Decomposing the Bulk Electrical Conductivity of Streamflow To Recover Individual Solute Concentrations at High Frequency

Paolo Benettin*†‡ and Boris M. van Breukelen‡

*Laboratory of Ecohydrology ENAC/IIE/ECHO, École Polytechnique Fédérale de Lausanne, 1004 Lausanne, Switzerland
†Department of Water Management, Faculty of Civil Engineering and Geosciences, Delft University of Technology, 2628 CN Delft, The Netherlands

ABSTRACT: The ability to evaluate stream hydrochemistry is often constrained by the capacity to sample streamwater at an adequate frequency. While technology is no longer a limiting factor, costs and sample management can still be a barrier to high-resolution water quality instrumentation. We propose a new framework for investigating the electrical conductivity (EC) of streamwater, which can be measured continuously through inexpensive sensors. We show that EC embeds information about individual ion content that can be isolated to retrieve solute concentrations at high resolution. The essence of the approach is the decomposition of the EC signal into its “harmonics”, i.e., the specific contributions of the major ions that conduct current in water. The ion contribution is used to explore water quality patterns and to develop algorithms that reconstruct solute concentrations starting from EC during periods where solute measurements are not available. The approach is validated on a hydrochemical data set from Plynlimon, Wales, showing that improved estimates of high-frequency solute dynamics can easily be achieved. Our results support the installation of EC probes to complement water quality campaigns and suggest that the potential of EC measurements in rivers is currently far from being fully exploited.

INTRODUCTION

River hydrochemistry is characterized by marked time variability that typically depends on hydrological and biogeochemical factors. In particular, one essential driver of water quality dynamics is streamflow, which is characterized by frequent transitions between low discharges (that typically reflect the composition of groundwater) and high discharges (that can flush large portions of soil water). It is hence well-known that capturing the structure of hydrochemical behavior requires a sampling frequency that is comparable to, and possibly higher than, the typical time scale of the hydrologic response. In most rivers, this is equivalent to a few hours, which makes the sufficient collection of water samples extremely challenging. Instead, weekly, biweekly, or monthly surveys are conducted, sometimes integrated by event-based high-resolution campaigns. These surveys are fundamental for first-order estimates of solute loads and long-term trends, but they may be insufficient for understanding solute dynamics and for a rigorous assessment of streamwater quality. Transport and water quality models are widely employed as a complementary tool, but they also strongly rely on high-resolution data to calibrate and validate model results. The availability of high-resolution hydrochemical data sets is thus crucial for both monitoring and modeling solute hydrochemistry.

In the past several years, high-resolution data sets have helped in the discovery complex hydrochemical patterns and unprecedented technological advances have now made continuous water quality measurements possible. In terms of costs and management, however, the collection of high-frequency hydrochemical data can still be a challenge. For this reason, reconstructing high-frequency solute behavior through inexpensive “surrogate measures” of solute concentration like electrical conductivity is a desirable opportunity.

Streamflow electrical conductivity (EC, also known as specific conductance) reflects the presence of ions in flowing water and can be easily measured along with temperature by relatively cheap and durable sensors. EC probes can acquire data at high frequency and have long been used to quantify the total amount of dissolved solids or as a quality check for water chemistry analyses. However, EC measurements are seldom used to support solute concentration measurements with only few applications based on linear regressions between the EC and solute concentration.

The main research question that is investigated here is whether EC measurements can be made useful for retrieving high-frequency water quality information. We propose a new way to interpret the EC signal in streamflow and use it to investigate the temporal evolution of major ion concentrations. The driving hypothesis is that the use of a continuous EC signal...
to integrate low-frequency solute measurements can provide improved estimates of high-frequency solute behavior.

## MATERIALS AND METHODS

The electrical conductivity of an aqueous solution is the capacity to transmit electrical current through the movement of charged ions. Various forms exist to express EC as the sum of the electrical conductivities of the individual ion species in water. In particular, Parkhurst and Appelo propose

$$EC = \sum_i EC_i = \sum_i (\Lambda^0_i \gamma_{EC,i})$$

(1)

where EC is expressed in siemens per meter and for each solute species (denoted by subscript $i$), $\Lambda^0_i$ is the molar conductivity [S/m/(mol/m^3)], $m$ is the molar concentration (mol/m^3), and $\gamma_{EC,i}$ is the electrochemical activity coefficient. To remove the effect of temperature on $\Lambda^0_i$ and $\gamma_{EC,i}$, EC is typically reported at a standard temperature of 25 °C. Further details about the terms of eq 1 are described in section S2. Equation 1 can be reformulated to stress the time variability of the individual terms. By denoting with $t_i$ the time at which a water sample is collected, we can express the relationship between EC and solute concentration $C$ (mg/L) as

$$EC(t_i) = \sum_i EC_i(t_i) = \sum_i a_i(t_i) C_i(t_i)$$

(2)

where the coefficients $a_i = (\Lambda^0_i \gamma_{EC,i}/M_i) [S/m/(g/mol)]$ indicate known chemical properties of the solutes. Coefficients $a_i$ have a mild dependence on the ionic strength of the solution, so they are not strictly constant or independent. However, in most environmental applications, the ionic strength is rather low and has limited variability, so coefficients $a_i$ can be effectively considered as independent and with only minor time variance.

For each solute species, we can define

$$f_i(t) = \frac{EC_i(t)}{EC(t)}$$

(3)

which represents the relative contribution of each solute to total EC. The terms $f_i$ can be seen as weights that describe how much an individual solute species influences the measured EC, because of its chemical properties and concentration. Besides allowing a rank of the solutes according to their contribution to EC, the knowledge of weights $f_i$ allows us to invert eq 2 and obtain the solute concentration starting from EC measurements as $C_i(t) = EC(t) f_i(t)/a_i(t)$. The key advantage of this inversion is that it can be extended to any time $t$ at which EC measurements and reliable estimates of coefficients $f_i$ and $a_i$ are available:

$$C_i(t) = \frac{f_i(t)}{a_i(t)} EC(t)$$

(4)

Given that EC probes can provide almost continuous measurements and that coefficients $a_i$ are rather constant, the ability to compute the high-frequency solute concentration through eq 4 translates into the capacity to properly estimate the individual contributions $f_i(t)$.

## PROOF OF CONCEPT

To show the validity of the approach, we applied it to the water quality data set publicly available for the Upper Hafren (UHF) river in the Plynlimon area, mid-Wales (U.K.). The data set includes 7 h frequency streamwater samples, analyzed for more than 40 elements of the periodic table and for additional parameters like EC (at 25 °C), pH, and alkalinity. We selected seven major ions (Na^+, Ca^{2+}, Mg^{2+}, K^+, Cl^-, SO_4^{2-}, and NO_3^-) and obtained H^+ from pH and HCO_3^- from speciation calculation with Gran Alkalinity as input. Some large gaps in the alkalinity series were filled through a linear regression with pH to allow the analysis to be extended to a larger number of samples (section S3).

EC Decomposition. The first goal of the analysis is the decomposition of the bulk EC signal into its ion contributions. To test the accuracy of eq 1, the estimated EC was first compared to the measured values (section S4.1). All computations refer to the standard temperature of 25 °C, for the sake of consistency with measurements. The result (Figure S2) is generally accurate, with 95% of the errors within ±10%.

The good match between measured and calculated EC indicates that the estimated contributions of the nine major ions are generally appropriate. Weights $f_i(t)$ were then computed through eq 3, where the $EC_i(t)$ terms were obtained as $a_i(t) C_i(t)$ and the $EC(t)$ term was set equal to the measured EC. The procedure is applied to each solute independently, so it can be used to compute $f_i(t)$ for the available ion measurements even in cases in which other major ion concentrations are missing. The time series of weights $f_i$ are shown in Figure 1, where solutes are ranked according to their mean contribution to EC. Note that given the differences between measured and computed EC (Figure S2), the sum of the weights can occasionally be different from 1. Figure 1 shows the “harmonics” of the EC signal. Cl^- and Na^+ are the most

Figure 1. Relative contributions ($f_i$) of individual ions to the electrical conductivity (EC) signal at UHF, computed through eq 3. Cl and Na account for more than 60% of EC and have the lowest relative variability.
influential solutes as together they provide more than 60% of EC. Although \( f_{\text{Cl}} \) and \( f_{\text{Na}} \) display some variability (especially after day 200), they are the weights with the lowest relative variability. Other ions have smaller contributions, \( \leq 10\% \), except for \( H^+ \) that has remarkably high and variable contributions related to acidic stream conditions during high flows.\(^{23}\)

Regardless of the particular dynamics, all solutes show potential for pattern exploration, including the dependence on stream discharge and the interdependence with other solutes. Because weights \( f_i \) represent the EC contribution of each solute compared to that of the whole solution, the variability in \( f_i \) follows from contrasting solute behavior. This is most visible in the second part of the data set, where most \( f_i \) values are characterized by sharp depressions that correspond to \( H^+ \) peaks during high flows. Computations also showed (Figure S3) that coefficients \( a_i \) exhibit only minor variations in time (maximum of \( \pm 1\% \)), so they could be effectively considered as a solute property.

Retrieving High-Frequency Solute Dynamics. For solutes whose weights can be reliably predicted, one can use eq 4 to obtain solute concentration estimates at the same frequency as EC. This can be especially useful for complementing long-term water quality surveys that are often conducted by environmental agencies. Indeed, in the absence of higher-frequency information, low-frequency solute concentrations are typically interpolated over the sampling interval to, e.g., estimate solute loads. The second goal of the analysis is then to assess whether the use of a continuous EC signal to integrate low-frequency solute measurements can provide an approximation of solute behavior that is significantly better than the simple interpolation of low-frequency concentration measurements. This is not a trivial hypothesis as by using EC one could induce an unrealistic behavior in the solute and ultimately obtain a worse approximation. To address this problem, we again used the UHF data set and selected the two ions with the largest contributions to EC, i.e., chloride and sodium. For both solutes, we first extracted low-frequency (e.g., weekly) “grab” subsamples of the data set, which may represent the low-frequency grab samples available from a water quality campaign. Then, instead of using the grab samples to interpolate their solute concentration, we used them to interpolate their ion contributions to EC and obtain high-frequency estimates of \( f_i(t) \) and \( a_i(t) \). Such estimates were finally coupled to the measured EC signal (eq 4) to obtain high-frequency estimates of solute concentration. This procedure was implemented for grab sample frequencies from 14 h to 31 days. As predictions are influenced by the choice of the first extracted sample, a different prediction was generated for each possible choice of the initial sample. To evaluate the quality of the EC-aided method, we computed a prediction error as the mean absolute difference between the measured and estimated high-frequency concentrations (excluding the data points corresponding to the grab samples, as their error is null by definition). For comparison, we computed the prediction error originating from the simple interpolation of the grab samples’ concentrations. Results are shown in Figure 2, where the errors are plotted as a function of the sampling frequency. All curves start from an error of \( \approx 2\% \) corresponding to the highest extractable frequency (14 h). For increasing sampling intervals, the errors grow, but the curves featuring the EC-aided methodology remain substantially lower than the ones corresponding to the linear interpolation, with an approximately 40% decrease in error. Figure 2 also shows that the error of the EC methodology with a 14 day frequency is the same as that from a linear interpolation with a 3 day frequency. For additional comparison, we also computed the error of a least-squares linear regression between solute concentration and EC (section S5.1). The error in the regression behaves as an asymptote for the EC-aided estimate, suggesting that these two methods approximately converge for very large (>1 month) sampling intervals. This is not surprising as by progressively increasing the sampling interval one tends to a single, mean solute contribution to EC, which in turn tends to the slope of the linear regression when the intercept is close to zero (section S5.2).

An example of chloride estimation using bimonthly samples is further shown in Figure 3. The plot shows that, compared to a linear interpolation, the estimated chloride concentration can reproduce most of the high-frequency fluctuations of the real signal. Indeed, the empirical distribution of the predicted concentration is very similar to that of measured chloride (Figure 3, inset). Figure 3 also shows that the use of a linear regression to estimate the chloride concentration can accurately reproduce the high-frequency fluctuations but cannot reproduce some seasonal trends like those between days 80 and 200; hence, the mean error of the performance (as shown in Figure 2) remains relatively high.

### DISCUSSION

The core and novelty of the approach are the interpretation of EC as a bulk signal of hydrochemistry to be decrypted. Regardless of the “decoding” technique, the opportunity to decompose the EC signal to trace the presence of different ions in the flowing water (as shown in Figure 1) is a new avenue that calls for additional research. One strength of the methodology is its mechanistic foundation that allows us to understand the complex dynamics of the EC signal. Indeed, EC
is not just correlated to Cl and Na; rather, it is caused by those solutes.

The proof-of-concept application to UHF provides some preliminary guidelines about cases in which the approach is expected to work. Solutes with small contributions to EC are more difficult to isolate in the EC decomposition and are prone to high relative errors on the estimation of weights $f_i$ (Figure S5). This directly translates into larger errors in the concentration estimate (section S4.3). The variability in weights $f_i$ (Figure 1) arises when solutes have contrasting behaviors and represents the major challenge to EC decomposition (section S4.3). Simplified techniques like the linear interpolation of low-frequency values are sufficient to show the potential of the approach and can provide valid approximations for the strongly contributing solutes [like Cl and Na at UHF (Figure 2)], but better algorithms are required to approximate the weights $f_i$ for poorly contributing solutes (like Ca and NO$_3$ at UHF). Further developments of the approach, hence, point to improved algorithms that explicitly take into account the integrated solute dynamics and incorporate hydrochemical knowledge that is available for the site. Moreover, other variables like water flow, temperature, and pH that typically have an influence on solute concentration and are often available at the same frequency as EC may be embedded.

The UHF stream is a natural environment where EC is low and mostly controlled by two solutes, but it is also characterized by acidic conditions at high flows that cause sharp variations in the $f_i$ contributions. Different systems are expected to have very different contributions to EC depending on their particular hydrochemistry, but the approach is general and can be explored in various ways, e.g., starting from the computation of the contributions to EC for the existing water quality data sets. The approach is also obviously related to the ability to accurately measure EC (which requires maintenance of the sensor) and can be influenced by several undesired factors, like road salting during snow seasons.

Long-term water quality campaigns are being conducted in many sites worldwide by research groups and water quality agencies. If EC probes are installed at such sites (if not already present), the methodology can be immediately applied at almost zero cost. For example, results suggest that continuous EC measurements at Plynlimon could be coupled to long-term chloride measurements to aid the estimation of the high-frequency chloride concentration. There is indeed enormous potential for deploying cheap networks of EC probes in streamflow (as also for precipitation or groundwaters) and identifying multiple signatures of hydrologic transport. This is not currently done because EC is traditionally treated as a qualitative indicator of total dissolved solids, but the key result of this research is that there is much more information that can be recovered from the EC signal.

Finally, results introduce a potential for using the EC signal in solute transport modeling. State-of-the-art models can provide outputs at high temporal resolutions and are often limited by the availability of data. Given the high information potential contained in EC and addressed in this paper, we envision the opportunity in the future to use information from a continuous EC signal to support the calibration of transport models.

### ASSOCIATED CONTENT

* Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.estlett.7b00472. Details about methods and results (PDF)

### AUTHOR INFORMATION

*E-mail: paolo.benettin@epfl.ch. Phone: +41 21 69 33773.

**ORCID**

Paolo Benettin: 0000-0001-7556-1417

---

*Figure 3. Example of chloride prediction based on biweekly grab samples. The inset reports the empirical distributions of the measured and estimated chloride signals. The interpolated concentration gives an incomplete picture of chloride behavior, while the regression with EC misses the seasonal chloride dynamics. The EC-aided methodology captures all main solute dynamics.*

---
Notes
The authors declare no competing financial interest.

ACKNOWLEDGMENTS
Water quality data from the Upper Hafren catchment are available from the Center for the Environment (CEH) through the data portal. B.M.v.B. thanks the Delft University of Technology (TU Delft) for direct funding. P.B. thanks the available from the Center for the Environment (CEH) through Water quality data from the Upper Hafren catchment are ACKNOWLEDGMENTS — What is Their Precise Relationship? Desalination 1989, 72, 275–292.

REFERENCES