0.002–10 mgC/L range. B) log scale representation of data points to show LOD. Error bars are the standard deviation from the normal distribution in Fig. S2.

**Table 2**

Error propagation for sum, difference, product, and quotient of measured quantities via quadratic sum (Taylor, 1997).

Formula	Error formula	Examples
$y = a + b$ $y = a - b$	$\Delta y_{abs} = \sqrt{\Delta a_{abs}^2 + \Delta b_{abs}^2}$	Error in mass adsorbed (based on mass balance)
$y = a \cdot b$ $y = \frac{a}{b}$	$\Delta y_{rel} = \frac{\Delta y_{abs}}{y} = \sqrt{\left(\frac{\Delta a_{abs}}{a}\right)^2 + \left(\frac{\Delta b_{abs}}{b}\right)^2}$	Error in solute retention and rate of disappearance

building materials, or operators' perfume.

### 3.4. Error quantification for calculated parameters

In water research and specifically membrane filtration, many performance parameters are not measured directly (e.g. concentration, permeability, solute retention and mass adsorbed), or the number of repeats is not sufficient to calculate standard deviation. In this case, the error is obtained by the propagation of the errors quantified from the measured quantities (i.e. pressure, mass, volume, activity, and peak area) (Taylor, 1997). Error propagation can be classified as an example of 'Type B' uncertainty, which includes error evaluations with non-statistical analysis (Woolliams, 2013). **Error propagation** is based on the quadratic sum, as shown in Eq. (6) (Taylor, 1997). This method is applied in this study to determine the error for calculated parameters, like the solute retention, mass adsorbed and rate of disappearance.

$$\Delta y_{abs} = \sqrt{\left(\frac{\partial y}{\partial a} \Delta a_{abs}\right)^2 + \left(\frac{\partial y}{\partial b} \Delta b_{abs}\right)^2 + \dots} \quad (6)$$

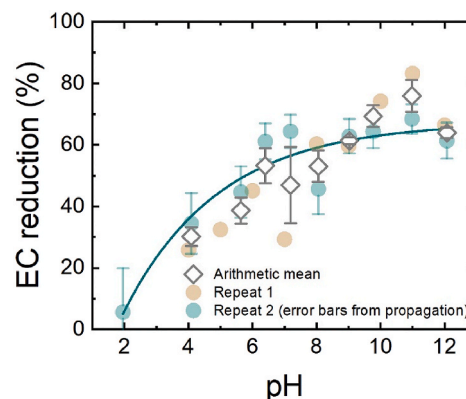
This equation can be adapted to any calculated parameter based on the relation between the experimental quantities ( $a$  and  $b$ ). An overview of quadratic sum applied for sum, difference, product and quotient of measured quantities is given in Table 2. Other equations to calculate absolute error in error propagation for sum, difference, product and quotient of measured quantities ( $a$ ,  $b$ ) are reported in Table S9.

Only the most contributing error sources (with more substantial weights) need to be considered in error propagation to simplify calculations and report comprehensible error analysis. Appropriate judgment and experience must be applied to evaluate the weighting of different parameters based on the variation of individual parameters between experiments. Error parameters considered in error propagation and added into the quadratic sum (for the selected examples) are: i) analytical error and error of feed preparation by the operator, ii) error due to water flux variation, iii) analytical error of permeate concentration. However, the weighting varies case by case, as this will be described in the case studies section. It must be remembered that analytical error is often very high when approaching the LOD and this plays an important role in water research.

### 3.5. Error validation

Errors calculated by error propagation need to be validated and consistent with the data quality. Error validation can be performed by repeating an experiment (for example, the 'core result' that appears in many graphs or the result that does not follow the visual data trend in a figure and is a suspected outlier) to evaluate if the error calculated covers the data points of the other repeated values. The number of repetitions is selected based on the relation between uncertainty ( $= 1 / \sqrt{n}$ ) on a repeated data point and the number of repetitions ( $n$ ), as reported in Fig. S6. From this relation, five repetitions are acceptable to reduce the error of a repeated result by 2.5 times.

An example of inconsistency between the experimental data scatter and the error bars calculated as standard deviation of two repetitions is given in Fig. 9. The data reported are the same as in Fig. 4, while the data quality of Repeat 1 is much worse than that of Repeat 2. The average



**Fig. 9.** Example of inconsistent data scatter and error bars. Error bars are calculated as standard deviation from two repetitions (EC = electrical conductivity related  $\text{NaHCO}_3$  concentration in water, NF270, 9.8 bar, fluoride conc. 50 mg/L and humic acid conc. 80 mgC/L). Data are adapted from (Owusu-Agyeman et al., 2017).

(grey open diamonds) was calculated from the two data sets. The grey error bars are calculated from the standard deviations between two data sets (which is an erroneous practice). In contrast, the green error bars on Repeat 2 were calculated using error propagation for the retention parameter (see the case studies later). The green curve indicates the trend and was plotted for Repeat 2 results.

Fig. 9 shows that the error bars calculated from averaging the two data sets and the standard deviation (grey error bars) are inconsistent with the data trend (green curve). Firstly, the error bar at pH 7 is very large because the value in the initial data set (Repeat 1), which is an outlier, is considered in the calculation. Secondly, the grey error bars at pH 4, 6, 10 and 11 estimated as standard deviations are too small compared to the green curve. This suggests that the error calculation method needs to be corrected. When error calculation is performed using the error propagation method and without the outliers in the calculation, the green error bars are more consistent with the data set. In conclusion, repetition of experiments helps identifying outliers in the data set and evaluate if the calculated error bars are consistent with the variability of the visual trend. Although outliers cannot be omitted in a data set, they can be identified as outliers and should not be considered in the error calculation. Data points suspected to be outliers need to be repeated. In summary, at least one experiment must be repeated multiple times to validate the error analysis.

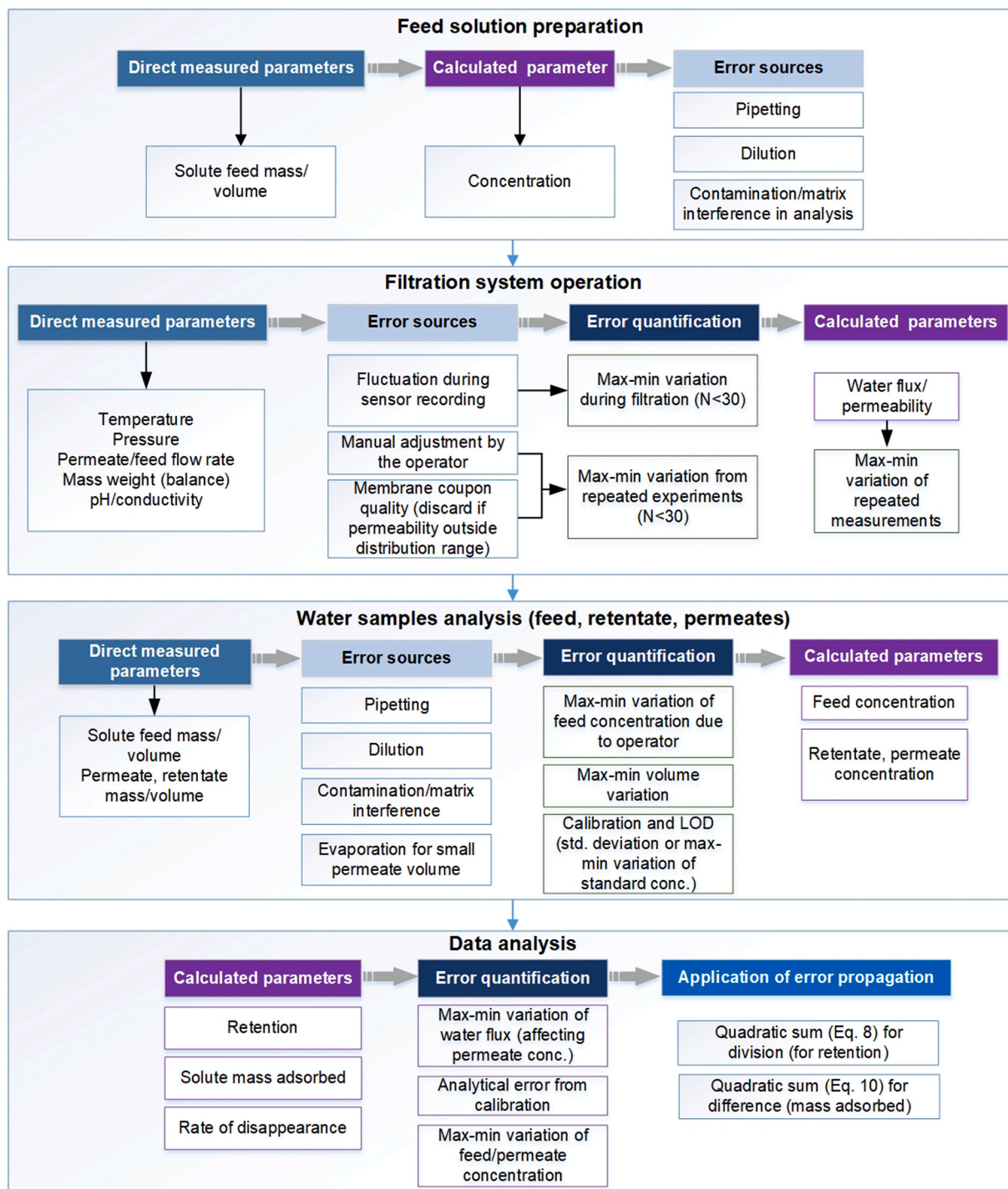
## 4. Case studies of error analysis in membrane filtration

The proposed error methodology is applied in different examples of membrane filtration experiments where error propagation is used to calculate the experimental error of selected example parameters: i) retention of solute (where a solute is typically a contaminant such as radiolabeled estradiol, organic matter), ii) solute mass adsorbed, and iii) rate of disappearance of a solute. The selection of specific examples and justification is summarized in Table 3.



**Table 3**  
Examples selected to calculate the error for retention, mass adsorbed and, rate of disappearance.

Calculated parameter	Example	Justification of example selection
Solute retention	Retention of organic tracers (measured with TOC) and estradiol (measured with LSC)	Contribution of instrument accuracy (high for LSC and low for TOC) on the propagated error
Solute mass adsorbed	Estradiol and humic acid mass adsorbed	Contribution of instrument accuracy and error by the operator on the feed solution preparation
Rate of disappearance of a solute	Estradiol rate of disappearance by a photocatalytic membrane operating in flow-through mode	Contribution from removal, feed concentration



**Fig. 10.** Schematic of measured and calculated quantities in membrane filtration experiments and related error sources/quantification considered in the error propagation.

#### 4.1. Overview of error source determination and quantification

In a membrane filtration experiment, the measured (from sensors and balance) and calculated parameters, as well as other sources considered in the error propagation stemming from feed solution preparation, filtration system operation and water sample analysis are shown in Fig. 10.

The starting point of error estimation is to identify relevant error sources for a specific experiment and parameter (such as flux, retention, solute mass adsorbed etc.). The contributing factors to the error of the filtration system parameters (for example, water flux) are the variation of measured quantities, such as pressure, temperature (which affects water viscosity and density), permeate mass, flow rate and the manual adjusting of the pressure or the flow rate by the operator. These errors can be reported as absolute or relative errors and are calculated as max-min variation, as reported in Table S8.

In the case of an error in the calculated parameters (such as retention, solute mass adsorbed, rate of disappearance), the error of permeate concentration due to variation of water flux, the analytical device accuracy (especially for an instrument with poor accuracy such as TOC) and the operator who prepares the feed solutions and calibrates the device are important error sources. The relative error of the concentration analysis is especially high when the values are close to the instrument detection limit. These errors are included in the error propagation method by applying specific equations according to the relationship between these parameters. Retention, solute mass adsorbed and rate of disappearance are some examples. Details on how to use the error propagation are given in the sections below.

#### 4.2. Error analysis for water flux variation

Water flux in a filtration experiment can be controlled with two different system designs, where either i) the transmembrane (or feed) pressure is adjusted manually to be constant over time ('constant pressure'), or ii) the permeate flow is kept constant by a pump and the pressure varies over time ('constant permeate flow'). Temperature variation (more likely if there is no water chiller to control temperature)

causes variation in water viscosity (which affects flux), the intrinsic variability of membrane coupon quality, and operative error sources (precision in pressure adjustment or variation in peristaltic pump tubing) can contribute to the experimental variation of water flux. In constant pressure mode, the manual adjustment of the pressure by the operator needs to be considered as an error source, while in the constant permeate flow mode, the pressure is not manually set, but the pump fluctuation and operation may cause an error in the permeate mass and pressure variation over time.

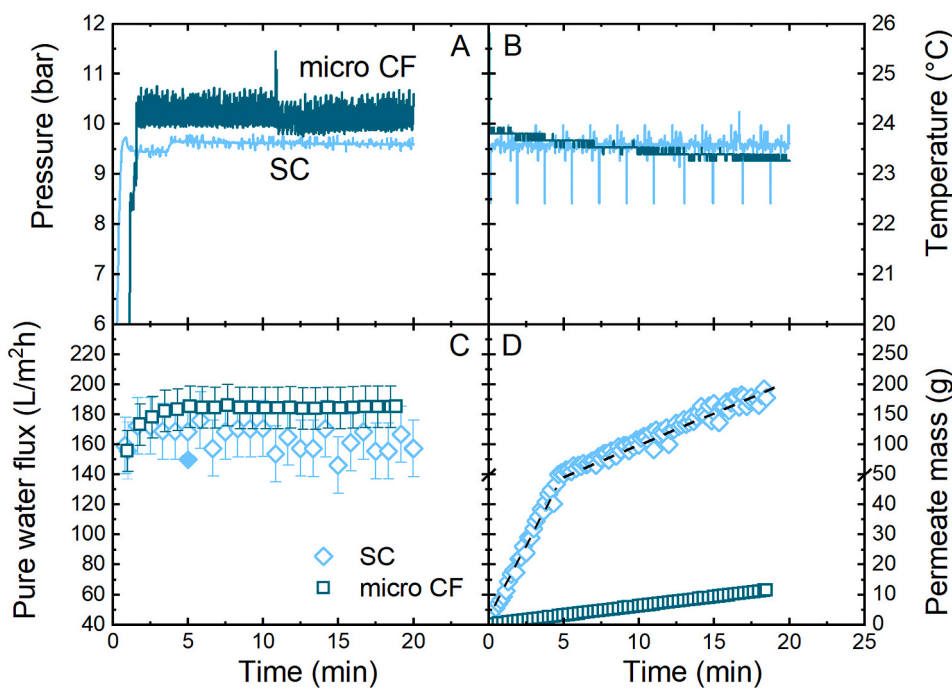
Water flow through a porous membrane under a pressure gradient is described by Darcy's law, where the water flow ( $J_v$ , m/s) is related to the dynamic water viscosity ( $\eta$ , Pa.s) and transmembrane pressure (TMP, Pa) as reported in Eq. (7) (Baker, 2004; Alvarez-Quintana et al., 2020);

$$J_v = \frac{k_D}{\eta} \cdot \frac{TMP}{h_m} \quad (7)$$

where  $k_D$  ( $m^2$ ) is the Darcy permeability constant and  $h_m$  is the membrane thickness (m). Darcy's law assumption is to have a laminar flow through the porous membrane (Reynolds number  $<1$ ), which is valid for most membranes.

An example of variations in operative conditions (e.g. pressure, temperature, permeate mass) over time and the calculated water flux for one filtration experiment performed in a pressurized stirred cell (SC) and micro cross-flow (micro CF) system in 'constant pressure mode' is reported in Fig. 11. The variation of the operative conditions over time is calculated as absolute error using the max-min variation to evaluate the main contributors to the error of water flux measured in systems with different pressure settings (flow control valve in the retentate side in micro CF and a pressure control valve in the pressured SC cell). Details of the system design and characteristics have been published previously (Imbrogno and Schäfer, 2019). Pure water flux was measured for 20 min after membrane compaction of 1 h at 10 bar to ensure that water flux and pressure were stable over time.

Several particularities are observed in Fig. 11. The pressure fluctuation is larger in the micro CF ( $10.1 \pm 0.9$  bar) compared to the SC ( $9.6 \pm 0.2$  bar) (Fig. 11A). This is a systematic error caused by the peristaltic pump pulsation on the feed/retentate site where the water is



**Fig. 11.** Variation of measured pressure (A), temperature (B), calculated pure water flux (C) and permeate mass (D) during water filtration using micro CF and SC systems (NF270 membrane, Milli-Q water, 10 and 9.6 bar in micro CF and SC, 23 °C). Error bars are calculated as max-min variation.

recirculated. This can be observed for most pumps if data is recorded with sufficient resolution. In the SC system, the cell is pressurized by supplying synthetic air so there is no fluctuation during filtration. Some peristaltic pumps may have an inherent fluctuation that is likely to affect the transport, even though the data acquisition resolution may hide such a fluctuation.

Naturally, larger pressure fluctuation is expected to cause a larger variation of water flux over time in the micro CF than in the SC system and it is expected that such an oscillation will disturb the boundary layer. Interestingly, the pure water flux (Fig. 11C) is constant over time in the micro CF, resulting from averaging by the volume acquisition.

A more variable trend is observed in the SC system, where the variation of the permeate mass weight (Fig. 11D) is larger in the SC system than in the micro CF. Through careful examination, some systematic error sources can contribute to this: i) frequency or setting of LabView recording that the user selects, ii) balance sensitivity, which causes a delay in the balance update of the mass when one water drop per second is added compared to more drops per second. This delay in the balance update would result in a stepwise increase, which is not observed in the permeate mass. This means that possible systematic errors in the micro CF is the LabView recording set by the operator that minimizes the water flux variation due to pressure fluctuation, resulting in a constant flux. Another systematic error caused by the temperature sensor recording is observed in Fig. 11B, which causes an unrealistic temperature drop of  $\pm 2$  °C during the measurement, which is not possible. Two primary learning outcomes from this example are: i) diligent observation of recorded parameters over time allows one to identify systematic errors during an experiment, such as pressure fluctuation (caused by the pump pulsation), sensor functioning (such as temperature recording) and balance sensitivity concerning the permeating water drop (this is especially important for low permeability in small systems), ii) the measurement recording (for instance LabView recording of permeate mass) needs to be carefully checked to avoid artifacts in the final measurements, such as the water flux variation over time.

To quantify random errors, such as pressure variation when the pressure valve is adjusted manually by the operator, and temperature variations in the laboratory on different days, repeated measurements of pressure, temperature and permeate mass are shown in Fig. 12 and the repeated water fluxes are reported in Fig. 13. In this example, the filtration system has no temperature control, while the temperature varies at  $\pm 2$  °C from 23 °C in the laboratory. Laboratory temperature variation is critical to enable meaningful experiments. It should be noted that humidity control is essential in a lot of membrane research, particularly when membrane materials are fabricated, while very few laboratories offer the possibility of humidity control. In such cases, monitoring and reporting the humidity during the experiment and evaluating the variability of results over the given range is essential.

When the water flux is measured multiple times (5 repeats) by the same operator, the random error caused by the manual pressure adjustment (the instrument air valve is adjusted by operator for each experiment) (Fig. 12A and B), and the laboratory temperature variation (Fig. 12C and D) are evident. The temperature variation is around  $23 \pm$

2 °C in both systems.

The larger fluctuation in the SC (Fig. 12D) is due to a systematic error in the sensor recording as the temperature cannot physically fluctuate between 20 and 23 °C in seconds. An alternative sensor with better resolution should be used to remove this error.

Repeated water flux measurements in the micro CF and the SC using different membrane coupons revealed a larger variation in the micro CF as shown in Fig. 13. Water viscosity is dependent on temperature and 1 °C temperature change results in a 2% viscosity variation corresponds to a 2% flux variation (Fig. S5) due to the linear relation in Darcy's law Eq. (7). The larger water flux variation in the micro CF (around 20%) cannot be explained with viscosity alone. An additional factor is the permeability of the membrane coupon. Seeing different membrane coupons are used for repeated measurements, an additional random error from the intrinsic variability of the membrane coupons (due to variable membrane morphology) is introduced.

This random error becomes significant especially in the micro CF, where the membrane area is almost 20 times smaller than in the SC system, resulting in a larger variation of the 5 repeats in the micro CF. Membrane coupons with more than 20% permeability variation due to possible defects should be discarded to reduce this random error (Fig. 5). Although membrane defects have a lower impact on the permeability of larger membrane areas, these defects can still influence solute retention results.

#### 4.3. Error analysis for solute retention

One experiment was performed at each specific operative condition to obtain a collection of data points and to estimate the error bars on solute retention. The error bars were calculated using the error propagation method in Eq. (8), where specific error sources were considered as reported in Table 4. The error sources contributing to the error propagation in retention ( $\Delta R$ ) are: i) error from the analytical measurements of feed, permeate and concentrate concentrations, ii) feed solution preparation by the operator, iii) variation of water flux (caused by pressure, temperature variation or membrane coupon quality), which affects permeate concentration by varying the convective transport of solute through the membrane. Error propagation is applied as a quadratic sum given in Eq. (8), where relative errors (dimensionless) are used to account for the error caused by the water flux variation. Eq. (8) is based on the general quadratic sum equation that is applied for division (full derivation of the equation is reported in Eq. (S3) and (S7)).

The error caused by the water flux variation reported as relative error ( $\Delta J_{v,abs}/J_v$ ) is included in the error of permeate concentration ( $\Delta c_{p,rel}$ ), while the error of the feed preparation by the operator ( $\Delta c_{f,prep}$ ) is included in both feed and permeate error. The flux variation is obtained as the difference between the maximum and minimum water fluxes measured before each filtration experiment for every membrane coupon.

where  $c_{p,b}$  are the concentrations in the permeate and bulk solution (feed or retentate), respectively. The relative error for bulk concentration ( $\Delta c_{b,rel}$ ), and permeate ( $\Delta c_{p,rel}$ ) is calculated from the analytical error ( $\Delta c_{b/f/p,anal}$ ) and the measured concentration ( $c_{b/f/p}$ ). A list of the

$$\begin{aligned} \Delta R &= 100 \cdot \frac{c_p}{c_b} \cdot \sqrt{(\Delta c_{b,rel})^2 + (\Delta c_{p,rel})^2} \\ &= 100 \cdot \frac{c_p}{c_b} \cdot \sqrt{\left[ \underbrace{\left( \frac{\Delta c_{b,anal}}{c_b} \right)^2 + \left( \frac{\Delta c_{f,prep}}{c_f} \right)^2}_{\text{Error on bulk concentration}} \right] + \left[ \underbrace{\left( \frac{\Delta c_{p,anal}}{c_p} \right)^2 + \left( \frac{\Delta J_{v,abs}}{J_v} \right)^2 + \left( \frac{\Delta c_{f,prep}}{c_f} \right)^2}_{\text{Error on permeate concentration}} \right]} \end{aligned} \quad (8)$$

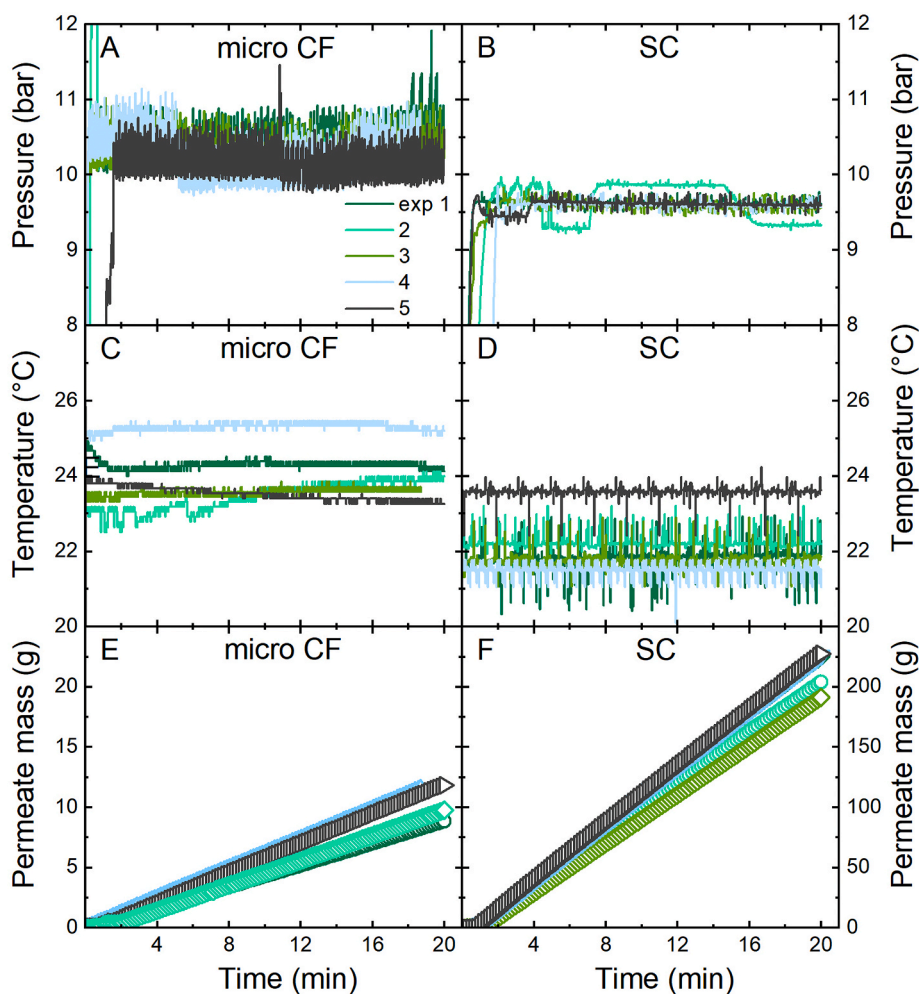


Fig. 12. Variation of A-B) pressure, C-D) temperature and E-F) permeate mass for 5 repeats (NF270 membrane, Milli-Q water, 10 and 9.6 bar in micro CF and SC system, target  $23 \pm 2$  °C, no experimental temperature control system).

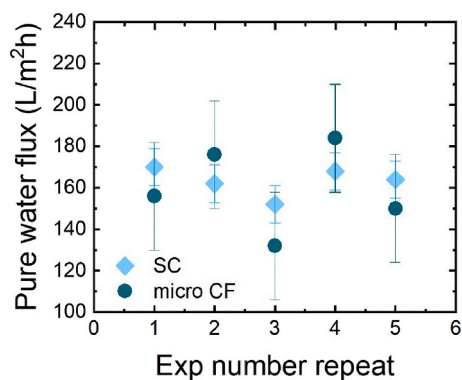


Fig. 13. Water flux of 5 repeated measurements (NF270 membrane, Milli-Q water, 10 and 9.6 bar in micro CF and SC system, target  $23 \pm 2$  °C, no experimental temperature control system). Error bar is calculated as max and min variation of 5 repeats.

error sources considered in the error propagation in Eq. (8) is shown in Table 4.

Two examples are selected to evaluate the contribution of random errors caused by water flux variation, the precision of the operator to prepare the feed solution and the systematic error from the accuracy of analytical instrument (high for LSC and low for TOC) on error propagation for retention; i) organic tracers concentration measured by TOC

Table 4

Type of errors, sources and quantification included in the error propagation for retention.

Error	Source	Quantification method
Analytical from feed/permeate concentration measurement	Instrument calibration/permeate concentration approaching LOD	Absolute error of concentrations ( $\Delta c_{f,anal}$ and $\Delta c_{p,anal}$ ) as standard deviation from calibration (discussed in section 4)
Feed concentration variation	Error caused by the operator on feed preparation ( $\Delta c_{f,prep}$ )	Max-min variation from repeated solution preparation
Variation in permeate concentration due to water flux variation	Variation in operative setting conditions	Max-min variation of water flux during filtration of repeated experiments ( $\Delta J_{abs}$ )

analysis in a wide range of concentrations from 1 to 25 mgC/L and a retention range from 20 to 84%, and ii) radiolabeled estradiol concentration measured by LSC and different membrane coupons with 15–20 % variable water flux (complete data in Fig. S7). Error contributions to feed, permeate concentration and water flux variation are shown in Table 5 as relative and absolute errors. One example of calculation is shown in Table S10. Relative errors are used to calculate the retention error with Eq (8).

**Table 5**

Relative and absolute errors for analytical accuracy, water flux variation and operator precision counted in the error propagation for solute retention.

Error term	TOC (organic tracers)		LSC (estradiol)		Quantification
	$\Delta y_{rel}$ (%)	$\Delta y_{abs}$	$\Delta y_{rel}$ (%)	$\Delta y_{abs}$	
$\Delta c_{f,anal}$	8	0.8 mgC/L	1	0.8 ng/L	Analytical error from calibration
$\Delta c_{p,anal}$	12	0.3–2.5 mgC/L	1–3	0.1–1.5 ng/L	
$\Delta c_p$	16	0.4–3.1 mgC/L	8–17	2–9 ng/L	Propagated with feed and flux variation
$\Delta c_{f,prep}$	9	2.4 mgC/L	7	6 ng/L	
$\Delta J_v$	4	5 L/m <sup>2</sup> .h	8–15	66 L/m <sup>2</sup> .h	Max-min (absolute)
$\Delta R$ (%)	2–5% (70% < R < 80%)		3–12% (UF10 kDa)		
	8–16% (20% < R < 70%)		10–16% (UF100 kDa)		Error propagation Eq. (8)

The contribution of analytical error for TOC is almost 8 times larger than LSC, resulting in a systematic error of 8–12% (for TOC) on feed and permeate concentrations analysis. When the random error of the feed preparation by the operator is considered, the relative error increases by a factor of 7 for LSC (from 1 to 7%), indicating that for accurate analytical equipment (LSC), the precision of the operator in preparing the feed solution plays a major role.

Filtration experiments with estradiol show a more significant variation of water flux with 8–15% relative error, which is almost five times higher than the water flux variation observed for the experiments performed with organic tracers. This is attributed to a different experimental design where ultrafiltration membranes are loaded with carbon nanotubes (SWCNT). The distribution of these in the membrane material may not be uniform due to nanoparticle aggregation, which causes random errors on the water flux measurements.

When the random errors of feed preparation and water flux variation are propagated to obtain the error of permeate concentration using the quadratic sum, the relative error increases up to six times for samples measured with LSC and up to two times for TOC. This suggests that analytical error is not a ‘strong-weighted’ error source of permeate concentration for accurate analytical equipment. In contrast, the random errors from the filtration performance and carefulness of the operator in preparing the solution have a major ‘weight’ on the error propagation. This is consistent with another study performed by Liang et al. (2022), where it was reported that the addition of random errors (such as bulk concentration variation at high salinity due to solution preparation) on the experimental measurement of water and salt permeability result in a significant increase of the error when these parameters were estimated by using an empirical method.

The error bars are plotted in Fig. 14A (TOC) and B (LSC) to evaluate if the calculated errors for solute retention reported in Table 5 are consistent with the experimental data set.

Error bars are larger in Fig. 14A (TOC) than B (LSC), which is consistent with the more significant analytical error (hence lower accuracy) reported in Table 5. This demonstrates that the analytical equipment accuracy needs to be considered in error propagation for

retention.

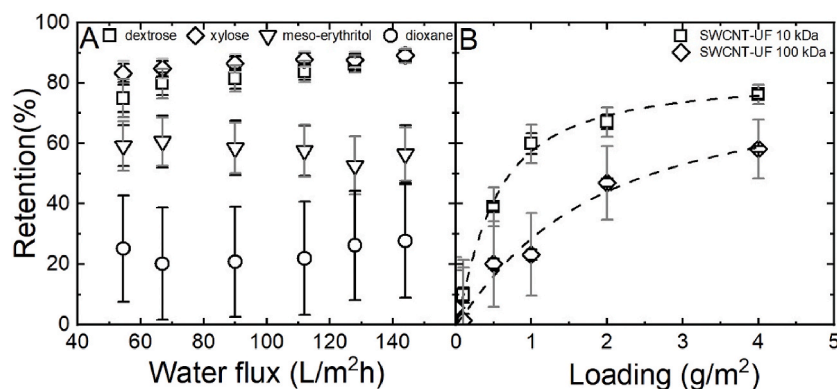
When only the analytical error is considered in the error propagation for solute retention (black error bar), the resulting retention error is underestimated in Fig. 14B, where the black error bars are too small and inconsistent with the experimental data points. When the random errors on water flux variation and feed preparation are considered, a larger grey error bar (in Fig. 14B) is obtained, which is more consistent with the experimental data and the errors in.

This suggests that random errors have more ‘weight’ in the propagation for retention error in the case of a highly accurate instrument. A similar result was found in another study where the imprecision in experimental measurements (for example, random error in feed concentration and pressure) plays a major role in the error increase of estimated membrane water and salt permeability and consequently transport parameters, such as the mass transfer coefficient (Liang et al., 2023).

The retention error increases when retention decreases (Fig. 14A), despite the data trend variability not reflecting the larger error bar. This is related to the larger variation of absolute error in feed and permeate concentration at lower retention (Fig. S8) and the mathematical arrangement in Eq. (8), where the retention error becomes larger proportionally to the increase in  $c_p/c_b$  ratio. This suggests that when retention is below 50%, the impact of operative conditions is unlikely to be distinguished from the error bar. The three primary learning outcomes from this example are: i) analytical error, which is an example of systematic error, has low weight on the error propagation of retention for highly accurate instruments (such as LSC), ii) random errors caused by variation of water flux and imprecision of operator to prepare the feed solution have higher weight in the error propagation for retention, iii) retention error increases with decrease of retention and can mask the impact of operative conditions when retention is very low (below 50%).

#### 4.4. Error analysis for solute mass adsorbed

During filtration, organic contaminants (such as steroid hormones or organic matter) can adsorb or deposit on the membrane, resulting in



**Fig. 14.** A) Organic tracer retention of NF270 as a function of water flux, data adapted from Imbrogno et al. (Imbrogno and Schäfer, 2019); B) Estradiol retention by SWCNT–UF membrane as a function of carbon nanotube (SWCNT) loading, data adapted from Nguyen et al. (Nguyen et al., 2021).

mass loss (or deposited mass adsorbed,  $m_{ads}$ ) from the feed or concentrate solution. This mass is calculated by mass balance, as reported in Eq. (9);

$$m_{ads} = m_f - m_c - \sum_{i=1}^n m_{p,i} \tag{9}$$

where  $m_{f/c/p}$  is the mass in the feed, retentate and permeate samples, respectively. The experimental absolute error of mass adsorbed ( $\Delta m_{ads}$ ) is calculated with the propagation method by using the quadratic sum applied for difference and division given in Eq. (10):

where  $V_{f/p}$  is the volume of feed solution and permeate samples,  $\Delta c_{b,rel}$  and  $\Delta c_{p,rel}$  are calculated as shown in Eq. (8),  $\Delta c_{c,rel}$  and  $m_c$  are the

$$\begin{aligned} \Delta m_{ads} &= \sqrt{\Delta m_{f,abs}^2 + \Delta m_{c,abs}^2 + \sum_{i=1}^n \Delta m_{p,i,abs}^2} \\ &= \sqrt{\underbrace{\left( m_f \cdot \left[ (\Delta c_{f,rel})^2 + \left( \frac{\Delta V_{f,abs}}{V_f} \right)^2 \right]^2}_{\text{Error on feed mass}} + \underbrace{\left( m_c \cdot (\Delta c_{c,rel})^2 \right)^2}_{\text{Error on concentrate mass}} + \underbrace{\left( m_p \cdot \left[ (\Delta c_{p,rel})^2 + \left( \frac{\Delta V_{p,i,abs}}{V_{p,i}} \right)^2 \right]^2 \right)^2}_{\text{Error on permeate mass}}} \end{aligned} \tag{10}$$

concentrate concentration relative error and the mass, respectively. The error sources considered in the error propagation are similar to the errors reported in Table 4 for retention. Relative errors are used in the propagation methods and multiplied by the mass in feed, concentrate and permeate to obtain the final error of mass adsorbed as an absolute value.

Examples of humic acid (HA, measured with TOC) and estradiol (E2, measured with LSC) are selected to evaluate the contribution of random errors, such as operator precision on feed preparation on mass adsorbed error. The absolute and relative errors on feed, permeate concentration and water flux variations are reported in Table 6 to evaluate the significant contributing error sources to the final error of mass adsorbed.

The main error contribution to estradiol feed concentration is the random error caused by the operator ( $\Delta c_{f,op}$ ), which is up to 9 times (6–9%) higher than the analytic error estimated from the calibration (1%). This requires training of the operator for preparing feed solutions at ng/L range concentrations to improve the precision of feed preparation. It is important for filtration experiments to report feed concentration variability in publications (as opposed to only writing concentration ratios) as this error can affect the results of mass adsorbed.

The major contributions to the error of the permeate concentration are the variation in the feed caused by the operator (6–9 %) and the

water flux variation (up to 12%) for LSC analysis, while for TOC, the analytical error is the dominant error source (up to 11%) up to 3 times higher than the relative errors estimated for water flux variation (4–6 %) and operator precision (3%). Volume errors are not considered in the propagation as relative errors are below 1%; hence they are considered negligible error sources.

The calculated absolute error bars for mass adsorbed, feed and permeate concentrations are added in Fig. 15A–C (for LSC) and B, D (for TOC) to evaluate if the estimated errors are consistent with the observed data set variability.

In Fig. 15A and B, the variability of feed concentration for the different experiments caused by operator imprecision is evident and consistent with the error percentage reported in Table 6. The larger the error of feed concentration, the larger the error in permeate concentration (see Fig. 15A and B), which is estimated using the error propa-

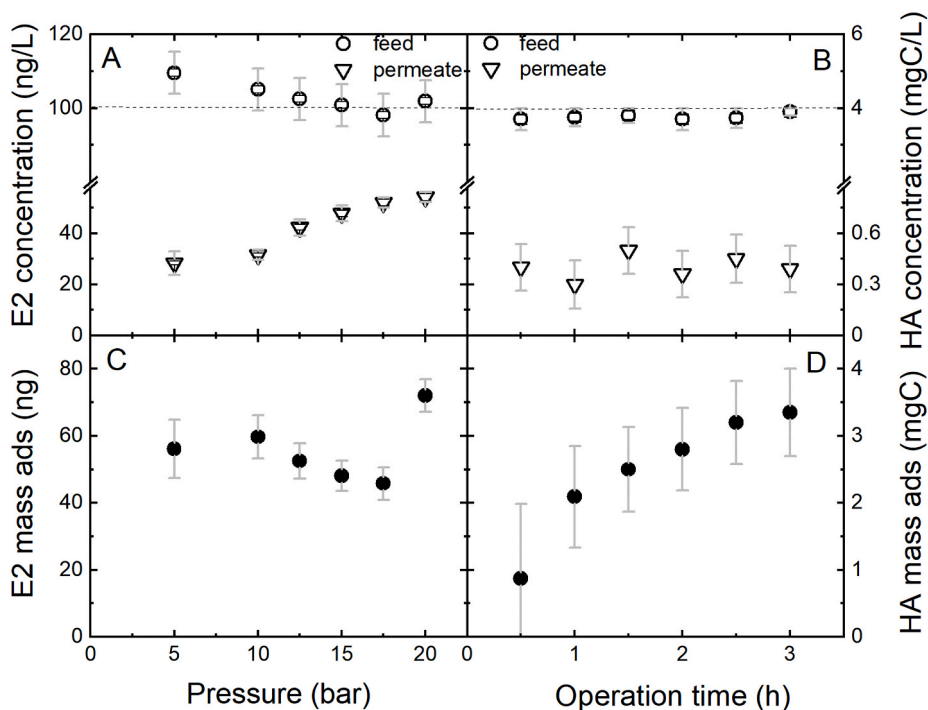
gation, where both systematic errors (e.g. analytical error) and random errors (e.g. feed preparation by the operator and water flux variation) are considered.

The error of mass adsorbed reported in Fig. 15C and D increases consistently with the error of feed preparation. This indicates that the feed concentration error is a significant error source that needs to be minimized to reduce the final error of mass adsorbed. The higher E2 mass adsorbed at 20 bar (Fig. 15C) is attributed to an error in the system operation, as the filtration experiment has been performed at the highest pressure suitable for the system. This makes the pressure control difficult and, consequently, the water flux varies over time, resulting in a mass loss in the feed almost one time lower compared to other pressures. The three main insights from these examples are: i) random error caused by the imprecision of the operator to prepare the feed has higher weight in the error propagation for mass adsorbed compared to analytical error of accurate instrument, ii) the imprecision of the operator to prepare the feed causes larger error in the permeate concentration and needs to be considered in the error propagation, especially when the analytical error is insignificant (such as LSC), iii) carefulness in the feed preparation helps to reduce the random errors of the feed and permeate concentration and consequently can increase the accuracy on the calculated error in the mass adsorbed.

**Table 6**

Relative and absolute errors for analytical accuracy ( $\Delta c_{b/p}$ ), water flux variation ( $\Delta J_v$ ) and operator precision ( $\Delta c_{f,prep}$ ) counted in the error propagation for solute mass adsorbed.

$\Sigma$ Error term	TOC (humic acid)		LSC (estradiol)		Quantification
	$\Delta y_{rel}$ (%)	$\Delta y_{abs}$	$\Delta y_{rel}$ (%)	$\Delta y_{abs}$	
$\Delta c_{f,anal}$	3	0.1–0.2 mgC/L	1	0.9 ng/L	Analytical error from calibration
$\Delta c_{p,anal}$	11	0.06 mgC/L	1–3	0.7–1.4 ng/L	$\Sigma$
$\Delta c_p$	27–30	0.14 mgC/L	4–16	2–5 ng/L	$\Sigma$ Propagated with feed and flux variation
$\Delta c_{f,prep}$	3	0.1 mgC/L	6–9	0.8–9 ng/L	Max-min (absolute)
$\Delta J_v$	4–6	1–2 L/m <sup>2</sup> .h	12	7–26 L/m <sup>2</sup> .h	
$\Delta m_{ads}$	0.6–1 mgC		5–9 ng		Error propagation Eq. (10)



**Fig. 15.** A, C) Estradiol (E2) and B, D) humic acid (HA) mass adsorbed on NF270 membrane. HA data adapted from Cai et al. (Cai et al., 2022), HA 12.5 mgC/L, 2.5 mM CaCl<sub>2</sub>, 1 mM NaHCO<sub>3</sub>, 10 mM NaCl, 9.6 bar, 400 rpm, 22 °C; E2 data adapted from Imbrogno & Schäfer (Imbrogno and Schäfer, 2021) E2 100 ng/L, 10 mM NaCl, 1 mM NaHCO<sub>3</sub>, 0.4 m/s).

#### 4.5. Error analysis for the rate of disappearance

In filtration experiments with reactive membranes (for example, photocatalytic membranes), solute removal may be dependent on the reaction kinetics at the membrane, which is expressed as the rate of disappearance,  $r''$  (Levenspiel, 1998) given in Eq. (11);

$$r'' = \frac{c_f - c_p}{t_R \cdot MW} \quad (11)$$

where  $t_R$  (s) is the hydraulic residence time of the solute in the membrane reactor, and  $MW$  (g/mol) is the molecular weight of the solute. Eq. (11) can be rewritten as shown in Eq. (12) (Lyubimenko et al., 2021);

$$r'' = \frac{R}{100} \cdot \frac{c_f}{t_R \cdot MW} \quad (12)$$

where  $R$  is the solute retention. Given that  $MW$  is fixed, the rate of disappearance error is the result of propagation of the errors in removal, feed concentration and hydraulic residence time, which are added in the quadratic sum for division, as shown in Eq. (13);

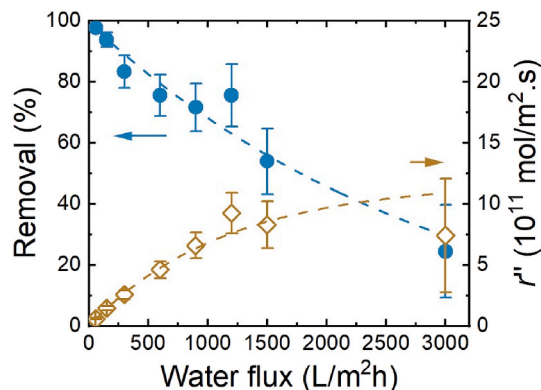
$$\frac{\Delta r''_{abs}}{r''} = \sqrt{\left(\frac{\Delta R_{abs}}{R}\right)^2 + \left(\frac{\Delta c_{f,abs}}{c_f}\right)^2 + \left(\frac{\Delta t_{R,abs}}{t_R}\right)^2} \quad (13)$$

Since the hydraulic residence time is inversely proportional to the flux  $t_R \propto \frac{1}{J_v}$ , Eq. (14) can be rewritten as follows;

$$\frac{\Delta r''_{abs}}{r''} = \sqrt{\left(\frac{\Delta R_{abs}}{R}\right)^2 + \left(\frac{\Delta c_{f,abs}}{c_f}\right)^2 + \left(\frac{\Delta J_v}{J_v}\right)^2} \quad (14)$$

Eq. (14) implies that, if the reaction is not effective (for example the removal is low) and the relative retention error ( $\Delta R_{abs}/R$ ) is high, the rate of disappearance will have a larger error.

An example, where the rate of disappearance is reported as a function of water flux and the estradiol concentration is measured with an ultra-high performance liquid chromatography-flow scintillation



**Fig. 16.** Estradiol (E2) removal and rate adsorption obtained with a photocatalytic membrane at varying flux. 100 ng/L E2, 1 mM NaHCO<sub>3</sub>, 10 mM NaCl, 23 °C. Adapted from Lyubimenko et al. (Lyubimenko et al., 2021).

(UHPLC-FSA) (Lyubimenko et al., 2020), is selected to evaluate the contribution of systematic error (such as analytical instrument accuracy) and random errors (such as water flux variation and variation on feed preparation) on the propagated error of rate of disappearance and removal. The error sources considered in the error propagation are summarized in.

A systematic analytical error of 12% is estimated on the feed concentration, which is more than twice higher than the random error caused by the feed preparation from the operator, reflecting that the analytical error provides the major contribution in this example. This is consistent with the lower accuracy of the UHPLC compared to the LSC, which provides more significant error of the standard measurements in the calibration (Lyubimenko et al., 2020).

The systematic analytical error of the permeate concentration increases up to 50% (relative error) when the concentration approaches the instrument LOD (range between 1.2 and 2.4 ng/L) (Lyubimenko

**Table 7**

Relative and absolute errors for analytical accuracy, water flux variation and operator precision counted in the error propagation for hormone removal and rate of disappearance.

Error term	UHPLC-FSA (radiolabelled estradiol)		Quantification
	$\Delta y_{rel}$ (%)	$\Delta y_{abs}$ ( $\pm$ ng/L)	
$\Delta c_{f,anal}$	12	12 ng/L	Analytical error from calibration
$\Delta c_{p,anal}$	16–50	1.2–13 ng/L	Propagated with feed and flux variation Max-min (absolute)
$\Delta c_p$	18–50	1.2–15 ng/L	
$\Delta c_{f,prep}$	5	5 ng/L	Error propagation Eq. (8)
$\Delta J_{V,abs}$	$\sim$ 5	Absolute values scale with flux	
$\Delta R$	2–15% (increasing with decreasing removal)		Error propagation Eq. (14)
$\Delta r_{abs}$	(0.07–4.7) $\bullet$ 10 <sup>11</sup> mol/m <sup>2</sup> .s (increasing with decreasing removal)		

et al., 2020), indicating a larger deviation at low permeate concentration. Primary error sources contributing to the propagated error of the rate of disappearance are the systematic analytical error of the feed and permeate concentrations. In contrast, the random errors caused by water flux variation and feed preparation from the operator are not significant (up to five times lower). The propagated errors calculated for removal and rate of disappearance are reported as error bars in Fig. 16 to evaluate if the calculated error is consistent with the data set variation. It is noted that the rate of disappearance reported by Lyubimenko et al. (2021) is equal to the expression in Eq. (12) multiplied by the membrane reactor thickness.

The error bar on the rate of disappearance follows the same trend as removal. At lower water flux, a full removal of 99% is reached, resulting in minimal error for both removal and the rate of disappearance. When the removal decreases to 20% at higher water flux, the error of the rate of disappearance increases consistently and the error bar variations are within the data set. The larger error in removal is mainly caused by systematic analytical errors, as reported in.

Table 7 and the larger increase in the  $c_p/c_b$  ratio.

#### 4.6. Error validation for solute retention and mass adsorbed

To validate the error propagation for solute retention and mass adsorbed, five filtration experiments were repeated at the same conditions using estradiol (E2 for LSC) and organic matter (OM, such as HA for TOC) using different membrane coupons. The resulting data points variation of the five repeats are compared with the propagated error bar. The calculated error bar is validated if the variation of the five data points repeated (highlighted with a grey background) is consistent with the error bars. The error sources considered in the calculations are the random errors from variations in the feed and permeate concentrations caused by the operator and analytical equipment and the errors caused by water flux variations. The error of permeate concentrations is propagated to include the error of feed and water flux variation. Absolute and relative errors are reported in Table 8 to evaluate the significant contributing errors of five repeats on the final error of solute retention and mass adsorbed.

For E2 experiments, the primary error sources contributing to the

final error of solute retention and mass adsorbed are the variation on water flux (8% see variation in Fig. S7) for the five membrane coupons and the precision to prepare the feed solution (up to 14%). The larger variation in water flux is attributed to the membrane coupon intrinsic variability and the coupons with water flux variation above 20% were not discarded (see Fig. S7).

For the OM experiments, the systematic error from the analysis is the most significant, as it causes an 11% error in feed/concentrate analysis. When random errors, such as feed water flux variation are propagated in the permeate concentration, the relative error increases slightly from 11 to 14%.

Overall, the errors reported in Table 8 highlight the relevance of operator training to reduce the random error of feed preparation and the careful selection of membrane coupons to minimize the random errors on concentrations and water flux measurements. The calculated errors on solute retention and mass adsorbed are added as error bars in Fig. 17 to validate the values that should lie within the grey box, indicating the area where the experimental data points are located.

The error bars calculated with the error propagation method reflect the data variation of the five repeats in the grey area, validating the error estimation method. The larger error of repeat number two in Fig. 17A and C is related to the larger variation in the feed concentration, resulting in an experimental error of solute retention and mass adsorbed almost three times larger than the error bars calculated for the other repeats.

The larger variability of the first experiment for the OM filtration is attributed to an error by the operator in the feed preparation (Fig. 17D), which causes a larger variation in retention and mass adsorbed, further highlighting the importance of diligence in the solution preparation. The estimated error bar covers this variability, which highlights that errors calculated with the propagation method can consistently reflect experimental data variation.

## 5. Conclusions

In this critical tutorial review on error methodology in water research, an error analysis methodology is proposed and demonstrated in the example of membrane filtration. The method involves: i) error

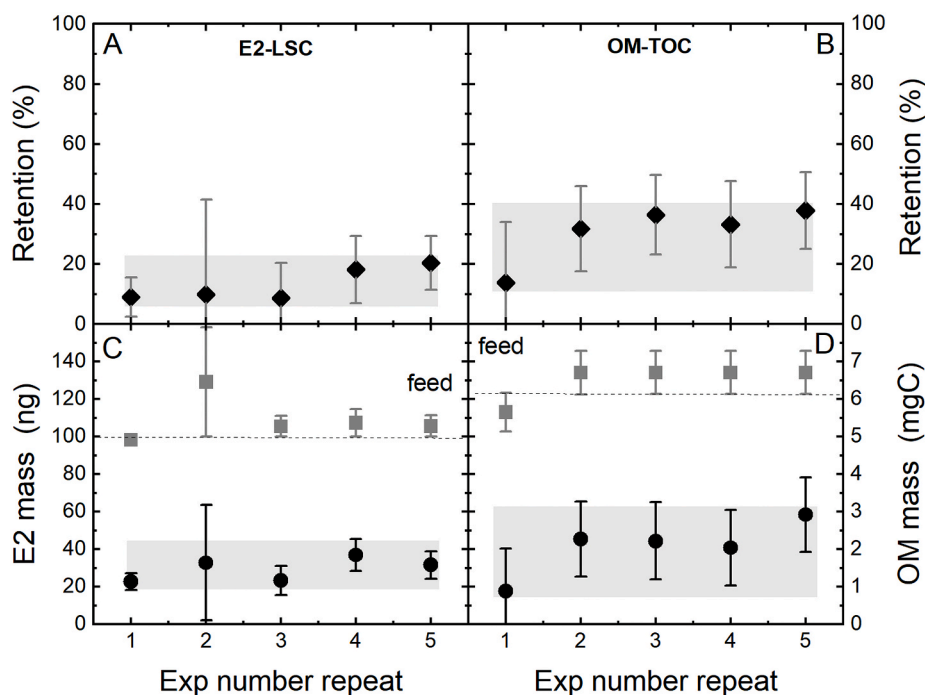
**Table 8**

Average measured values, relative and absolute errors for water flux, feed and permeate concentration for five repeated experiments of E2 and OM filtration.

Parameter	E2 filtration			OM filtration		
	Average	$\Delta y_{abs}$	$\Delta y_{rel}$ (%)	Average	$\Delta y_{abs}$	$\Delta y_{rel}$ (%)
$\Delta c_b$	108 ng/L	0.9 ng/L	1	9.5 mgC/L	1.0 mgC/L	11
$\Delta c_p$	97 ng/L	7 ng/L	7	7 mgC/L	1 mgC/L	14
$\Delta c_{f,prep}$	108 ng/L	15 ng/L	14	9.5 mgC/L	0.4 mgC/L	4
$\Delta J_v^a$	175 L/m <sup>2</sup> .h	13 L/m <sup>2</sup> .h	8	154 L/m <sup>2</sup> .h	2 L/m <sup>2</sup> .h	1
$\Delta R$	(13 $\pm$ 10) %			(35 $\pm$ 13) %		
$\Delta m_{ads}$	30 $\pm$ 7 ng			1.4 $\pm$ 0.9 mgC		

<sup>a</sup> Average is calculated from the values reported in Fig. S7 without considering the outliers.





**Fig. 17.** A) Estradiol (E2) retention, B) organic matter (OM) retention, C) E2 and D) organic matter mass loss of 5 repeated experiments (Hydracore 10,  $5.0 \pm 0.1$  bar,  $22.0 \pm 0.7$  °C, SC system, 300 rpm, 70 % recovery, E2 100 ng/L, FP 12 mgC/L, pH 8, 1 mM NaCl, 10 mM NaHCO<sub>3</sub>).

source determination, ii) error contribution evaluation, iii) error quantification, and iv) error validation with five repeats. It is demonstrated that standard deviation as experimental error is valid only when a minimum of 30–40 repetitions are performed, which is normally only achieved in the regular analytical instrument calibration or the pure water flux measurements.

An error propagation method is proposed to quantify the experimental error for calculated parameters, such as retention and mass balance, as it is impractical to repeat filtration experiments 30–40 times with the same conditions. This method accounts for the analytical error contribution, variation of operative conditions and operator settings. The repetition of one experiment five times under the same conditions is a practical method to validate the error bars determined with the error propagation method. Based on the presented case studies, it was demonstrated that the operator imprecision in the feed solution preparation, failure to select membrane coupons with similar permeability and lack of temperature control in the filtration systems were error sources that contribute to data variability and quality. Such errors can be reduced by operator training (although improvement is difficult to quantify) and careful experimental design. The application of error propagation for quantification can raise operator awareness of the need to reduce experimental errors to improve data quality.

This tutorial review aims to provide a method that will enable researchers from different disciplines to identify error sources in experimental design, minimize and quantify such errors with the error propagation method based on “sound judgment” evaluation of the contributing factors. Most importantly, this will hopefully inspire dealing with the issue of data quality and validity, which requires critical engagement with the topic. Explanation of error analysis include: i) the method applied to calculate the error and verification of assumptions (such as standard deviation, max-min variation, and/or error propagation), ii) repeated experiments clearly visible in the figures and, iii) variability of experimental conditions mentioned in figure captions, should be more transparent in scientific publications. The procedure proposed in this study will help to improve reproducibility in science and make error analysis more consistent with data variability and the reality of water research.

#### CRediT authorship contribution statement

**Alessandra Imbrogno:** Writing – review & editing, Writing – original draft, Validation, Investigation, Data curation. **Minh N. Nguyen:** Writing – review & editing, Visualization, Investigation, Formal analysis. **Andrea I. Schäfer:** Writing – review & editing, Supervision, Resources, Funding acquisition, Conceptualization.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Data availability

Data will be made available on request.

#### Acknowledgements

Helmholtz Recruitment Initiative and Helmholtz ERC Recognition Award NAMEPORED are acknowledged for IAMT funding. The Ministry of Science, Research, and the Arts of Baden-Württemberg is thanked for co-funding of A.I. (Brigitte-Schlieben-Lange Program). DAAD-BMBF provided Ph.D. scholarship of M.N.N. DuPont Water Solutions FilmTec and Nitto Hydranautics supplied the membranes. Yang-Hui Cai, Youssef-Amine Bousouga, Ramatisa Ladeia Ramos, Roman Lyubimenko and James Joseph provided feedback on the error methodology. Himma Nurul Faiqotul and Majda Breida performed filtration experiments for error validation. Frank-Dieter Kopinke (UFZ Leipzig) engaged in a discussion on the weighting of errors in the error propagation method. Bryce S. Richards (IMT, KIT) and Tyler Malkoske (IAMT, KIT) are thanked for providing critical feedback for the manuscript.

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.chemosphere.2024.141833>.

[org/10.1016/j.chemosphere.2024.141833](https://doi.org/10.1016/j.chemosphere.2024.141833).

## Supporting information

The SI includes an overview of statistical software packages used in error analysis (section 1), random and systematic errors in membrane filtration and remedies to minimize them (section 2), guideline to report significant digits and decimals (section 3), normal distribution of the standards measured by LSC, TOC and LC-OCD, methods to determine LOD and LOQ and water matrix interference (sections 4,5 and 6), calculation of absolute and relative errors for operative parameters in a filtration system (section 7), equations applied in the error propagation and validation (section 8), derivation of error propagation for solute retention with an error calculation example (section 9), raw data variation for the case of retention and the validation experiments (sections 10,11).

## References

- Alvarez-Quintana, S., Carmona, F.J., Palacio, L., Hernandez, A., Pradanos, P., 2020. Water viscosity in confined nanoporous media and flow through nanofiltration membranes. *Microporous Mesoporous Mater.* 303, 110289.
- Baker, R.W., 2004. Membrane transport and theory. In: W S, J. (Ed.), *Membr. Technol. App.*, pp. 29–31. England.
- Baker, M., 2016. 1,500 scientists lift the lid on reproducibility. *Nature* 533, 452–454.
- Beck, J.W., Scholer, A.A., Schmidt, A.M., 2017. Workload, risks, and goal framing as antecedents of shortcut behaviors. *J. Business Psychol.* 32, 421–440.
- Beck, M.W., de Valpine, P., Murphy, R., Wren, I., Chelsky, A., Foley, M., Senn, D.B., 2022. Multi-scale trend analysis of water quality using error propagation of generalized additive models. *Sci. Total Environ.* 802, 149927.
- Bisutti, I., Hilke, I., Raessler, M., 2004. Determination of total organic carbon—an overview of current methods. *TrAC Trends Anal. Chem.* 23, 716–726.
- Böcking, A., Koleva, V., Wind, J., Thiermeyer, Y., Blumenschein, S., Goebel, R., Skiborowski, M., Wessling, M., 2019. Can the variance in membrane performance influence the design of organic solvent nanofiltration processes? *J. Membr. Sci.* 575, 217–228.
- Box, G.E., Hunter, W.H., Hunter, S., 1978. *Statistics for Experimenters*. John Wiley and Sons, New York.
- Cai, Y.-H., Burkhardt, C.J., Schäfer, A.I., 2022. Renewable energy powered membrane technology: impact of osmotic backwash on organic fouling during solar irradiance fluctuation. *J. Membr. Sci.* 647, 120286.
- Committee, A.M., 1987. Recommendations for the definition, estimation and use of the detection limit. *Analyst* 112, 199–204.
- Cox, D.R., 1961. Design of experiments: the control of error. *J. R. Stat. Series A (General)* 124, 44–48.
- Crawford, L.A., Kosciński, D., Keyghobadi, N., 2012. A call for more transparent reporting of error rates: the quality of AFLP data in ecological and evolutionary research. *Mol. Ecol.* 21, 5911–5917.
- Di Pasquale, V., Miranda, S., Iannone, R., Riemma, S., 2015. A simulator for human error probability analysis (SHERPA). *Reliab. Eng. Syst. Saf.* 139, 17–32.
- Dietrich, C.F., 2017. In: Hilger, A. (Ed.), *Uncertainty, Calibration and Probability: the Statistics of Scientific and Industrial Measurement*. Routledge, Bristol, Philadelphia, New York, p. 39.
- Dym, C., 2004. Approximating and validating models. In: *Principles of Mathematical Modeling*, second ed. Elsevier Academic Press, pp. 71–113.
- Edwards, M.S., Ashkanasy, N.M., 2018. Emotions and failure in academic life: normalising the experience and building resilience. *J. Manag. Organ.* 24, 167–188.
- Gauch Jr, H.G., 2006. Winning the accuracy game: three statistical strategies—replicating, blocking and modeling—can help scientists improve accuracy and accelerate progress. *Am. Sci.* 94, 133–142.
- Goodman, S.N., Fanelli, D., Ioannidis, J.P., 2016. What does research reproducibility mean? *Sci. Transl. Med.* 8, 1–12.
- Greenwood, J.A., Sandomire, M.M., 1950. Sample size required for estimating the standard deviation as a per cent of its true value. *J. American Stat. Ass.* 45, 257–260.
- Haddaway, N.R., Verhoeven, J.T., 2015. Poor methodological detail precludes experimental repeatability and hampers synthesis in ecology. *Ecol. Evol.* 5, 4451–4454.
- Horowitz, A.J., Elrick, K.A., Colberg, M.R., 1992. The effect of membrane filtration artifacts on dissolved trace element concentrations. *Water Res.* 26, 753–763.
- Huber, S.A., Frimmel, F.H., 1991. Flow injection analysis for organic and inorganic carbon in the low-ppb range. *Anal. Chem.* 63, 2122–2130.
- Huber, S.A., Balz, A., Abert, M., Pronk, W., 2011. Characterisation of aquatic humic and non-humic matter with size-exclusion chromatography—organic carbon detection—organic nitrogen detection (LC-OCD-OND). *Water Res.* 45, 879–885.
- Imbrogno, A., Schäfer, A.I., 2019. Comparative study of nanofiltration membrane characterization devices of different dimension and configuration (cross flow and dead end). *J. Membr. Sci.* 585, 67–80.
- Imbrogno, A., Schäfer, A.I., 2021. Micropollutants breakthrough curve phenomena in nanofiltration: impact of operational parameters. *Sep. Purif. Technol.*, 118406
- International Organization for Standardization, 2006. (ISO 3534-1:2006) *Statistics — Vocabulary and Symbols — Part 1: General Statistical Terms and Terms Used in Probability*, p. 105.
- Jochum, C., Jochum, P., Kowalski, B.R., 1981. Error propagation and optimal performance in multicomponent analysis. *Anal. Chem.* 53, 85–92.
- Joint Committee for Guides in Metrology (JCGM), 2008. Evaluation of measurement data—guide to the expression of uncertainty in measurement. *Int. Organ. Stand. Geneva ISBN 134*.
- Kim, Y., Park, J., Jung, W., Jang, I., Hyun Seong, P., 2015. A statistical approach to estimating effects of performance shaping factors on human error probabilities of soft controls. *Reliab. Eng. Syst. Saf.* 142, 378–387.
- Kotek, L., Mukhametzianova, L., 2012. Validation of human error probabilities with statistical analysis of misbehaviours. *Procedia Eng.* 42, 1955–1959.
- Krzywinski, M., Altman, N., 2013. Error bars. *Nat. Methods* 10, 921–922.
- Ku, H.H., 1966. Notes on the use of propagation of error formulas. *J. Res. Natl. Bur. Stand. - C. Engineering and Instrumentation* 70C, 263–273.
- Kuselman, I., Penneccchi, F., Fajgelj, A., Karpov, Y., 2013. Human errors and reliability of test results in analytical chemistry. *Accred. Qual. Assur.* 18, 3–9.
- Lalia, B.S., Kochkodan, V., Hashaikhah, R., Hilal, N., 2013. A review on membrane fabrication: structure, properties and performance relationship. *Desalination* 326, 77–95.
- Le Phuong, H.A., Blanford, C.F., Szekely, G., 2020. Reporting the unreported: the reliability and comparability of the literature on organic solvent nanofiltration. *Green Chem.* 22, 3397–3409.
- Lee, S.H., Chen, W., 2008. A comparative study of uncertainty propagation methods for black-box-type problems. *Struct. Multidiscip. Optim.* 37, 239.
- Levenspiel, O., 1998. Chapter 2. Kinetics of homogeneous reactions. In: *Chemical Reaction Engineering*, third ed. John Wiley & Sons, pp. 13–37.
- Liang, Y., Dudchenko, A.V., Mauter, M.S., 2022. Inadequacy of current approaches for characterizing membrane transport properties at high salinities. *J. Membr. Sci.* 121246
- Liang, Y., Dudchenko, A.V., Mauter, M.S., 2023. Novel method for accurately estimating membrane transport properties and mass transfer coefficients in reverse osmosis. *J. Membr. Sci.* 679, 121686.
- Liu, L., Chen, W., Li, Y., 2016. An overview of the proton conductivity of nafion membranes through a statistical analysis. *J. Membr. Sci.* 504, 1–9.
- Lyubimenko, R., Richards, B.S., Turshatov, A., Schäfer, A.I., 2020. Separation and degradation detection of nanogram-per-litre concentrations of radiolabelled steroid hormones using combined liquid chromatography and flow scintillation analysis. *Sci. Rep.* 10, 1–13.
- Lyubimenko, R., Cardenas, O.I.G., Turshatov, A., Richards, B.S., Schäfer, A.I., 2021. Photodegradation of steroid-hormone micropollutants in a flow-through membrane reactor coated with Pd (II)-porphyrin. *Appl. Catal. B Environ.* 291, 120097.
- McMahon, J.W., 1973. Membrane filter retention—a source of error in the <sup>14</sup>C method of measuring primary production. *Limnol. Oceanogr.* 18, 319–324.
- Millar, N., Budgell, B., Salager-Meyer, F., 2020. Hype in reports of clinical research: the authors' perspectives. *Engl. Specif. Purp.* 60, 53–64.
- Miller, J., Miller, J.C., 2018. *Statistics and chemometrics for analytical chemistry*. Pearson Education 5, 19–142.
- Moffat, R.J., 1985. Using uncertainty analysis in the planning of an experiment. *J. Fluid Eng.* 2, 173–178.
- Nguyen, M.N., Weidler, P.G., Schwaiger, R., Schäfer, A.I., 2020. Interactions between carbon-based nanoparticles and steroid hormone micropollutants in water. *J. Hazard Mater.*, 122929
- Nguyen, M.N., Hervás-Martínez, R., Schäfer, A.I., 2021. Organic matter interference with steroid hormone removal by single-walled carbon nanotubes - ultrafiltration composite membrane. *Water Res.* 199, 117148.
- Ort, C., Lawrence, M.G., Rieckermann, J., Joss, A., 2010. Sampling for pharmaceuticals and personal care products (PPCPs) and illicit drugs in wastewater systems: are our conclusions valid? A critical review. *Environ. Sci. Technol.* 44, 6024–6035.
- Owusu-Agyeman, I., Jehanipour, A., Luxbacher, T., Schäfer, A.I., 2017. Implications of humic acid, inorganic carbon and speciation on fluoride retention mechanisms in nanofiltration and reverse osmosis. *J. Membr. Sci.* 528, 82–94.
- Peng, W., Escobar, I.C., 2005. Evaluation of factors influencing membrane performance. *Environ. Prog.* 24, 392–399.
- Prager, E.M., Chambers, K.E., Plotkin, J.L., McArthur, D.L., Bandrowski, A.E., Bansal, N., Martone, M.E., Bergstrom, H.C., Bernal, A., Graf, C., 2019. Improving transparency and scientific rigor in academic publishing. *J. Neurosci. Res.* 97, 377–390.
- Rao, B.K.N., 1993. *An engineer's view of human error*. Profitable Condition Monitoring. Springer, Dordrecht, The Netherlands, pp. 179–191.
- Reilly, F., 2007. Understanding accuracy, repeatability and reproducibility. *Met. Finish.* 10, 537–538.
- Sekharan, S., Samal, D.R., Phuleria, H.C., Chandel, M.K., Gedam, S., Kumar, R., Sethi, V., Supate, A.R., Karmakar, S., 2022. River pollution monitoring over an industrial catchment in urban ecosystem: challenges and proposed geospatial framework. *Environ. Challenges* 7, 100496.
- Sholl, D.S., 2019. Five easy ways to make your research more reproducible. *Langmuir* 35, 13257–13258.
- Shrivastava, A., Gupta, V.B., 2011. Methods for the determination of limit of detection and limit of quantitation of the analytical methods. *Chronicles Young Sci.* 2, 21–25.
- Stark, P.B., 2018. Before reproducibility must come preproducibility. *Nature* 557, 613–614.
- Taylor, J., 1997. In: *Introduction to Error Analysis, the Study of Uncertainties in Physical Measurements*. University Science Books, Sausalito, California.
- Tellinghuisen, J., 2001. Statistical error propagation. *J. Phys. Chem. A* 105, 3917–3921.

- Van Belle, G., 2011. *Statistical Rules of Thumb*. John Wiley & Sons.
- Vaux, D.L., Fidler, F., Cumming, G., 2012. Replicates and repeats—what is the difference and is it significant? A brief discussion of statistics and experimental design. *EMBO Rep.* 13, 291–296.
- Vecchi, R., Valli, G., Fermo, P., D'Alessandro, A., Piazzalunga, A., Bernardoni, V., 2009. Organic and inorganic sampling artefacts assessment. *Atmos. Environ.* 43, 1713–1720.
- Winchester, C., 2018. Give every paper a read for reproducibility. *Nature* 557, 281–282.
- Woolliams, E.R., 2013. Determining the uncertainty associated with integrals of spectral quantities. EMRP Joint Research Project Technical Report.
- Xu, W., Wang, X., Cai, Z., 2013. Analytical chemistry of the persistent organic pollutants identified in the Stockholm Convention: a review. *Anal. Chim. Acta* 790, 1–13.