

## Uncertainty estimation of end-member mixing using generalized likelihood uncertainty estimation (GLUE), applied in a lowland catchment

Joost R. Delsman,<sup>1,2</sup> Gualbert H. P. Oude Essink,<sup>1</sup> Keith J. Beven,<sup>3,4</sup> and Pieter J. Stuyfzand<sup>2,5</sup>

Received 3 April 2013; revised 28 May 2013; accepted 29 May 2013; published 7 August 2013.

[1] End-member mixing models have been widely used to separate the different components of a hydrograph, but their effectiveness suffers from uncertainty in both the identification of end-members and spatiotemporal variation in end-member concentrations. In this paper, we outline a procedure, based on the generalized likelihood uncertainty estimation (GLUE) framework, to more inclusively evaluate uncertainty in mixing models than existing approaches. We apply this procedure, referred to as G-EMMA, to a yearlong chemical data set from the heavily impacted agricultural Lissertocht catchment, Netherlands, and compare its results to the “traditional” end-member mixing analysis (EMMA). While the traditional approach appears unable to adequately deal with the large spatial variation in one of the end-members, the G-EMMA procedure successfully identified, with varying uncertainty, contributions of five different end-members to the stream. Our results suggest that the concentration distribution of “effective” end-members, that is, the flux-weighted input of an end-member to the stream, can differ markedly from that inferred from sampling of water stored in the catchment. Results also show that the uncertainty arising from identifying the correct end-members may alter calculated end-member contributions by up to 30%, stressing the importance of including the identification of end-members in the uncertainty assessment.

**Citation:** Delsman, J. R., G. H. P. Oude Essink, K. J. Beven, and P. J. Stuyfzand (2013), Uncertainty estimation of end-member mixing using generalized likelihood uncertainty estimation (GLUE), applied in a lowland catchment, *Water Resour. Res.*, 49, 4792–4806, doi:10.1002/wrcr.20341.

### 1. Introduction

[2] Using mixing model approaches to separate the different components of a hydrograph has been instrumental in the development of hydrological science, as environmental tracers provide a unique view of the catchment-integrated response of hydrological flow paths. The use of mixing models has evolved from two-component mixing models [e.g., Johnson *et al.*, 1969; Pinder and Jones, 1969; Sklash and Farvolden, 1979], mostly aimed at separating event and preevent water, to the now commonly used multitracer end-member mixing analysis (EMMA) outlined by Christophersen *et al.* [1990] and Christophersen and Hooper [1992]. EMMA has in recent years been applied in various geographical settings and across

spatial scales [Barthold *et al.*, 2010; Burns *et al.*, 2001; Guinn Garrett *et al.*, 2012; James and Roulet, 2006; Long and Valder, 2011; Soulsby *et al.*, 2003b]. Mixing model approaches are not limited to hydrology, they are also extensively used in other geosciences as geology [Keay *et al.*, 1997; Weltje, 1997], sedimentology [Jmker *et al.*, 2012], and ecology [Rasmussen, 2010].

[3] Mixing model approaches rely on the assumptions that (1) stream water can be explained as a linear mixture of extreme source solutions or end-members, (2) solutes used as tracers in the analysis are conservative, and (3) chemical signatures of end-members are invariant in time and space (at least for single events) and can be reliably characterized [Hooper *et al.*, 1990; Sklash and Farvolden, 1979]. As noted by various authors, these assumptions are commonly violated in real-world applications, giving rise to uncertainty in the resulting hydrograph separations [e.g., Hooper *et al.*, 1990; Soulsby *et al.*, 2003a; Uhlenbrook and Hoeg, 2003]. Two separate uncertainty components can be distinguished. First, the end-members contributing to the stream water mixture have to be properly identified. EMMA theory requires end-members to be chosen that best bound stream water tracer data, given a conceptual understanding of the catchment functioning [Hooper *et al.*, 1990]. Problems in duplicating hydrograph separations using different sets of tracers, however, point to the difficulty in identifying the complete set of relevant end-

<sup>1</sup>Department of Soil and Groundwater, Deltares, Utrecht, Netherlands.

<sup>2</sup>Critical Zone Hydrology Group, Department of Earth Sciences, VU University Amsterdam, Amsterdam, Netherlands.

<sup>3</sup>Lancaster Environment Center, Lancaster University, Lancaster, UK.

<sup>4</sup>Geocentrum, Uppsala University, Uppsala, Sweden.

<sup>5</sup>KWR Watercycle Research Institute, Nieuwegein, Netherlands.

Corresponding author: J. R. Delsman, Department of Soil and Groundwater, Deltares, Utrecht, Netherlands, PO Box 85467, NL-3508 AL, Utrecht, Netherlands. (joost.delsman@deltares.nl)

members using a limited number of tracers [Barthold *et al.*, 2011; Rice and Hornberger, 1998]. By applying more tracers than mathematically necessary, the EMMA approach avoids this problem to a certain extent [Christophersen and Hooper, 1992; Christophersen *et al.*, 1990], and the diagnostic tools developed by Hooper [2003] provide a means to investigate the number of contributing end-members as evidenced from the stream water data set. Nevertheless, Barthold *et al.* [2011] show that the appropriate choice of end-members varies considerably over varying tracer set sizes and composition, resulting in a significant uncertainty. In this paper, we term this type of uncertainty “identification uncertainty.”

[4] Second, in addition to the analytical error always associated with reported concentrations, spatial and temporal variability in end-member concentrations is ubiquitous at the scales considered, and is nigh impossible to characterize adequately using inevitably sparse sampling [Beven, 1989; Burns *et al.*, 2001; Hoeg *et al.*, 2000; Hooper *et al.*, 1990; James and Roulet, 2006; Kendall *et al.*, 2001]. Moreover, even when the variability of a suggested end-member is adequately characterized from sampling, it cannot be assumed that the characterized variability is mirrored in the flux-weighted contribution to the stream water [Kendall *et al.*, 2001; Rinaldo *et al.*, 2011]. And although the authors have argued that spatiotemporal variability may smooth out at larger scales, resulting in “emergent” end-members [Soulsby *et al.*, 2003b], similar characterization problems will apply. We use the term “characterization uncertainty” for this type of uncertainty.

[5] Various authors have quantified characterization uncertainty in mixing models. For instance, Hooper *et al.* [1990] and later Genereux [1998] and Uhlenbrook and Hoeg [2003] mathematically propagated the uncertainty in end-member concentrations, Soulsby *et al.* [2003a] developed a hierarchical Bayesian approach, while other authors applied a Monte Carlo approach to propagate the uncertainty in the chemical signatures of end-members [Bazemore *et al.*, 1994; Durand and Torres, 1996]. Joerin *et al.* [2002] extended the latter approach by allowing for non-normal end-member concentration distributions, and by taking uncertainty in the applied model hypotheses regarding spatial and temporal variation into account, albeit in a simple manner. Iorgulescu *et al.* [2005, 2007] tried to allow for time changing end-members over a sequence of events using a data-based hydrochemical model within a GLUE framework. Barthold *et al.* [2011] proposed an iterative methodology to explore, though not quantify, the identification uncertainty in EMMA.

[6] However, none of the existing approaches account for both identification and characterization uncertainty quantitatively. In addition, none can be applied to end-member mixing analyses using more solutes than mathematically necessary to solve the mixing equations (i.e., overdetermined), even though this is a central property of the widely used EMMA approach [Christophersen and Hooper, 1992]. In this paper, we therefore propose a new method to quantify uncertainty in end-member mixing models, one that specifically considers uncertainty in both identification and characterization of end-members, and allows for overdetermined mixing models. We based our

approach on the generalized likelihood uncertainty estimation (GLUE) methodology of Beven and Binley [1992], which recognizes that given the fundamental limitations of models as descriptors of environmental systems, multiple models and parameter sets may exhibit equifinality in that they all acceptably describe the available observational data [Beven and Binley, 1992; Beven, 1989, 2006].

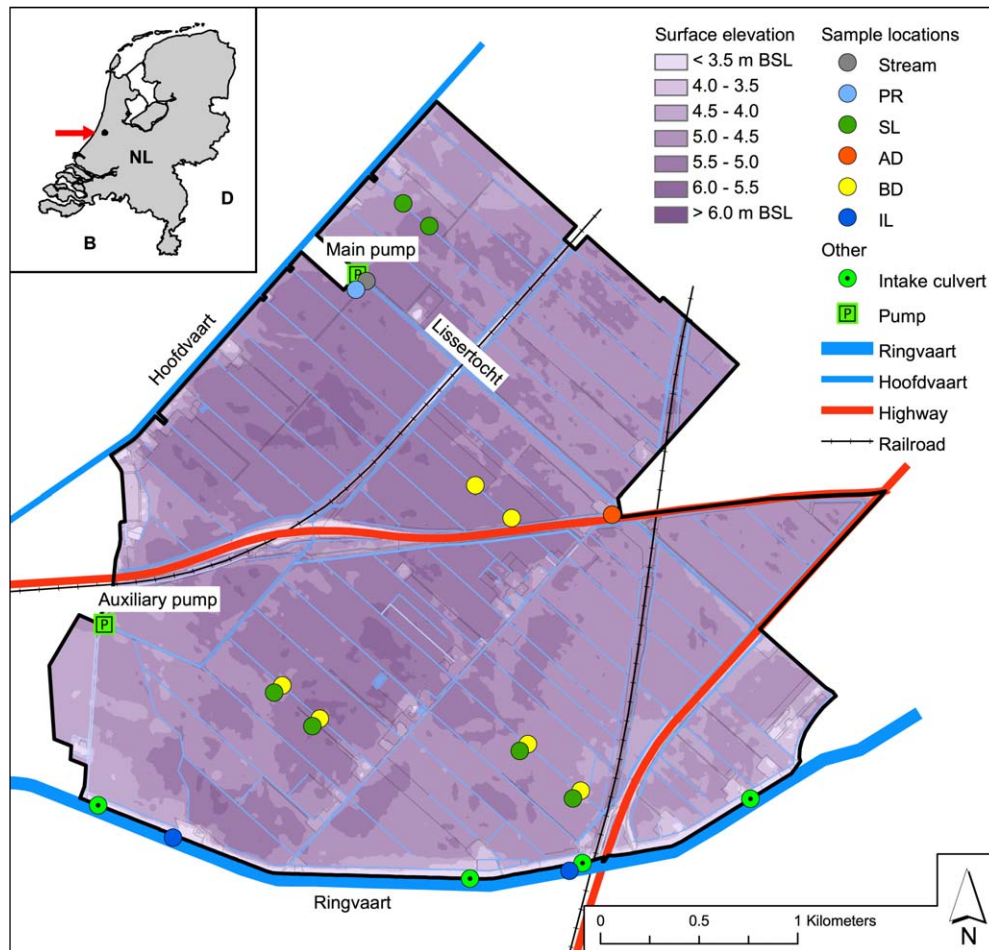
[7] We apply the proposed approach to a small (10 km<sup>2</sup>), heavily impacted agricultural catchment in the coastal region of Netherlands. The catchment provides a difficult test case for our approach, as heavily impacted catchments pose specific challenges to the application of end-member mixing models, with agricultural activities and active water management causing marked changes in hydrology and chemistry [Durand and Torres, 1996]. In addition, the “open boundary” nature of this particular catchment, receiving extraneous fluxes of both regional groundwater flow and water intake, further hampers the application of mixing models. Interest in the hydrological functioning of this catchment is motivated by a projected increase in saline seepage [Oude Essink *et al.*, 2010], that would render the surface water in the catchment unfit for agricultural use.

## 2. Materials and Methods

### 2.1. Lissertocht Catchment

[8] The artificial Lissertocht canal drains a 10 km<sup>2</sup> intensively drained agricultural catchment (Figure 1). The catchment is part of the former lake Haarlemmermeer, reclaimed in 1852, and is located 25 km southwest of the city of Amsterdam in the Netherlands (52°13' latitude, 4°36' longitude). Relief in the catchment is all but flat, with an altitudinal range of 6–3.5 m below mean sea level (BSL). Mean annual precipitation amounts to 840 mm, mean annual potential evapotranspiration to 590 mm [Royal Netherlands Meteorological Institute, 2010]. Excess precipitation is quickly drained through an extensive system of tile drains and ditches. A pumping station at the end of the Lissertocht maintains water levels throughout the catchment at a relatively constant 6.55 m BSL in winter (October–April) and 6.4 m BSL in summer (April–October). An auxiliary pumping station at the western end of the catchment is used only during extreme discharge events. Water is let into the catchment through four culverts from April to October, to maintain surface water levels and improve water quality. Additional fresh water can be taken into the catchment at the location of the auxiliary pump.

[9] The catchment is underlain by an aquifer of Pleistocene fluvial sands (with transmissivity of  $4600 \pm 150$  m<sup>2</sup>/d; data from the Netherlands Hydrological modeling Instrument (NHI) model, available at <http://www.nhi.nu/>). The aquifer is covered by a 6.7 m ( $\pm 0.7$  m) thick layer of heterogeneous Holocene estuarine clays, loamy sands, and peat deposits on top of a thin (5–10 cm) layer of compressed peat deposits, referred to as basal peat [Stafleu *et al.*, 2009]. This Holocene layer presents a considerable hydraulic resistance (i.e., thickness/vertical hydraulic conductivity;  $2400 \pm 750$  d; NHI model data) to vertical groundwater flow. Aquifer hydraulic heads exceed shallow groundwater levels (mostly within 2 m below ground surface) throughout the catchment, causing a permanent



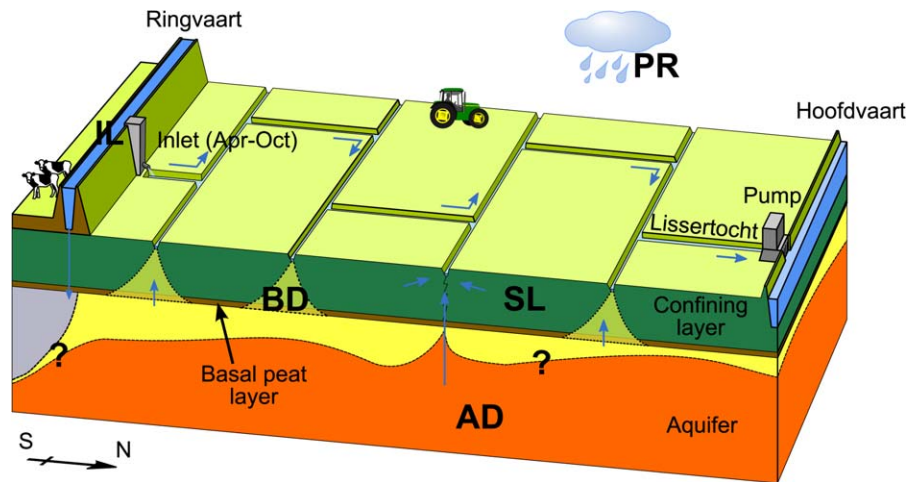
**Figure 1.** Topographic features, surface elevation and sample locations in the Lissertocht catchment. PR: precipitation, SL: shallow, phreatic groundwater, AD: deep aquifer groundwater, BD: groundwater below ditches, IL: inlet water, BSL: below mean sea level.

upward seepage flux [Oude Essink *et al.*, 2010]. Part of this seepage is concentrated in boils, which form preferential flow paths between aquifer and surface water [De Louw *et al.*, 2010, 2011].

[10] We hypothesize five end-members to represent flow path contributions to stream water at the catchment outlet, two of which are external inputs to the catchment: (1) precipitation, entering the stream with minimum interaction with the soil (denoted as PR) and (2) inlet water, extraneous water taken into the catchment through inlet culverts (IL). The other three end-members represent different local groundwater stores, each with a characteristic flow path contributing to stream water: (3) deep aquifer groundwater, discharged through boil seepage (AD), (4) groundwater below ditches, representing diffuse seepage (BD), and (5) shallow, phreatic groundwater discharged mostly through tile drains (SL) (Figure 2, chemistry in Table 2).

[11] Resulting from differences in geologic history, lithology, palaeohydrology, water management, and agricultural activities, the three local groundwater types show distinct chemical signatures. Groundwater type AD infiltrated the aquifer when a marine transgression approximately 8 – 3.8 kyr B.P. flooded the area [Post *et al.*, 2003]. This brackish groundwater type has a salinized, deeply

anoxic, calcite saturated facies, indicated by a negative base exchange index (BEX) [Stuyfzand, 1999], and significantly lower-than-expected  $\text{SO}_4$  concentrations, given the admixing of sea water (Table 2). AD exfiltrates directly into the surface water through boils [De Louw *et al.*, 2010], thus preventing any subsequent chemical interaction altering its signature. The brackish AD water type is overlain in the aquifer by a layer of fresh groundwater, infiltrating after coastal barriers started to form from 5.5 kyr B.P. onward and extensive marshlands developed behind them, covering the study area. This fresh groundwater has a different facies: freshened, deeply anoxic, and calcite saturated, demonstrated by a positive BEX and calcite saturation. This groundwater type seeps upward through a reactive layer of basal peat before exfiltrating into the stream. We therefore opted to sample this water type directly below ditches, just before exfiltration, as water type BD. BD shows the highest concentrations of  $\text{HCO}_3^-$ ,  $\text{SiO}_2$ , B, and Li, testifying of peat interaction, dissolution of diatom skeletons and desorption of marine components after fresh water intrusion [Stuyfzand, 1993] (Table 2). Shallow phreatic groundwater (SL) is even fresher (Table 2), but bears the chemical signature of agricultural activities including fertilizer application and drainage, leading to raised levels of  $\text{SO}_4$  by pyrite oxidation [Pons and Van der Molen, 1973].



**Figure 2.** Schematic representation of expected flow path contributions to Lissertocht stream water. IL: inlet water, PR: precipitation, BD: groundwater below ditches, SL: shallow, phreatic groundwater, AD: deep aquifer groundwater.

## 2.2. Sampling and Analytical Methods

[12] Stream water was sampled at the catchment's main pumping station, from 11 October 2011 until 4 October 2012. Samples were automatically obtained at the end of each pumping cycle (Teledyne ISCO automatic sampler) and collected within 3 weeks of sampling. Pumping cycles occurred approximately daily, resulting in a total of 362 samples. We investigated the response of solutes in representative stream water in a sample bottle for a worst-case collection scenario of 4 weeks waiting time. The sample collection test showed significant responses of EC, alkalinity, Ca, and  $\text{NO}_3$ , while the response of other possible tracers was in the order of analytical uncertainty. We therefore discarded EC, alkalinity, Ca, and  $\text{NO}_3$  in subsequent data interpretation. Catchment discharge was obtained by multiplying pumping times, logged at 10 min intervals by the automated water management system of the local water authority, with pumping capacity, measured in three repetitions using a boat-mounted acoustic Doppler current profiler (TeleDyne RD) [Mueller and Wagner, 2009]. Maximum discharge of the various intake culverts was determined in three repetitions by measuring the time necessary to fill a 100 L polyethylene bag.

[13] The five end-members were sampled with varying frequency either before or throughout the stream water

**Table 1.** Sampling Locations, Frequency, and Period of Stream Water and End-Members in Lissertocht Catchment<sup>a</sup>

	Number of Locations	Frequency	Period
Stream	1	End of pumping cycle	Oct 2011–Oct 2012
PR	1	Three-weekly bulk samples	Dec 2011–Oct 2012
IL	2; 1	Monthly; three-weekly	Mar 2011–Nov 2011; Mar 2012–Oct 2012
AD	1; 5	Twice; sporadic	Jun 2011, Nov 2012; 1993–Oct 2012
BD	6	Monthly	Jun 2011–Nov 2011
SL	6	Monthly	Jun 2011–Nov 2011

<sup>a</sup>PR: precipitation, IL: inlet water, AD: deep aquifer groundwater, BD: groundwater below ditches, SL: shallow, phreatic groundwater.

sampling period, depending on their observed temporal variance (Table 1). Shallow groundwater end-members (SL and BD) were sampled with a peristaltic pump, in piezometers screened approximately 1–2 m below the ground surface or ditch bottom respectively. AD was sampled with a peristaltic pump in one existing well, screened at 30, 40, and 60 m below surface level. Historic data from this and four additional nearby wells (< 9 km) was obtained from the Dutch database on subsurface data (DINO, available at <http://www.dinoloket.nl>). IL was sampled by grab sampling, while PR was sampled using a bulk collector connected to a rain gauge, constructed to minimize evaporation [Gröning *et al.*, 2012].

[14] All samples were filtered through a 0.45  $\mu\text{m}$  membrane filter and stored in the dark at 4°C on the day of collection. Alkalinity was determined by end-point titration (Titralab) on the day of sample collection. Anions were analyzed using a DIONEX DX-120 ion chromatograph within 2 days after sample collection. A vial for cations was acidified with 65%  $\text{HNO}_3$  suprapure (0.7 mL/100 mL) on the day of sampling, for preservation until analysis by a VARIAN 730-ES ICP-OES. Analytical uncertainty was determined by analysis of internal calibration standards and set to at least 3% (relative standard deviation) to account for dilution errors.

## 2.3. Generalized Likelihood Uncertainty Estimation

[15] The GLUE methodology was developed by Beven and Binley [1992] as an extension of the regionalized sensitivity analysis (RSA) of Spear and Hornberger [1980]. Given uncertainties and errors in model structure, model parameterization, and observational data, GLUE recognizes that multiple models or model parameterization will be equally good descriptors of the modeled system and thus exhibit equifinality [Beven, 2006]. GLUE therefore, rather than trying to optimize a single parameter set for a given model structure, retains multiple model structures or model parameterizations that adequately fit the observational data and are consequently deemed behavioral. Instead of just accepting or rejecting a parameter set (or more precisely a model structure—parameter set combination) as in the

**Table 2.** Mean and Range of the Chemical Composition of Stream Water and End-Members<sup>a</sup>

	Unit		Stream	PR	IL	AD	BD	SL
Cl <sup>b</sup>	mg/L	mean	1440	6.1	136	5453	336	75
		range	290–3956	1.1–22.1	113–167	4534–6590	134–840	34–169
SO <sub>4</sub> <sup>b</sup>	mg/L	mean	359	4.1	90	297	299	394
		range	140–578	2.1–9.1	64–112	106–665	0–872	138–837
HCO <sub>3</sub>	mg/L	mean	454	10.1	263	707	1008	522
		range	165–691	0.1–35.1	209–317	383–1122	658–1374	212–830
NO <sub>3</sub>	mg/L	mean	13.1	3.1	8.1	1.1	2.1	2.1
		range	0.1–69.1	0.1–7.1	0.1–32.1	0.1–4.1	0.1–18.1	0.1–7.1
PO <sub>4</sub>	mg/L	mean	0.02	0.12	0.12	2.52	13.62	0.32
		range	0.02–2.02	0.02–0.52	0.02–1.72	1.22–3.62	0.02–49.52	0.02–5.42
Na <sup>b</sup>	mg/L	mean	762	4.1	89	2688	278	59
		range	171–1950	1.1–12.1	72–102	2300–3289	164–490	19–142
K	mg/L	mean	31	2.1	13	71	49	19
		range	11–65	0.1–10.1	10–15	20–180	39–57	4–45
Ca	mg/L	mean	262	2.1	92	530	179	258
		range	127–377	1.1–5.1	79–107	288–860	83–306	188–345
Mg <sup>b</sup>	mg/L	mean	91	1.1	18	266	108	34
		range	32–206	0.1–2.1	16–20	210–410	72–128	9–69
Fe	mg/L	mean	0.02	0.02	0.02	10.12	0.12	0.92
		range	0.02–0.22	0.02–0.02	0.02–0.12	0.22–40.02	0.02–0.92	0.02–14.32
Mn	mg/L	mean	0.32	0.02	0.12	0.62	1.72	1.72
		range	0.02–0.92	0.02–0.12	0.12–0.22	0.12–1.92	0.72–2.52	0.92–2.82
SiO <sub>2</sub>	mg/L	mean	17	1.1	8	17	77	52
		range	5–27	0.1–2.1	4–12	11–22	22–99	34–77
B <sup>b</sup>	μg/L	mean	345	18	128	511	1039	336
		range	148–587	9–31	104–152	470–551	642–1480	137–526
Ba	μg/L	mean	188.1	12.1	43.1	1103.1	15.1	27.1
		range	21.1–461.1	6.1–25.1	32.1–54.1	1060.1–1141.1	5.1–46.1	12.1–89.1
Br <sup>b</sup>	μg/L	mean	5082	22	543	16868	1615	640
		range	1186–13,449	0–85	268–1623	15,663–18,134	469–3207	260–1039
F	μg/L	mean	0.62	0.02	0.32	0.12	0.52	0.72
		range	0.02–3.22	0.02–0.12	0.22–0.52	0.02–0.22	0.22–0.72	0.32–1.22
Li <sup>b</sup>	μg/L	mean	30.1	1.1	13.1	28.1	68.1	50.1
		range	18.1–39.1	0.1–5.1	11.1–17.1	25.1–35.1	35.1–92.1	24.1–83.1
Mo	μg/L	mean	2.1	0.1	2.1	0.1	1.1	5.1
		range	1.1–4.1	0.1–0.1	1.1–2.1	0.1–0.1	0.1–3.1	1.1–15.1
Sr <sup>b</sup>	μg/L	mean	1329	8.1	465	2516	1210	736
		range	574–2270	3.1–16.1	403–524	2240–2940	675–1870	266–1060
EC <sub>20</sub>	μS/cm	mean	4924	65	963	15306	2675	1480
		range	1728–11,270	30–112	835–1022	12,900–17,113	2050–3670	1065–1860
BEX <sup>c</sup>	meq/L	mean	–2.1	0.1	2.1	–24.1	12.1	4.1
		range	–16.1–3.1	0.1–0.1	1.1–2.1	–36.1–(–6.1)	5.1–18.1	0.1–7.1
SI <sub>calcite</sub> <sup>d</sup>		mean	0.82	–5.12	0.42	0.62	0.72	0.32
		range	–0.62–1.62	–7.12–(–2.92)	–0.52–1.02	0.32–0.92	0.22–1.32	–0.82–1.02

<sup>a</sup>PR: precipitation, IL: inlet water, AD: deep aquifer groundwater, BD: groundwater below ditches, SL: shallow, phreatic groundwater.

<sup>b</sup>Solute used as tracer.

<sup>c</sup>Base Exchange index: Na + K + Mg – 1.0716 Cl (all in meq/L). Negative BEX indicates salinization, positive BEX freshening of facies [Stuyfzand, 1999].

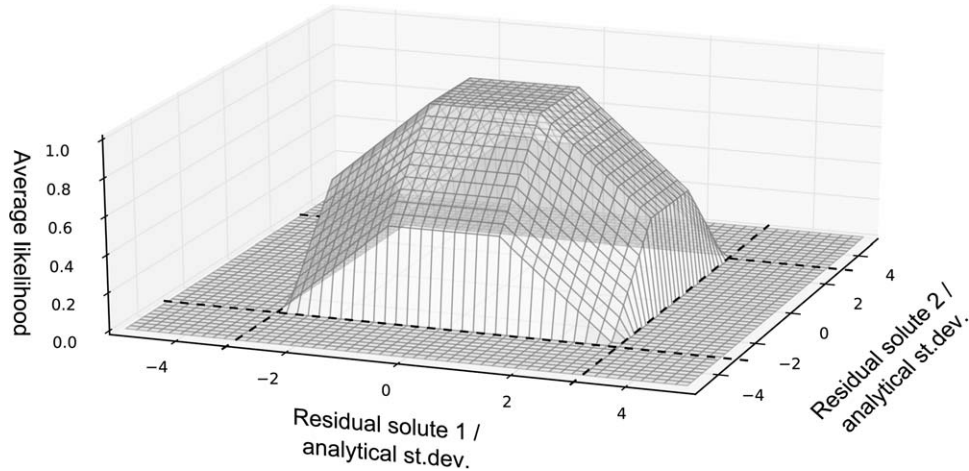
<sup>d</sup>Saturation Index of calcite, calculated following [Stuyfzand, 1989].

original RSA, a likelihood measure is used to express a degree of confidence in the parameter set. All behavioral parameter sets are used to predict a likelihood-weighted distribution of model response(s). Interaction between parameters is implicitly accounted for by GLUE focusing on parameter sets rather than individual parameters. The collection of behavioral parameter sets is obtained by Monte Carlo sampling of prior parameter ranges, running model simulations, and evaluating the simulated result against a likelihood measure to accept or reject the parameter set. A more complete description of GLUE is presented by *Beven and Binley* [1992], *Beven* [2009, 2006], and *Freer et al.* [1996].

#### 2.4. A GLUE Approach to End-Member Mixing Analysis (G-EMMA)

[16] An end-member mixing model, explaining stream water chemistry as a conservative mixture of end-

member concentrations, is a very simple conceptual description of the origin of stream water. Unsurprisingly, mixing models suffer from similar issues with model equifinality due to uncertainty and errors in model structure, parameters, and observations as the rainfall-runoff models GLUE was first applied to. GLUE minimizes the need for prior assumptions about model structure and structure of errors, and is therefore especially suited to quantify the uncertainty in mixing models pertaining to end-member characterization, that is, the variability in end-member concentrations. Additionally, as GLUE permits different model structures to be simultaneously evaluated as adequate system descriptors, uncertainty in end-member identification can be quantified by testing different sets of end-members against the available stream chemistry. Note that what we term identification uncertainty in this paper is paralleled by “structural uncertainty” in GLUE



**Figure 3.** Calculation of the fuzzy likelihood measure for two fictitious solutes. Dashed lines denote limits of acceptability ( $-3$  to  $+3$  analytical std. dev.) for the individual solutes.

terminology, and characterization uncertainty by “parameter uncertainty.”

[17] Our GLUE approach to end-member mixing (G-EMMA) starts with a definition of possible end-members. In EMMA, the Euclidean distance between end-members and their projection in the mixing space is used as a measure of the ability of the end-member to explain stream water concentrations [Barthold *et al.*, 2011; Christophersen and Hooper, 1992; James and Roulet, 2006]. This procedure might, however, obscure end-members that are not characterized properly by their median observed tracer concentrations. Instead, our approach minimizes the necessary prior assumptions by allowing for different end-member combinations during different periods, while relying on the time-variant data to reject invalid end-members.

[18] Subsequently, appropriate tracers must be identified, subject to two of the usual conditions prescribed by mixing model theory: (1) tracers must mix conservatively and (2) tracers must differ in concentration between end-members [Hooper *et al.*, 1990; Sklash and Farvolden, 1979]. A usual third condition: end-member concentrations must be invariant in time and space, does, however, not apply to the G-EMMA approach, which explicitly accounts for end-member variation. The diagnostic tools of Hooper [2003] can aid in defining appropriate tracers. All identified end-members are characterized by a prior concentration distribution for each tracer. Although we were able to characterize concentration distributions of end-members, either stored in the system or as direct inputs, we decided against using these distributions as priors in the procedure. Instead, as we lack information on how these concentrations are convoluted to observed concentration distributions, conditional on the sampling time at the catchment outlet [Rinaldo *et al.*, 2011], we adopted a minimal-assumption approach and used a uniform distribution over the full range of observed concentrations in samples belonging to an end-member. The G-EMMA methodology then allows the posterior effective concentrations for different end-members to be identified, conditional on this minimal prior assumption for effective end-member concentrations.

[19] A G-EMMA mixing model consists of: (1) a combination of end-members as a subset of all possible end-

members, (2) end-member fractions, and (3) end-member tracer concentrations, and is, following the notation of Christophersen and Hooper [1992], represented in matrix notation by (1):

$$\mathbf{I}_i \mathbf{B} = \mathbf{x}_i, \quad (1)$$

where  $\mathbf{I}_i$  represents the  $k$  sized row vector of end-member fractions,  $\mathbf{B}$  the  $k \times p$  sized matrix of end-member concentrations, and  $\mathbf{x}_i$  the  $p$  sized vector of tracer concentrations in the stream water sample, with  $k$  and  $p$  as the number of end-members and tracers, respectively. End-member fractions are sampled from a uniform Dirichlet distribution, yielding a uniform distribution of mixtures while ensuring mass balance closure (end-members always sum to one). Note that we opted to sample end-member fractions, rather than infer them from a least-squares regression technique, so as to retain a direct dependence of the results on the chosen likelihood measure (see below).

[20] For each separate stream water sample, a large number of mixing models is generated by uniform Monte Carlo sampling and evaluated against the observed stream water concentrations in terms of a fuzzy likelihood measure. A fuzzy measure, after Zadeh [1965], can be used to express a “degree of belief” in the model as a valid simulator of the system [Beven and Binley, 1992] and has been used in various previous GLUE applications [Blazkova and Beven, 2002, 2009; Freer *et al.*, 2004; Liu *et al.*, 2009; Page *et al.*, 2003, 2007; Pappenberger *et al.*, 2007]. It can be a useful approach to model evaluation when there is an expectation of epistemic (nonrandom), rather than aleatory (random) errors in the modeling process and observational data [Beven, 2006, 2012]. We define our fuzzy likelihood measure as the average over all tracers of individual trapezoids around the analytical values for each tracer, with a relative likelihood of one for calculated values within one standard deviation of the analytical value, decreasing linearly to zero at three standard deviations. Simulations are considered behavioral only if calculated values for all tracers fall within their respective trapezoids (Figure 3). The repetition of this procedure for each stream water sample allows for time-varying end-member fractions and end-

member concentrations, as a reflection of catchment processes. Likelihoods are rescaled to sum to unity over the ensemble of behavioral models identified for each time step independently. All software and source code written to facilitate the G-EMMA procedure are available for download at <http://g-emma.deltares.nl/>.

### 2.5. Application to the Lissertocht Data Set

[21] We compared applications of both the G-EMMA and original EMMA approaches to the Lissertocht data set to assess the significance of accounting for uncertainty in mixing models in a challenging catchment. We first used the diagnostic tools of Hooper [2003] to identify appropriate tracers, complemented by expert knowledge on the chemical stability of solutes in the catchment. Following the EMMA procedure outlined by Christophersen and Hooper [1992], we constructed a correlation matrix, by standardizing the stream water samples to zero mean and a standard deviation of one, before performing a principal components analysis (PCA) on the correlation matrix using all appropriate tracers. We investigated the dimensionality in the data set by analysis of both the eigenvalues (“the rule of one”) and the apparent structure in the residuals for increasing dimensionality, and calculated relative RMS errors (RRMSE) for all residuals [Hooper, 2003]. We subsequently used the methodology proposed by Barthold *et al.* [2011] to evaluate all possible combinations (minimum of three) of end-members for all possible combinations (minimum of four) of tracers on the three criteria: (1) the Euclidean distance between end-members in solute space and their projections in the mixing space is less than 15% [James and Roulet, 2006], (2) smallest deviations of the calculated end-member fractions from the plausible 0%–100% range and (3) the smallest Euclidean distance between end-members and the median of stream water in the mixing space. We calculated end-member fractions for the best performing end-member combination for comparison with G-EMMA results.

[22] In the G-EMMA procedure, we retained all possible end-members and tracers. We identified behavioral end-member fractions using the G-EMMA procedure outlined above, using the full range of observed concentrations for our five end-members (Table 2). The number of end-members was allowed to vary randomly between three and five, and we used  $1 \times 10^9$  Monte Carlo runs for each stream sample. We set the uncertainty of stream water samples to their respective analytical uncertainty and calculated the likelihood of each run following the procedure outlined above. G-EMMA results were evaluated by comparing modeled stream water chemistry with observed stream water chemistry (a valid test because the likelihood is averaged over all tracers), by determining the identification of the end-member fractions and by evaluating the calculated catchment response in terms of its physical plausibility. To explore the relative contribution of identification and characterization uncertainty, we investigated the variety of end-member combinations that yielded behavioral results. In addition, we compared the uncertainty calculated for all possible end-member combinations to that for the end-member combination most likely based on conventional EMMA criteria, and investigated the time-variant response of behavioral end-member concentrations.

## 3. Results

### 3.1. Catchment Hydrometry and Chemistry

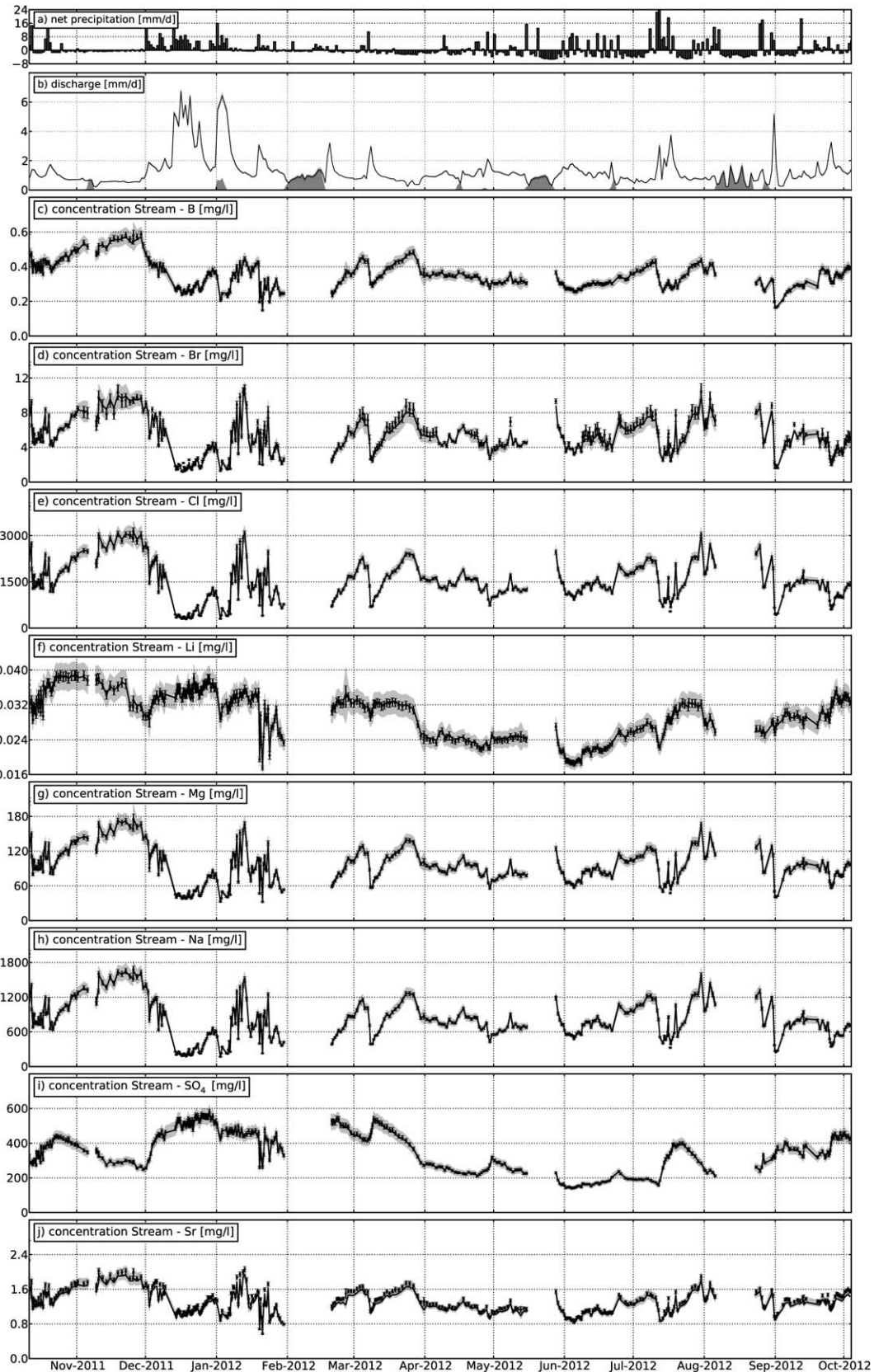
[23] Measured chemical composition of the catchment and end-members is summarized in Table 2. Concentration ranges for the end-members SL and BD were relatively wide, reflecting their high spatial variability. The chemical composition of the stream water was highly variable and showed a distinct response to precipitation events (Figure 4). Generally, solutes B, Br, Cl, Mg, Na, and Sr showed a decrease, whereas Li and  $\text{SO}_4$  concentrations rose with increasing discharge. April 2012 signified a marked drop in all solute concentrations, coinciding with the start of intake of inlet water into the catchment. Maximum capacity of the four intake culverts together was measured at  $95.7 \pm 4.3$  l/s, which equals  $0.83 \pm 0.04$  mm/d. Pumping capacity of the main pump was measured at  $1.01 \pm 0.02$  m<sup>3</sup>/s and  $1.35 \pm 0.05$  m<sup>3</sup>/s in normal and maximum operation, respectively.

### 3.2. EMMA, Hooper’s Diagnostic Tools, and Evaluation of Possible End-Members

[24] After investigation of bivariate solute-solute plots, we selected B, Br, Cl, Li, Mg, Na,  $\text{SO}_4$ , and Sr as suitable tracers. Other possible tracers showed no significant linear correlation with other solutes and were therefore discarded. After performing a PCA on the Lissertocht stream samples, the rank of the data set was analyzed by studying the structure in the residuals of the solute concentrations in the reduced model space. The “rule of one” suggested a two-dimensional model space explaining 96% of the variance in the stream concentrations, a result corroborated by visual inspection of the residuals and calculated RRMSEs (average 5.6%). Some structure was, however, still apparent for solute B, which disappeared in a three-dimensional model space (average RRMSE 3.5%). The evaluation of possible end-member combinations, following Barthold *et al.* [2011], resulted in end-members AD, SL, and IL (100%, 98%, and 95%, respectively) featuring in nearly all and BD (74%) in the majority of plausible combinations, while PR featured in markedly less (13%). Differences between tracers were small, all tracers were present in between 55% and 65% of plausible results. The combination of IL, SL, BD, and AD was by far the most prominent, making up 57% of plausible results. Calculated end-member fractions using this combination and all tracers are shown in Figure 5. The fractions of all end-members except AD often fall outside the plausible 0–1 range, most notably during the high discharge period of December 2011 to January 2012.

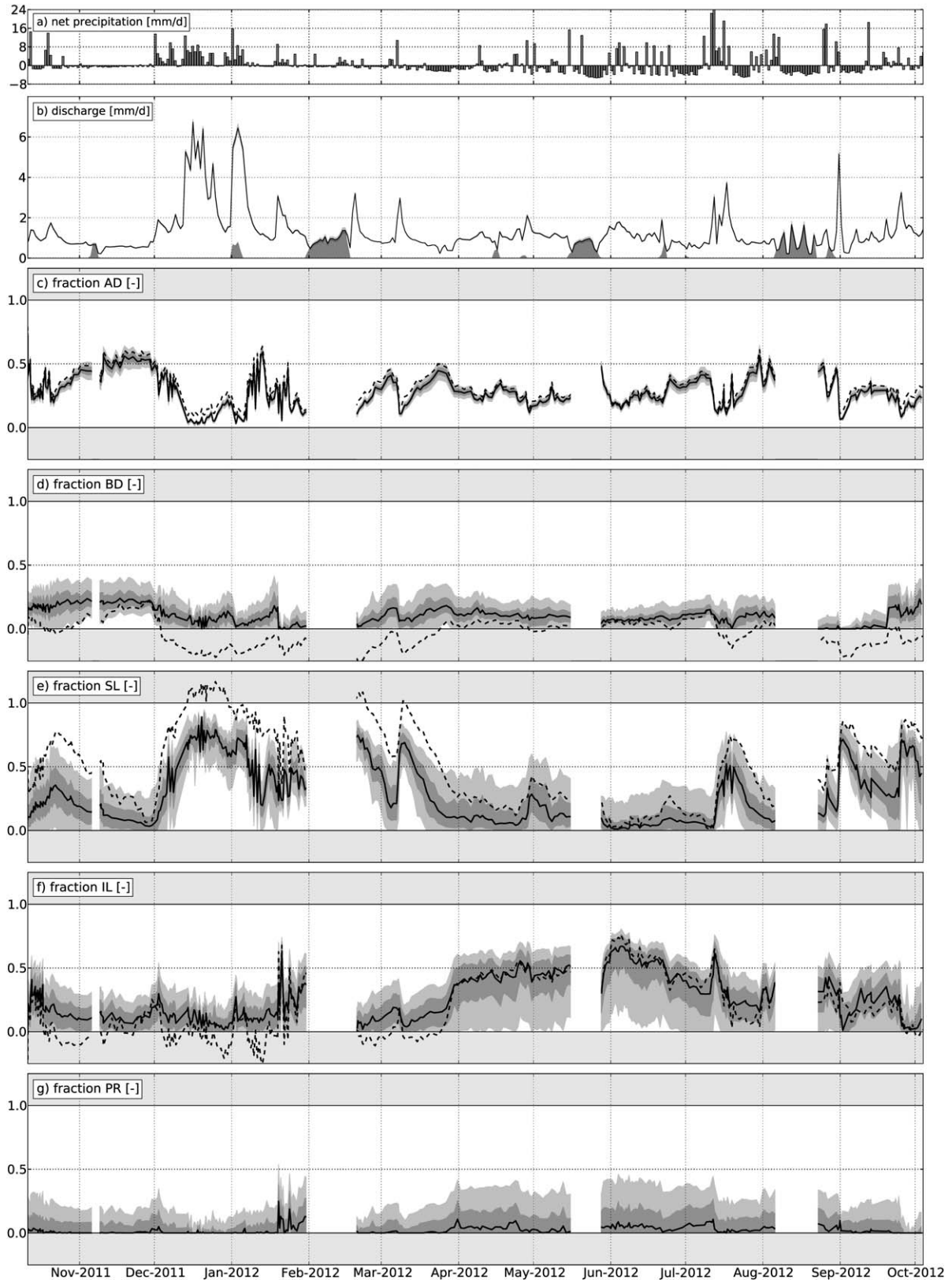
### 3.3. GLUE End-Member Mixing Analysis (G-EMMA)

[25] GLUE analysis of the 362 stream samples resulted in a median value and 25–75 percentile range of  $3.8 \times 10^3$  ( $3.1 \times 10^2$ – $1.2 \times 10^4$ ) behavioral runs (with positive fuzzy membership for all eight tracers) out of a possible  $1 \times 10^9$ . Two samples (on 17 December 2011 and 17 July 2012) yielded no behavioral runs (i.e., all of the tried combinations of fractions failed to match the defined fuzzy support for one or more tracers). Measured stream water concentrations could, with these two exceptions, consistently be explained by mixtures of our chosen end-members, as is reflected in the excellent agreement of modeled and



**Figure 4.** (a) Net precipitation, (b) discharge, and (c–j) measured and G-EMMA modeled stream water concentrations for tracers B, Br, Cl, Li, Mg, Na, SO<sub>4</sub>, and Sr. In Figure 4b, dark shaded area represents operation of auxiliary pump, light shaded area represents  $\pm 1$  standard deviation. In Figures 4c–4j, dots represent measured values, with error bars denoting  $\pm 1$  analytical standard deviation, and solid line and shaded area represent median and 5–95 percentile range of modeled values.





**Figure 5.** (a) Net precipitation, (b) discharge, and (c–g) EMMA result (dashed line) and median (solid line), 25–75 percentile range (dark-shaded band), and 5–95 percentile range (light-shaded band) of G-EMMA calculated fractions of AD, BD, SL, IL, and PR in stream water. The area in Figures 5c–5g outside the plausible 0–1 range of end-member fractions is indicated in light gray.

**Table 3.** Frequencies of End-Member Combinations Producing Behavioral Results

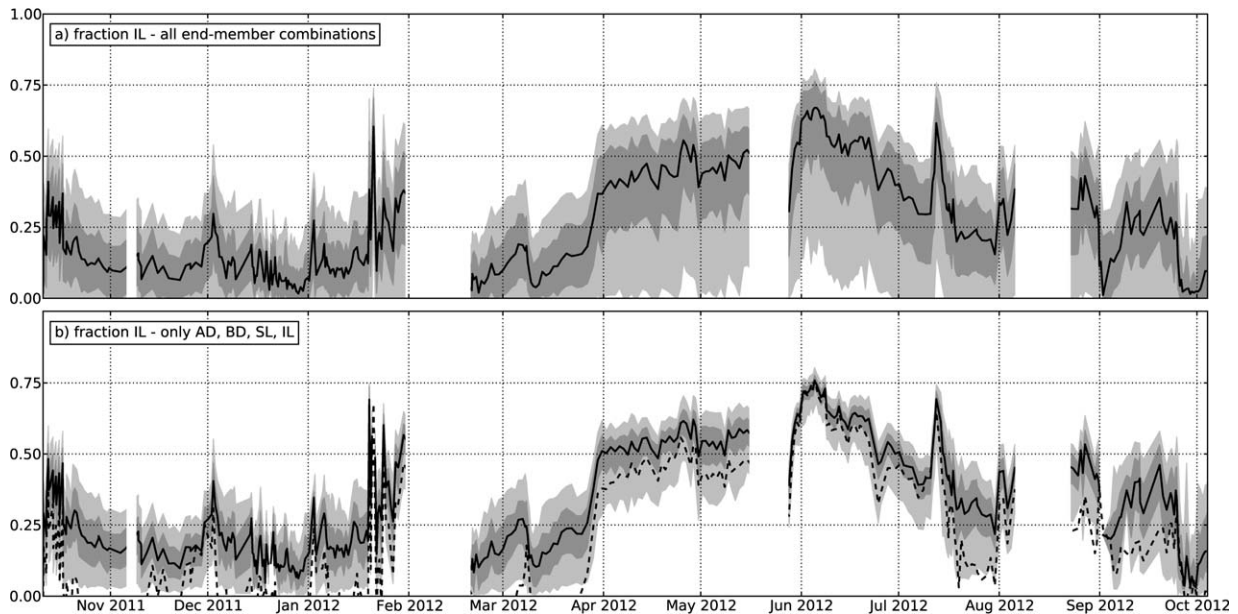
Combination of End-Members <sup>a</sup>	IL, SL	IL, SL, BD	IL, SL, BD, AD	IL, SL, BD, AD, PR	IL, SL, BD, AD, PR	IL, SL, BD, AD, PR	IL, SL, BD, AD, PR	IL, SL, BD, AD, PR	IL, SL, BD, AD, PR	IL, SL, BD, AD, PR	IL, SL, BD, AD, PR	IL, SL, BD, AD, PR	IL, SL, BD, AD, PR	IL, SL, BD, AD, PR	IL, SL, BD, AD, PR	IL, SL, BD, AD, PR	IL, SL, BD, AD, PR
Frequency (%)	30	29	9	8	7	5	5	4	2	0	0	0	0	0	0	0	0

<sup>a</sup>IL: inlet water, SL: shallow, phreatic groundwater, BD: groundwater below ditches, AD: deep aquifer groundwater, PR: precipitation.

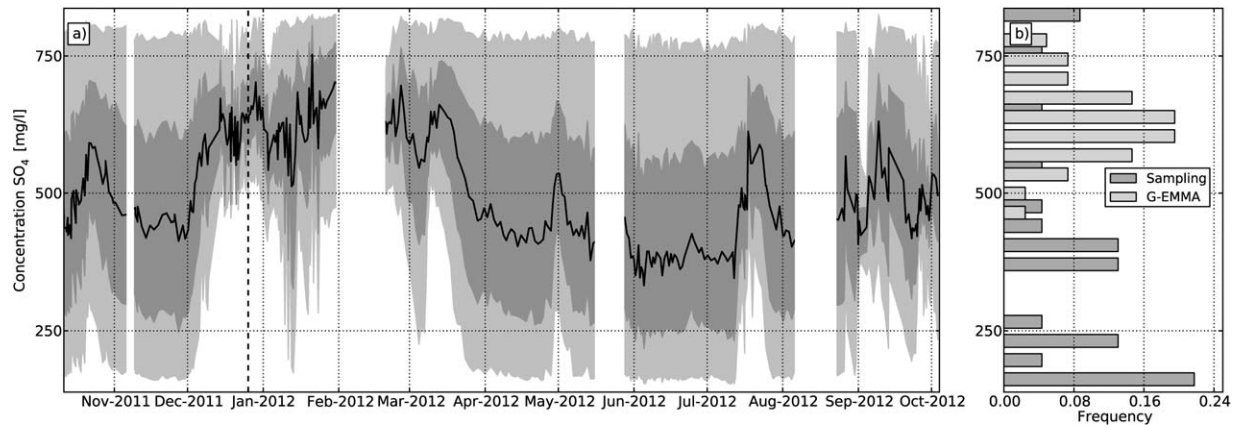
measured stream water concentrations of tracers B, Br, Cl, Li, Mg, Na, and SO<sub>4</sub>. Only Sr is consistently under predicted, albeit slightly (Figure 4). The possibilistic distributions of end-member fractions that yielded behavioral results for the different samples are plotted in Figure 5. This plot can be regarded as a time-variant version of the well-known “dotty-plots” of GLUE applications [e.g., Beven, 2006], showing the likelihood-weighted marginal distributions of behavioral model parameters changing over time, as each sample is represented by a separate Monte Carlo calculation. The calculated uncertainty in end-member fractions, indicated by the 5–95 and 25–75 percentile ranges (shaded bands) in Figure 5, varied over time and between end-members. The complete marginal distributions of all end-member fractions lay (necessarily) within the 0–1 range, and are asymmetrical. While there was considerable uncertainty in the fractions of all end-members except AD, all end-member contributions were sensitive parameters in the GLUE sense and could therefore be adequately identified throughout the time series. Except for AD, behavioral end-member fractions differed markedly from fractions calculated with conventional EMMA. G-EMMA calculated SL fractions were lower than those calculated with EMMA, which at times exceeded a fraction of 1. Contrastingly, G-EMMA calculated BD and IL fractions were higher than the equivalent EMMA fractions, which at times fell below 0.

[26] We took a closer look at the distribution of end-members, end-member combinations, and end-member concentrations in the posterior parameter set, that is, the models and parameters that make up the behavioral runs. Averaged over the entire time series, frequencies of end-members occurring in behavioral end-member combinations were: AD: 100%, SL: 90%, IL: 86%, BD: 82%, and PR: 52%. Results resemble those obtained through the criteria of Barthold *et al.* [2011], although the contribution of PR is much more prominent in the G-EMMA analysis. The end-member combination of IL, SL, BD, and AD yielded the most behavioral runs, closely followed by the combination of all five end-members (Table 3). Results from G-EMMA include more combinations, and frequencies are spread out more evenly over the different combinations.

[27] The effect of including the identification uncertainty in G-EMMA was investigated by comparing the behavioral end-member fractions for all possible combinations, to the subset of behavioral fractions for the combination IL, SL, BD, AD, the most dominant combination from the criteria of Barthold *et al.* [2011]. Maximum effects were seen in IL, the median fraction of IL resulting from all possible end-member combinations is consistently lower (average  $-30 \pm 8\%$ ) than from the subset of one possible combination, and its uncertainty (5–95 percentile range) is consistently larger (average  $64 \pm 90\%$ ) than from the subset (Figure 6). Smallest effects were observed for AD, but



**Figure 6.** Difference of median (solid line), 25–75 percentile range (dark-shaded band), and 5–95 percentile range (light-shaded band) of IL between (a) all possible end-member combinations and (b) only combination AD, BD, SL, and IL. Dashed line in Figure 6b denotes calculated EMMA result for IL.



**Figure 7.** (a) Posterior distribution of  $\text{SO}_4$  concentrations of end-member SL and (b) frequency distribution of sampled SL– $\text{SO}_4$  concentrations (dark) versus G-EMMA behavioral SL– $\text{SO}_4$  concentrations on 25 December 2011 (light). The day 25 December 2011 is indicated in Figure 7a by a vertical dashed line. Solid line in Figure 7a represents median values, 25–75 and 5–95 percentile ranges are represented by dark- and light-shaded bands, respectively.

effects were still on average  $-4 \pm 24\%$  on median fractions (uncertainty range  $5 \pm 7\%$  larger). For comparison, the EMMA result (which also pertained to this end-member combination) for IL is also shown in Figure 6b. Even with identical end-member combinations, results differ markedly between EMMA and G-EMMA.

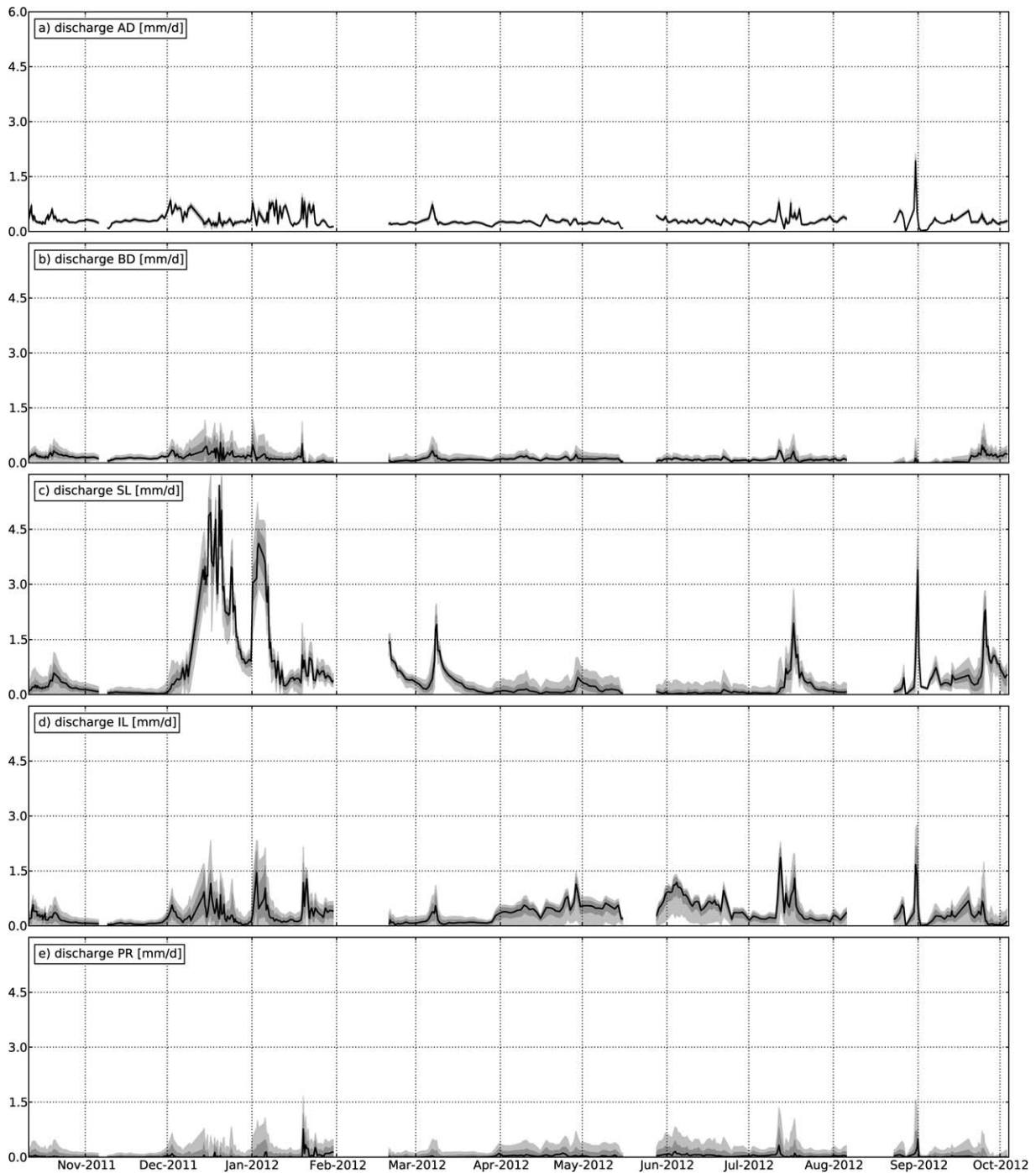
[28] Analysis of the likelihood weighted marginal distributions of end-member tracer concentrations revealed a general insensitivity of the likelihood of modeled stream water concentrations to IL and PR concentrations, and limited sensitivity to most AD and BD concentrations, as behavioral simulations were found throughout the respective parameter distributions (not shown). Note that model likelihood is associated with a combination of a model structure (end-member combination) and model parameters, rather than a single model parameter [Beven, 2006]. So while the model likelihood (i.e., the fit of stream water concentrations) may be insensitive to end-member concentrations, end-member fractions do not necessarily have to be. Most SL tracer concentrations were, however, constrained to part of their initial range during discharge events, when the fraction of SL in stream water is highest ( $\text{SO}_4$  shown in Figure 7a). Behavioral results were limited to lower concentrations of B, Li, and Mg, and to higher concentrations of  $\text{SO}_4$  and Sr during discharge events. Subsequent indicative calculations using these constrained concentrations of SL, instead of the full range, clearly lessened the sensitivity of SL concentrations, while hardly affecting modeled stream concentrations. The resulting median SL fraction was slightly lower ( $-8 \pm 14\%$ ) than using the full range, while its uncertainty decreased ( $-20 \pm 15\%$ ).

### 3.4. Catchment Response

[29] Calculating the discharge for each end-member by multiplying the fraction with the discharge provides a comprehensive view of the catchment's response to rainfall events and enables a, subjective, plausibility check of G-EMMA results (Figure 8). Generally, the observed patterns in catchment response are physically plausible and are consistent with our previously formed perceptual model of the

hydrologic functioning of the catchment. The catchment response showed a relatively constant flux of AD, consistent with the relatively constant head difference between the aquifer and the tightly managed surface water levels, and in agreement with results for a similar catchment [De Louw *et al.*, 2011]. Precipitation events resulted in a dominant contribution of SL to discharge, a behavior exhibited by numerous catchments over a range of different geographical settings (overview in Weiler *et al.* [2005]). A long dry period before the onset of precipitation delayed the response of SL and BD considerably, indicating a thorough depletion of shallow groundwater stores.

[30] Active water management in the catchment is evidenced in the hydrograph by a rising contribution of IL (and PR to a lesser extent) around 1 April 2012, coinciding with the start of intake of fresh water into the catchment. The discharge of IL rose to a relatively constant value of about 0.5 mm/d, which was in the order of the measured maximum capacity of the intake culverts. Additional intake of water at the auxiliary pump had started on 29 May 2012, lasting approximately 1 week (M. Riethoff, Rijnland Water Authority, personal communication, 2012), coinciding well with the temporary rise in IL discharge in June 2012. The contribution of IL during the winter months and during summer precipitation events was unexpected however, as its input is controlled by actively managed hydraulic structures. This unexpected result may be caused by (1) the lack of separation between IL and PR, the unexpected IL contribution in fact being PR, (2) an unidentified source of water with similar chemical properties as IL, most likely subterranean inputs from the canal supplying IL water, or (3) storage in the extensive surface water system flushed out at discharge events. Due to the uncertainty associated with the proximate locations of IL and PR in mixing space, an additional tracer that better distinguishes between the two would be necessary to better separate the two end-members. Gadolinium has proved successful in a similar setting [Rozemeijer *et al.*, 2012] and also  $^{18}\text{O}$  could yield better contrasts [Stuyfzand, 1993]. The contribution of PR appears small even during the larger precipitation events,



**Figure 8.** Median (solid line), 25–5 percentile range (dark-shaded band) and 5–95 percentile range (light-shaded band) of discharge of (a) AD, (b) BD, (c) SL, (d) IL, and (e) PR.

indicating the absence of significant fast flow routes like overland flow, although the lack of separation between PR and IL necessitates caution when drawing this conclusion.

#### 4. Discussion

[31] Given the well-established problems in identifying and characterizing end-members, end-member mixing models are, at best, simple hypotheses about catchment functioning. Nevertheless, they can still offer valuable

insights, assuming uncertainty is adequately accounted for [Soulby *et al.*, 2003a; Uhlenbrook and Hoeg, 2003]. This paper presents G-EMMA, a novel method of quantifying uncertainty, both in identifying and characterizing end-members, in end-member mixing models, based on the GLUE methodology of Beven and Binley [1992]. An additional advantage is that our method allows for using more tracers than necessary, a central feature of EMMA, but lacking in existing quantitative uncertainty assessments. We showed that G-EMMA is able to adequately model

stream water concentrations and identify contributions of five different end-members, albeit with varying uncertainty. Therefore, as was also shown by *Soulsby et al.* [2003a], even in agricultural catchments, heavily impacted by agricultural activities and intricate water management, mixing models can help to better understand catchment functioning.

[32] Several existing approaches have quantified the uncertainty resulting from an inability to adequately characterize end-member concentrations [*Bazemore et al.*, 1994; *Genereux*, 1998; *Hooper et al.*, 1990; *Joerin et al.*, 2002; *Soulsby et al.*, 2003a]. Uncertainty in end-member concentrations was inferred from sampling of stored water, and in most approaches, approximated by a Gaussian distribution. However, as *Joerin et al.* [2002] recognize, and illustrated by Figure 7b, end-member concentrations do not always follow a Gaussian distribution, so that this approximation may lead to incorrect uncertainty estimations. Furthermore, an accurate characterization of stores of end-member water in a catchment does not necessarily equate to a proper characterization of the flux-weighted input to the stream [*Rinaldo et al.*, 2011], although implicitly assumed by these approaches. Instead, recognizing the impossibility of adequate characterization of the flux-weighted input to the stream, we adopted a minimal assumption approach in G-EMMA and assumed a uniform prior distribution over the complete range of sampled end-member concentrations. As evidenced from Figure 7b, the posterior distribution of behavioral end-member concentrations can indeed differ markedly from the distribution obtained through sampling, signifying our inability to adequately a priori characterize end-member concentrations.

[33] We did not explicitly include the temporal variation of end-member tracer concentrations in our mixing models, as temporal variance was relatively low in measured end-member concentrations. Furthermore, adequate quantification of the effect of temporal variance in catchment inputs on stream concentrations would require a theoretical framework that accounts for both nonlinearity and nonstationarity in travel times [*Jorgulescu et al.*, 2005, 2007; *Rinaldo et al.*, 2011], which is outside the scope of this research. Temporal variation is, however, implicitly accounted for in G-EMMA, as every stream water sample is independently modeled using the full range of observed end-member tracer concentrations. If a temporal signal is significant enough to be expressed in stream water concentrations despite all uncertainty, end-member concentrations should be sensitive parameters in GLUE. Our model results were, however, generally insensitive to end-member concentrations, with the exception of SL. While we cannot exclude temporal variation in SL concentrations, the constraining of SL concentrations during discharge peaks is more likely a result of the high proportion of SL in stream water, increasing the sensitivity to SL concentrations. The constrained concentrations of SL during discharge events may therefore be a closer representation of “real” SL water (i.e., the flux-weighted input to the stream) than the range obtained from sampling.

[34] Explicitly including time-variant patterns in end-member fractions and concentrations in G-EMMA potentially offers several advantages and is an important direction for future research. First, extending the work of

*Jorgulescu et al.* [2005, 2007], combining G-EMMA with the recent progress made in research on transit time distributions [*Heidbüchel et al.*, 2012; *Rinaldo et al.*, 2011; *van der Velde et al.*, 2012] may be a way to shed more light on the time-variant behavior of end-member concentrations or their convolution to stream chemistry through instationary transit times. Second, combining results for successive samples in a time-filtered way may reduce the uncertainty of end-member fractions as opposed to the current independent simulation of successive samples.

[35] In mixing model analyses, the choice of end-members is often a translation of the researcher’s hypothesis of catchment functioning and therefore, by definition, also an uncertain one. The GLUE approach of G-EMMA quantifies this identification uncertainty by simultaneously evaluating different possible end-member combinations. A comparison between results for a selected end-member combination and the complete result set (Figure 6) illustrated the possible significance of identification uncertainty, in this particular case amounting to a maximum 30% difference in median calculated end-member fractions. End-member IL occupies a proximate location to PR in the mixing space of the Lissertocht catchment, resulting in interference and hence a relatively high uncertainty of both end-members. Similar uncertainty due to interference has, to our knowledge, not been reported, as conventional EMMA guidelines [e.g. *Christophersen and Hooper*, 1992; *Christophersen et al.*, 1990] recommend the use of end-members that are sufficiently different to each other. We would argue, however, that even if adequate separation is simply impossible based on the available measurement data, retaining proximal end-members presents a more realistic notion of the uncertainty in our understanding of catchment functioning.

[36] The heavily impacted Lissertocht catchment is, due to the significant spatial variation in end-member concentrations and extraneous inputs of regional groundwater and fresh water intake, a difficult test case for applying end-member mixing models. Indeed, conventional EMMA suffered from repeated excursions of end-member fractions outside the plausible 0–1 range (Figure 5). As these excursions predominantly occurred during discharge events with a large fraction of SL water, this end-member is probably not well represented by its sampled concentration median. The skewed constraining of behavioral SL concentrations in the G-EMMA analysis also points in this direction. Contrastingly, G-EMMA application was not significantly affected by the uncertainty in end-member concentrations, and was still able to identify the (uncertain) contributions of five different end-members to the Lissertocht. Therefore, in addition to quantifying uncertainty in end-member mixing models, G-EMMA can potentially be applied over a wider range of catchments than conventional EMMA, while still yielding meaningful results. Moreover, application of G-EMMA is not limited to hydrology, but may be successfully applied to end-member mixing problems in other (earth) sciences.

## 5. Conclusion

[37] Using a GLUE-based approach to end-member mixing models allowed a more complete investigation of end-

member mixing uncertainty than existing methods, as the approach includes both characterization and identification uncertainty. Despite this uncertainty, G-EMMA was able to characterize end-member contributions to the Lissertocht, where conventional EMMA results suffered from repeated excursions outside the plausible 0–1 range. We therefore recommend using G-EMMA to more robustly test hypotheses about catchment functioning, especially in complex catchments with considerable concentration ranges. In spite of the well-rehearsed difficulties in applying end-member mixing models to agricultural catchments, our approach enabled us to improve our understanding of the functioning of an actively managed Dutch polder catchment throughout the course of a year.

[38] **Acknowledgments.** We thank J. Visser for laboratory assistance, Rijnland Water Authority for helping with the measurement setup and supplying additional data, and two anonymous reviewers for providing valuable suggestions to this paper. This work was carried out within the Dutch “Knowledge for Climate” program.

## References

- Barthold, F. K., J. Wu, K. B. Vaché, K. Schneider, H.-G. Frede, and L. Breuer (2010), Identification of geographic runoff sources in a data sparse region: Hydrological processes and the limitations of tracer-based approaches, *Hydrol. Processes*, *24*(16), 2313–2327, doi:10.1002/hyp.7678.
- Barthold, F. K., C. Tyralla, K. Schneider, K. B. Vaché, H.-G. Frede, and L. Breuer (2011), How many tracers do we need for end member mixing analysis (EMMA)? A sensitivity analysis, *Water Resour. Res.*, *47*, W08519, doi:10.1029/2011WR010604.
- Bazemore, D., K. Eshleman, and K. Hollenbeck (1994), The role of soil water in stormflow generation in a forested headwater catchment: Synthesis of natural tracer and hydrometric evidence, *J. Hydrol.*, *162*(1–2), 47–75, doi:10.1016/0022-1694(94)90004-3.
- Beven, K. J. (1989), Changing ideas in hydrology—The case of physically-based models, *J. Hydrol.*, *105*(1–2), 157–172, doi:10.1016/0022-1694(89)90101-7.
- Beven, K. J. (2006), A manifesto for the equifinality thesis, *J. Hydrol.*, *320*(1–2), 18–36, doi:10.1016/j.jhydrol.2005.07.007.
- Beven, K. J. (2009), *Environmental Modelling: An Uncertain Future?*, Routledge, London.
- Beven, K. J. (2012), Causal models as multiple working hypotheses about environmental processes, *C. R. Geosci.*, *344*(2), 77–88, doi:10.1016/j.crte.2012.01.005.
- Beven, K. J., and A. Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Processes*, *6*(3), 279–298, doi:10.1002/hyp.3360060305.
- Blazkova, S., and K. J. Beven (2002), Flood frequency estimation by continuous simulation for a catchment treated as ungauged (with uncertainty), *Water Resour. Res.*, *38*(8), 1139, doi:10.1029/2001WR000500.
- Blazkova, S., and K. J. Beven (2009), A limits of acceptability approach to model evaluation and uncertainty estimation in flood frequency estimation by continuous simulation: Skalka catchment, Czech Republic, *Water Resour. Res.*, *45*, W00B16, doi:10.1029/2007WR006726.
- Burns, D. A., J. J. McDonnell, R. P. Hooper, N. E. Peters, J. E. Freer, C. Kendall, and K. J. Beven (2001), Quantifying contributions to storm runoff through end-member mixing analysis and hydrologic measurements at the Panola Mountain Research Watershed (Georgia, USA), *Hydrol. Processes*, *15*(10), 1903–1924, doi:10.1002/hyp.246.
- Christophersen, N., and R. P. Hooper (1992), Multivariate analysis of stream water chemical data: The use of principal components analysis for the end-member mixing problem, *Water Resour. Res.*, *28*(1), 99–107, doi:10.1029/91WR02518.
- Christophersen, N., C. Neal, R. P. Hooper, R. Vogt, and S. Andersen (1990), Modelling streamwater chemistry as a mixture of soilwater end-members—A step towards second-generation acidification models, *J. Hydrol.*, *116*(1–4), 307–320, doi:10.1016/0022-1694(90)90130-P.
- De Louw, P. G. B., G. H. P. Oude Essink, P. J. Stuyfzand, and S. E. A. T. M. van der Zee (2010), Upward groundwater flow in boils as the dominant mechanism of salinization in deep polders, The Netherlands, *J. Hydrol.*, *394*(3–4), 494–506, doi:10.1016/j.jhydrol.2010.10.009.
- De Louw, P. G. B., Y. van der Velde, and S. E. A. T. M. van der Zee (2011), Quantifying water and salt fluxes in a lowland polder catchment dominated by boil seepage: A probabilistic end-member mixing approach, *Hydrol. Earth Syst. Sci.*, *15*(7), 2101–2117, doi:10.5194/hess-15-2101-2011.
- Durand, P., and J. L. J. Torres (1996), Solute transfer in agricultural catchments: The interest and limits of mixing models, *J. Hydrol.*, *181*, 1–22, doi:10.1016/0022-1694(95)02922-2.
- Freer, J., K. J. Beven, and B. Ambrose (1996), Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach, *Water Resour. Res.*, *32*(7), 2161–2173, doi:10.1029/95WR03723.
- Freer, J. E., H. McMillan, J. J. McDonnell, and K. J. Beven (2004), Constraining dynamic TOPMODEL responses for imprecise water table information using fuzzy rule based performance measures, *J. Hydrol.*, *291*(3–4), 254–277, doi:10.1016/j.jhydrol.2003.12.037.
- Genereux, D. (1998), Quantifying uncertainty in tracer-based hydrograph separations, *Water Resour. Res.*, *34*(4), 915–919, doi:10.1029/98WR00010.
- Gröning, M., H. O. Lutz, Z. Roller-Lutz, M. Kralik, L. Gourcy, and L. Pölsenstein (2012), A simple rain collector preventing water re-evaporation dedicated for  $\delta^{18}\text{O}$  and  $\delta^2\text{H}$  analysis of cumulative precipitation samples, *J. Hydrol.*, *448–449*, 195–200, doi:10.1016/j.jhydrol.2012.04.041.
- Guinn Garrett, C., V. M. Vulava, T. J. Callahan, and M. L. Jones (2012), Groundwater-surface water interactions in a lowland watershed: Source contribution to stream flow, *Hydrol. Processes*, *26*(21), 3195–3206, doi:10.1002/hyp.8257.
- Heidbüchel, I., P. Troch, and S. Lyon (2012), The master transit time distribution of variable flow systems, *Water Resour. Res.*, *48*, W06520, doi:10.1029/2011WR011293.
- Hoeg, S., S. Uhlenbrook, and C. Leibundgut (2000), Hydrograph separation in a mountainous catchment—Combining hydrochemical and isotopic tracers, *Hydrol. Processes*, *14*(7), 1199–1216, doi:10.1002/(SICI)1099-1085(200005)14:7<1199::AID-HYP35>3.0.CO;2-K.
- Hooper, R. P. (2003), Diagnostic tools for mixing models of stream water chemistry, *Water Resour. Res.*, *39*(3), 1055, doi:10.1029/2002WR001528.
- Hooper, R. P., N. Christophersen, and N. E. Peters (1990), Modelling streamwater chemistry as a mixture of soilwater end-members—An application to the Panola Mountain catchment, Georgia, USA, *J. Hydrol.*, *116*(1–4), 321–343, doi:10.1016/0022-1694(90)90131-G.
- Ijmer, J., G. Stauch, E. Dietze, K. Hartmann, B. Diekmann, G. Lockot, S. Opitz, B. Wünnemann, and F. Lehmkuhl (2012), Characterisation of transport processes and sedimentary deposits by statistical end-member mixing analysis of terrestrial sediments in the Donggi Cona lake catchment, NE Tibetan Plateau, *Sediment. Geol.*, *281*, 166–179, doi:10.1016/j.sedgeo.2012.09.006.
- Iorgulescu, I., K. J. Beven, and A. Musy (2005), Data-based modelling of runoff and chemical tracer concentrations in the Haute-Mentue research catchment (Switzerland), *Hydrol. Processes*, *19*(13), 2557–2573, doi:10.1002/hyp.5731.
- Iorgulescu, I., K. J. Beven, and A. Musy (2007), Flow, mixing, and displacement in using a data-based hydrochemical model to predict conservative tracer data, *Water Resour. Res.*, *43*, W03401, doi:10.1029/2005WR004019.
- James, A. L., and N. T. Roulet (2006), Investigating the applicability of end-member mixing analysis (EMMA) across scale: A study of eight small, nested catchments in a temperate forested watershed, *Water Resour. Res.*, *42*, W08434, doi:10.1029/2005WR004419.
- Joerin, C., K. J. Beven, I. Iorgulescu, and A. Musy (2002), Uncertainty in hydrograph separations based on geochemical mixing models, *J. Hydrol.*, *255*(1–4), 90–106, doi:10.1016/S0022-1694(01)00509-1.
- Johnson, N., G. Likens, F. H. Bormann, D. W. Fisher, and R. S. Pierce (1969), A working model for the variation in stream water chemistry at the Hubbard Brook Experimental Forest, New Hampshire, *Water Resour. Res.*, *5*(6), 1353–1363.
- Keay, S., W. Collins, and M. McCulloch (1997), A three-component Sr-Nd isotopic mixing model for granitoid genesis, Lachlan fold belt, eastern Australia, *Geology*, *25*(4), 307–310, doi:10.1130/0091-7613(1997)025<0307.

- Kendall, C., J. J. McDonnell, and W. Gu (2001), A look inside “black box” hydrograph separation models: A study at the Hydrohill catchment, *Hydrol. Processes*, 15(10), 1877–1902, doi:10.1002/hyp.245.
- Liu, Y., J. Freer, K. J. Beven, and P. Matgen (2009), Towards a limits of acceptability approach to the calibration of hydrological models: Extending observation error, *J. Hydrol.*, 367(1-2), 93–103, doi:10.1016/j.jhydrol.2009.01.016.
- Long, A. J., and J. F. Valder (2011), Multivariate analyses with end-member mixing to characterize groundwater flow: Wind cave and associated aquifers, *J. Hydrol.*, 409(1-2), 315–327, doi:10.1016/j.jhydrol.2011.08.028.
- Mueller, D., and C. Wagner (2009), *Measuring Discharge With Acoustic Doppler Current Profilers From a Moving Boat*, U.S. Geol. Surv., Reston, Va.
- Oude Essink, G. H. P., E. S. van Baaren, and P. G. B. De Louw (2010), Effects of climate change on coastal groundwater systems: A modeling study in the Netherlands, *Water Resour. Res.*, 46, W00F04, doi:10.1029/2009WR008719.
- Page, T., K. J. Beven, J. Freer, and A. Jenkins (2003), Investigating the uncertainty in predicting responses to atmospheric deposition using the model of acidification of groundwater in catchments (MAGIC) within a generalised likelihood uncertainty estimation (GLUE) framework, *Water Air Soil Pollut.*, 142(1-4), 71–94.
- Page, T., K. J. Beven, J. Freer, and C. Neal (2007), Modelling the chloride signal at Plynlimon, Wales, using a modified dynamic TOPMODEL incorporating conservative chemical mixing (with uncertainty), *Hydrol. Processes*, 307(3), 292–307, doi:10.1002/hyp.
- Pappenberger, F., K. Frodsham, K. J. Beven, R. Romanowicz, and P. Matgen (2007), Fuzzy set approach to calibrating distributed flood inundation models using remote sensing observations, *Hydrol. Earth Syst. Sci.*, 11(2), 739–752, doi:10.5194/hess-11-739-2007.
- Pinder, G. F., and J. F. Jones (1969), Determination of the ground-water component of peak discharge from the chemistry of total runoff, *Water Resour. Res.*, 5(2), 438–445, doi:10.1029/WR005i002p00438.
- Pons, L., and W. Van der Molen (1973), Soil genesis under dewatering regimes during 1000 years of polder development, *Soil Sci.*, 116(3), 228–235.
- Post, V., H. Plicht, and H. Meijer (2003), The origin of brackish and saline groundwater in the coastal area of the Netherlands, *Geol. Mijnbouw*, 82(2), 133–147.
- Rasmussen, J. B. (2010), Estimating terrestrial contribution to stream invertebrates and periphyton using a gradient-based mixing model for delta13C., *J. Anim. Ecol.*, 79(2), 393–402, doi:10.1111/j.1365-2656.2009.01648.x.
- Rice, K. C., and G. M. Hornberger (1998), Comparison of hydrochemical tracers to estimate source contributions to peak flow in a small, forested, headwater catchment, *Water Resour. Res.*, 34(7), 1755–1766, doi:10.1029/98WR00917.
- Rinaldo, A., K. J. Beven, E. Bertuzzo, L. Nicotina, J. Davies, A. Fiori, D. Russo, and G. Botter (2011), Catchment travel time distributions and water flow in soils, *Water Resour. Res.*, 47, W07537, doi:10.1029/2011WR010478.
- Royal Netherlands Meteorological Institute (2010), *Schiphol, langjarige gemiddelden, tijdvak 1981–2010*, De Bilt, Netherlands.
- Rozemeijer, J., C. Siderius, M. Verheul, and H. Pomarius (2012), Tracing the spatial propagation of river inlet water into an agricultural polder area using anthropogenic gadolinium, *Hydrol. Earth Syst. Sci.*, 16(8), 2405–2415, doi:10.5194/hess-16-2405-2012.
- Sklash, M. G., and R. N. Farvolden (1979), The role of groundwater in storm runoff, *J. Hydrol.*, 43(1-4), 45–65, doi:10.1016/0022-1694(79)90164-1.
- Soulsby, C., J. Petry, M. J. Brewer, S. M. Dunn, B. Ott, and I. A. Malcolm (2003a), Identifying and assessing uncertainty in hydrological pathways: A novel approach to end member mixing in a Scottish agricultural catchment, *J. Hydrol.*, 274(1-4), 109–128, doi:10.1016/S0022-1694(02)00398-0.
- Soulsby, C., P. Rodgers, R. Smart, J. Dawson, and S. Dunn (2003b), A tracer-based assessment of hydrological pathways at different spatial scales in a mesoscale Scottish catchment, *Hydrol. Processes*, 17(4), 759–777, doi:10.1002/hyp.1163.
- Spear, R., and G. Hornberger (1980), Eutrophication in Peel Inlet—II. Identification of critical uncertainties via generalized sensitivity analysis, *Water Res.*, 14(1), 43–49.
- Stafleu, J., F. S. Busschers, D. Maljers, and J. L. Gunnink (2009), Three-dimensional property modeling of a complex fluvio-deltaic environment: Rhine-Meuse Delta, The Netherlands, in *Extended Abstracts of the Annual Meeting*, edited by R. Berg, H. Russell, and L. Thorleifson, pp. 47–50, Geol. Soc. of Am., Portland, Ore.
- Stuyfzand, P. J. (1989), An accurate, relatively simple calculation of the saturation index of calcite for fresh to salt water, *J. Hydrol.*, 105(1-2), 95–107, doi:10.1016/0022-1694(89)90098-X.
- Stuyfzand, P. J. (1993), Hydrochemistry and hydrology of the coastal dune area of the Western Netherlands, PhD thesis, 362 pp., Vrije Univ. Amsterdam, Amsterdam.
- Stuyfzand, P. J. (1999), Patterns in groundwater chemistry resulting from groundwater flow, *Hydrogeol. J.*, 7(1), 15–27, doi:10.1007/s100400050177.
- Uhlenbrook, S., and S. Hoeg (2003), Quantifying uncertainties in tracer-based hydrograph separations: A case study for two-, three- and five-component hydrograph separations in a mountainous catchment, *Hydrol. Processes*, 17(2), 431–453, doi:10.1002/hyp.1134.
- van der Velde, Y., P. J. J. F. Torfs, S. E. A. T. M. van der Zee, and R. Uijlenhoet (2012), Quantifying catchment-scale mixing and its effect on time-varying travel time distributions, *Water Resour. Res.*, 48, W06536, doi:10.1029/2011WR011310.
- Weiler, M., J. J. McDonnell, I. Tromp-Van Meerveld, and T. Uchida (2005), Subsurface stormflow, in *Encyclopedia of Hydrological Sciences*, edited by M. Anderson, pp. 1–14, John Wiley, Sussex, U. K.
- Weltje, G. J. (1997), End-member modeling of compositional data: Numerical-statistical algorithms for solving the explicit mixing problem, *Math. Geol.*, 29(4), 503–549, doi:10.1007/BF02775085.
- Zadeh, L. A. (1965), Fuzzy sets, *Inform. Contr.*, 8(3), 338–353, doi:10.1016/S0019-9958(65)90241-X.