**Transportable automated HRMS platform enables insights into water quality dynamics in real time**

Michael A. Stravs†, Christian Stamm, Christoph Ort* and Heinz P. Singer*

1Department of Environmental Chemistry, Swiss Federal Institute of Aquatic Science and Technology Eawag, Dübendorf, Switzerland
2Department of Urban Water Management, Swiss Federal Institute of Aquatic Science and Technology Eawag, Dübendorf, Switzerland
† Current address: Institute of Molecular Systems Biology, Eidgenössische Technische Hochschule (ETH) Zurich, Zurich, Switzerland
* Correspondence to: christoph.ort@eawag.ch, heinz.singer@eawag.ch

**Abstract**

Tracking the occurrence of a plethora of chemicals in the aquatic environment at high temporal resolution over extended periods is a huge challenge. Here, we present a transportable high-resolution mass spectrometry platform including a fully automated workflow for advanced data processing. It measured several thousand concentration data points at 20-min intervals over several weeks, providing unprecedented insights into pollution dynamics, e.g. acute pesticide toxicity peaks in a small creek, intra-day variation of illicit drugs in raw wastewater and identifying contamination clusters of unknowns. This enabling technology has potential for researching and managing chemicals in natural and technical environments beyond current possibilities, e.g. real-time control in process engineering and sewer operation (water management and environmental toxicology), industrial surveillance (law enforcement) and wastewater-based epidemiology (public health).

**Introduction**

Synthetic chemicals are an integral part of human society in the Anthropocene. They play an ever-increasing role in most economic sectors, households, medicine, and science. Chemicals such as pesticides can both benefit and threaten many aspects of human life and may also exceed the planetary boundaries. Many chemicals are released to the environment without proper understanding of their fate or potential impacts. The quantification of numerous chemicals at sufficiently high temporal resolution to determine whether targeted action is warranted presents a huge challenge. The development of high-resolution mass spectrometry (HRMS) has enabled thousands of chemicals to be quantified, unknown chemicals to be detected, and digital archives to be generated. However, multiple bottlenecks from sampling to analysis still limit the acquisition of comprehensive datasets and prevent online monitoring with HRMS. Overcoming these challenges would result in game-changing breakthroughs for environment monitoring and management.

Mobile sensors have been developed in recent years for air and water quality variables such as water isotopes, nutrients, microorganisms, and sediments. They provide new insight into important environmental processes and benefit both science and environmental management. However, studying the occurrence and fate of synthetic chemicals in aquatic matrices still relies on collecting samples and taking them to laboratories for analysis. Unfortunately, traditional sample collection, preparation, measurements, and evaluation require substantial outlays of time and personnel, resulting in a limited number of samples and/or compounds that can be analyzed within any project or program budget. For example, a typical laboratory liquid-chromatography mass spectrometry (LC-MS) sample preparation and analysis requires approximately one week for 250 samples, not including sampling, transportation, and storage, and these may require compound-specific preservation to avoid transformation or losses. These practical limitations result in a tradeoff between achievable temporal resolution and sampling duration and, hence, often impede maximal information gain. The ISO recommendation that the “best … solution … would be … an on-line automatic instrument” has long resembled a pipe dream.

Whereas existing approaches solve some of these limitations individually, our MS2FIELD is a general-purpose platform that integrates automated high-frequency sampling, sample preparation, and measurement; facilitates preliminary target quantification; and generates large time series datasets for highly time-resolved targeted and nontargeted evaluation both in real time and for retrospective analysis. These comprehensive datasets include positive- and negative-mode full scan mass spectra (MS1) and data-dependent and data-independent tandem mass spectra (MS2) at 20-minute intervals spanning weeks to months. They enable novel insights by providing data suitable for high-dimensional data mining with both supervised and unsupervised statistical approaches.

**Materials and Methods**

To overcome previous limitations, we developed a workflow that encompasses sampling, filtration, analytical measurements, data processing, and data visualization yet requires no personnel. Key to success was the selection, combination, and installation of high-end laboratory equipment in a trailer to operate it directly in the field while unattended, stable, and safe.
A trailer (L=4m, W=1.85m, H=2m) houses the MS$^2$field (Fig 1c, S1, S2). An eccentric screw pump delivers water – approx. 10L/min, adjustable – through a self-cleaning filtration device (Collins 9150, 2µm mesh stainless steel disk, TWP Inc.). Sample preparation and analysis are performed by a programmable autosampler (PAL RTC, CTC Analytics) and HPLC pump (Rheos 2000, Flux Instruments) connected to a HRMS (Q-Exactive HF, Thermo Fisher). Briefly, the filtrate (500-750µL) is automatically diluted and spiked with isotope-labeled internal standard (ILIS) before being analyzed by online solid phase extraction (oSPE)-LC-HRMS. Custom-packed SPE cartridges are used for enrichment and fine filtering and the filtrate is subsequently eluted over a conventional LC column (XBridge BEH C18, 3µm, 2.1mm x 5cm). A six-cartridge selector increases lifetime between servicing visits. For every sample, positive and negative ionization data is acquired in polarity-switching mode. MS2 acquisition alternates by sample between data-dependent (Top2) and data-independent (2×2 windows) mode. This achieves comprehensive MS1 and MS2 coverage. A 4-minute sample preparation followed by a 16-minute SPE enrichment and LC-MS run results in a 20-minute measurement cycle (analytical details in Supporting Information (SI) Materials and Methods).

The bypass loop, comprising pump, filter, and valves, is controlled by a programmable logic controller (PLC) system (750-8204, WAGO), and sensors for pump flow, pressure, filtrate flow, and autosampler flow are used for feedback. Smoke, temperature, and water sensors trigger safety stops and system shut downs. The autosampler and LC-HRMS system operate on conventional software using custom autosampler scripts and are synchronized with the PLC using two digital signals (24V). A 4G/LTE modem and a VPN connection allow remote control, and two cameras are installed for additional observation. All PLC data and the results of an automated quantification procedure for target analytes are transferred to an online dashboard (InfluxDB, Grafana), facilitating a live view of process parameters and preliminary results.

**Results and discussion**

**Field sites and performance.** Over a 10-month period, we operated MS$^2$field at three sites to demonstrate potential applications and stable operation in the field and to evaluate analytical performance. Subsequently, we refer to three unique datasets, all with a temporal resolution of dt=20min: 1) seven weeks in a small creek, 2) two weeks at a wastewater treatment plant’s raw influent after the sand trap WWTP$^{raw}$, and 3) seven weeks in a large river (detailed information in Table S12). Field site requirements are an even surface of approx. 5×3m and access to power supply (3-phase, 400V). Transportation and installation typically required two days. While two maintenance trips were required per week for raw wastewater, longer periods were achieved for treated wastewater and creek and river water (up to 21 days).

During all field campaigns, a complex mixture of >200 isotope-labeled internal standards (ILIS) was spiked into every
sample, and >400 standards were measured for a calibration row and as daily quality control. The use of oSPE-LC-HRMS allows sensitive and selective quantification of a broad range of compounds. The system was tested for a concentration range of 1–5,000 ng/L (Tables S5 and S6, Fig. S11 and S12). In summary, for 32 pharmaceuticals, pesticides and industrial chemicals in raw wastewater, LOQ (limit of quantification) was ≤10 ng/L for 5 and ≤100 ng/L for 29 compounds. In surface water, LOQ was ≤10 ng/L for 12 and ≤100 ng/L for 18 pesticides and urban compounds. Recovery of spiked quantities for compounds with matching ILIS generally ranged between 75% and 120%. Quantification of compounds without matching ILIS requires more caution due to varying matrix effects, which are also reflected in diurnal variations for ILIS areas particularly in the WWTP\textsubscript{raw} dataset (Fig. S4); quantification without ILIS often suffers from these effects, not just with this platform\textsuperscript{20,21}.

**Unprecedented insights into pollutant dynamics.** The MS\textsuperscript{2}field platform can characterize the dynamics of organic micropollutants in unprecedented ways at multiple timescales. We illustrate this potential with three examples: 1) pesticide dynamics in an agricultural catchment (the creek dataset), 2) wastewater-based epidemiology (the WWTP\textsubscript{raw} dataset) and 3) long-term surveillance (the river dataset).

1) **Creek.** Pesticide concentrations are known to vary greatly over time, but highly temporally resolved long-term data has been lacking due to the prohibitive effort required to obtain such data. This limits the quantitative analysis of peak concentrations in rivers, which may be of crucial ecotoxicological importance, and the understanding of sources, transport pathways, and travel times. We quantified 20 analytes in the creek dataset, yielding almost 3,000 data points (41-day time series at 20-min intervals). For the insecticide thiacloprid, this dataset reveals up to tenfold exceedances of the acute environmental quality standard after rain events. The 3.5-day composite sampling approach, which is the current standard in Swiss water quality monitoring, would have missed this entirely (Fig. 2a). Only the high temporal resolution sampling correctly reports the acute risk to water organisms posed by agricultural pesticides. Furthermore, time profiles of five pesticides appear with individual delays after a rain event (Fig. 2b). The differing delay times indicate the spatial separation of their sources in the catchment or different wash-off behaviors or a combination of the two, resulting in a variety of transport patterns. The sequence of pulsed exposure is relevant to macroinvertebrates\textsuperscript{22}, and MS\textsuperscript{2}field is able to provide realistic scenarios for toxicokinetic-toxicodynamic studies\textsuperscript{23}.

2) **WWTP\textsubscript{raw}.** Quantifying the use of illicit drugs is one area of wastewater-based epidemiology. Fig. 2c shows three time series: cocaine (COC), benzoylecgonine (BE), and their ratio (COC/BE). After consumption, COC is excreted as its main metabolite BE, which is stable under most environmental conditions\textsuperscript{24}, and in smaller amounts as unmetabolized COC, which is less stable\textsuperscript{25}. Therefore, COC consumption is typically estimated from BE concentrations. Abnormal COC/BE ratios in 24-hour composite samples may indicate COC disposal\textsuperscript{26}. MS\textsuperscript{2}field provides reliable COC/BE ratios at high temporal resolution due to minimal COC in-sample degradation. It thus provides insight into the COC fraction that was excreted after consumption and into sources of unconsumed COC resulting from any kind of disposal. At least two such events that exceed the 95% quantile (Fig. 2c: >3.5-fold ratio compared to median, red) were observed. Such short events would not be reflected in daily averages (Fig. 2c, blue) even if in-sample stability of COC was guaranteed. Highly temporally resolved time
series combined with knowledge of pharmacokinetics and sewer system properties will benefit wastewater-based epidemiology in identifying whether levels of health indicators are generally elevated or only episodic.

The 20-minute interval time series from the WWTPraw dataset also allows for the detection and characterization of pollution patterns that may go undetected or remain poorly quantified. This includes unexpected weekday loads of the herbicide diuron (potentially from industrial discharge), episodic events such as rain wash-off of the herbicide mecoprop from flat roofs, and surprisingly systematic diurnal patterns in pharmaceuticals and household chemicals (valsartan, candesartan, diclofenac, benzotriazole; see Fig. S7).

3) River. The river dataset confirms that robust long-term operation is possible under variable hydrological conditions without expert personnel onsite. High-quality acquisition can be achieved without weekly services, with LC and HRMS running at high accuracy over at least seven weeks without maintenance (<3ppm mass drift over 2 weeks, <0.3min retention time drift over 6 weeks, see Figs. S5, S6). Concentrations as low as <100ng/L were measured over this time without loss in sensitivity or precision. Despite several rain events, the 7-week dataset shows no elevated pollutant concentrations in this karst-dominated catchment with four WWTPs upstream. This dataset provides information about potential pollution sources that is important to water managers yet is unobtainable with traditional means.

Time series analysis reveals contamination clusters. The LC-HRMS dataset’s >10^4 unknown signals enable unsupervised pattern discovery using time series and clustering methods. These methods can be used to i) discover chemical compounds with characteristic time patterns; ii) identify outlier clusters pointing to singular events, such as spills; and iii) relate data to external variables.

In an extended retrospective analysis, the WWTPraw dataset was preprocessed with a pipeline extending an established workflow (see SI Materials and Methods) to yield a dataset of 7,885 normalized consolidated chemical signals. Subsequently, each time series was transformed into the frequency and phase domain by the Lomb periodogram. The centered, scaled feature matrix was grouped into 107 clusters using hierarchical clustering. Fig. 3 shows a two-dimensional embedding, frequency profiles, and time profiles of three clusters. Cluster A (red) contains 163 periodic diurnal signals with characteristic peaks after 6 am, pointing to compounds relating to morning excretion. In addition to 3 pharmaceuticals quantified as targets (e.g., valsartan), 8 additional pharmaceuticals (e.g., mycophenolic acid) and 16 excretion-related compounds (cholic acid and other bile acids, urobilin, stercobilin and steroids) could be tentatively identified in this cluster using spectral libraries and computational methods. Cluster B (green) contains 66 features with a diurnal pattern on weekdays and no occurrence on weekends, possibly stemming from industrial emissions. This cluster contains the target compound diuron, whose behavior was noted above, and interestingly a structurally similar active ingredient, cyclanilide, which is not authorized for use in either Switzerland or Europe. We speculate that this compound, used in cotton harvesting, is released from imported products in textile finishing industries. Cluster C (blue, 91 features) is associated with a major and a minor rainfall event. The features in this cluster include atrazine, possibly leaching from soil, and bisphenol S (tentative), which has recently been recognized as a water contaminant. Additional clusters (Fig. S8) show diurnal patterns for pharmaceuticals and weekday/weekend patterns for industrial homologous series. Tables S7-S11, S13-S16 de-

![Figure 3. Frequency cluster analysis of WWTPraw dataset.](image-url)
scribe compound identifications in detail. While further elucidation of environmental pathways is beyond the scope of this work, we demonstrate that the time series analysis of comprehensive environmental datasets can identify novel and unexpected compounds of urban, industrial, and episodic origins that could not previously be detected.

**On-site measurement enables live monitoring.** MS²field extends the application of HRMS to a new realm of possibilities. It offers capabilities unique to mobile, on-site analysis by integrating sophisticated chemical analysis with Internet of Things (IoT) functionality. Automatic processing makes results available minutes after measurement, uploaded to a time series database and accessed through a dashboard via a smartphone or computer. During the campaign at the creek site, the dashboard was configured to track four pesticides online (Fig. S9). These values, here used to assess the exceedance of environmental quality standards, can be applied directly to trigger emergency mitigation measures, for instance by detecting toxic spills in rivers that would compromise safe drinking water production or by managing industrial discharges for in-house pretreatment.

**Modifications and perspectives.** The MS²field transportable platform provides a blueprint that can easily be adapted to a broad range of specific purposes.

The capabilities of the system can be further extended with small modifications. For example, multiplexing between two inlets (e.g., to monitor up-/downstream of a point source) can be achieved by adding a single valve. Furthermore, adding a second HPLC pump for a column switching system could increase temporal resolution to 10min.

The current configuration of MS²field can be combined with orthogonal technologies to provide even more comprehensive analysis of systems, such as measuring chemical and microbiological parameters simultaneously. Fig. S10a shows the attachment of an online flow cytometer to the current MS²field.

In Fig. S10b,c,d, we demonstrate an alternative setup that allows high-throughput analysis of less polar compounds without modifying the unit. A dielectric barrier discharge ionization (DBDI) source is combined with a solid-phase microextraction (SPME) device for liquid or headspace extraction followed by thermal desorption directly in the source.

Beyond these adaptations, more substantial future technical developments would broaden the range of applications of the current platform. A smaller overall footprint and reduced energy usage (e.g., by avoiding the N-generator) would allow MS²field’s use in more remote locations. Integration into a sensor and monitoring network would add a further spatial dimension to the extensive sampling of organic pollutants enabled by the current platform. In combination with other sources of big data such as remote sensing and social media, the chemical data lend themselves to analysis with deep learning to unravel the fate and behavior of chemicals in the environment in previously unimaginable ways.

**Data availability**
The processing code for this manuscript will be made available through GitHub. Additional data is available from the corresponding authors upon request.

**Acknowledgements**
This work was supported by an interdisciplinary Eawag discretionary seed-funding project, and by the Swiss Federal Office for the Environment. We thank all colleagues and partners for their essential contributions and invaluable discussions. Provision of instruments and technical support from industrial partners: Th. Möhring and N. Zehethofer (Thermo Fisher Scientific); G. Böhm and R. Bolliger (CTC Analytics). Construction of the MS²field platform, installation and maintenance at field sites, and data evaluation: R. Fankhauser, A. Joss, C. Ebi, M. Kipf, P. Longrée, P. Becker, A. Dax, A.-K. McCall (Eawag). External partners at field sites: R. Kern (WWTP), M. Koster (Canton Thurgau), C. Badertscher (Canton Jura).

**Author information**

**Contributions**
C.S., H.P.S. and C.O. conceived and supervised the study. All authors provided specific input for the setup and construction of the MS²field trailer. M.A.S. realized the concept, including sensors, analytical workflow, coding of remote control functionalities, and live view. M.A.S, H.P.S. and C.O. were actively involved in the setup and maintenance of at least two field sites. M.A.S. and H.P.S. performed data quality checks of all raw datasets. M.A.S. processed the data (target, non-target, and cluster analysis). All authors discussed the results and shared their expertise in data interpretation. M.A.S wrote the manuscript. All authors contributed towards improving the final manuscript.

**Competing interests**
The authors declare no competing financial interest.
References

(1) UNEP. Global Chemicals Outlook: Towards Sound Management of Chemicals; 2013.


