Supplemental Material: Annu. Rev. Anal. Chem. 2023. 16:401-428
https://doi.org/10.1146/annurev-anchem-091222-112115

SUPPLEMENTAL MATERIALS:

EFFICIENT VALIDATION STRATEGIES IN ENVIRONMENTAL ANALYTICAL CHEMISTRY: A FOCUS ON ORGANIC MICROPOLLUTANTS IN WATER SAMPLES

Félix Hernández[1]*, David Fabregat-Safont[1,2], Marina Campos Mañas[1], José Benito Quintana[3]

[1] Environmental and Public Health Analytical Chemistry, Research Institute for Pesticides and Water, University Jaume I, 12006 Castellon, Spain

[2] Applied Metabolomics Research Laboratory, IMIM-Hospital del Mar Medical Research Institute, 88 Doctor Aiguader, 08003 Barcelona, Spain

[3] Department of Analytical Chemistry, Nutrition and Food Sciences. Institute of Research on Chemical and Biological Analysis (IAQBUS), Universidade de Santiago de Compostela, R. Constantino Candeira SN, 15782 Santiago de Compostela, Spain

Email addresses:

- Félix Hernández: hernandf@uji.es.
- David Fabregat-Safont: fabregad@uji.es.
- Marina Campos Mañas: manasm@uji.es.
- José Benito Quintana: jb.quintana@usc.es.

Corresponding Author:

- Prof. Félix Hernández
- Email: hernandf@uji.es
- Tel: +34 964 387366
- Fax: +34 964 387368

INDEX OF SUPPLEMENTAL MATERIALS

Supplemental Table 1. Scoring criteria for evaluating the reviewed papers.

Compounds	*Pharmaceuticals*	*Pesticides*	*DOA*
Matrix	*Surface water*	*Ground water*	*EWW / IWW*
Sample treatment	*SPE*	*DI*	*LLE*
Chromatography	*LC*	*GC*	
Mass spectrometry	*MS*	*MS/MS*	*HRMS/ HRMS/MS*
Validation strategy (2-6 p.)	1) Type of sample used in validation		
	Same as sample (3 p.)	*Simulated wastewater / Diluted sample(2 p.)*	*Tap water/ Milli-Q water (1 p.)*
	2) Validation type		
	CRM (3 p.)	*Interlaboratory (2 p.)*	*Spiked samples (1 p.)*
Validation data (1-3 p.)	*Included in main text (3 p.)*	*Supplementary Information (2 p.)*	*Referenced work (1 p)*
Analytical Evidences (1-3 p.)	*Chromatograms and spectra (3 p.)*	*Detailed numeric results (2 p.)*	*Summarized results (1 p.)*
Analytical discussion (1-3 p.)	*Detailed results interpretation (3 p.)*	*Detailed comments (2 p.)*	*Generic comments (1 p.)*
Quantification strategy (1-3 p.)	*ILIS / Standard addition (3 p.)*	*Matrix-matched calibration (2 p.)*	*Solvent calibration (1 p.)*
Identification criteria (1-3 p.)	*Based on guidelines (3 p.)*	*Detailed arbitrary criteria (2 p.)*	*No criteria (1 p.)*
LOQ stablishment (1-3 p.)	*Lowest validated level (3 p.)*	*Statistically determined from an spiked level close to LOQ (2 p.)*	*Statistically determined from an high spiked level (1 p.)*
QCs in batch (1-3 p.)	*Shows results and discuss them (3 p.)*	*QC are used but data is not provided (2 p.)*	*QCs are not used (1 p.)*
Score (9-27)			

Supplemental Table 3. Full proposal of chromatographic-MS methods validation for OMPs determination in water samples, including critical issues and remarks

Accuracy/trueness and precision				
Critical questions	**Minimum validation requirements**	**Optimal validation**	**Acceptability criteria**	**Remarks**
Is a typical validation in a given sample (e.g. n=5) sufficient to ensure the accuracy/trueness of the method, considering that the sample matrix composition is highly variable in environmental samples, and the difficulties to obtain a representative sample for validation purposes? Which is the acceptable recovery range in this type of analysis? Should different acceptability ranges be applied as a function of the specific analysis performed? Which strategy can be used when the recovery is out of the range (either below or above the acceptability values)? How can the recoveries be measured at (near) LOQ, when samples contain the analytes at similar or even higher concentrations?	Perform validation by recovery experiments in real-world samples of the same type than those that will be subsequently analysed, each spiked at two analyte concentrations (low - between 1-10 x LOQ; and high -around 10-50 times the low level). To grant method robustness and performance with varying matrix composition, validation should be performed with at least 3 different samples of the same type and the total number of analyses should at least be 6. This allows different combinations depending on particular circumstances (e.g. 6 different spiked samples analysed only once, 3 different samples analysed at in duplicate (6 analyses) or triplicate (9 analyses), etc.)	Include another spiking level (medium concentration, between low and high) (in total 3 spiking levels). Increase the number of different spiked samples to at least 5, and the total number of analyses to at least 10.	Recoveries between 70 and 120%, and overall RSD below 30%. In exceptional cases, average recovery outside 70-120% could be accepted if they are consistent (RSD ≤ 30%) and are ≥30% or ≤140%. In such cases, a correction factor as a function of the validation QCs recovery and supported by QCs recovery might be applied to the concentrations measured in samples.	As a typical validation with several replicates of just one sample does not seem enough, the strategy proposed include the analysis of different spiked samples. After initial method validation, if it is going to be subsequently applied to water samples from different type, at least 3 more samples from each type should be used in validation experiments (e.g. 3 wastewater samples; 3 surface water samples; 3 groundwater samples). The "blank" samples used for validation experiments must be analysed together with the spiked samples, and the analyte concentrations reported. If any of the samples contains the analytes under study, the concentration in the "blank" sample must be subtracted to that measured in the spiked sample to calculate recoveries. When analyte concentration in a "blank" sample is similar, or even higher, than the spiked level, the recovery calculated may be discarded and the reasons for not including such value in the validation data must be specified. If due to these limitations, few recovery data are available, particularly at low analyte concentration, another sample as similar as possible should be used at that spike level as a proxy. The validation criteria might have been previously established in studies applied for regulatory purposes. In such cases, the criteria set-up by the regulation must be prioritized.

LOD and LOQ				
Critical questions	**Minimum validation requirements**	**Optimal validation**	**Acceptability criteria**	**Remarks**
Are the LOD/LOQs calculated in method validation representative and realistic enough? In other words, can it be ensured that such values are achievable in daily analysis? Is it appropriate to estimate LOQ based on the S/N? Is it more appropriate to establish the LOQ as the lowest concentration tested and satisfactorily validated (as e.g. proposed in the SANTE guideline for pesticide residue analysis)? Is the use of the lowest calibration level (taking into account the pre-concentration or dilution experimented by the samples in the analytical procedure) acceptable for estimation of LOQ? The most sensitive transition is commonly used for quantification, and therefore it is also used for estimation of LOD/LOQ. However, for detection and quantification of a given compound it is necessary to guarantee its identification too. Is therefore the identification ensured at the LOD or LOQ levels reported in the literature? Is it possible to estimate the LOD/LOQ in analyte/sample combinations where the compound is nearly always present in samples, and then is not useful to spike the samples at low concentrations?	Estimate LOQ and LOD in a water sample (from the same type of those that will be monitored later) spiked at analyte concentrations near the LOQ, maximum 10-times the LOQ finally proposed. Confirm the identity of the compound at the level tested by acquiring at least two transitions (3 ions in single MS methods) (see acceptability criteria). Periodically test that, at least, the LOQ is attainable in the daily work, analysing QC samples spiked at a level near (maximum 10 times higher) the LOQ.	Estimate the LOQ/LOD in 5 different samples and calculate the average value finally proposed as LOQ/LOD, also indicating the range. The ion ratio must be accomplished ensuring the reliable identification of the analyte not only at the LOQ but also at the LOD level (maximum deviation 30%).	The estimation must be made from the chromatograms of spiked samples, corresponding to the quantification transition, based on S/N =10 (LOQ) or S/N=3 (LOD). For the LOQ level, the identification of the analyte must be ensured as well. So, at least one qualification/ confirmation transition must be also observed, and the ion ratio deviation criterion accomplished (maximum deviation with respect to a reference standard ± 30%). For the LOD, the chromatographic peak corresponding to the second transition must be observed (minimum) and the ion ratio accomplished (optimal).	The LOQs and LODs should be as realistic as possible, and reachable in the daily analysis. The trend to report such parameters as low as possible, even when it is not actually necessary as a function of the analysis purpose, should be avoided. The appropriate maintenance of the instrument is essential to reach good sensitivity, and so these parameters are highly dependent of the state of cleanliness of the equipment. The great variety in the characteristics of the matrix-samples analysed has a predominant influence on the LOQDs/LOQs reached too. For this reason, the LOD and particularly the LOQ, must be periodically tested using real-world samples. In specific combinations analyte/sample is difficult to find real-world blank samples. In these cases, the LOQ/LOD should de estimated directly in samples at low analyte concentration, without spiking the sample, and using the same criteria as above. For the LOD, the chromatographic peak corresponding to the second transition will be near the chemical background as the second ion selected is usually less abundant than the that used for quantification. So, the accomplishment of ion ratios may be troublesome. These parameters might have been established in studies applied for regulatory purposes. In such cases, the criteria set-up by the regulation must be prioritized.

Calibration				
Critical questions	**Minimum validation requirements**	**Optimal validation**	**Acceptability criteria**	**Remarks**
Is it a realistic approach the use of matrix-match calibration in environmental analysis, considering the lack of representative matrices free of analytes and the high variation in sample composition? Although the standard additions method is theoretically a good option, it seems rather problematic from a practical point of view by several reasons: notable increase in the number of samples injected; need to adjust the concentrations added to each sample as a function of the analyte concentration in the sample, which however is unknown; calculation subjected to notable errors when calibration is not well adjusted. Is therefore, this method practical in environmental analysis?	Perform calibration with standards in solvent (including the same ILIS than in samples) with at least one point below the concentration corresponding to the LOQ. The calibration should include at least 5 levels and the standards concentrations corresponding to the (two or three) levels validated, and should be extended up to concentrations commonly found in the samples.		Report the value of R2 and either some visual evidence or data on residuals, lack-of-fit or other tests. Non-linear calibrations can be used, but it should be clearly mentioned and supported by appropriate information as in linear calibration. One-point calibration can be used for estimative purposes ("semi-quantification") as long as this is clearly indicated and the concentration in the samples does not differ more than 30% than the calibrator.	The applicability of matrix-matched calibration in environmental analysis is questionable, as the ideal blank sample, representative of those that will be analysed, is hard to be found. In certain cases (e.g. different samples collected along the same river), a "blank-representative" sample might be found from less polluted sites to perform match-calibration. In general, performing calibration with standards in solvent is recommended, after carefully testing matrix effects in different matrices and the appropriate ways for an efficient correction. The use of ILIS is essential at this point. Thus, absolute responses or relative responses analyte/ILIS will be used as a function of the ILIS availability. When the concentration in sample is above the highest calibration point, the water sample or extract must be diluted, but the same composition regarding ILIS concentration and organic solvent composition must be maintained for standards and sample extracts.

Evaluation of matrix effects. Use of ILIS				
Critical questions	**Minimum validation requirements**	**Optimal validation**	**Acceptability criteria**	**Remarks**
Can be ensured the appropriate correction/minimization/elimination of ME in all samples analysed, considering the high variability in sample-matrix composition in environmental analysis? There is an increasing trend to use ILIS for ME correction and/or correction of potential errors associated to sample treatment and injection processes. To this aim, using the own analyte ILIS is the most efficient approach, although it becomes problematic in multi-residue methods including many analytes (high cost and limited commercial availability of reference ILIS standards). A key question is whether an analogue ILIS, different to the analyte ILIS, can be efficiently used for ME correction? Can be a simple dilution (e.g. with ultrapure water) an efficient way to deal with ME, considering the excellent sensitivity of the modern MS/MS instrumentation? Even if ME is corrected (e.g. by ILIS or by simply dilution), it is necessary to assess its impact into method LOD/LOQs, since strong signal suppression will significantly impair these parameters. For instance, the SWGTOX guidelines stablish that this needs to be investigated with 10 different matrices and the impact on LOD/LOQ evaluated if signal suppression/enhancement is higher than 25% [a]. Is this issue correctly addressed in the literature?	Spike 3 different samples (in DI-based methods) or 3 different sample extracts (e.g. in SPE-based methods) at a medium concentration level and compare the measurement with a reference standard in solvent at the same concentration. Inject spiked samples/extracts and standards by quintuplicate and obtain the average response. Pay attention to the blank measurement, in order to subtract its response in case that the analyte under study is present in the "blank" sample. Another way, less useful in environmental studies, is the comparison of matrix-match calibration and calibration in solvent. Here, the difference in the slopes will give an indication of the matrix effects. Matrix effects, expressed in % enhancement or suppression can be evaluated according to the following equation: *Matrix effects [%] = 100 * peak area or slope (matrix)/peak area or slope (solvent) – 100*		Matrix effects are considered significant if they exceed ±20%. Thus, if ME is less than ± 20%, no correction is in principle necessary. Nevertheless, the final recovery of the method, considering all aspects affecting the overall procedure, including ME, will allow to know whether some correction is required.	Although ME are implicitly considered when a method is validated by means of recovery experiments, it is recommended to perform experiments to specifically evaluate ME as a part of the validation process. It will provide a better knowledge of the problems that may be encountered in subsequent analysis. The use of ILIS is highly useful in chromatographic-MS methods, particularly in LC-MS/MS. ILIS can be used for matrix effects correction but also to correct potential analytical errors associated to sample treatment when used as surrogates. Although different ILIS could be used, the use of the own analyte ILIS is the best approach to ensure appropriate correction. This seems reachable for most laboratories in individual methods or in methods including few analytes. However, in multi-residue multiclass methods containing dozens of analytes, the acquisition of many isotope-labelled reference standards, if available, requires a notable investment not affordable for all laboratories. Each laboratory should find the right way in terms of acquiring of ILIS reference standards, taking into account the obvious pros and cons. A multi-residue method, consisting of direct injection of water samples (e.g. groundwaters, surface water) is perfectly possible, even after previous dilution of more concentrated/dirty samples (e.g. wastewater), given the excellent sensitivity of modern instrumentation. This strategy is clearly favoured when using a notable amount of analyte-ILIS. It is worth noticing that not all analyte-ILIS are equally efficient for matrix effects correction. It is important to verify that there is no cross-talk between ILIS and analyte, as it could occur when using minimal-labelled ILIS and compounds in which the isotope pattern present isotopes with higher abundance [b]. ME can also be evaluated from samples spiked with ILIS, instead of analytes, as ILIS are, in principle, affected by the matrix sample in the same way than analytes. A more complete ME evaluation should include its potential correction by using ILIS, or by other ways tested by the laboratory (e.g. clean-up steps).

Identification of compounds				
Critical questions	**Minimum validation requirements**	**Optimal validation**	**Acceptability criteria**	**Remarks**
Should a potential positive be discarded if the ion ratio or retention time in sample slightly falls out of the acceptability rage (commonly ±30%, or ±0.1 min deviation, respectively)? If more ions than the minimum required in the guidelines are acquired, is it necessary that all comply the ion ratio, or is it sufficient just one ion ratio? What can be done when there is only one intense ion/transition available, as it may occur when a compound is hardly fragmented or when it has low m/z fragment ions, not enough specific (e.g. when analysing low molecular weight molecules)? It is widely recognized the high potential of HRMS for identification/elucidation due to the large amount of information provided by this accurate-mass full-spectrum acquisition technique. Tentative identifications are based on data acquired (accurate mass measurements/mass errors, isotope pattern), together with fragmentation interpretation of spectra, comparison with databases (e.g. Mass Bank), or in-silico fragmentation models. A critical issue arises when using HRMS: what is the confidence level of tentative identifications (i.e. without reference standards) reported in the literature?	Acquire at least 3 MS/MS transitions in tandem MS methods and obtain the ion ratios (normally using peak areas). One transition (named Q) will be used for quantification, and the rest (named q1, q2, …) will be used as confirmatory transitions. Obtain the average q/Q ratios (q1/Q, q2/Q, …) for the standards included in the calibration, and use them as a reference when analysing samples. Compare the ion ratios in samples with those of the reference standard and calculate the deviation. Acquire at least 3 ions in single MS methods, obtain the ion ratios and compare with the reference standards. Acquire at least 2 accurate mass ions in HRMS-methods. Calculate the Rt deviation with respect to the reference standard.	Acquire the maximum number of MS/MS transitions or ions, if feasible, to improve the identification in problematic cases (see Remarks).	At least one ion ratio (q1/Q or q2/Q) in the sample must not exceed a deviation of ± 30% with respect to the reference standard (e.g. average of the standards included in the calibration). Maximum error for accurate-mass measured ions in HRMS methods < 5 ppm (<1 mDa for m/z < 200). Identification criteria also includes the deviation in the chromatographic retention time, normally ±0.1 min or ±0.5% if relative to an ILIS.	The acquisition of more than 2 ions, if feasible, is highly recommendable to ensure adequate identification in problematic cases, where some transitions may be shared/interfered for other coeluting compounds present in the samples. In such cases, the transitions interfered, or those with more background or chemical noise, may be discarded, using more specific transitions instead. If only one suitable ion is available, the compound should either be marked as "tentatively identified" or measured by an orthogonal separation technique or by HRMS or IMS, in order to provide enough confidence. Molecular ions, (de) protonated molecules, or adduct ions are characteristic for the analyte and are recommended to be included in the measurements and identification whenever possible. High m/z ions are more selective; so, low m/z ions (e.g. < 100) should be avoided if possible. However, high m/z ions arising from loss of water or loss of common moieties may be problematic as they may be more easily interfered [c]. The use of ILIS and/or spiked samples (e.g. QC) is recommended to evaluate ion ratios or retention time deviations exceeding the maximum acceptable value (see Reporting concentration data). Identifications without the use of reference standards (e.g. using spectral libraries match or based on the presence of ions reported in the literature) cannot be considered as definitive but as tentative. This must be clearly specified in the report.

Quality Control				
Critical questions	**Minimum validation requirements**	**Optimal validation**	**Acceptability criteria**	**Remarks**
Which strategy can be applied when there are difficulties to find authentic "blank" samples for QCs preparation (e.g. many pharmaceuticals and some illicit drugs in urban wastewater) How can QCs recoveries be calculated when spiking at low concentrations (e.g. close to the LOQ) samples that contain the analyte at similar concentrations? Which should be the acceptability criteria for QCs recoveries? How many QCs should be analysed together with the sample batch, and specially how many different samples should be used for QCs preparation? Which strategy could be applied when a QC recovery is out of range? Should the data corresponding to such analyte be discarded? Would it be possible to apply a correction factor in these cases?	Prepare QCs at two concentration levels (low and high) in selected samples that will be analyzed later (see Remarks for the number of QCs to be prepared). When the samples used for QCs preparation are not true blank samples (i.e. they contain the analytes at concentrations similar or higher than the spiked), the recovery calculation is compromised, and such QCs might be discarded. Subtracting the "blank" concentration to the spiked QC is compulsory but this approach may not be successful in such cases. For this reason, it may occur that only some QCs, (normally those at high analyte concentrations) are useful to support the quality of data.	Include a third concentration level (i.e. low, medium, high), and increase the number of samples used for QCs preparation (see Remarks).	Acceptability proposed for individual QCs recovery is 60-140%, which is in the line of the SANTE guideline. When recoveries are out of this range, the quantification is compromised. In those cases, concentration data might be reported as "estimated", indicating the QCs recoveries obtained for the compound, or that the samples should be re-analyzed. When robust and reproducible recoveries are obtained, with low RSD (e.g. < 20%), even if they are out of the acceptability range, a correction factor might be applied, but indicating this circumstance in the report.	The best way to guarantee the method reliability when analyzing a variety of samples from different origin and composition is the analysis of QC samples included in the sample batch. QCs must be prepared with the same samples that will be analyzed (ideally all, but for practical reasons only some selected samples), spiking them with the analytes at appropriate concentrations. The recovery data obtained for QCs should be included in the analytical report as a support of the quality of analyses. As for the number of samples used for QCs preparation, there is not any guideline, but using a balanced criterion is recommended trying to find an equilibrium between quality control and effort/cost of analysis. For example, selecting between 10-50% of the samples analyzed is a possibility, as a function of the analytical difficulties and the number of samples included in the study. For example, including 20% of QCs in a study of 10 IWW, 10 EWW and 10 SW, would imply to prepare QCs from two samples of each type, leading to 4 QCs (minimum) or 6 QCs (optimal) for each IWW, EWW and SW, with a total of 12 or 18 QCs samples. This means that the analytical work will pass from 30 original samples to 42 (minimum) or 48 (optimal) samples including QCs. Although a notable effort must be done, the support given to the data reported is out of question.

[a] Scientific Working Group for Forensic Toxicology (SWGTOX) Standard Practices for Method Validation in Forensic Toxicology. 2013. *J. Anal. Toxicol.* 37(7):452–74

[b] Campos-Mañas M, Fabregat-Safont D, Hernández F, de Rijke E, de Voogt P, et al. 2022. Analytical research of pesticide biomarkers in wastewater with application to study spatial differences in human exposure. *Chemosphere*, p. 135684

[c] EU Reference Laboratories for Residues of Pesticides. 2019. SANTE/12682/2019. Analytical Quality Control and Method Validation Procedures for Pesticide Residues Analysis in Food and Feed