



Supplement of

Complex interactions of in-stream dissolved organic matter and nutrient spiralling unravelled by Bayesian regression analysis

Matthias Pucher et al.

Correspondence to: Matthias Pucher (matthias.pucher@wcl.ac.at)

The copyright of individual parts of the supplement might differ from the article licence.

S1 Mathematical details on the INSBIRE approach

In-stream samples were taken in a longitudinal series which is formally identical to a time series problem. We re-arranged the equations of the nutrient spiralling concept (Stream Solute Workshop, 1990) so that differences are replaced by current (e.g. C_x) and past (e.g. C_{x-1}) values of series. These equations conform to a time series including past values of the same

5 variable as well as current and past values of other variables and are a form of non-linear autoregressive exogenous models (NARX; e.g. Billings, 2013). Several studies used the original equations of the Stream Solute Workshop protocol (1990) and solved them via variable transformation. Still, the results from a linear regression using transformed data and those of a direct non-linear fit differ (e.g. Stedmon et al., 2000). Therefore, we regard a non-linear solving algorithm superior in terms of accuracy. We proceeded differently, transforming all equations into a NARX form. Additionally, we extended the
10 equations to incorporate interactions (Eqs. S7 and S8).

For the parameter determination we used a non-linear Bayesian fitting algorithm from the R package *brms* (Bürkner, 2017). We choose priors to approximately fit knowledge from other studies (e.g. Mineau et al., 2016) while keeping them broad, so they do not dominate the results. Priors and especially their limits were also adjusted to deliver converging models.

15 During a plateau addition experiment, concentration changes in a conservative tracer due to dilution effects. This was transformed into a NARX form (Eq. S1). We used this equation to determine the dilution factors and to correct measured
DOC and nutrient concentrations as well as DOM components by the measured changes in conductivity.

$$C_{x,t} = C_{amb,x,t} + (C_{x-1,t} - C_{amb,x,t}) \frac{dil_x}{dil_{x-1}} \quad (S1)$$

x ... index of longitudinal sampling points

t ... index of addition date

$C_{x,t}$... concentration at point x and date t (variable)

20 $C_{amb,x,t}$... ambient concentration at point x and date t (variable)

dil_x ... dilution factor at point x (once calculated fixed values)

A reactive substance can be modelled using Eq. (S2). Variable x from the original equation (Stream Solute Workshop, 1990) was replaced by $(d_{x-1} - d_x)$ to address a case with several sampling points with different distances from the addition point.

$$C_{x,t} = C_{amb,x,t} + (C_{x-1,t} - C_{amb,x,t}) \frac{dil_x}{dil_{x-1}} e^{\frac{d_{x-1} - d_x}{s_w}} \quad (S2)$$

s_w ... nutrient uptake length (parameter)

25 prior: $s_w \sim Lognormal(400,200)$, $s_w \in [0.01,10000]$

d_x ... distance of point x from origin (fixed)

Using the flow velocity and the water depth, the nutrient uptake velocity can be calculated from s_w (Eq. S3). This is useful to reduce flow-dependent effects.

$$\frac{1}{s_w} = v_f (uz)^{-1} \quad (S3)$$

$$C_{x,t} = C_{amb,x,t} + (C_{x-1,t} - C_{amb,x,t}) \frac{dil_x}{dil_{x-1}} e^{(d_{x-1} - d_x) v_f (uz)^{-1}} \quad (S4)$$

v_f ... nutrient uptake velocity (parameter)

30 prior: $v_f \sim Lognormal(0.7,3)$, $v_f \in [0.01,35]$

u ... flow velocity (calculated by Hec-RAS, then fixed)

z ... water depth (calculated by Hec-RAS, then fixed)

The areal uptake rate can then be modelled using Eqs. 5 and 6:

$$v_f = U C_{x,t}^{-1} \quad (S5)$$

$$C_{x,t} = C_{amb,x,t} + (C_{x-1,t} - C_{amb,x,t}) \frac{dil_x}{dil_{x-1}} e^{(d_{x-1} - d_x) U C_{x,t}^{-1} (uz)^{-1}} \quad (S6)$$

U ... areal uptake rate (parameter)

35 prior: $U \sim Lognormal(2,3)$, $U \in [0.01,40]$

A linear relation between uptake velocity and concentration is needed to properly calculate U. In other cases, uptake functions such as the Michaelis-Menten formulation can be used to describe the observed uptake-concentration relation (Stream Solute Workshop, 1990). An uptake efficiency loss, mathematically described by a power function, was shown in experiments with N-NO₃ (Dodds et al., 2002; O'Brien et al., 2007). We additionally tested a linear function, an asymptotic regression function and an exponential function. A mechanistic argumentation for either of these functions is difficult (Stream Solute Workshop, 1990), but testing the suitability with the Bayes factor (BF) leads to good empirical fits.

40 To include interactions, we added a product of power functions for relevant compounds and nutrients (Eqs. S7 and S8). Where beneficial, the wetted width w was added to analyse influences of the stream bed surface on retention processes. The added value l was added as a degree of freedom to allow curves, that do not go through the origin. The relevance of these 45 effects was tested in the modelling process by comparing different combinations of compounds in models using the BF, which can also be a measurement of variable importance.

$$v_f = kw \left(l + \prod_i C_{i,x,t}^{mi} \right) \quad (S7)$$

$$C_{x,t} = C_{amb,t} + (C_{x-1,t} - C_{amb,t}) \frac{dil_x}{dil_{x-1}} e^{(d_{x-1} - dx)kw \left(l + \prod_i C_{i,x,t}^{m_i} \right) / uz} \quad (S8)$$

k ... uptake rate factor (parameter)

$$\text{prior: } k \sim \text{Lognormal}(0.7, 3), k \in [0.01, 35]$$

w ... wetted width, constant 1 to represent no influence (calculated by Hec-RAS, then fixed)

50 1 ... additive value (parameter)

i ... index of nutrient or DOM component

$C_{i,x,t}$... concentration of compound i at point x and date t (variable)

m_i ... exponent determining the strength of the relations (parameter)

$$\text{prior: } m_i \sim \text{Normal}(-0.2, 0.4), m_i \in [-1, 1] \text{ if a dampening influence was assumed from literature}$$

55 $m_i \sim \text{Normal}(0.2, 0.4), m_i \in [-1, 1] \text{ if a stimulating influence was assumed}$

Since we had no prior information for m_i from previous studies, it was important to test the influence of the prior on the final results by using a uniform distribution and normal distributions with different parameters. In the presented models, the priors for any parameters did not dominate the results, but were important to ensure a convergent of the fitting algorithm. Due to the double-exponential structure of Eq. (S8) in m_i , the limits were essential.

60 To set up the models, we used the difference of concentrations (Eq. S9) as the dependent variable and restructured the equations above accordingly. We assumed a normal error distribution for the differences of concentrations and the differences of fluorescence. The nature of the measurements would also allow a log-normal error distribution, but our data clearly deviated from that assumption, so we used a normal error distribution.

$$D_{x,t} = C_{x,t} - C_{amb,t} \quad (S9)$$

$D_{x,t}$... concentrations (DOC, SRP, N-NO₃) or fluorescence (DOM PARAFAC components) deviation from ambient

65 conditions

$$\text{model error assumptions: } D_{x,t} \sim \text{Normal}(\mu_{x,t}, \sigma^2)$$

$\mu_{x,t}$... calculated difference from Eqs. (S2), (S4) and (S8) restructured to suffice Eq. (S9)

70 The accuracy of the model can be compared to expected measurement errors (e.g. lab instrument errors, errors from sampling procedure) and show the point where no additional information can be expected from the data (for proper error propagation analysis see Haefner, 2012, chapter 9). Using the simulated probability density of the residuals, which is in the same units as the measured values, we get an impression if further information can be expected from the data.

The 95% probability interval of the residuals can be a meaningful metric of the model accuracy. This approach makes it easier to distinguish between signal and noise compared to an approach where Eqs. (S2), (S4) and (S8) are applied step-wise

75 and error propagation is not considered. It can also help in planning the experimental scheme to improve the signal-to-noise ratio because amongst others, the error depends on the instruments, sample handling, concentrations and concentration difference of consecutive samples.

S2 Additional resources

80 **Table S1: Samplings, dates, discharge, concentrations of nutrients and fluorescences (additionally in %) of DOM PARAFAC components. Additionally introduced amounts are calculated from expected ambient and measured values at the listed dates.**

source	date	code	discharge l s ⁻¹	DOC						P-PO ₄		N-NO ₃		N-NO ₂		N-NH ₄		Hum-mic (C1)		Hum-ter (C2)		Hum-micter (C3)		Qui (C4)		Trp (C5)		Tyr (C6)	
				µg l ⁻¹	Fmax RU	%	Fmax RU	%	Fmax RU	%	Fmax RU	%	Fmax RU	%	Fmax RU	%	Fmax RU	%	Fmax RU	%	Fmax RU	%							
ambient																													
none	2018-07-16	A	0.93	1301	9.4	2548	11.1	10.9	0.53	44	0.21	17	0.24	20	0.08	6	0.13	10	0.02	2									
none	2018-07-23	C	0.73	1459	13.9	1890	12.9	10.2	0.56	42	0.26	20	0.27	20	0.09	7	0.11	8	0.03	2									
none	2018-07-30	F	0.69	1279	11.2	1963	10.4	9.5	0.54	44	0.23	19	0.26	21	0.09	7	0.11	9	0.01	0									
none	2018-08-06	I	0.67	1420	18.8	2194	11.3	13.4	0.57	43	0.25	19	0.27	20	0.10	7	0.13	10	0.01	1									
none	2018-08-13	L	0.47	1257	11.3	1893	7.8	10.5	0.54	43	0.23	18	0.25	20	0.09	7	0.13	10	0.01	1									
none	2018-08-20	O	0.41	1229	3.3	1640	5.4	0.9	0.51	44	0.21	18	0.24	20	0.08	7	0.12	10	0.01	1									
material additionally introduced during leachate additions																													
corn	2018-08-07	J	0.58	403	4.9	75	0.2	0.0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.07	59	0.05	41							
corn	2018-08-14	M	0.82	1534	19.5	25	0.8	0.0	0.03	19	0.01	9	0.01	4	0.02	12	0.04	25	0.05	31									
cow dung	2018-07-19	B	0.80	2284	62.0	0	5.9	60.4	0.20	28	0.09	12	0.08	11	0.11	16	0.12	17	0.12	16									
cow dung	2018-07-26	E	0.64	1235	31.5	710	1.1	0.0	0.15	19	0.02	3	0.02	3	0.04	5	0.12	15	0.45	56									
leaves	2018-07-24	D	0.60	272	2.6	193	3.8	1.5	0.28	26	0.00	0	0.00	0	0.03	3	0.57	54	0.18	17									
leaves	2018-07-31	G	0.41	1225	25.2	707	0.0	0.4	0.59	67	0.00	0	0.00	0	0.08	9	0.03	4	0.17	20									
nettles	2018-08-16	N	0.44	148	2.3	92	51.4	23.0	0.00	0	0.00	0	0.00	0	0.01	15	0.01	35	0.02	49									
nettles	2018-08-21	Q	0.38	188	8.5	252	33.8	26.9	0.04	18	0.01	5	0.01	3	0.02	10	0.05	26	0.07	37									
pig dung	2018-08-02	H	0.46	469	37.6	157	5.5	45.1	0.09	25	0.06	16	0.03	9	0.01	3	0.06	17	0.11	2	9								
pig dung	2018-08-09	K	0.72	251	49.0	193	5.0	62.3	0.05	25	0.02	10	0.02	10	0.02	9	0.05	27	0.03	1	9								

Table S2: Correlation of nutrient and DOM fraction uptake velocity v_f ; Bayes factor in brackets; only shown, if Bayes factor > 1.

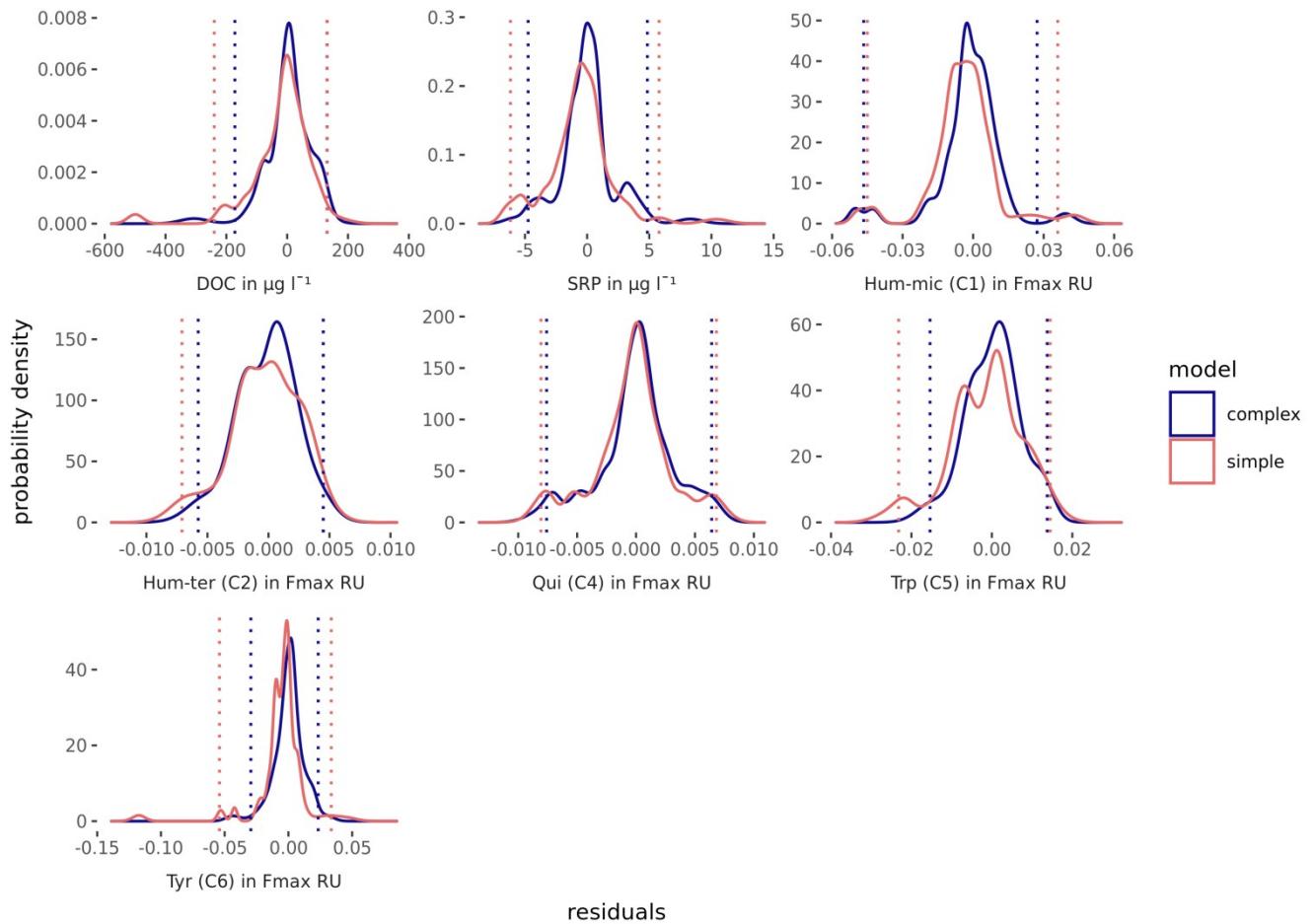
	Hum-micter (C3)	Qui (C4)	Tyr (C6)
Trp (C5)	0.78 (2.98)	0.50 (1.31)	
Tyr (C6)		0.66 (6.10)	
DOC		0.58 (4.63)	0.41 (1.36)

85 **Table S3: Model comparison v_f with and without random effects (mixed models, MM) of source and addition date. The Bayes R^2 shows the absolute model performance and the BF indicates whether the addition of the random effects leads to a model improvement.**

model	Original model	Source effects		Date effects		comment	
		Bayes R^2	Bayes R^2	BF	Bayes R^2	BF	
Hum-mic (C1)		0.51	0.48	0.17	0.50	4.61	The date has a substantial impact on v_f .
Hum-ter (C2)		0.34	0.49	0.7	0.49	0.65	Neither date nor source improved the model.
Hum-micter (C3)		0.52	0.51	0.21	0.54	0.22	Neither date nor source improved the model.
Qui (C4)		0.46	0.46	0.09	0.45	0.12	Neither date nor source improved the model.
Trp (C5)		0.29	0.29	1.39	0.48	134.23	The date has a decisive influence on v_f .
Tyr (C6)		0.29	0.34	10.68	0.70	1.2e8	Both, source and date improved the model, but the effects of date were stronger.
DOC		0.26	0.46	1563	0.46	146	Both, source and date improved the model, but the effects of source were stronger.
NO ₃		0.16	0.29	0.41	0.29	0.65	Neither addition nor source improved the model.
SRP		0.56	0.57	0.17	0.56	0.11	Neither addition nor source improved the model.

Table S4: BFs for inclusion of model parameters and probability intervals of the parameter estimates. The Table provides additional information to Table 7. v_f : uptake velocity, k : uptake rate factor, w : wetted width, C_i : fluorescence of PARAFAC components, m_i : exponent of relation, l : additive parameter

fraction/ nutrient	most probable model (Eq. 1)	BF in favour of variable inclusion	estimates, [95% probability interval]
Hum-mic (C1)	$v_f = k w C1^{mc1}$	$P(w \neq 1): BF = 7.34$ $P(mc1 \neq 0): BF = 1.4$	$k = 2.11, [1.65, 2.59]$ $mc1 = -0.38, [-0.93, 0.28]$
Hum-ter (C2)	$v_f = k (l + DOC^{mc})$	$P(l \neq 0): BF = 2.36$ $P(mc \neq 0): BF = 7.69$	$k = 0.11, [0.01, 0.61]$ $l = 3.16, [0.23, 8.01]$ $mc = 0.32, [-0.42, 0.60]$
Hum- micter (C3)	$v_f = v_f$	-	-
Qui (C4)	$v_f = k C1^{mc1} C4^{mc4}$	$P(mc1 \neq 0): BF = 2.54$ $P(mc4 \neq 0): BF = 2.44$	$k = 0.71, [0.14, 2.23]$ $mc1 = -0.25, [-0.89, 0.39]$ $mc4 = -0.35, [-1.05, 0.38]$
Trp (C5)	$v_f = k C2^{mc2} C5^{mc5}$	$P(mc2 \neq 0): BF = 2.71$ $P(mc5 \neq 0): BF = 3.13$	$k = 0.85, [0.10, 3.20]$ $mc2 = -0.44, [-1.23, 0.35]$ $mc5 = -0.55, [-1.31, 0.22]$
Tyr (C6)	$v_f = k C2^{mc2} C6^{mc6}$	$P(mc2 \neq 0): BF = 2.34$ $P(mc6 \neq 0): BF = 1.46e7$	$k = 0.27, [0.06, 0.76]$ $mc2 = -0.23, [-0.98, 0.52]$ $mc6 = -0.96, [-1.25, -0.69]$
DOC	$v_f = k C6^{mc6}$	$P(mc6 \neq 0): BF = 10.50$	$k = 0.30, [0.10, 0.75]$ $mc6 = -0.62, [-0.95, -0.18]$
NO_3	$v_f = v_f$	-	-
SRP	$v_f = k w SRP^{mp}$	$P(w \neq 1): BF = 31.93$ $P(mp \neq 0): BF = 6.21$	$k = 26.18, [10.17, 39.20]$ $mp = -0.31, [-0.45, -0.07]$



95 **Figure S1: Probability density of the residuals for DOC, SRP and five of the six PARAFAC components. Comparison** of two models: the simple model is according to Eq. (4), the complex one is the best performing model found, which is a mixed model including the leachate source (DOC), a mixed model including the sampling (Trp (C5) and Tyr (C6)) or the interaction models from table 5 (SRP, Hum-mic(C1), Hum-ter (C2), Qui (C4)). The dotted vertical lines are the boundaries of the 95% probability density interval. The probability density interval is smaller for better performing models.

100

References

Billings, S. A.: Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains, John Wiley & Sons, 605 pp., 2013.

Bürkner, P.-C.: brms: An R Package for Bayesian Multilevel Models Using Stan, *J. Stat. Soft.*, 80, 1–28, <https://doi.org/10.18637/jss.v080.i01>, 2017.

Dodds, W. K., López, A. J., Bowden, W. B., Gregory, S., Grimm, N. B., Hamilton, S. K., Hershey, A. E., Martí, E., McDowell, W. H., Meyer, J. L., Morrall, D., Mulholland, P. J., Peterson, B. J., Tank, J. L., Valett, H. M., Webster, J. R., and Wollheim, W.: N uptake as a function of concentration in streams, *J. N. Am. Benthol. Soc.*, 21, 206–220, <https://doi.org/10.2307/1468410>, 2002.

Haefner, J. W.: Modeling Biological Systems: Principles and Applications, Springer Science & Business Media, 486 pp., 2012.

Mineau, M. M., Wollheim, W. M., Buffam, I., Findlay, S. E. G., Hall, R. O., Hotchkiss, E. R., Koenig, L. E., McDowell, W. H., and Parr, T. B.: Dissolved organic carbon uptake in streams: A review and assessment of reach-scale measurements, *J. Geophys Res.-Biogeo.*, 121, 2019–2029, <https://doi.org/10.1002/2015JG003204>, 2016.

O’Brien, J. M., Dodds, W. K., Wilson, K. C., Murdock, J. N., and Eichmiller, J.: The saturation of N cycling in Central Plains streams: ^{15}N experiments across a broad gradient of nitrate concentrations, *Biogeochemistry*, 84, 31–49, <https://doi.org/10.1007/s10533-007-9073-7>, 2007.

Stedmon, C. A., Markager, S., and Kaas, H.: Optical Properties and Signatures of Chromophoric Dissolved Organic Matter (CDOM) in Danish Coastal Waters, *Estuar. Coast. Shelf S.*, 51, 267–278, <https://doi.org/10.1006/ecss.2000.0645>, 2000.

Stream Solute Workshop: Concepts and Methods for Assessing Solute Dynamics in Stream Ecosystems, *J. N. Am. Benthol. Soc.*, 9, 95–119, <https://doi.org/10.2307/1467445>, 1990.