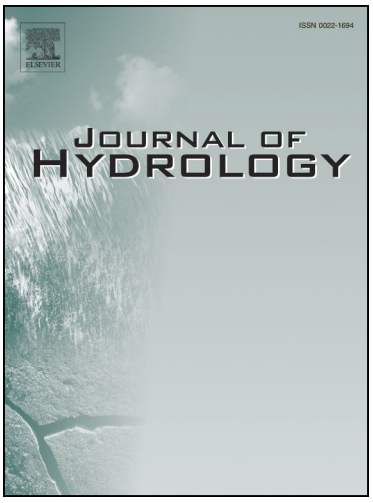
Worth of hydraulic and water chemistry observation data in terms of the reliability of surface water-groundwater exchange flux predictions under varied flow conditions

Daniel Partington, Matthew J Knowling, Craig T Simmons, Peter G Cook, Yueqing Xie, Takuya Iwanaga, Camille Bouchez

Journal Pre-proofs



Research papers

Worth of hydraulic and water chemistry observation data in terms of the reliability of surface water-groundwater exchange flux predictions under varied flow conditions

Daniel Partington, Matthew J. Knowling, Craig T. Simmons, Peter G. Cook, Yueqing Xie, Takuya Iwanaga, Camille Bouchez

PII: S0022-1694(20)30901-X
DOI: https://doi.org/10.1016/j.jhydrol.2020.125441
Reference: HYDROL 125441

To appear in: *Journal of Hydrology*

Received Date: 7 November 2019
Revised Date: 19 March 2020
Accepted Date: 18 August 2020

Please cite this article as: Partington, D., Knowling, M.J., Simmons, C.T., Cook, P.G., Xie, Y., Iwanaga, T., Bouchez, C., Worth of hydraulic and water chemistry observation data in terms of the reliability of surface water-groundwater exchange flux predictions under varied flow conditions, *Journal of Hydrology* (2020), doi: https://doi.org/10.1016/j.jhydrol.2020.125441

Worth of hydraulic and water chemistry observation data in terms of the reliability of surface water-groundwater exchange flux predictions under varied flow conditions

Daniel Partington[1], Matthew J. Knowling[2], Craig T. Simmons[1], Peter G. Cook[1], Yueqing Xie[1,3], Takuya Iwanaga[4], Camille Bouchez[5]

[1]National Centre for Groundwater Research and Training, & College of Science and Engineering, Flinders University, Adelaide, Australia

[2]GNS Science, Lower Hutt, New Zealand

[3]School of Earth Sciences and Engineering, Nanjing University, Nanjing, China

[4]Integrated Catchment Assessment and Management (iCAM) Centre, The Fenner School of Environment and Society, The Australian National University, Canberra , Australia

[5]Univ Rennes, CNRS, Géosciences Rennes, UMR 6118, 35000 Rennes, France

Corresponding author: Daniel Partington (daniel.partington@flinders.edu.au)

Abstract

This study assesses the worth of routinely collected hydraulic data (groundwater head, stream stage and streamflow) and lesser collected water chemistry data (Radon-222, Carbon-14, electrical conductivity (EC)) in the context of making regional-scale surface water-groundwater (SW-GW) exchange flux predictions. Using integrated SW-GW flow and transport numerical models, first-order, second-moment (FOSM) analyses were employed to assess the extent of the uncertainty reduction or lack thereof in SW-GW exchange flux predictions following acquisition of hydraulic and water chemistry observation data. With a case study of the Campaspe River in the Murray-Darling Basin (Australia), we explored the apparent information content of these data during low, regular and high streamflow conditions. Also, a range of spatial and temporal prediction scales were considered: catchment-wide and reach-based spatial scales and annual and monthly temporal scales. Generally, the data worth evaluations showed significant variability across predictions that were dependent on the spatiotemporal scale of the SW-GW exchange, the magnitude and direction of the SW-GW exchange flux and the prevailing streamflow conditions. These dependencies serve to emphasise the importance of prediction specifity with respect to SW-GW exchange. Among existing data, the most worth was found in Radon-222, groundwater hydraulic head, EC, and streamflow data showing average reductions in uncertainty of 41%, 38%, 32%, and 23% respectively. Assessment of type and spatiotemporal locations of potential data showed Radon-222 to be the next most important observation type across many predictions in locations with data paucity of all data types. Hydraulic observation data types were found to inform SW-GW exchange flux best under high- and regular- streamflow conditions when the magnitude of exchange fluxes were largest, whereas the water chemistry data was of highest

39 value for low- and regular- streamflow conditions where groundwater is discharging to the
40 stream.

1 Introduction

Observation data underpins effective water resource management, and managers have significant responsibility in decisions around data collection strategies. Conjunctive surface water (SW) and groundwater (GW) resource management requires having a quantitative insight into SW-GW exchange flux. Furthermore, the models that support such management are often required to provide the SW-GW exchange across multiple spatiotemporal scales.

Numerous methods exist for SW-GW exchange flux measurement and estimation (Fleckenstein et al. 2010; Kalbus et al. 2006). Means of estimating SW-GW exchange include direct methods (e.g., seepage meters), aimed at measuring the actual SW-GW exchange flux in-stream at a point. Direct methods are limited to very small spatial scales less than a few square metres and cannot be extrapolated to reach or regional scale SW-GW exchange (Cook 2015). Furthermore, such methods are limited by difficulty in identifying net SW-GW exchange – as opposed to hyporheic flow – and thus have limited utility at larger spatial scales (e.g. kilometre scale and greater).

The estimation of SW-GW exchange along stream reaches, as opposed to at a point, include methods based on stream water balance, hydraulic head gradient, river chemistry and ground water chemistry (see review by Cook 2015). The stream's water balance (through differential gauging) can be used where the SW-GW exchange is a significant component of the stream's water balance, greater than any uncertainties associated with other components of the balance. SW-GW exchange can be estimated by Darcian flux based on the average hydraulic head gradient across the stream and the average hydraulic conductivity of the streambed/aquifer. The exchange can also be estimated with stable and radioactive geochemical tracers (Cook 2013), requiring information on features such as the flow in the stream, stream geometry, and

hyporheic cycling. Electrical conductivity (EC) is one stable tracer that, given significant differences between the GW and SW EC can be used with end-member mixing analysis to estimate inflow of GW to a stream (Barthold et al. 2011). The presence of Radon-222 (^{222}Rn) in SW is indicative of GW discharge (Ellins et al. 1990), in the absence of significant hyporheic flow. When the concentration of ^{222}Rn in the GW is measured at multiple locations along a stream along with adequate sampling in the SW, then areas of GW discharge at the time of sampling can be pinpointed, and exchange fluxes estimated. As ^{222}Rn is a gas with a half-life of 3.8 days, it will only remain in the stream for short periods of time. The conservative nature of EC allows for differentiation of river water that briefly enters the streambed for a period before returning to the stream, which with accumulation of ^{222}Rn could be otherwise misinterpreted as regional GW discharge. Thus it is possible that these two observation data types contain unique information (i.e. uncorrelated) with respect to SW-GW exchange fluxes, which will be tested herein.

The use of physically based numerical modelling of flow in SW-GW systems, which is commonly applied to support managing water resources, allows for groundwater hydraulic head, stream stage and streamflow data to be integrated (e.g., Schilling et al. 2018; Wöhling et al. 2018). Furthermore, coupling of transport to such a flow model affords further integration of various stream chemistry and geochemical data, e.g. EC, ^{222}Rn or ^{14}C. Numerical models simulating SW-GW exchange that support water resources management provide an important basis for assessing the extent to which observations can build confidence in the prediction of SW-GW exchange, i.e., data worth (Fienen et al. 2010).

The level of confidence in regional scale predictions of SW-GW exchange flux obtained from various models is partly limited by the quality (measurement noise), quantity and types of

87 available data that inform such model predictions. Predictions of surface water-groundwater
88 exchange at a regional scale are critical to support conjunctive management of surface and
89 groundwater resources within strongly connected SW-GW systems, i.e. systems whereby change
90 to management of a river has a notable impact on the underlying aquifer and vice versa.
91 Determining the most informative observational data types and spatiotemporal quantities of such
92 data is an ever increasing need for water resource management practitioners (Kikuchi 2017).

93 The use of numerical models as a tool to formally assess the benefit of different data
94 types and optimal data acquisition/experimental design within a formal "data worth" assessment
95 framework is continually growing in popularity (Kikuchi 2017). The problem of "data worth", in
96 the context of water resources modelling, can be defined in terms of the reduction or lack thereof
97 in the uncertainty of any key prediction of management interest that is afforded through the
98 acquisition of observation data. A popular method used is first-order, second-moment (FOSM)
99 analysis, which assumes the model behaves linearly with respect to its input parameters and
100 simulated outputs. This approach is commonly used because of its suitability to be applied in
101 combination with complex models (which are often used to support environmental management)
102 owing to its computational efficiency (e.g. Dausman et al. 2010; Fienen et al. 2010; Moore and
103 Doherty 2005). Brunner et al. (2012) used this approach to explore the worth of groundwater
104 hydraulic head, ET and soil moisture observations in informing regional scale groundwater
105 models. Wallis et al. (2014) demonstrated the utility of a FOSM-based data worth analysis of
106 bromide, temperature, methane and chloride in the context of aquifer injection trials following
107 coal seam gas-related water production. Schilling et al. (2014) investigated the utility of novel
108 tree ring data in reducing predictive uncertainty of SW-GW exchange. More recently, Zell et al.
109 (2018) used a similar approach to analyse groundwater hydraulic head, stream discharge, SF_6,

CFCs and ^{3}H in GW transport times. Finally, Knowling et al. (2019b), explored the worth of tritium-derived mean-residence time data for forecasts of spring discharge.

In the current study, the worth of existing and potential different hydraulic and water chemistry data are quantitatively investigated in the context of SW-GW exchange flux predictions over monthly and annual timescales and over a range of length scales (whole of river (141 km) vs reach (0.8 - 40.4 km)), for a field site in south-eastern Australia (Campaspe River catchment). We specifically consider the worth of: groundwater hydraulic head, streamflow, stream stage/depth, stream EC, stream ^{222}Rn, and groundwater ^{14}C for such predictions. To the best of the authors' knowledge, the benefit or otherwise of these tracer methods have not been quantitatively evaluated compared to more traditional and routinely collected data types in the context of SW-GW exchanges at the regional scale. This paper aims to answer:

Q1. To what degree, if at all, does the addition/omission of existing hydraulic observation and/or chemical observation data reduce/increase the uncertainty of SW-GW exchange fluxes and what is the spatiotemporal variability of such reductions/increases during low, regular and high streamflow conditions?

Q2. Through consideration of potential future sampling locations and times, which hydraulic and/or chemical data should be targeted in the future to yield the best reductions in uncertainty of SW-GW exchange flux during low, regular and high streamflow conditions?

2 Case Study: Campaspe River

The Campaspe River, located in north-central Victoria, lies within the Murray-Darling Basin, shown in Figure 1. The river runs for 220 km, beginning in the hilly terrain of the Great Dividing Range and flowing down through undulating foothills to the wide flat riverine plain in

the north before it joins the Murray River; the river provides 0.9% of the inflow for the basin. The Campaspe River overlies a series of alluvial aquifers, namely the Coonambidgal Formation, Shepparton Formation, Calivil Formation and Renmark Group (the latter two also commonly referred to collectively as the Deep Lead aquifer) which interact with the river along its length. As shown in Figure 1, the hydrogeological units of the Lower Campaspe valley area are made up of a Palaeozoic basement of fractured and faulted rocks, overlain by the Renmark Group which contains a blanket of thinly bedded carbonaceous sand, silt, clay and peaty coal, then overlying this is the Calivil Formation comprising coarse grained quartzose sand and gravel sheet with minor kaolonite clay, and atop of the Calivil Formation the Shepparton Formation is made up of fine-grained clastics and polymitic sand and gravel. Finally, incised into the Shepparton Formation, the Coonambidgal Formation is made up of primarily light grey or brown silty clay, with sand beds often at the base (Arad and Evans 1987). The main groundwater resource in the Lower Campaspe Valley is the Deep Lead aquifer. Conjunctively managing both the Campaspe River and Deep Lead aquifer necessitates estimation of SW-GW exchange flux.

Figure 1. Campaspe River study area (a), location of Campaspe River catchment within the Murray-Darling Basin (b), and 3D model of hydrogeological units within the study area (c).

Average annual rainfall ranges from 424 to 746 mm, with the higher rainfall occurring at higher elevations above Lake Eppalock, and lower rainfall occurring in the Lower Campaspe Valley. On average, the highest rainfall occurs between June and August, with the driest months being January to March. As well as providing sustained flow, the Lake Eppalock dam (completed in 1964) allows enhanced recharge to the Lower Campaspe Valley, due to its use in providing irrigation water in the area. The focus of this study is the area downstream of Lake

Eppalock and all the way to where the Campaspe River joins the Murray River. Large diversions from the Campaspe River until recent years were made through offtakes from the Campaspe Weir (constructed in the late 1800s) in the Campaspe Irrigation District (CID). Significant groundwater pumping developed in the Lower Campaspe area in the 1960's.

2.1 Existing and potential observation data

Observation data types considered in this study include routinely collected and publicly available hydraulic data, i.e., streamflow (daily), stream stage (daily) and groundwater hydraulic head data (variable, usually quarterly) (Australian Bureau of Meteorology (http://www.bom.gov.au), Victorian Government (http://data.water.vic.gov.au)). We also used existing EC data from the Victorian Government at stream gauges (http://data.water.vic.gov.au). As part of this study we collected surface (multiple times) and groundwater (once) ^{222}Rn data (spot sampling), and new groundwater ^{14}C to supplement existing ^{14}C data collected in previous studies (Cartwright et al. 2012; Cartwright et al. 2006; Cartwright et al. 2010) (spot sampling). All considered observation data sampling locations are shown in Figure 2, where it can be seen that the majority of groundwater head observations are located in the north of the study area. All data collected previously and part of this study is herein referred to as "existing data", as opposed to "potential data" which herein refers to as yet uncollected data. A dense network of future potential observation locations (i.e. data not yet collected) was considered spanning the entire study area, which includes GW sampling locations for groundwater hydraulic head and ^{14}C and SW sampling locations for stream stage, streamflow, stream ^{222}Rn and stream EC. The potential locations were chosen with the aim of filling the spatial gaps in data, e.g. hydraulic head in the south of the study area.

Figure 2. Locations for observation data, including a) existing hydraulic (groundwater hydraulic head, stream stage and streamflow) and b) chemical (stream ^{222}Rn, ^{14}C, stream EC) data. Potential future observation data (c) collection locations considered are also shown. The potential GW observations cover the extent of the aquifers, with gaps existing in the south of the study area due to the presence of only bedrock. The number of locations for each data type is shown in brackets in the legend.

3 Methodology

3.1 Integrated SW-GW model setup

The integrated SW-GW numerical models described below collectively serve as a tool for quantifying regional-scale SW-GW exchange flux prediction uncertainty and its reduction (or lack thereof) through the collection of various types of hydraulic and chemical data. A requirement for effective model usage in this context is that the models provide a robust basis for representing the primary processes and parameters on which predictions of interest may depend. For example, Fienen et al. (2010) showed that spatially distributed parameterisation schemes are necessary for effective predictive uncertainty estimation and to avoid corrupted data worth interpretations that may arise when adopting more parsimonious parameterisation schemes. As such, the numerical models employed here are physically based and highly parameterised (> 1500 parameters), to allow simulation of surface and subsurface hydraulics (hydraulic head, stream stage and flow) and transport (^{222}Rn, ^{14}C and EC), and to robustly express uncertainty in regional-scale SW-GW exchange flux predictions (e.g., Hunt et al. 2007; Knowling et al. 2019a), respectively.

198 The integrated models considered herein comprise a series of SW-GW flow models and a
199 series of SW-GW and SW solute transport models (Figure 3). The flow modelling in this study
200 was carried out using MODFLOW-NWT (Niswonger et al. 2011). Assimilation of the hydraulic
201 data (groundwater hydraulic head, streamflow, stream stage) is achieved through the integrated
202 SW-GW flow models. The flow models simulate 3D saturated groundwater flow (ignoring
203 unsaturated flow) and 1D surface flow routing through rivers (by the kinematic wave equation;
204 SFR2 (Niswonger and Prudic 2005)). The flow model focuses representation of surface flow on
205 the Campaspe River, ignoring some of the small tributaries that feed into the main river
206 downstream of Lake Eppalock. This simplification is made as little flow arises from these
207 tributaries other than in large rainfall events.

208 The flow solutions obtained from the steady and transient MODFLOW-NWT models
209 were subsequently used to simulate transport of ^{14}C using MT3D-USGS (Bedekar et al. 2016a;
210 Bedekar et al. 2016b). Due to existing limitations in simulating radioactive decay and
211 evapoconcentration with the stream flow transport (SFT) package of MT3D-USGS, the
212 simulation of EC and ^{222}Rn stream concentrations was carried out with an analytical steady-state
213 transport model which accounts for evapoconcentration, decay and hyporheic exchange (similar
214 to that of Cook et al. (2006) but rearranged to solve for concentration as shown in the Appendix)
215 using the MODFLOW-simulated streamflows and SW-GW exchange fluxes as inputs. With this
216 SW transport model, GW ^{222}Rn and EC concentrations were treated as a static boundary (see
217 Table 2). It is assumed that over the monthly time-step used in the flow model that the river is
218 completely flushed and that all inflows (and corresponding concentrations) are steady, hence the
219 use of the steady-state transport model.

The series of flow and transport models, shown in Figure 3, were used as a basis for representing the different hydraulic and chemical data types. Firstly, groundwater hydraulic head, streamflow, and stream depth/stage, are simulated under pre-clearance conditions in a steady-state flow model (SS) (MODFLOW-NWT). Secondly, transient SW-GW flow (TR) is simulated spanning the period 1840 to 2018. In the transient flow model, it is assumed that clearance of native trees and shrubs was immediate (1840) and that irrigation was static (using long term average). ^{14}C was simulated in two models, firstly using the output from the SS flow model but simulating transport for 40,000 yrs ($TR1_{C14}$) with an initial concentration of ^{14}C set to 0 PMC across the model domain; subsequently the final concentration of ^{14}C in the $TR1_{C14}$ simulation was passed as the initial conditions for the post-clearance to present day simulation of ^{14}C ($TR2_{C14}$). Inflow from recharge was assigned as 100 PMC in both ^{14}C simulations. The Campaspe river flows and SW-GW exchange fluxes simulated at each stream reach in the period of interest from the transient flow solution (TR) (June 2016-May 2017) were passed to the 1D transport model for simulation of ^{222}Rn (SS_{Rn}) and EC (SS_{EC}) at each of the months within this period of interest. Groundwater concentrations for ^{222}Rn and EC were assigned as static.

Figure 3. a) The series of flow and transport models employed and the associated data types simulated by each, b) a 3D model schematic, including spatial grid and boundary conditions: drains (DRN), Campaspe River (SFR), general head (GHB), Murray River (RIV), pumping (WEL), and recharge (RCH), which is depicted in the overlaid and elevated surface with 11 time invariant recharge zones.

The numerical grid was discretised into 1 km x 1 km cells (Figure 3b) with 7 layers of variable thickness covering the 6 hydrogeological units shown in Figure 1c. The mean, minimum and maximum values of each unit are shown below in Table 1.

Table 1. Summary of hydrogeological layer thicknesses including, mean, minimum, maximum thickness and percent volume. Hydrogeological units are abbreviated as Coonambidgal (co), Shepparton (sh), Calivil (ca), Renmark (re), Newer Volcanics (nv), Basement (ba).

Variable time steps are employed that are 40 yrs (1840-1880), 84 yrs (1881-1965), 20 yrs (1966-2005), 10 yrs (2006-2015) and then monthly from January 2015 to March 2018. SW-GW exchange flux predictions of interest considered herein for the purposes of the current data worth analysis are made over a one year period of simulation between the start of June 2016 and the end of May 2017. The exchange fluxes are considered along the entire river from Lake Eppalock to the Murray River (141 km), and for reaches between river gauges along this length of river (11 reaches ranging from 0.8 to 40.4 km in length). The outputs of SW-GW exchange flux from the transient flow model at each reach were considered at the monthly (important for river operations and ecological assessment) and yearly resolution (important for water allocations). Also considered was the spatial sum of the SW-GW exchange flux along the whole river at monthly and annual time scales (important for groundwater use management strategies). These different spatiotemporal predictions give rise to a total of 156 predictions of interest.

The MODFLOW-NWT models were forced by recharge using the RCH (specified flux) package, by groundwater pumping using the WEL (specified flux) package (applied from 1966 onwards), and by rivers and drains using the SFR (streamflow routing and head-dependent exchange flux, applied to the Campaspe River, which is the focal point of this study), RIV (head-dependent exchange flux applied to the Murray River) and DRN (head-dependent flux)

packages. Stream diversions in-to and out-of the Campaspe River are also captured using the SFR package. The locations of the boundary conditions are shown in Figure 3. There is no pumping or artificial drainage in the pre-clearance steady-state (SS) flow model. Zonal rainfall reduction parameters (11 zones) are set to 1% (i.e. assuming very low recharge at a time when the land was densely covered in vegetation) and are multiplied by the spatially varying map of temporal long-term average rainfall over this period for specifying the recharge boundary condition. As rainfall reduction parameters were used for recharge, evapotranspiration was not explicitly modelled in this study. In the post-clearance transient flow model (TR) the zonal values for each of the 11 zones are modified to reflect the land use, soil type and mean annual rainfall and multiplied by temporally varying spatial maps of rainfall to provide the temporal recharge input maps for the model. In the post-clearance transient flow model (TR) irrigation is embedded in the recharge factor for requisite zones as reflected by the land use. Temporally static aquifer and stream properties were spatially parameterised using pilot points. The locations of pilot points, which also correspond with the locations of potential observations, were automatically generated (see SI).

Model history matching was performed on the basis of groundwater hydraulic head, streamflow, stream stage, stream ^{222}Rn, groundwater ^{14}C and stream EC, at the locations shown in Figure 2. Model parameters (Table 2) for stream and aquifer properties were subject to estimation through history matching. History matching was carried out using the parameter estimation suite PEST (Doherty 2016) using Tikhonov regularisation.

Table 2. Summary of model parameterization. For the aquifer property parameters, the hydrogeological units are abbreviated as follows: Coonambidgal (co), Shepparton (sh), Calivil (ca), Renmark (re), Newer Volcanics (nv), Basement (ba).

The flow and transport models were built utilising FloPy (Bakker et al. 2016). The in-stream transport model for ^{222}Rn and EC was implemented in Python (see SI).

3.2 Assessment of predictive uncertainty in SW-GW exchange flux

The worth of data is considered herein as the reduction or lack thereof of the uncertainty of the prediction of interest (SW-GW exchange) with the addition of various acquired or potential observation data. A brief overview of the key theoretical aspects underlying the approach adopted for the quantification of predictive uncertainty and data worth is now provided.

The posterior (i.e., post-history matching) parameter covariance matrix ($\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}$) can be estimated with Schur's complement as (Christensen and Doherty 2008; Doherty 2015; Tarantola 2005; White et al. 2016):

$$\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}} = \boldsymbol{\Sigma}_{\boldsymbol{\theta}} - \boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{J}^T[\mathbf{J}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{J}^T + \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}]^{-1}\mathbf{J}\boldsymbol{\Sigma}_{\boldsymbol{\theta}} \qquad (1)$$

Where, $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ is the prior parameter covariance matrix, $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ is the epistemic noise covariance matrix (i.e. accounting for both measurement and model error), and $\mathbf{J}$ is the Jacobian matrix of partial first derivatives of model outputs (for which there are corresponding observations) with respect to parameters $\boldsymbol{\theta}$. The second term on the RHS of (1) expresses the reduction in uncertainty surrounding parameters as a result of conditioning the model on the information contained in the observations.

The prior and posterior uncertainty variance for a prediction s, $\bar{\sigma}_s^2$, respectively can be estimated via uncertainty propagation:

$$\sigma_s^2 = \mathbf{y}^T \boldsymbol{\Sigma}_{\boldsymbol{\theta}} \mathbf{y} \qquad (2)$$

And

$$\overline{\sigma}_s^2 = \mathbf{y}^T \overline{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}} \mathbf{y} \qquad (3)$$

Where **y** is the sensitivity vector for prediction *s* with respect to the parameters $\boldsymbol{\theta}$ (a row extracted from the **J** matrix).

This study assumes the parameter covariance matrix ($\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$) is a diagonal matrix (does not contain non-zero off-diagonal elements). This assumption means that zero correlation exists between parameters. The assumption of zero correlation between spatially distributed parameters, and in particular pilot point aquifer and river property parameters was considered appropriate given the limited spatial coverage of both aquifer and river property information as well as the large distance separating pilot points (on average, 6 km x 6 km). The lower and upper bounds of the parameter range (see Table 2) are specified based on hydrological and geological expert knowledge (in this case conservative) and are assumed to represent the 5^{th} and 95^{th} percentile of the (assumed Gaussian) parameter distribution.

The Jacobian matrix **J** was populated using 1% two-point derivative increments. These derivatives with respect to a parameter set were obtained following a history matching undertaking that is unrelated to the implementation of the data worth analyses presented herein (see SI for details).

The epistemic noise covariance matrix ($\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$) is also assumed diagonal. It is specified practically by assigning observation weights (inversely proportional to the standard deviation of noise) (Doherty 2015). Observation weights are specified on the basis of a subjective assessment of the measurement noise standard deviation (Table 3), before being adjusted in accordance with

model-to-measurement residuals to account for model error. Weight adjustment is undertaken in such a way that the weight of any observation cannot be increased (thereby respecting the contribution to epistemic noise from measurement noise), and that the contribution of each observation group to the model-to-measurement objective function is equal to the number of non-zero weighted observations in that group (Doherty 2016). The use of model-to-measurement residuals to approximate model error is deemed appropriate given that this quantity can never be known, and that the residuals constitute the only information available for reflecting model error with respect to different types of observation and simulated outputs. For potential observations, i.e., where no measurements exist for which to undertake the above-described weight adjustment, weights are specified based on the average (adjusted) weight assigned to existing observations of the same group/type. The observation noise variance for all but stream flow was fixed as stream flow measurement error is known to become larger at higher flows, particularly when extrapolating the rating curve (Di Baldassarre and Montanari 2009) and an assumed 40% error accounts for a worse case evaluation of this error. The posterior parameter covariance matrix and ensuing data worth analysis was calculated using the Python package pyEMU (White et al. 2016).

Table 3. Assumed observation noise (used together with model-to-measurement residual information to populate $\boldsymbol{\Sigma_{\epsilon}}$).

The worth of different observation data is evaluated in different ways in this study. Firstly, we consider the worth with observation data groups by themselves, i.e. the ability of an individual observation data group to reduce uncertainty on a prediction of SW-GW exchange. To do this, predictive uncertainty with that particular observation data group is estimated by evaluating equation 1 and 3 twice, once where the particular observation data does not appear in

$\boldsymbol{\Sigma_\epsilon}$, which we term the 'base' uncertainty ($\overline{\sigma}_{s\ base}$), and again where $\boldsymbol{\Sigma_\epsilon}$ contains the particular observation data group, which gives the 'group' predictive uncertainty ($\overline{\sigma}_{s\ group}$). The calculation of reduction in uncertainty through adding the observation data type group (DW_{add}%) is calculated as:

$$DW_{add}\% = \left(1 - \frac{\overline{\sigma}_{s\ group}}{\overline{\sigma}_{s\ base}}\right) \times 100\% \qquad (4)$$

Secondly, to evaluate the mutually exclusive information that exists in each observation data group we consider the difference between predictive uncertainty reduction using all observation data groups and omitting an observation data group from all groups. To do this, we again evaluate equation 1 and 3 twice, once where in $\boldsymbol{\Sigma_\epsilon}$ all observation data groups are considered together ($\overline{\sigma}_{s\ all}$), and again with all observation data groups except for the one of interest considered ($\overline{\sigma}_{s\ all-group}$). Then we compare the two against the 'base' uncertainty:

$$DW_{remove}\% = \left(\frac{\overline{\sigma}_{s\ all-group} - \overline{\sigma}_{s\ all}}{\overline{\sigma}_{s\ base}}\right) \times 100\% \qquad (5)$$

Finally, in the context of the potential observations, we evaluate the next most important observations (using built in functions in pyEMU; White et al. (2016)) by iterating over each of the potential observations alone (for select predictions) and in groups (for all predictions within a group) to find the best reduction in uncertainty, then we add that observation or group of observations to the list of existing observations and repeat the process. The addition of the previously identified best observation or observation group accounts for any correlation between observations or groups of observations.

In order to answer the first question posed (**Q1**) regarding the degree to which water chemistry data and hydraulic data reduces the uncertainty surrounding the 156 predictions of SW-GW exchange, each of the observation data types (or "groups") were first considered

individually for existing observation data with the "base" containing no observation data (Table 4). Then the potential data (**Q2**) were evaluated with the "base" consisting of all existing observation data. The individual contribution of each potential observation to the whole of river exchange at three different times covering low, regular and high streamflow conditions was examined to determine the worth of particular potential data types and quantities of value.

Table 4. Data worth assessments employed for the range of SW-GW exchange predictions from the Campaspe transient flow model (TR).

4 Results and Discussion

4.1 Simulated SW-GW exchange behaviour

The behaviour of SW-GW exchange flux is first examined at each of the TR model's 122 stream segments on a monthly basis as a means of establishing an understanding of dynamics and spatiotemporal variability of the exchanges before considering the data worth analysis.

Throughout the results section we adopt the convention of denoting gaining conditions in a stream by negative values and losing conditions by positive values. The monthly flows are herein subjectively categorised as either low (<35th percentile), regular (between 35th and 80th percentile), or high (>80th percentile). Each of the flow categorisations is also associated with clearly differing SW-GW exchange patterns.

The post-clearance transient flow simulation generally showed that SW-GW exchanges along the length of the river exhibit gaining behaviour (Figure 4). The exception to this being during high flow events, when the river transitions to a largely losing river (0.23 m^2/d) as significant inflows from Lake Eppalock raise the stream stage and reverse the hydraulic gradient along the majority of the river. The exchange fluxes per unit length of stream at any point along

396 the stream during the simulation period range from losing by 9 m^2/d during the October 2016
397 high flow event to gaining at -9 m^2/d as the system recovers in the following month. The
398 spatially and temporally averaged SW-GW exchange along the entire river is gaining at
399 approximately -1.1 m^2/d. Reach r3 shows the strongest variance in exchange flux along its length
400 through differing inflow conditions followed by r5, r6 and r8, while r10 and r11 show the least
401 variance and are always gaining.

402

403 **Figure 4. Simulated SW-GW exchange fluxes per unit length of stream at each reach along**
404 **the Campaspe River. Coloured lines depict exchange fluxes under differing low, regular or**
405 **high flow conditions for each month. The black line depicts the temporally averaged**
406 **exchange flux. The dotted grey vertical lines indicate gauge locations with the 11 between-**
407 **gauges-reaches annotated at the top of the graph.**

408 The simulated SW-GW exchange fluxes for the whole river and for each of the 11 river
409 reaches for annual average and monthly average time scales (i.e. 156 predictions) range from -
410 3.08 to 2.23 m^2/d (shown in Figure 5). The strongest losing flux appears along reach 3 during the
411 high flow event in October 2016. The simulated SW-GW exchange flux was highest along reach
412 3. Reach 6 (r6) just upstream of the Campaspe weir exhibits losing conditions, while all other
413 reaches show gaining conditions. For the examined year, there is a clear link between the inflow
414 from Lake Eppalock and the pattern of exchange fluxes (Figure 5).

415

416 **Figure 5. Heatmap of simulated SW-GW exchange per unit length along the whole of the**
417 **river and along each of the 11 reaches at annual and monthly time scales, which comprises**

418 **the 156 predictions of interest. Red cells in the heatmap indicate losing conditions and blue**
419 **cells indicate gaining conditions. The left panel shows the location and length of each of the**
420 **reaches for reference. The bottom panel indicates the annual average (mean) inflow from**
421 **Lake Eppalock to the system as well as the monthly inflow with the colours of the bars**
422 **indicating whether the flow is low, regular or high.**

423 *4.2 Worth of existing hydraulic and water chemistry data types (Q1)*

424 Assessment of the worth of individual observational data types alone (i.e., DW_{add}%) for
425 the spatial and temporally aggregated whole of river annual exchange showed that hydraulic
426 head followed by EC, ^{222}Rn and flow observation groups had sizeable worth of > 40% (the green
427 dots in Figure 6a). Similar relative trends across data types were seen for the uncertainty
428 increases *without* individual data groups when compared to all data groups (DW_{remove}%). The
429 lower values in DW_{remove}% as compared to DW_{add}% arise because the former yields the unique
430 information contained in an individual data group; this allows for assessment of the extent of
431 correlation and redundancy of individual data groups. It was evident from the analysis of
432 DW_{remove}% for the spatially and temporally averaged SW-GW exchange prediction, that the head,
433 streamflow, ^{222}Rn and EC data contain unique information.

434 Across all 156 predictions for SW-GW exchange flux predictions along the Campaspe
435 River, the median worth obtained from both DW_{add}% and DW_{remove}% showed that head data were
436 significant whereas the streamflow data were less informative. Furthermore, there appears to be
437 redundancy in the flow data. Also, the median worth of ^{222}Rn was significant whereas the EC
438 data was much lower (Figure 6a). Stream stage and ^{14}C were both poor, with the distribution of
439 DW_{add}% for stream stage data close to zero. The large range in worth of head, flow, ^{222}Rn and
440 EC data types and the degree of information redundancy across the 156 predictions highlights the

nature of the local information content that particular data types have for specific predictions. The local information content is highlighted by the outliers in DW_{remove}% showing unique information in head, ^{14}C and ^{222}Rn for some predictions of SW-GW exchange.

Figure 6. Analysis of worth for the whole of river annual exchange (green dots) shown by a) % reduction in predictive uncertainty for SW-GW exchange associated with each particular observation group (DW_{add}%), and b) difference in reduction between using all observation data types and using all except for a particular group from the combination (DW_{remove}%). Furthermore, boxplots are shown for the distributions across all 156 SW-GW exchange predictions.

The worth for the 156 SW-GW exchange predictions shows distinct patterns that are associated with inflow to the Campaspe River from Lake Eppalock (Figure 7). During the large flow event during October 2016, across all reaches and the whole river there was consistently poorer worth for predictions within that month. The relative lack of information resident in the data types continues in the following months associated predictions, while the system recovers. The whole of river predictions show an expected dampened variability and higher average worth due to its spatially integrated nature. At the end of the river system in reaches r10 and r11 all data types are seen to have quite low worth, except ^{222}Rn which appears to be the only data-type to show value in reach r10. The poor performance at the end of system could be due to a "boundary effect", related to the fact that both the Murray River boundary condition and underlying groundwater general head boundary conditions possibly influence the model to such an extent that they overpower the information content of any existing observational data. Potential boundary effects could be avoided by extending the northern boundary of the model past the

Murray River and converting the Murray River to a flow routing representation rather than the fixed head representation that was implemented.

Figure 7. Heatmaps of percentage reduction in SW-GW exchange uncertainty obtained with all data (a), and with each of the observation data types alone (b-g). The reduction is shown for each SW-GW exchange flux considered, i.e. for each of the stream reaches (r1-r11) and whole of river (nrf), and at each of the temporal scales of annual and monthly between the start of June 2016 and end of May 2017. The inflow to the system is shown under the first heatmap (a) from which the high flow event in October can be seen. The locations and lengths of reaches (r1-r11) are shown to the left of the first heatmap for reference.

The analysis of existing observation data types, firstly, identified that the temporal integration in annual predictions (both whole of river and reach scale) generally led to better reductions in uncertainty than in the monthly predictions. These are plausible given that the annual signal is smoothed. For hydraulic head, ^{14}C and ^{222}Rn, there were not data available in every month of the year, but particular data were at least present enough in the seemingly dependent months (during large exchanges) required to inform the annual SW-GW exchange prediction.

4.3 Worth of individual potential future hydraulic and water chemistry data points (Q2)

We analysed the worth from individual potential in-stream observations to explore the extent to which the whole of river SW-GW exchange prediction reliability could be improved through data acquisition at new sampling locations. The worth of individual potential observations of stage, flow, ^{222}Rn and EC at select times when flow conditions were low (July

2016), high (October 2016) and regular (November 2016) are shown in Figure 8; the corresponding SW-GW exchange along the stream at each of these times is also shown. The value of both stage and EC potential observations are shown to be poor across these flow conditions, with only ^{222}Rn (low and regular flow conditions) and flow (high and regular flow conditions) showing considerable value. Under low flow conditions, ^{222}Rn data displayed the highest utility where existing data were not present and where the stream exhibits strong gaining conditions in the first 40 km of the stream. Also, ^{222}Rn showed considerable value in the slightly gaining areas at around 130 km downstream, likely due to the paucity in local existing data for all data types. However, ^{222}Rn data were seen to be of low utility when high flow conditions prevail with a corresponding reversal of hydraulic gradient resulting in a mostly losing river, as simulated during the high flow event in October 2016. In the high flow event instance, there is a trend of increasing flow data worth with distance downstream.

Figure 8. Further reduction in uncertainty for net SW-GW exchange in October 2016, November 2016 and May 2017 obtained through potential observations of stream stage, flow, ^{222}Rn and EC at 80 locations along the stream. Each potential observation is considered alone but added to the existing observations across all observation data types. Underlying each uncertainty reduction plot is the pattern of exchange along the river during each of the months with the scatter showing the exchange rate (m^2/d) and the colours representing the exchange flux along the reach at that location; reds indicate losing and blues indicate gaining conditions. The mean exchange rate is shown in light grey on each of these plots to give context to the exchange conditions relative to the mean.

To explore the extent of reduction in uncertainty obtained through potential addition of subsurface data to the existing dataset, we analysed the whole of river SW-GW exchange flux during low, high and regular flow conditions through both head and ^{14}C in the shallow and deep aquifers at the potential sampling locations (Figure 9). The SW-GW exchange along the Campaspe River in the highest (southern) parts of the model domain is very sensitive to the GW level which is strongly connected at the top of the catchment to the heads in the narrow alluvial channels and hence it is observational data located here that appears to inform the whole of river SW-GW exchanges the most, although the reductions in predictive uncertainty are only marginal. This generally highlights opportunity for more value from targeted observations in the southern part of the catchment in the narrow alluvial valleys. The deep heads are seen to hold the least value of the potential observations across flow conditions. As expected, due to the lesser variance of subsurface data as compared to stream data, spatial patterns of worth are more consistent across the low, regular and high flow conditions. Furthermore, there is clear crossover of high utility potential sampling locations for shallow and deep ^{14}C and shallow head too in the southern part of the catchment, although shallow head also appears of some value along the entire length of the stream.

Figure 9. Percent reduction in uncertainty for annual whole of river SW-GW exchange fluxes of the Campaspe River obtained through hydraulic head and ^{14}C at potential sampling locations (identical to pilot points) in the shallow and deep aquifers.

4.4 Worth of potential future hydraulic and water chemistry data types (Q2)

For all of the potential data and also each of the data type groups, we examined the benefit of using all potential data (Figure 10), which highlights the extent to which particular

predictions can be improved (as compared to existing data) under comprehensive sampling of all data types and individual data types. The potential hydraulic head data showed improvement in terms of SW-GW prediction reliability in the middle section of the river prior to the largest of the high-flow events and a few months after the recovery of this event (Figure 10b). The stage data again shows little value, no matter the location or time of sampling (Figure 10c). The ubiquitous poor worth of stage data was surprising as it could be assumed that the intrinsic link between hydraulic head and stage in the calculation of the exchange flux would result in stage data containing prediction-relevant information. We posit a potential reason for this is the stream stage as simulated by the numerical model showed less variance (minimum and maximum of 1.21E-4 and 1.11E-2) then the observed data owing to the representation of a rectangular channel and associated parameterisation. For example, the effective stream width may have been overestimated in the model and hence the associated error with the simulated stage was potentially overrated. This led to a poor rating in stage data but is likely more linked to the modelling assumptions, e.g. a perfectly known riverbed elevation, and structure for the Campaspe River, rather than the "true" worth of the data itself.

Further reductions in predictive uncertainty up to around 25% are seen with the addition of the flow, head and EC data. The addition of potential ^{222}Rn data (Figure 10f) shows clear improvements for reaches r2–r9. Especially of interest are those improvements in worth at the time prior to the high flow in October 2016 and just after. Assessment of potential ^{14}C data showed significant improvement of a further 40% reduction in most predictions along reaches r1–r5 (Figure 10e), suggesting that the existing spatiotemporal ^{14}C data locations were suboptimal in this context of SW-GW exchange. However, the comprehensive sampling of both

552 ^{222}Rn and ^{14}C through space and time is likely impractical due to the costs, especially for ^{14}C
553 which would incur drilling costs.

554 **Figure 10. Difference in uncertainty reductions between existing and potential observation**
555 **data for all data (a) and for each of the individual observation groups (b-g).**

556 The analysis of the ten "next best" most important observations to collect on top of all
557 existing data during low, high and regular streamflow conditions showed that the optimal
558 location set is different under differing flow conditions. ^{222}Rn was the most beneficial data type
559 to collect next with head, ^{14}C, and flow also in the top ten (Figure 11). Interestingly, during low
560 and high streamflow conditions, within-stream observations of ^{222}Rn and flow were distributed
561 in the upper, middle and lower reaches of the stream, whereas for regular flow conditions, the
562 observation data were in the middle and lower reaches only. The role of ^{14}C observations in the
563 top left of the maps (Figure 11a-c) are at first perhaps counter-intuitive, however, this location
564 represents the longest flow path through the aquifers before exiting through the northern
565 boundary, informing the velocity of flow and its variation and hence the hydraulic conductivity
566 of the aquifers and its effective porosity, the former of which in turn informs the SW-GW
567 exchange along the river.

568 **Figure 11. Locations and rank of 10 next most important potential observations to add to**
569 **the existing observations for whole of river exchange during months of low, high and**
570 **regular streamflow conditions (a-c). Further reduction in uncertainty (%) due to each of**
571 **the 10 next most important potential observations (d-f).**

572 Finally, we examined the next most important observation groups for each of the 156
573 SW-GW exchange predictions (Figure 12), i.e. using all potential data within each data type.

This assessment showed that potential ^{222}Rn was the most important group across the majority of predictions, followed by shallow head, ^{14}C, EC, flow, deep head and lastly stage, which was ranked 7^{th} for the majority of predictions (Figure 12).

Figure 12. Ranking (x-axis) of best uncertainty reduction across each of the 156 predictions of interest in a "next most important" type analysis. The uncertainty reduction is based on each observation group for potential observations when added to the existing data.

4.5 General observations

The above results have demonstrated that the worth of different hydraulic and chemical observations in the context of making SW-GW predictions is dependent on the prevailing streamflow conditions, the magnitude and direction of the SW-GW exchange and the spatial and temporal scale of the exchange considered. It has also demonstrated the large variability in worth across different SW-GW exchange predictions as a result of these dependencies. These findings are in addition to previously reported dependence of data worth on prediction specificity more generally (e.g. Dausman et al. 2010; White et al. 2016).

4.5.1 Influence of streamflow conditions and magnitude and direction of SW-GW exchange

The prediction-specificity of data worth with respect to prevailing stream flow conditions can be explained by the sensitivities of the different SW-GW exchange predictions to uncertain model parameters that are conditioned on the basis of both existing and potential hydraulic and chemical observational data. During high-flow conditions, generally lower data worth is apparent. This is because SW-GW exchange predictions under high-flow conditions depend on a larger portion of uncertain model parameters (e.g., recharge and aquifer properties for the TR model); this results in lower worth given the limit on the ability of information to "spread" from

596 data in space and time. That is, there are a number of prediction-parameter sensitivities that are
597 heightened as the Campaspe flow system is perturbed firstly by large losing SW-GW exchanges,
598 and secondly by the presence of distributed above-average recharge which also propagates
599 through the subsurface. Furthermore, the uncertainty in the flow observations increases with high
600 flows. As the stream is largely losing during high flows, the SS_{EC} and SS_{Rn} transport parameters
601 become less sensitive with respect to the corresponding SW-GW predictions. During low-flow
602 conditions, where rainfall recharge is often relatively small and the river is generally weakly
603 gaining along the Campaspe River, the SW-GW exchange predictions are generally less sensitive
604 to the flow model parameters. This allows for the information contained in the ^{222}Rn and EC
605 observations to be used more effectively through the SS_{EC} and SS_{Rn} transport parameters.

606 For example, in the annual whole of river SW-GW exchange flux prediction, it was
607 apparent (Figure 8) that in-stream sampling of ^{222}Rn can lead to a further reduction in uncertainty
608 (up to 10%) during low flows, with some value during regular flow conditions (up to 6%) but
609 with reduced utility (<0.5%) in high streamflow conditions. This is because the predictions
610 during lower flow show higher sensitivity to the parameters that are conditioned by the
611 information contained in the ^{222}Rn observations. It is thus necessary to target the particular time
612 and location carefully for sampling water chemistry data due to the transient and local
613 information content. This is further evidenced by the improvements through all potential data for
614 each data type which showed the theoretically possible improvements when comprehensive
615 sampling takes place in space and time. This differs in comparison to the hydraulic data, which
616 seems to show similar patterns across sampling times in worth for groundwater hydraulic head
617 and streamflow data points as explained above (Figure 8 and Figure 9). Despite the similar

patterns, the "flow of information" from hydraulic observation data appears to be larger under regular and high flow conditions.

As was explained at the start of the results (4.1), the general patterns between flow conditions and SW-GW exchange are clear. The apparent information content in all observation data appears linked to the magnitude and direction of the SW-GW exchange flux in many predictions. It was evident that for the very weakest exchanges, the poorest worth was found, no matter the data type; however, the opposite was not true for the strongest SW-GW exchanges which exhibit more complex worth patterns across data type.

4.5.2 Influence of spatial and temporal scale of SW-GW exchange

The simulated variability in monthly reach-scale SW-GW exchange in the Campaspe River was clear, and so were the corresponding reductions in predictive uncertainty due to data collection. When averaging the SW-GW exchange over the whole of the river, the worth of data was reasonably consistent on a monthly basis for each data type alone and for all data types, with a clear trend in variability being linked to the flow conditions (discussed above). Furthermore, the consistent data worth across months was also consistently close to the best uncertainty reductions from the reaches, i.e. the reductions were not a simple average of the individual reach uncertainty reductions, but more closely linked to the best reductions. This is expected due to the spatial integration of information contained in the hydraulic and chemical data. A similar pattern with respect to temporal integration of information is present in comparing monthly to annual SW-GW exchanges.

There was no clear relationship between the length of the reach and annual reductions in predictive uncertainty. However, the lowest uncertainty reductions were apparent in the shortest

reach r10 (0.8 km). The lack of a clear relationship is likely due to a combination of the above-mentioned factors of prevailing flow conditions and magnitude of the SW-GW exchange fluxes.

4.5.3 Model simplifications and assumptions

Interestingly, even though the water chemistry data provide more indirect means to calculate SW-GW exchange flux than the hydraulic data (i.e., chemistry data serve as proxy for flux), these data types alone showed greater worth in many predictions. For the cases in which the river is not experiencing low-flow conditions and gaining, we would posit that this is partly a result of the simplified 1D transport models (SS_{Rn} and SS_{EC}) used to map the ^{222}Rn and EC observations to SW-GW exchange predictions through the SS_{Rn} and SS_{EC} transport parameters. The parameterisation, spatial scale of river segments, and process assumptions that were applied to the simplified 1D transport model (e.g., the uniform fixed groundwater concentrations of ^{222}Rn and EC, and the monthly steady-state assumption) likely inflates their sensitivity for these data types and hence their worth to SW-GW exchange predictions (Fienen et al. 2010). Future modelling of in-stream ^{222}Rn and EC transport would benefit from testing this by further evaluating the worth of these data types in a transient transport model with spatiotemporally varying parameters, including the inputs of both ^{222}Rn and EC. With appropriate data collection of times series data of near-stream GW ^{222}Rn and EC, future modelling of the Campaspe system may benefit from explicit modelling of groundwater transport of ^{222}Rn and EC to help demonstrate that current the boundary simplification of ^{222}Rn and EC to static GW conditions does not introduce large impacts to the data worth analysis.

The simulation of SW-GW exchange is of course subject to some simplifying assumptions that were employed to develop a tractable regional-scale model of the Campaspe system for the data worth analysis. Such simplifications include, but are not limited to, that of

ignoring unsaturated zone flow processes, ignoring representation of overland flow during flooding and mostly non-contributing small tributaries, and the exclusion of a hyper-resolution model grid for solving the governing equations. For example, in the case of ignoring unsaturated flow, it has been shown previously by Brunner et al. (2010) that the violation of the assumption of a hydraulically connected losing-gaining system will lead to underestimation of infiltration of GW. More generally, model simplifications (e.g., 1D steady-state transport, surface flow representation, ignoring unsaturated flow processes, numerical discretization errors, etc.) are likely to lead to uncertainty variance under-estimation (e.g., Knowling et al. 2019b; White et al. 2014); however, our relativistic (i.e., concerning changes in second moments) analysis of worth is expected to be somewhat immune to the impact of such model simplifications. As the Campaspe system becomes better characterised in the future and the model employed herein is refined, exploration of such simplifications could potentially benefit the interpretation of worth.

4.5.4 Choice of data types

This study focused on the worth of particular observational data in the context of SW-GW exchange. It was not exhaustive of all possible data types, and didn't include, e.g. other stream chemistry data, such as stream ^{14}C, dissolved organic carbon (DOC) or total inorganic carbon (TIC), due to the complexity of additional carbon processes required to model these. It is recognised that, e.g., hydrometeorological data such as evapotranspiration and precipitation data, physical stream property measurements, aquifer property measurements and data informing the 3D hydrostratigraphic model including its geometry and internal facies distribution may also be of worth and warrant investigation in future studies.

4.5.5 Choice of FOSM

Our use of FOSM techniques involves consideration of only relative quantities (i.e., changes in uncertainty following parameter conditioning). A number of works have demonstrated its particular robustness in this context (e.g., Dausman et al. 2010; Herckenrath et al. 2011). The computational efficiency of FOSM ultimately allowed for a detailed exploration of a number of different predictions in this study; far more than would have been possible within reasonable time constraints on the basis of a less approximate but more computationally demanding non-linear uncertainty quantification method (Nowak et al. 2012; Wöhling et al. 2018; Wu et al. 2014). However, in highly non-linear models (Herckenrath et al. 2011; Kunstmann et al. 2002), FOSM has been shown to yield similar results to the Monte Carlo type methods. It would nevertheless be of benefit in the future to quantify the impact of the linearity assumption in this case study through a Monte Carlo sampling-type approach.

5 Conclusions

As water resource management and the understanding of river ecosystem functioning both rely on estimation of SW-GW exchange fluxes, meaningful estimation of the exchange flux at appropriate scales must be accompanied by a corresponding quantitative assessment of the exchange flux uncertainty, which, ideally is minimised through smart data collection. The FOSM analysis of spatiotemporally varying SW-GW exchange flux predictions presented in this study provides useful insight into the worth of various hydraulic and water chemistry observation data types in isolation, in various combinations and under individual and comprehensive sampling strategies during low, regular and high streamflow conditions.

The worth of particular data types is dependent on streamflow conditions, the magnitude and direction of the SW-GW exchange flux and the spatiotemporal resolution of the SW-GW

exchange prediction of interest. The unique information in different data types is evidenced by the significant spread of uncertainty reductions across the different predictions. For the spatiotemporal averaging of whole of river annual SW-GW exchange flux predictions, reductions in predictive uncertainty were generally higher than for the finer scale reach and monthly predictions. For the finer scale, the necessity of local scale (time and space) observations is more pertinent for obtaining considerable uncertianty reductions.

With the large variability in worth of varying data types for different specific predictions of SW-GW exchange, we have shown where and when particular data might be of most worth. Hydraulic groundwater head and stream flow were found to inform SW-GW exchange flux best under high- and regular- streamflow conditions. ^{222}Rn and EC were of highest value for low- and regular- streamflow conditions where the stream is gaining.

6 Acknowledgments

This work was funded by the Murray-Darling Basin Authority through the MDBA-NCGRT Strategic Groundwater Research Partnership. The authors thank Grace Lin for the development of figures. All data pre-processing including links to input data, model building, running and post-processing scripts along with key data outputs are provided at https://github.com/daniel-partington/CampaspeModel/tree/amal_speed/CampaspeModel/models/Campaspe_Cascade.
Links for all data used in this study are detailed in the supplementary material, except for surface water extractions and private bore extractions data due to privacy constraints.

728 ## References

729 Arad, A. & R. Evans, 1987. The hydrogeology, hydrochemistry and environmental isotopes of
730 the Campaspe River aquifer system, north-central Victoria, Australia. Journal of
731 Hydrology 95(2):63-86.

732 Bakker, M., V. Post, C. D. Langevin, J. D. Hughes, J. T. White, J. J. Starn & M. N. Fienen, 2016.
733 Scripting MODFLOW Model Development Using Python and FloPy. Groundwater
734 54(5):733-739 doi:10.1111/gwat.12413.

735 Barthold, F. K., C. Tyralla, K. Schneider, K. B. Vaché, H. G. Frede & L. Breuer, 2011. How
736 many tracers do we need for end member mixing analysis (EMMA)? A sensitivity
737 analysis. Water Resources Research 47(8) doi:doi:10.1029/2011WR010604.

738 Bedekar, V., E. D. Morway, C. D. Langevin & M. Tonkin, 2016a. MT3D-USGS version 1.0.0:
739 Groundwater Solute Transport Simulator for MODFLOW: U.S. Geological Survey
740 Software Release, 30 September 2016. USGS.

741 Bedekar, V., E. D. Morway, C. D. Langevin & M. Tonkin, 2016b. MT3D-USGS version 1: A
742 U.S. Geological Survey release of MT3DMS updated with new and expanded transport
743 capabilities for use with MODFLOW: U.S. Geological Survey Techniques and Methods.
744 69.

745 Brunner, P., J. Doherty & C. T. Simmons, 2012. Uncertainty assessment and implications for
746 data acquisition in support of integrated hydrologic models. Water Resources Research
747 48(7):n/a-n/a doi:10.1029/2011WR011342.

748 Brunner, P., C. T. Simmons, P. G. Cook & R. Therrien, 2010. Modeling Surface Water-
749 Groundwater Interaction with MODFLOW: Some Considerations. Ground Water
750 48(2):174-180 doi:10.1111/j.1745-6584.2009.00644.x.

751 Cartwright, I., T. R. Weaver, D. I. Cendón, L. K. Fifield, S. O. Tweed, B. Petrides & I. Swane,
752 2012. Constraining groundwater flow, residence times, inter-aquifer mixing, and aquifer
753 properties using environmental isotopes in the southeast Murray Baasin, Australia.
754 Applied Geochemistry 27(9):1698-1709 doi:10.1016/j.apgeochem.2012.02.006.

755 Cartwright, I., T. R. Weaver & L. K. Fifield, 2006. Cl/Br ratios and environmental isotopes as
756 indicators of recharge variability and groundwater flow: An example from the southeast
757 Murray Basin, Australia. Chemical Geology 231(1):38-56
758 doi:https://doi.org/10.1016/j.chemgeo.2005.12.009.

759 Cartwright, I., T. R. Weaver, C. T. Simmons, L. K. Fifield, C. R. Lawrence, R. Chisari & S.
760 Varley, 2010. Physical hydrogeology and environmental isotopes to constrain the age,
761 origins, and stability of a low-salinity groundwater lens formed by periodic river
762 recharge: Murray Basin, Australia. Journal of Hydrology 380(1):203-221
763 doi:https://doi.org/10.1016/j.jhydrol.2009.11.001.

764 Christensen, S. & J. Doherty, 2008. Predictive error dependencies when using pilot points and
765 singular value decomposition in groundwater model calibration. Advances in Water
766 Resources 31(4):674-700 doi:http://dx.doi.org/10.1016/j.advwatres.2008.01.003.

767 Cook, P. G., 2013. Estimating groundwater discharge to rivers from river chemistry surveys.
768 Hydrol Process 27(25):3694-3707 doi:doi:10.1002/hyp.9493.

769 Cook, P. G., 2015. Quantifying river gain and loss at regional scales. Journal of Hydrology 531,
770 Part 3:749-758 doi:https://doi.org/10.1016/j.jhydrol.2015.10.052.

771 Cook, P. G., S. Lamontagne, D. Berhane & J. F. Clark, 2006. Quantifying groundwater discharge
772 to Cockburn River, southeastern Australia, using dissolved gas tracers 222Rn and SF6.
773 Water Resources Research 42(10):n/a-n/a doi:10.1029/2006WR004921.

774 Dausman, A. M., J. Doherty, C. D. Langevin & M. C. Sukop, 2010. Quantifying Data Worth
775 Toward Reducing Predictive Uncertainty. Ground Water 48(5):729-740
776 doi:10.1111/j.1745-6584.2010.00679.x.

777 Di Baldassarre, G. & A. Montanari, 2009. Uncertainty in river discharge observations: a
778 quantitative analysis. Hydrol Earth Syst Sci 13(6):913-921 doi:10.5194/hess-13-913-
779 2009.

780 Doherty, J., 2015. Calibration and Uncertainty Analysis for Complex Environmental Models.
781 PEST: complete theory and what it means for modelling the real world. Watermark
782 Numerical Computing.

783 Doherty, J. E., 2016. PEST, Model-Independent Parameter Estimation User Manual Part I:
784 PEST, SENSAN and Global Optimisers.

785 Ellins, K. K., A. Roman-Mas & R. Lee, 1990. Using 222Rn to examine groundwater/surface
786 discharge interaction in the Rio Grande de Manati, Puerto Rico. Journal of Hydrology
787 115(1):319-341 doi:https://doi.org/10.1016/0022-1694(90)90212-G.

788 Fienen, M. N., J. E. Doherty, R. J. Hunt & H. W. Reeves, 2010. Using prediction uncertainty
789 analysis to design hydrologic monitoring networks: Example applications from the Great
790 Lakes water availability pilot project Scientific Investigations Report. - edn.

791 Fleckenstein, J. H., S. Krause, D. M. Hannah & F. Boano, 2010. Groundwater-surface water
792 interactions: New methods and models to improve understanding of processes and
793 dynamics. Advances in Water Resources 33(11):1291-1295
794 doi:http://dx.doi.org/10.1016/j.advwatres.2010.09.011.

795 Herckenrath, D., C. D. Langevin & J. Doherty, 2011. Predictive uncertainty analysis of a
796 saltwater intrusion model using null-space Monte Carlo. Water Resources Research 47(5)
797 doi:10.1029/2010wr009342.

798 Hunt, R. J., J. Doherty & M. J. Tonkin, 2007. Are Models Too Simple? Arguments for Increased
799 Parameterization. Ground Water 45(3):254-262 doi:10.1111/j.1745-6584.2007.00316.x.

800 Kalbus, E., F. Reinstorf & M. Schirmer, 2006. Measuring methods for groundwater –
801 surface water interactions: a review. Hydrol Earth Syst Sci 10(6):873-887
802 doi:10.5194/hess-10-873-2006.

803 Kikuchi, C., 2017. Toward Increased Use of Data Worth Analyses in Groundwater Studies.
804 Groundwater 55(5):670-673 doi:10.1111/gwat.12562.

Knowling, M. J., J. T. White & C. R. Moore, 2019a. Role of model parameterization in risk-based decision support: An empirical exploration. Advances in Water Resources 128:59-73 doi:https://doi.org/10.1016/j.advwatres.2019.04.010.

Knowling, M. J., J. T. White, C. R. Moore, P. Rakowski & K. Hayley, 2019b. On the assimilation of environmental tracer observations for model-based decision support. Hydrol Earth Syst Sci Discuss 2019:1-19 doi:10.5194/hess-2019-436.

Kunstmann, H., W. Kinzelbach & T. Siegfried, 2002. Conditional first-order second-moment method and its application to the quantification of uncertainty in groundwater modeling. Water Resources Research 38(4):6-1-6-14 doi:10.1029/2000wr000022.

Moore, C. & J. Doherty, 2005. Role of the calibration process in reducing model predictive error. Water Resources Research 41(5):n/a-n/a doi:10.1029/2004WR003501.

Niswonger, R. G., S. Panday & M. Ibaraki, 2011. MODFLOW-NWT, A Newton formulation for MODFLOW-2005: U.S. Geological Survey Techniques and Methods. 44.

Niswonger, R. G. & D. E. Prudic, 2005. Documentation of the Streamflow-Routing (SFR2) Package to include unsaturated flow beneath streams—A modification to SFR1: U.S. Geological Survey Techniques and

Methods 6-A13, 50 p.

Nowak, W., Y. Rubin & F. P. J. de Barros, 2012. A hypothesis-driven approach to optimize field campaigns. Water Resources Research 48(6):n/a-n/a doi:10.1029/2011WR011016.

Schilling, O. S., J. Doherty, W. Kinzelbach, H. Wang, P. N. Yang & P. Brunner, 2014. Using tree ring data as a proxy for transpiration to reduce predictive uncertainty of a model simulating groundwater–surface water–vegetation interactions. Journal of Hydrology 519, Part B:2258-2271 doi:http://dx.doi.org/10.1016/j.jhydrol.2014.08.063.

Schilling, O. S., C. Gerber, D. J. Partington, R. Purtschert, M. S. Brennwald, R. Kipfer, D. Hunkeler & P. Brunner, 2018. Advancing Physically-Based Flow Simulations of Alluvial Systems Through Atmospheric Noble Gases and the Novel 37Ar Tracer Method. Water Resources Research:n/a-n/a doi:10.1002/2017WR020754.

Tarantola, A., 2005. Inverse problem theory and methods for model parameter estimation. SIAM.

Wallis, I., C. Moore, V. Post, L. Wolf, E. Martens & H. Prommer, 2014. Using predictive uncertainty analysis to optimise tracer test design and data acquisition. Journal of Hydrology 515:191-204 doi:http://dx.doi.org/10.1016/j.jhydrol.2014.04.061.

White, J. T., J. E. Doherty & J. D. Hughes, 2014. Quantifying the predictive consequences of model error with linear subspace analysis. Water Resources Research 50(2):1152-1173 doi:10.1002/2013WR014767.

White, J. T., M. N. Fienen & J. E. Doherty, 2016. A python framework for environmental model uncertainty analysis. Environmental Modelling & Software 85:217-228 doi:http://dx.doi.org/10.1016/j.envsoft.2016.08.017.

Wöhling, T., M. J. Gosses, S. R. Wilson & P. Davidson, 2018. Quantifying River-Groundwater Interactions of New Zealand's Gravel-Bed Rivers: The Wairau Plain. Groundwater:n/a-n/a doi:10.1111/gwat.12625.

Wu, B., Y. Zheng, Y. Tian, X. Wu, Y. Yao, F. Han, J. Liu & C. Zheng, 2014. Systematic assessment of the uncertainty in integrated surface water-groundwater modeling based on the probabilistic collocation method. Water Resources Research 50(7):5848-5865 doi:10.1002/2014WR015366.

Zell, W. O., T. B. Culver & W. E. Sanford, 2018. Prediction uncertainty and data worth assessment for groundwater transport times in an agricultural catchment. Journal of Hydrology doi:https://doi.org/10.1016/j.jhydrol.2018.02.006.

Appendix A: 1D steady-state transport model for ^{222}Rn and EC

The 1D steady-state model presented in Cook et al. (2006) for determining SW-GW flows with ^{222}Rn observation data, used a mass balance of streamflow and radon similar to the following, but here we have added in inflows from tributaries and precipitation:

$$\frac{\partial Q(x)}{\partial x} = I_{GW}(x) - O_{GW}(x) + I_{TR}(x) - L(x) + (P(x) - E(x))w(x) \quad \text{(A.1)}$$

$$\frac{\partial Q(x).c_S}{\partial x} = I_{GW}(x).c_{GW} - O_{GW}(x).c_S + I_{TR}(x).c_{TR} - L(x).c_S + P(x).w(x).c_R - k.w(x).c_S - \lambda.d.w(x).c_S + F \quad \text{(A.2)}$$

Where Q is the rate of streamflow [L^3/T], x is the distance along the stream [L]. Varying along the x-axis, $I_{GW}(x)$, $O_{GW}(x)$, $I_{TR}(x)$ and $L(x)$ are the groundwater inflow rates, stream losing rates, tributary and diversion inflows, and losses through direct pumping [L^3/L/T]. $E(x)$ is the evaporation rate [L/T] $P(x)$ is the precipitation rate [L/T], and $w(x)$ is the width of the river [L] along x. c_S, c_{GW}, c_{TR}, and c_R are the concentrations [M/L^3] of the stream, groundwater, tributaries and rainfall respectively. k is the gas transfer velocity across the water surface [L/T], λ is the radioactive decay constant [T^{-1}], d is the mean stream depth [L] and F is the flux of radon through the hyporheic zone [M/L/T].

$$F = \frac{\gamma h w \theta}{1+\lambda t_h} - \frac{\lambda h w \theta}{1+\lambda t_h} c_S \quad \text{(A.3)}$$

Where γ is the production rate within the hyporheic zone [M/L^3/T], h is the mean depth of the hyporheic zone [L], and θ is its porosity [-]. t_h is the mean residence time of water within the hyporheic zone. Expanding the partial derivative in (2) and substituting in (1) with the chain rule and rearranging yields:

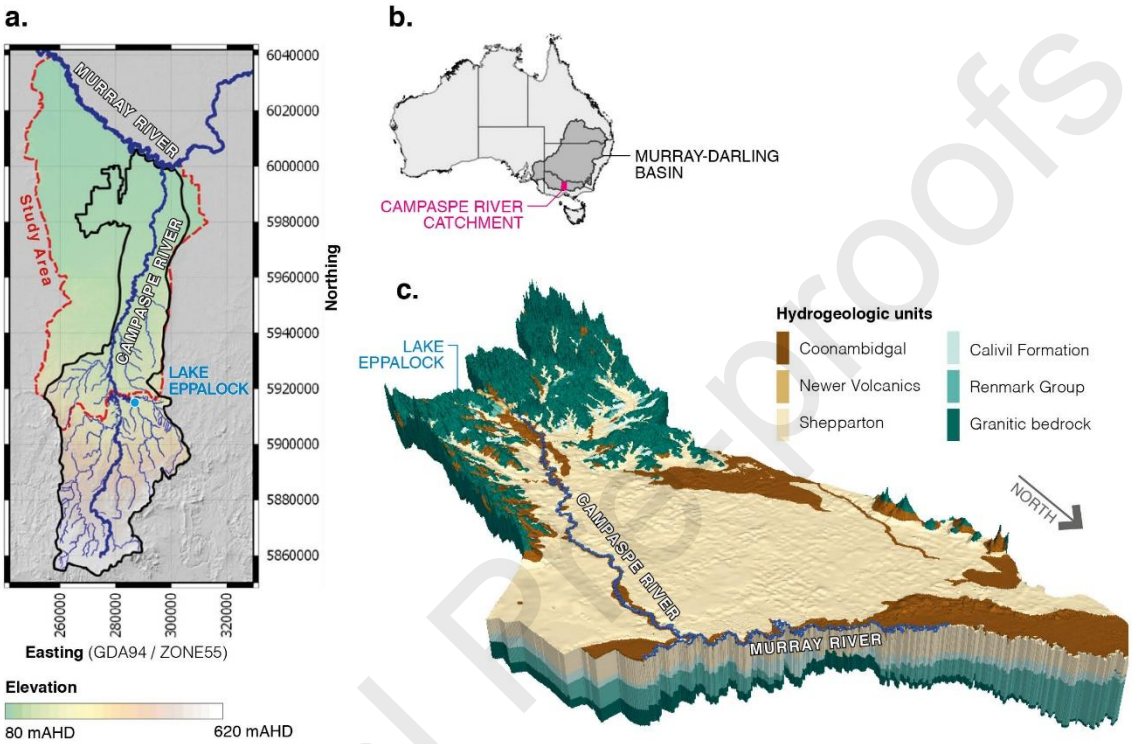
$$Q(x)\frac{\partial c_S}{\partial x} = I_{GW}(x).(c_{GW} - c_S) + I_{TR}(x).(c_{TR} - c_S) + P(x).w(x).(c_R - c_S) + E(x).w(x).c_S - k.w(x).c_S - \lambda.d.w(x).c_S + F \quad \text{(A.4)}$$

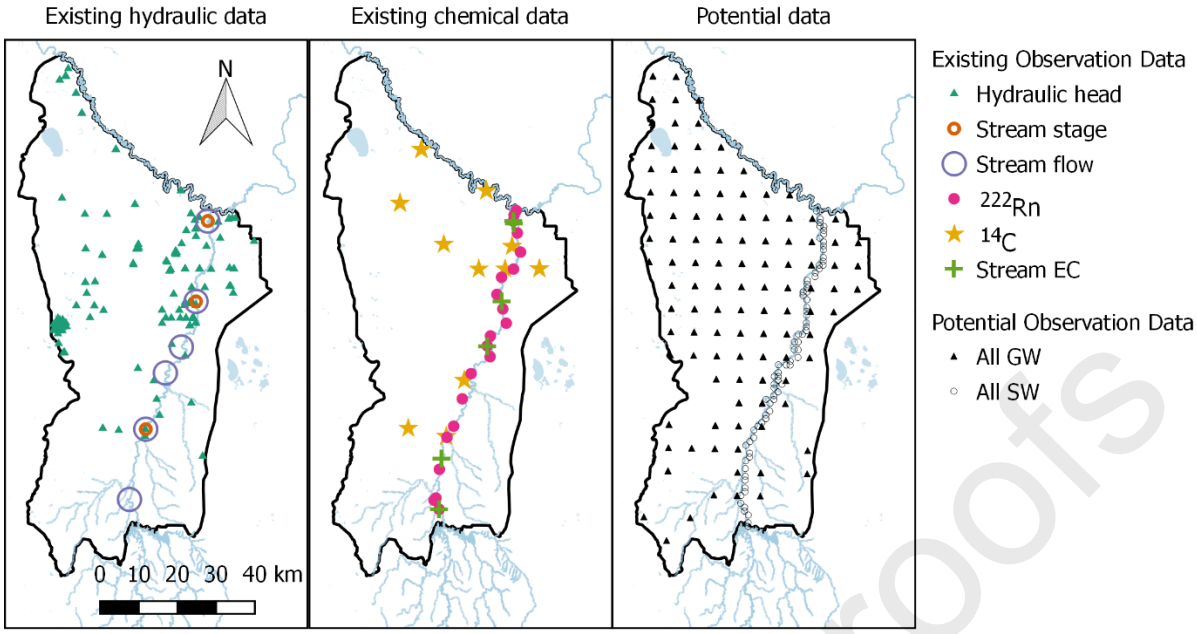
For a given stream length *dx* with a boundary upstream concentration of c_S we can solve for the downstream concentration $c_{S/DS}$ with:

$$^{222}\boldsymbol{Rn}: c_{S/DS} = c_S - \frac{dx}{Q(x)}(I_{GW}(x).(c_{GW} - c_s) + I_{TR}(x).(c_{TR} - c_s) + P(x).w(x).(c_R - c_s) + E(x).w(x).c_s - k.w(x).c_s - \lambda.d.w(x).c_s + F) \quad \text{(A.5)}$$

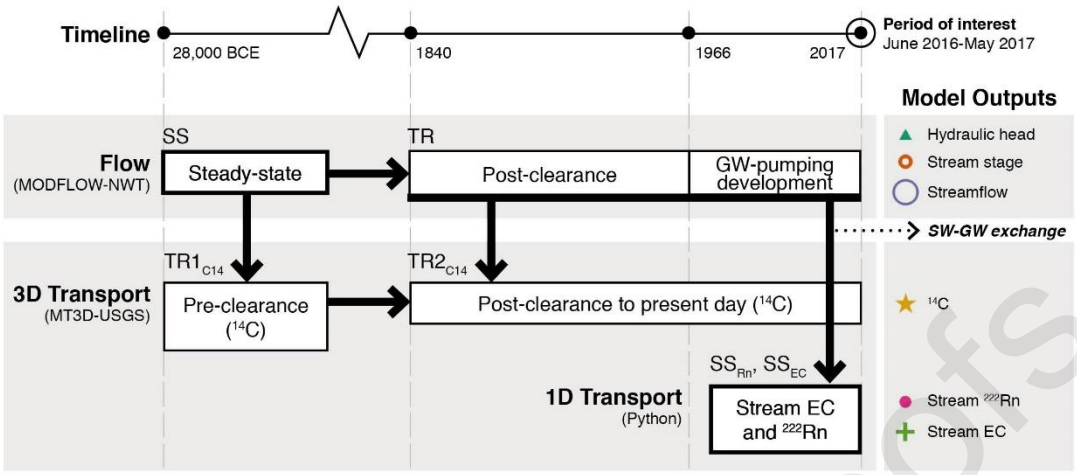
Noting the conservative nature of EC (we have no degassing, decay or production in the hyporheic zone) we can write a similar expression for EC as:

$$\boldsymbol{EC}: c_{S/DS} = c_S - \frac{dx}{Q(x)}(I_{GW}(x).(c_{GW} - c_s) + I_{TR}(x).(c_{TR} - c_s) + P(x).w(x).(c_R - c_s) + E(x).w(x).c_s) \quad \text{(A.6)}$$

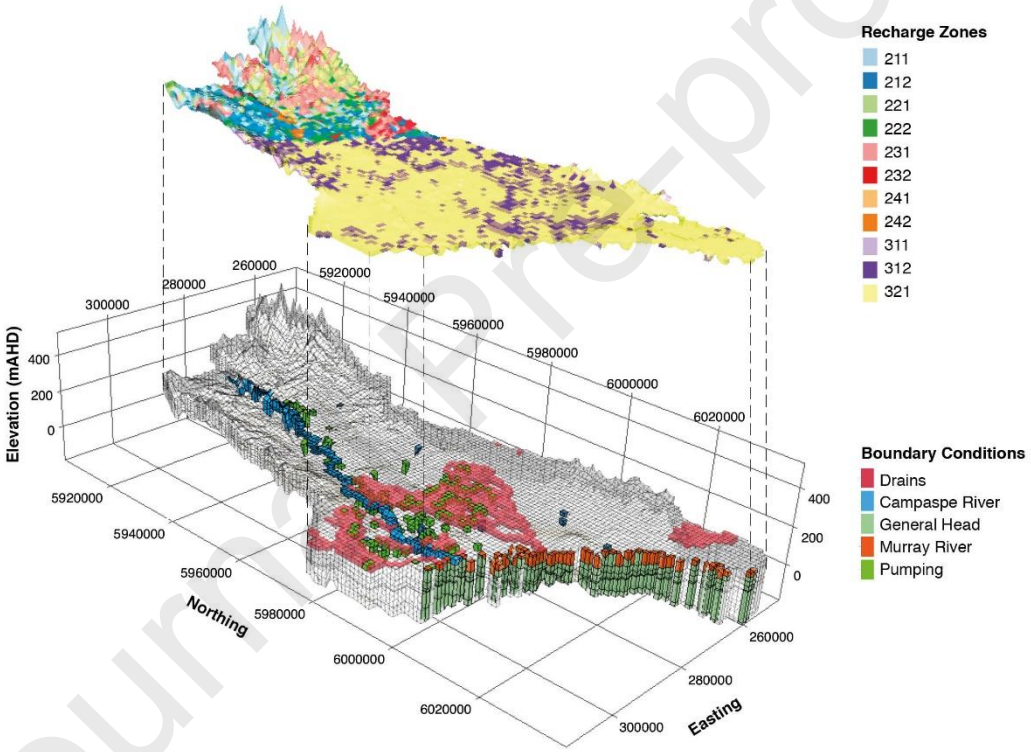


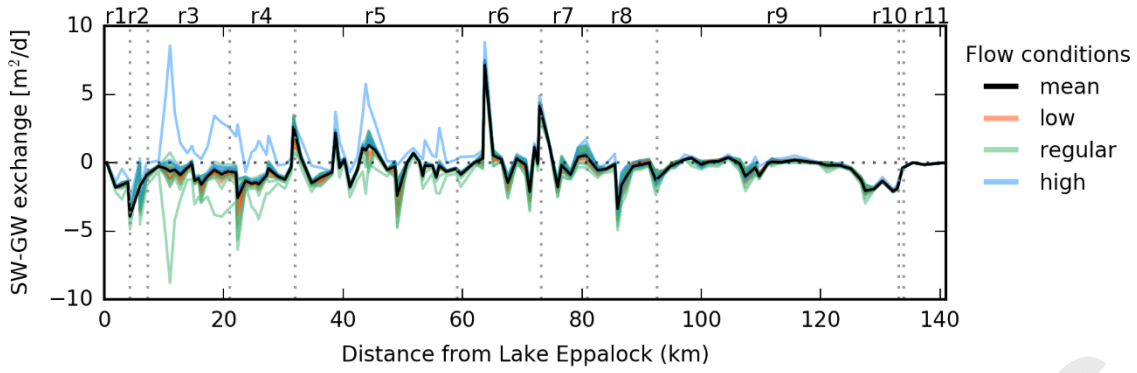


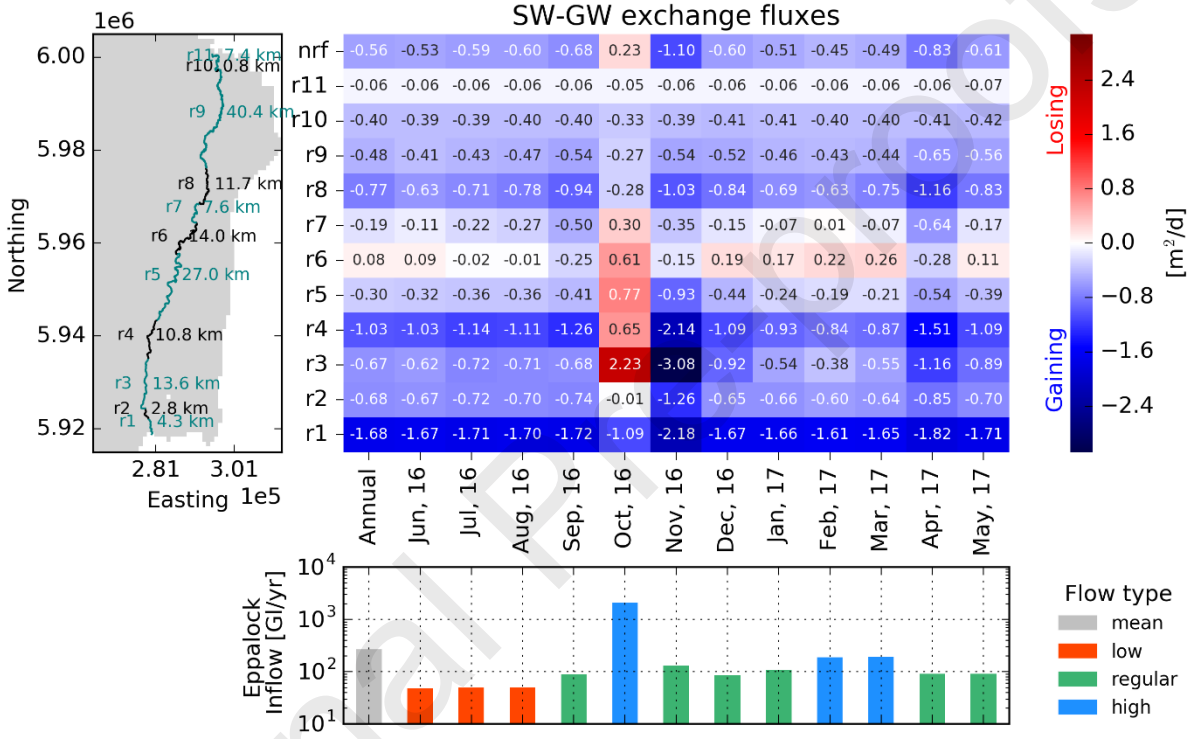
a.

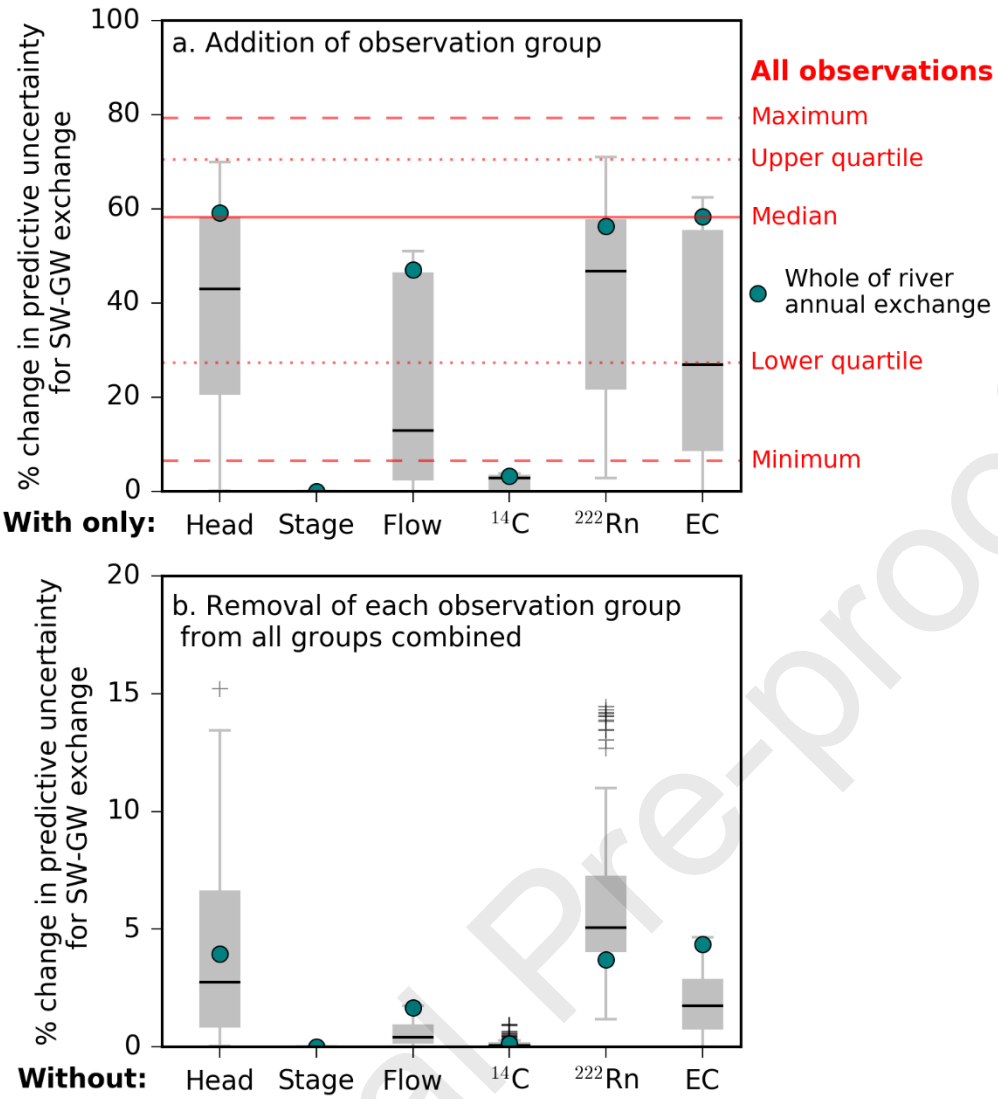


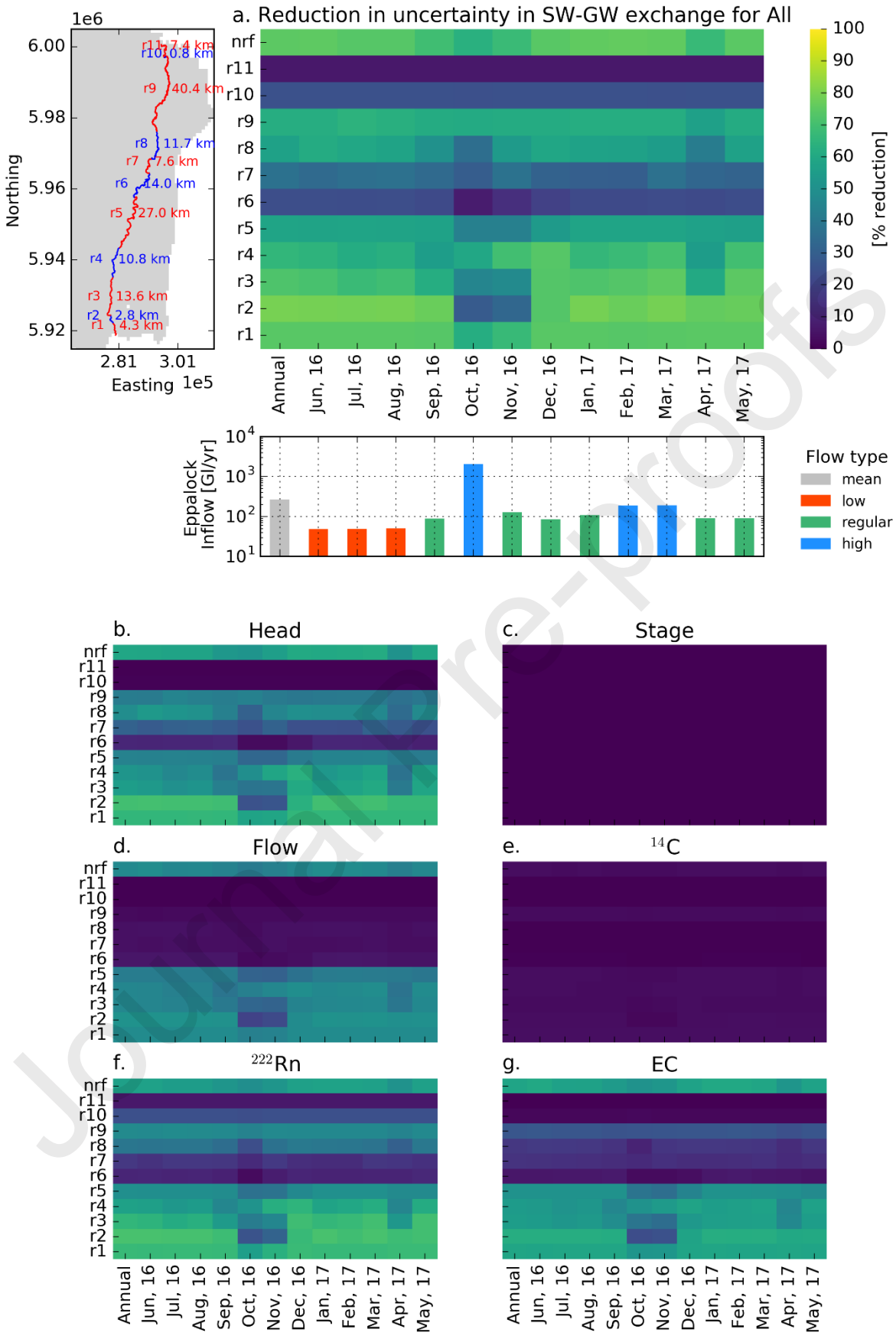
b.

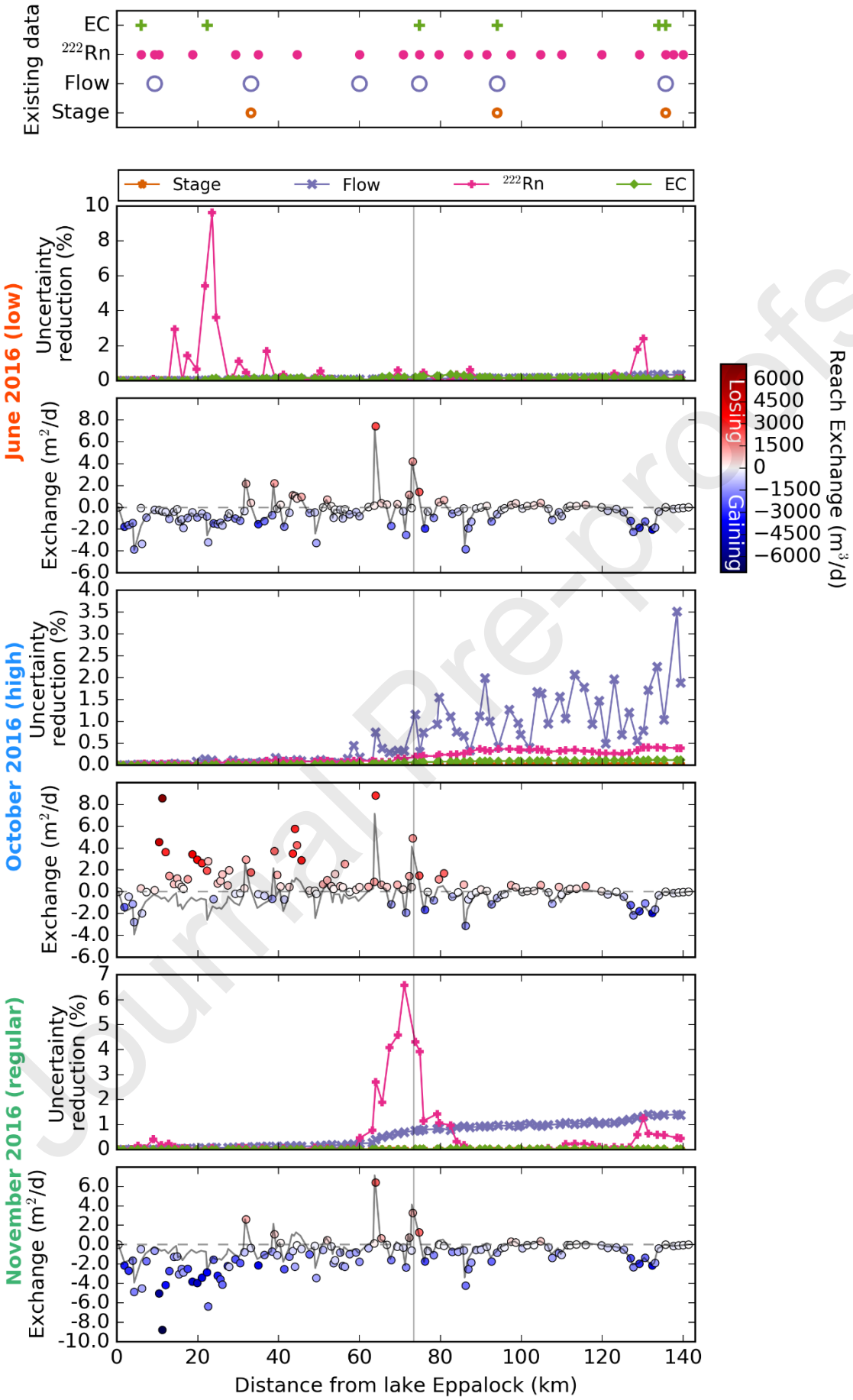


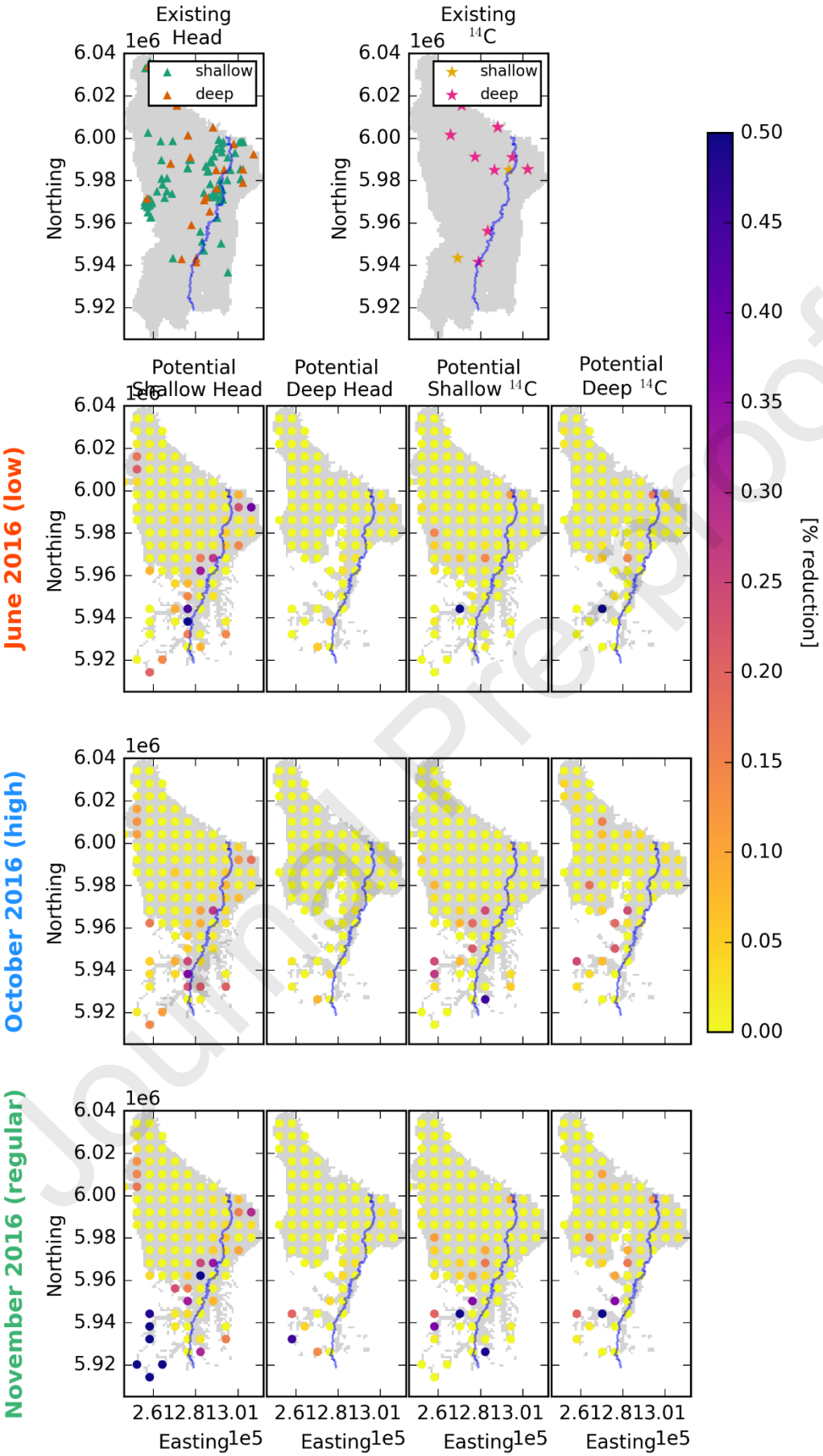


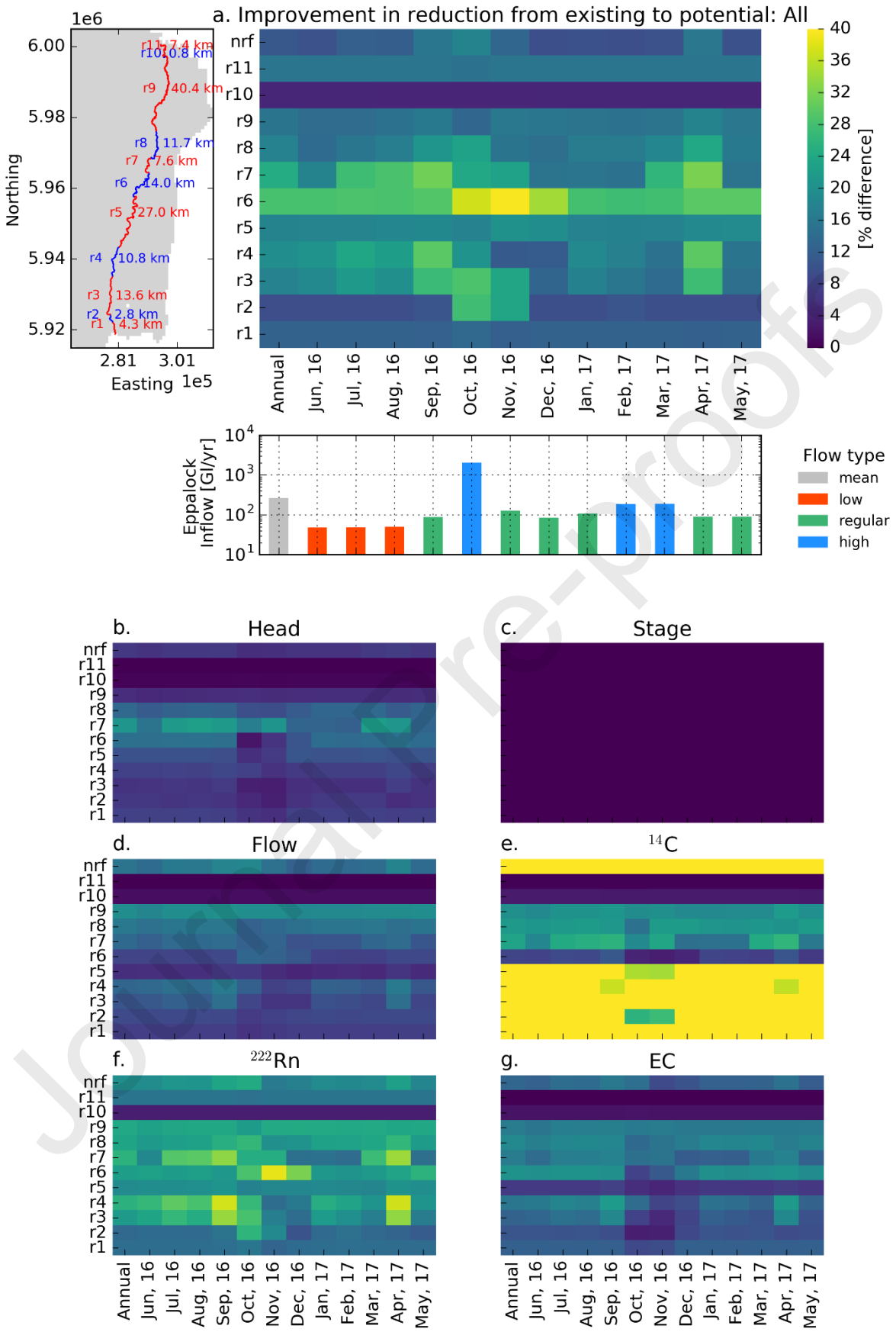


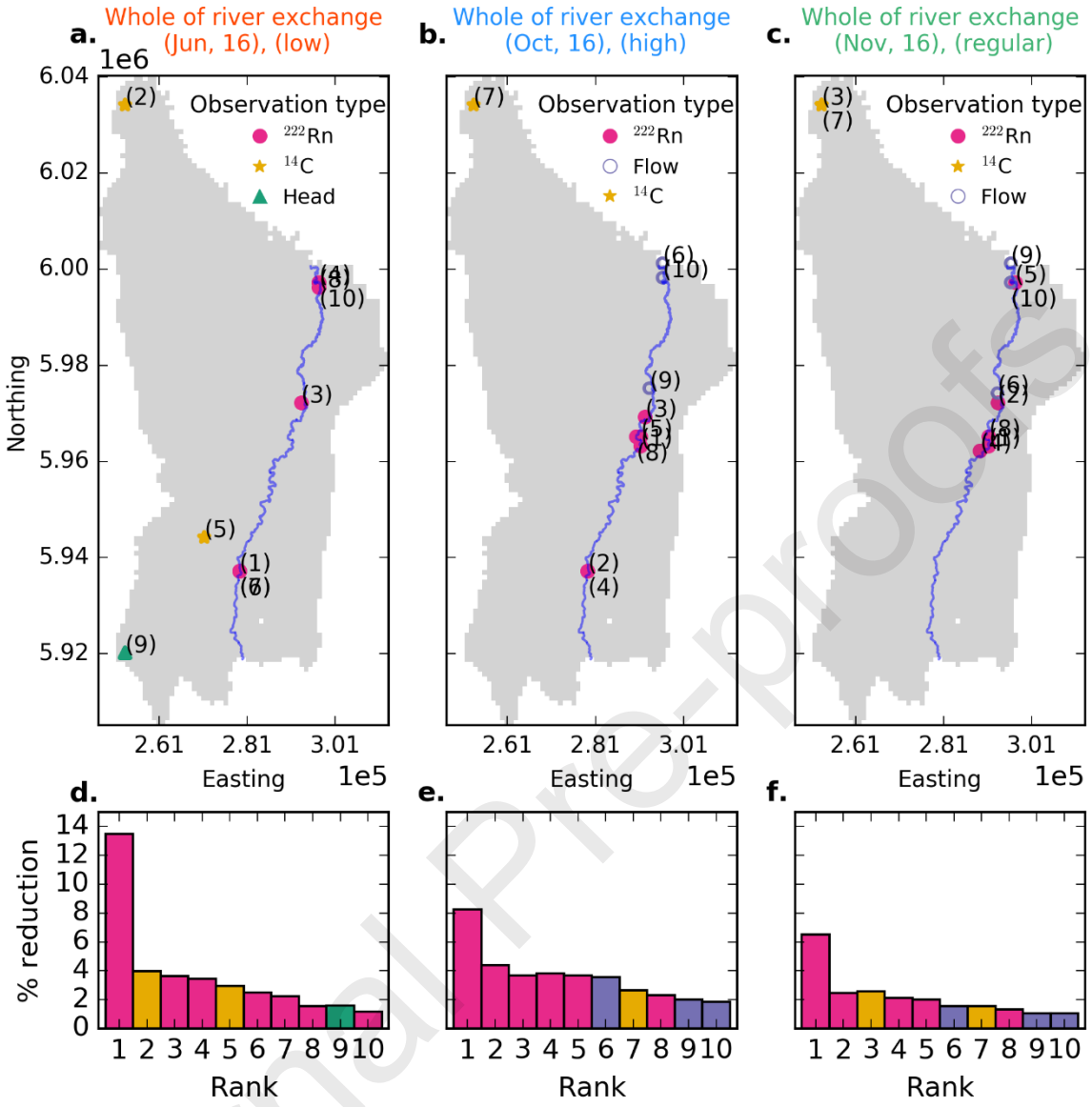












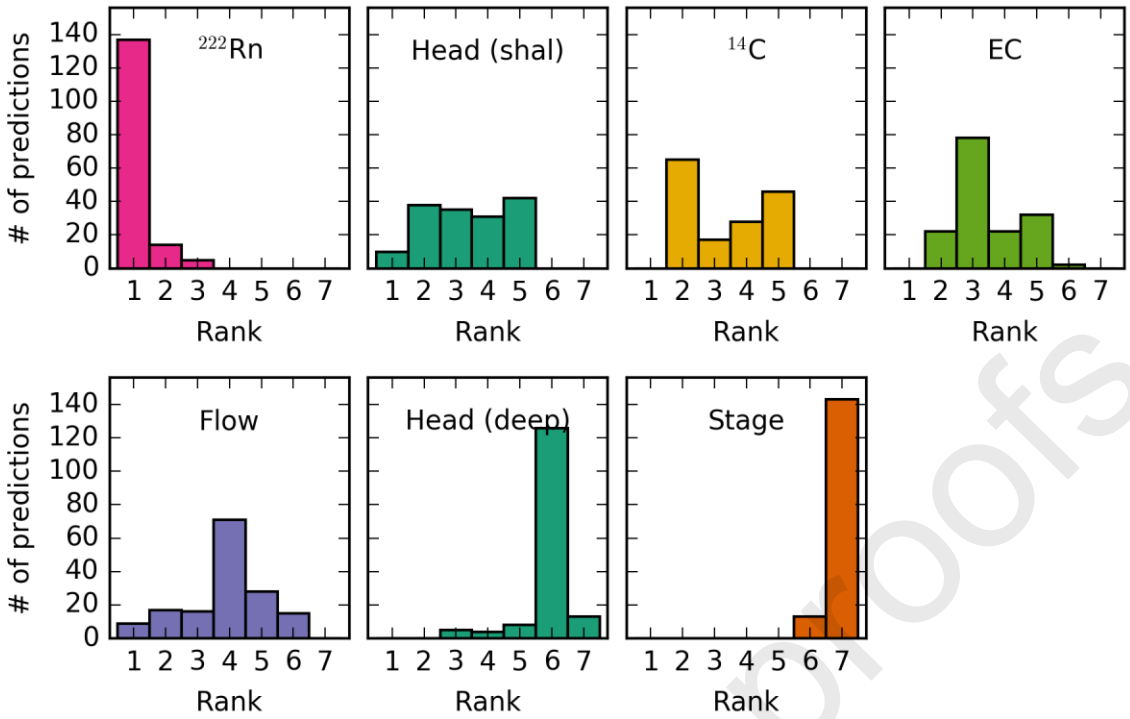


Table 1. Summary of hydrogeological layer thicknesses including, mean, minimum, maximum thickness and percent volume. Hydrogeological units are abbreviated as Coonambidgal (co), Shepparton (sh), Calivil (ca), Renmark (re), Newer Volcanics (nv), Basement (ba).

Hydrogeological Unit	Thickness			Percentage of active model volume (%)
	Mean (m)	Minimum (m)	Maximum (m)	
Coonambidgal	5.9	1.0	31.8	0.67
Shepparton	52.3	0.3	122.3	26.58
Calivil	26.1	1.0	82.7	9.76
Renmark	40.4	0.4	163.4	11.90
Newer Volcanics	4.6	0.1	14.3	0.05
Basement	74.4	9.7	109.5	50.93

Table 2. Summary of model parameterization. For the aquifer property parameters, the hydrogeological units are abbreviated as follows: Coonambidgal (co), Shepparton (sh), Calivil (ca), Renmark (re), Newer Volcanics (nv), Basement (ba).

Parameter set	Model	Type	Number	Initial Range	Log Variance (σ_θ^2)
Aquifer properties (7 HGUs)					
Horizontal hydraulic conductivity (K_h)	SS, TR	Pilot points (PP) (416)	co: 31 sh: 107 ca: 77 re: 58 nv: 4 ba: 137	co: 4.44 – 444 m/d sh: 0.1 – 10 m/d ca: 4.25 – 425 m/d re: 6.75 – 675 m/d nv: 5 – 500 m/d ba: 0.1 – 10 m/d	0.25 $(m/d)^2$ 0.25 $(m/d)^2$ 0.25 $(m/d)^2$ 0.25 $(m/d)^2$ 0.25 $(m/d)^2$ 0.25 $(m/d)^2$
Vertical hydraulic conductivity	SS, TR	-	1	-	-
Specific storage	SS, TR	PP (416)	see K_h	1E-6 – 1E-4 (-)	0.25 (-)
Specific yield	SS, TR	PP (416)	see K_h	1E-2 – 0.44 (-)	0.53 (-)
Porosity (7 HGUs)	$TR1_{C14}$, $TR2_{C14}$	Zonal	7	0.05 - 0.4 (-)	5.1E-2 (-)
Dispersivity (7 HGUs)	$TR1_{C14}$, $TR2_{C14}$	Zonal	7	1E-5 – 100 m	3.06 m^2
Recharge					
Pre-clearance rainfall Proportion for recharge	TR	Zonal	16	1E-3 – 0.5 (-)	0.46 (-)
Rainfall proportion for recharge	SS	Zonal	16	1E-4 – 0.9 (-)	2E-3 - 0.55(-)
GW boundary under Murray					
GHB level adjuster	SS, TR	-	1	0.01 – 70.0 m	9.9E-3 m^2
GHB conductivity	SS, TR	-	1	1E-8 – 50.0 m/d	5.88 $(m/d)^2$
Murray River properties					
Murray River streambed K	SS, TR	-	1	1E-8 – 20.0 m/d	5.41$(m/d)^2$
Drain properties					
Drain bed K	TR	-	1	1E-8 – 20.0 m/d	5.41$(m/d)^2$
Drain bed adjust	TR	-	1	0.001 – 0.1 m	0.25 m^2
Campaspe River properties					
Stream dispersivity	$TR1_{C14}$, $TR2_{C14}$	-	1	1E-5 – 1000.0 m	4.0 m^2
Riverbed hydraulic cond.	SS, TR	PP	80	1E-4 – 10.0 m/d	1.56 $(m/d)^2$
River width	SS, TR	PP	80	4.0 – 40.0 m	6.25E-2 m^2
Riverbed roughness	SS, TR	PP	80	0.001 – 0.1 $d/m^{1/3}$	0.25 $(d/m^{1/3})^2$
GW EC concentration	SS_{EC}	-	1	1E3 – 5E3 µS/cm	0.03 $(\mu S/cm)^2$
GW ^{222}Rn concentration	SS_{Rn}	-	1	1E4 – 5E4 mBq/l	0.03 $(mBq/l)^2$
Gas transfer velocity	SS_{Rn}	-	1	0.6 – 1.4 m/s	8.46E-3 $(m/s)^2$
Hyporheic zone porosity	SS_{Rn}	-	1	0.1 – 0.4 (-)	2.27E-2 (-)
Hyporheic zone production	SS_{Rn}	-	1	1E3– 1E4 mBq/l/d	0.0625 $(mBq/l/d)^2$
Hyporheic zone residence time	SS_{Rn}	-	1	0.05 – 5.0 d	0.25 d^2
Hyporheic zone depth	SS_{Rn}	-	1	0.0 - 1.0 m	0.26 m^2
TOTAL			**1546**		

Table 3. Assumed observation noise (used together with model-to-measurement residual information to populate Σ_ϵ).

Observation data type	Observation noise (σ_ϵ)
Hydraulic head	1 m
Stream stage	1 m
Streamflow	40% of observation value [m^3/d]
^{222}Rn	50 mBq/l
^{14}C	10 PMC
Electrical conductivity	100 µS/cm

Table 4. Data worth assessments employed for the range of SW-GW exchange predictions from the Campaspe transient flow model (TR).

	$\overline{\sigma}_s$	Combinations	Predictions
Q1	Adding individual data types with existing data	6	All (156)
	Removing individual data types with existing data	6	All (156)
Q2	Added benefit of each potential observation for annual whole river exchange: SW observations	316	Whole river exchange (4)
	Added benefit of each potential observation for annual whole river exchange: GW observations	376	Whole river exchange (4)
	Next best observation group (potential observations)	7	All (156)
	Next best observation (potential observations) for low, regular and high streamflow conditions	10	Whole river exchange (4)

Daniel Partington: Conceptualisation, methodology, software, validation, formal analysis, investigation, data curation, writing – original draft, writing – review and editing, visualisation

Matthew J. Knowling: Conceptualisation, methodology, formal analysis, writing – original draft, writing – review and editing

Craig T. Simmons: Conceptualization, supervision, project administration, funding acquisition, writing – review and editing

Peter G. Cook: Conceptualization, supervision, project administration, funding acquisition, writing – review and editing

Yueqing Xie: Conceptualization

Takuya Iwanaga: Conceptualization, methodology, software, data curation, validation, writing – review and editing

Camille Bouchez: Conceptualization, writing – review and editing

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Highlights

- Worth of data depend on spatiotemporal scale, flow and exchange conditions
- Hydraulic data informs SW-GW exchange flux best under high/regular flow conditions
- Radon-222 and EC data hold value for low/regular flow conditions where gaining