

# A hydrochemically guided landscape classification system for modelling spatial variation in multiple water quality indices: Process-attribute mapping

C.W.F. Rissmann<sup>a,b,\*</sup>, L.K. Pearson<sup>a</sup>, M. Beyer<sup>a</sup>, M.A. Couldrey<sup>a</sup>, J.L. Lindsay<sup>a</sup>, A.P. Martin<sup>c</sup>, W.T. Baisden<sup>d</sup>, T.J. Clough<sup>e</sup>, T.W. Horton<sup>f</sup>, J.G. Webster-Brown<sup>b</sup>

<sup>a</sup> Land and Water Science, Invercargill, New Zealand

<sup>b</sup> Waterways Centre for Freshwater Management, University of Canterbury and Lincoln University, Christchurch, New Zealand

<sup>c</sup> GNS Science, Dunedin, New Zealand

<sup>d</sup> University of Waikato, Hamilton, New Zealand

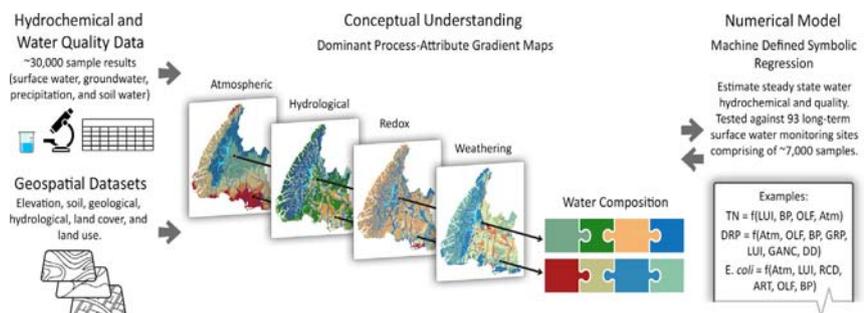
<sup>e</sup> Lincoln University, Lincoln, New Zealand

<sup>f</sup> University of Canterbury, Christchurch, New Zealand

## HIGHLIGHTS

- Highly accurate and transparent water quality modelling framework presented
- Landscape driven hydrological, redox, and chemical weathering gradients mapped
- Hydrochemical signatures of dominant processes used to guide classification
- Coupled with land use, process-attribute gradients explain water quality variation.
- New approach to water quality modelling proposed

## GRAPHICAL ABSTRACT



## ARTICLE INFO

### Article history:

Received 23 January 2019

Received in revised form 27 March 2019

Accepted 31 March 2019

Available online 02 April 2019

Editor: Jay Gan

### Keywords:

Dominant processes  
Landscape gradients  
Physiographic  
Redox  
Hydrology  
Chemical weathering

## ABSTRACT

Spatial variation in landscape attributes can account for much of the variability in water quality relative to land use on its own. Such variation results from the coupling between the dominant *processes* governing water quality, namely hydrological, redox, and weathering and gradients in key landscape *attributes*, such as topography, geology, and soil drainage. Despite the importance of ‘*process-attribute*’ gradients (PAG), few water quality models explicitly account for their influence. Here a processes-based water quality modelling framework is presented that more completely accounts for the role of landscape variability over water quality – Process-Attribute Mapping (PoAM). Critically, hydrochemical measures form the basis for the identification and mapping of effective landscape attributes, producing PAG maps that attempt to replicate the natural landscape gradients governing each dominant process. Application to the province of Southland (31,824 km<sup>2</sup>), New Zealand, utilised 12 existing geospatial datasets and a total of 28,626 surface water, groundwater, spring, soil water, and precipitation analyses to guide the identification and mapping of 11 individual PAG. The ability of PAGs to replicate regional hydrological, redox, and weathering gradients was assessed on the accuracy with which the hydrochemical indicators of each dominant process (e.g. hydrological tracers, redox indicators) were estimated across 93 long-term surface water monitoring sites (cross-validated R<sup>2</sup> values of 0.75–0.95). Given hydrochemical evidence that PAGs replicate actual landscape gradients governing the dominant processes, they were combined with a land use intensity

\* Corresponding author at: Land and Water Science, Invercargill, New Zealand.  
E-mail address: [clint@landwatersci.net](mailto:clint@landwatersci.net) (C.W.F. Rissmann).

layer and used to estimate steady-state surface water quality. Cross-validated  $R^2$  values ranged between 0.81 and 0.92 for median total nitrogen, total oxidised nitrogen, total phosphorus and dissolved reactive phosphorus. Models of particulate species *E. coli* and total suspended sediment, although reasonable ( $R^2$  0.72–0.73), were less accurate, suggesting finer-grained land use, landscape attribute, and/or flow normalised measures are required to improve estimation.

© 2019 Published by Elsevier B.V.

## 1. Introduction

Water quality can vary spatially across the landscape, even when there are similar environmental pressures. These differences occur because of natural spatial variation in landscape attributes (e.g. soil and geology), which alter water composition through coupled physical, chemical, and biological processes (Wright, 1988; Moldan and Černý, 1994; Clark and Fritz, 1997; Giller and Malmqvist, 1998; Kendall and McDonnell, 1998; Krantz and Powars, 2000; Lydersen et al., 2004; Doctor et al., 2008; McMahon and Chapelle, 2008; Gray et al., 2011; Inamdar, 2011; Rissmann, 2011; Rissmann et al., 2015). Previous research has shown that spatial variation in landscape attributes may account for the majority of the variability in water quality relative to land use on its own (Johnson et al., 1997; Thomas et al., 1999; Hale et al., 2004; King et al., 2005; Kratzer et al., 2006; Becker et al., 2014).

Despite the importance of landscape attributes over variation in water quality, there have been few attempts to explicitly account for their influence within most water quality models. As a result, most water quality modelling approaches fail to incorporate the effective properties of the landscape at relevant scales, seldom consider the important role of groundwater over surface water quality, are often restricted to the estimation of a small number of contaminants (e.g.  $\text{NO}_3^-$ ), may lack transparency (e.g. 'black-box' models and the approximation of decay constants to represent attenuation), and are typically of limited accuracy. For these reasons, water quality models are often not relevant to land users and their communities, with a growing number of land-based practitioners, policy makers, and researchers calling for a shift towards a more transparent and process-based understanding of environmental systems (Grayson and Bloschl, 2000; Sivakumar, 2004, 2008; National Research Council, 2007; Tsakiris and Alexakis, 2012; Bracken et al., 2013; Parliamentary Commissioner for the Environment, 2018).

One approach to overcoming such limitations has been the application of landscape based controlling factor classifications to estimate spatial variability in water quality or ecosystem health (Krantz and Powars, 2000; Snelder and Biggs, 2002; Snelder et al., 2004; Scott et al., 2006; Hume et al., 2007; Dymond, 2010). Landscape based classifications recognise an organisational hierarchy of 'controlling landscape factors,' e.g. the role of attributes such as topography, soil type, and rock type over spatial variation in water quality. These classifications rank the landscape according to the similarity of controlling factors via a top-down approach that results in a mosaic of geographically independent categories (Snelder and Biggs, 2002; Hume et al., 2007; Rissmann, 2011). The approaches start with the development of a basic thesis or model of the 'assumed' causes of spatial variation to provide a pragmatic simplification of reality with the benefit of not being constrained by nil or limited measurement data (Snelder and Biggs, 2002; Krantz and Powars, 2000).

Despite evidence for strong correlations between assumed controlling factors and hydrochemical signatures at large scales (>100–1000 km<sup>2</sup>), the assumption of spatial linkages between controlling factors and governing processes at finer scales may not always hold true (Troy et al., 2008; Matott et al., 2009; Bracken et al., 2013). This can be due to:

- (i) a lack of process-level understanding of the relationships between landscape attributes and resultant processes (e.g. model structure), and;

- (ii) the use of geospatial datasets typically created for purposes other than water quality, which may lack, or fail to represent at adequate scales, the 'effective' properties of the landscape that are most critical.

For example, if the tendency of a local soil series to 'crack' in response to soil moisture deficit is not recognised as an important controlling factor, or if the soil survey being used for controlling factor classification lacks representation of cracking soils, significant misalignment in the resultant hydrochemical and water quality signatures may occur (Jamieson et al., 2002; Beven and Germann, 2013; Beyer et al., 2016a; Kurtzman et al., 2016; Hughes et al., 2016). In situations such as these, the controlling factor relationship breaks down resulting in a loss of resolution at the scale of the effective property (e.g. cracking soil units), although the classification may still work at larger scales (Rissmann et al., 2016a).

To incorporate the most important, hereafter 'effective,' landscape attributes and partially overcome issues of scale, we propose a new, hybrid controlling factor classification that incorporates bottom-up measurements of hydrochemical signatures within the classification target (i.e. surface water and hydrologically connected aquifers), to provide a point of reference for testing and refining the hypotheses that underpin the top-down controlling factor classification of the landscape for water quality. We term this hybrid controlling factor approach Process-Attribute Mapping (PoAM) to reflect the use of hydrochemical *process* signals within the classification target (i.e. water) to guide the identification and subsequent classification of the effective *attributes* of the landscape. The PoAM approach is based on the following key assumptions:

- (i) There are three dominant processes that control the hydrochemical evolution of water (Moldan and Černý, 1994; Clark and Fritz, 1997; Langmuir, 1997; Kendall and McDonnell, 1998; Güler et al., 2002):
  - a) hydrological;
  - b) microbially-mediated redox, and;
  - c) chemical weathering
- (ii) Gradients in each dominant *process* (e.g. redox succession) arise in response to gradients in particular landscape *attributes* (e.g. electron donor abundance), and therefore;
- (iii) Hydrochemical signatures are best suited to guide landscape classifications for water quality as they provide direct evidence of the evolutionary history of water, including its hydrological origins and interactions with soil and geological materials (e.g. redox and weathering reactions).

The ultimate aim is to produce a series of process-attribute gradient (PAG) maps, within a geographic information system (GIS) framework, that replicate as faithfully as possible the natural landscape gradients that control spatial variation in steady-state hydrological, redox, and weathering response. If the PAGs successfully estimate spatial variation in the concentration of key hydrochemical indicators (e.g., hydrological tracers, redox sensitive species, pH, alkalinity) the model is assumed fit

to be combined with a land use intensity gradient, to estimate steady-state surface water quality.

The purpose of this paper is to summarise the general conceptual framework for PoAM and provide a case study of the application and performance of the model for the estimation of spatial variation in steady-state water quality for the province of Southland, New Zealand. A detailed examination of the underlying process-attribute relationships governing hydrochemical evolution and water quality across Southland is beyond the scope of this manuscript, but is available elsewhere (Beyer et al., 2016a, 2016b; Rissmann et al., 2016a, 2016b, 2016c, 2016d, 2016e, 2016f; Rodway et al., 2016; Rissmann et al., 2018a, 2018b).

## 2. Materials and methodology

### 2.1. Environmental setting

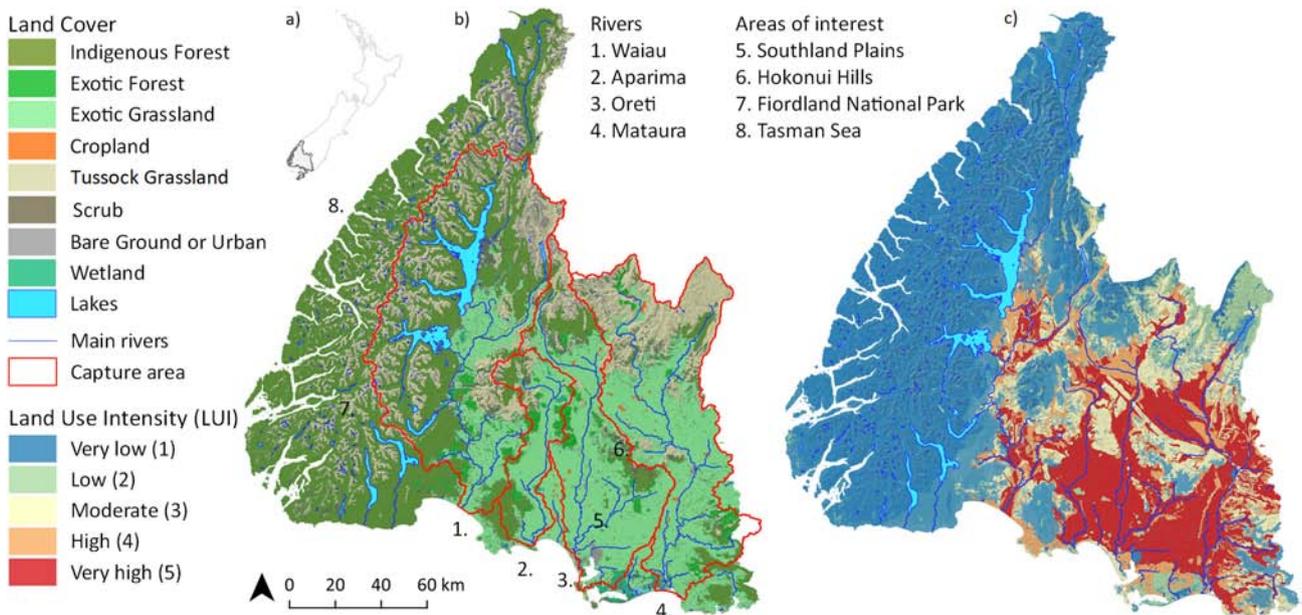
Southland is the southernmost province of the South Island, New Zealand, comprising an area of 31,824 km<sup>2</sup> (Fig. 1). It is characterised by a complex geological history including a succession of glacial and interglacial cycles throughout the Quaternary that have strongly influenced the surficial environment (Turnbull and Allibone, 2003). Alpine regions, with a maximum elevation of 2450 m relative to sea level (RSL), are prominent features of the western and northern most extent of the province. Subalpine hill country and hill country (<800 m RSL) makes up 56% of the area, followed by extensive lowland plains (21%) associated with mantling of shallow basement rock by fluvio-glacial outwash, and a series of marine terraces at the coast. Four main river catchments, the Waiau, Aparima, Oreti, and Mataura originate in the northern mountains, receive drainage from hill and lowland catchments, and ultimately discharge at the coast via significant estuarine systems (Fig. 1).

The region has a cool, humid, temperate climate with a pronounced easterly rain shadow associated with prevailing westerly airflows off the Tasman Sea and interception by the Southern Alps in Fiordland (Macara, 2013). Precipitation in Fiordland exceeds 6000 mm annually. East of Fiordland precipitation ranges between 800 mm for some northern inland basins and 1300 mm at the southern coast. Snow pack

accumulates across the northern and western mountains over the winter months, with melt occurring during the Austral spring.

Alluvial source rock and bedrock lithology is varied, ranging from felsic sedimentary and metamorphic rock, limestone, carbonaceous mudstones, through to ultramafic assemblages (Turnbull and Allibone, 2003). Unconfined aquifers occur as a mosaic of small discontinuous units restricted to the lowland plains and constitute a thin mantle (c. 0–60 m; median 12.5 m;  $n = 1750$  bore logs) of alluvium overlying predominantly poorly permeable basement rock (Hughes, 2001). The groundwater table is shallow (2.3 m below ground level;  $n = 1750$  wells) and due to the structural grain of the region, regional-scale groundwater flow paths within the overlying alluvial cover sequence are absent (Hughes, 2001; Environment Southland unpublished data). Short mean residence times of  $\leq 1$ –10 years for groundwater hosted by alluvial aquifers systems (Burberry, 2012; Daughney et al., 2015), reflects limited aquifer storage and rapid drainage via a dense surface water network, and both surficial and subsurface artificial drainage, which in lower lying areas intersects the local water table during the wetter months of the year (Ledgard, 2013; Pearson, 2015a).

Soils are mostly fine textured and slowly permeable, reflecting the fine texture of fluvio-glacial outwash and associated loess accumulation. In addition to fine texture, the majority of lowland soils (>70%) are imperfectly to poorly drained, although well drained and/or highly permeable soils occur along the riparian margins of main stem rivers, and as high terrace remnants close to the northern mountains (Topoclimate South, 2001; Turnbull and Allibone, 2003). Since colonisation by Europeans in the 1800s, the majority of lowland vegetation has been cleared and wetlands drained for pastoral farming (Ledgard, 2013; Pearson and Couldrey, 2016; Moran et al., 2017). Large areas of hill country have also been cleared of tussock grasslands, scrub, and forest although just over half of the region, mainly in Fiordland and Stewart Island, is preserved as conservation estate land. A recent shift from sheep farming to dairy farming over the last 25 years has further increased the intensity of land use across lowland and hill country areas and resulted in a decline in surface water and estuarine ecosystem health (Stevens and Robertson, 2012 and references therein; Ledgard, 2013; Moreau and Hodson, 2015; Pearson and Couldrey, 2016; Moran et al., 2017).



**Fig. 1.** Maps showing a) the location of the province of Southland within New Zealand, b) the land cover classification with the main rivers and their associated catchments (red lines), and c) the land use intensity (LUI). Rakiura Stewart Island is not shown in b) or c). The land cover and land use classifications are from 2012 and 2015 respectively. The areas of interest referred to in the text are numbered. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

## 2.2. Input data

### 2.2.1. Hydrochemical and water quality samples

PoAM utilised 20,226 surface water, 8023 groundwater, 28 spring, 67 soil matrix water, 102 soil artificial drainage (mole-pipe type) waters, and 210 precipitation (rain, snow, hail) hydrochemical and water quality analyses to guide the identification and mapping of regional process-attribute gradients for the Southland province (Rissmann et al., 2016a). Table 1 provides a summary of the range of hydrochemical and water quality measures collected for the period 2000 to 2016, and Fig. 2 shows the sampling locations. Groundwater samples from the unconfined aquifer systems of the region provided important constraint across lowland areas (Rissmann et al., 2016a).

Detailed information as to the sampling methodologies, field measures, laboratory analysis, calculation of hydrochemical metrics (e.g. major ion facies, saturation indices), and quality assurance and quality control of the samples used in this project are provided in Rissmann et al. (2016a) and were collected according to the New Zealand Groundwater Sampling Protocol (Ministry for the Environment, 2006). Major and trace species and water quality measures were analysed by R. J. Hill Laboratories Limited, New Zealand. All samples were required to have a charge balance error of less than  $\pm 5\%$ . The stable isotopes of water  $\delta^{18}\text{O}\text{-H}_2\text{O}$  [‰, Vienna Standard Mean Ocean Water (V-SMOW)],  $\delta^2\text{H}\text{-H}_2\text{O}$  [‰, V-SMOW] and dissolved inorganic carbon [ $\delta^{13}\text{C}\text{-DIC}$  ‰, Vienna Pee Dee Belemnite] were sampled according to the recommendations of the International Atomic Energy Association and analysed at the Geological Sciences Department, University of Canterbury, New Zealand.

**Table 1**  
Hydrochemical and water quality analytes (gray highlight) used in this study. The number of samples with hydrochemical measures is much smaller than those with standard water quality measures. The 93 surface water sites retained for model analysis have all the parameters listed below.

Type	Analyte	Units	Parameter name	
Major inorganic constituents	Ca <sup>2+</sup>	mg/L <sup>a</sup>	Dissolved calcium <sup>b</sup>	
	Cl <sup>-</sup>	mg/L	Dissolved chloride	
	DIC	mg/L	Dissolved inorganic carbon <sup>c</sup>	
	HCO <sub>3</sub> <sup>-</sup>	mg CaCO <sub>3</sub> /L <sup>d</sup>	Dissolved bicarbonate alkalinity <sup>d</sup>	
	K <sup>+</sup>	mg/L	Dissolved potassium	
	Mg <sup>2+</sup>	mg/L	Dissolved magnesium	
	Na <sup>+</sup>	mg/L	Dissolved sodium	
	SO <sub>4</sub> <sup>2-</sup>	mg/L	Dissolved sulphate	
	SiO <sub>2</sub>	mg/L	Dissolved reactive silica	
	B	mg/L	Dissolved boron	
Minor inorganic constituents	Br <sup>-</sup>	mg/L	Dissolved bromide	
	F	mg/L	Dissolved fluoride	
	Fe <sup>2+</sup>	mg/L	Dissolved iron	
	I	mg/L	Dissolved iodide	
	Mn <sup>2+</sup>	mg/L	Dissolved manganese	
	TON	mg N/L <sup>e</sup>	Total oxidised nitrogen <sup>f</sup>	
	NO <sub>3</sub> <sup>-</sup>	mg N/L	Dissolved nitrate	
	NO <sub>2</sub> <sup>-</sup>	mg N/L	Dissolved nitrite	
	NH <sub>4</sub> <sup>+</sup>	mg N/L	Dissolved ammoniacal nitrogen	
	TKN	mg N/L	Total Kjeldahl nitrogen	
Nutrients	TN	mg N/L	Total nitrogen	
	DRP	mg P/L <sup>g</sup>	Dissolved reactive phosphorus	
	TDP	mg P/L	Total dissolved phosphorus	
	TP	mg P/L	Total phosphorus	
	DOC	mg/L	Dissolved organic carbon	
	TOC	mg/L	Total organic carbon	
	Isotopes	$\delta^2\text{H}\text{-H}_2\text{O}^{\text{h}}$	‰ V-SMOW <sup>i</sup>	<sup>2</sup> H to <sup>1</sup> H isotope ratio in water
		$\delta^{18}\text{O}\text{-H}_2\text{O}$	‰ V-SMOW	<sup>18</sup> O to <sup>16</sup> O isotope ratio in water
		$\delta^{13}\text{C}\text{-DIC}$	‰ V-PDB	<sup>13</sup> C to <sup>12</sup> C isotope ratio in dissolved inorganic carbon (DIC)
	Biological	<i>E. coli</i>	cfu/100 mL <sup>j</sup>	<i>Escherichia coli</i>
FC		cfu/100 mL	Faecal coliforms	
Clarity		m	Visual clarity	
Other parameters	DOField	mg/L	Dissolved oxygen (field measured)	
	EC	µS/cm <sup>k</sup>	Electrical conductivity, temperature corrected	
	ORP	mV	Oxidation-reduction potential	
	pHField	pH units	pH (field measured)	
	pHLab	pH units	pH (lab measured)	
	TSS	mg/L	Total suspended solids	
	VSS	mg/L	Volatile suspended solids	
	Temp	°C	Water temperature	
	Turb	NTU <sup>l</sup>	Turbidity	

<sup>a</sup> mg/L = milligrams per litre (equivalent to grams per cubic meter and parts per million).

<sup>b</sup> Most samples collected by Environment Southland are filtered (0.45 µm) prior to analysis, and hence analytical results reflect dissolved rather than total concentrations.

<sup>c</sup> DIC is the sum of dissolved CO<sub>2</sub>, H<sub>2</sub>CO<sub>3</sub>, HCO<sub>3</sub><sup>-</sup>, and CO<sub>3</sub><sup>2-</sup>.

<sup>d</sup> Alkalinity quantifies the acid-neutralising capacity of a water sample, which is typically reported in mg CaCO<sub>3</sub> per litre. Multiply by 0.82 to convert concentration from mg CaCO<sub>3</sub>/L to mg HCO<sub>3</sub><sup>-</sup>/L.

<sup>e</sup> mg N/L means that the reported concentrations only reflect the weight of nitrogen within the compound.

<sup>f</sup> TON = NO<sub>3</sub><sup>-</sup> + NO<sub>2</sub><sup>-</sup>.

<sup>g</sup> mg P/L means that the reported concentrations only reflect the weight of phosphorus within the compound.

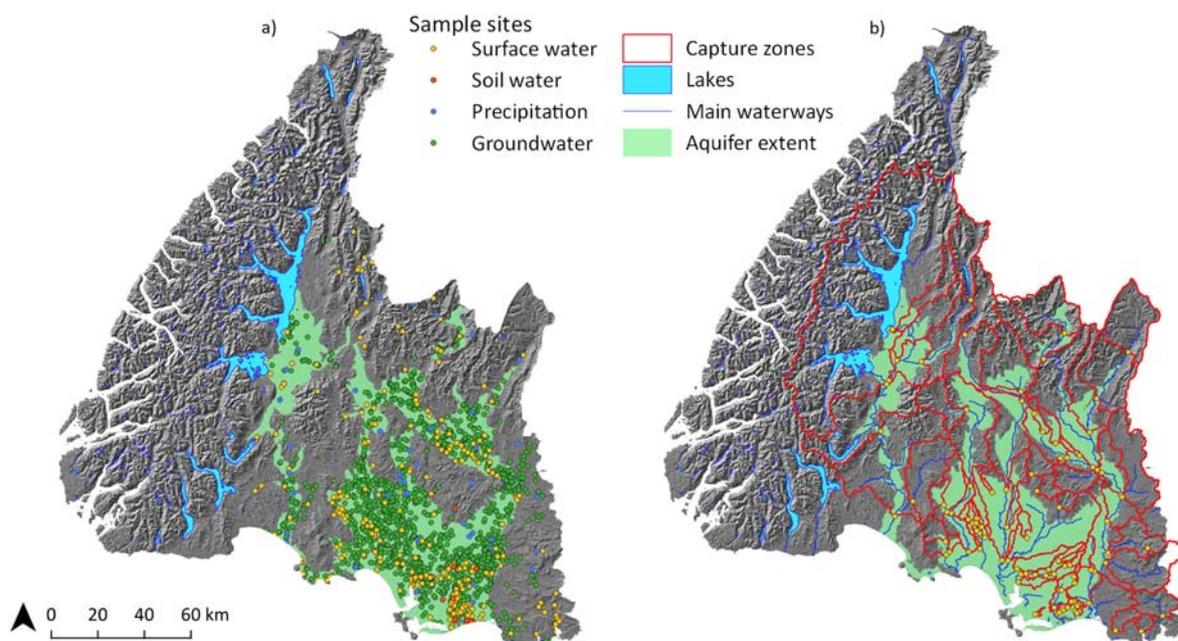
<sup>h</sup> The  $\delta$  notation relates the isotope ratio of the sample to the isotope ratio of a standard and is reported in parts per thousand (‰), e.g.  $\delta^2\text{H} = \left( \frac{(^2\text{H}/^1\text{H})_{\text{sample}}}{(^2\text{H}/^1\text{H})_{\text{standard}}} - 1 \right) \times 1000$ .

<sup>i</sup> Calibrated standards used for isotope ratios are Vienna Standard Mean Ocean Water (V-SMOW), Vienna Pee Dee Belemnite (V-PDB), and boric acid Standard Reference Material 951 (SRM 951).

<sup>j</sup> Colony forming units per 100 mL, which provides an indication of the number of viable cells.

<sup>k</sup> Microsiemens per centimetre.

<sup>l</sup> Nephelometric turbidity units.



**Fig. 2.** Sample site locations used in this study overlain on an 8 m Digital Elevation Map of the Southland province, New Zealand. a) Location site and type of sample taken across Southland. Most groundwater samples are from unconfined aquifer systems. b) The long-term surface water monitoring sites (yellow circles) and associated surface water capture zones (red outlines) used for model evaluation. Lakes, aquifers, and the main waterways are also shown for reference. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

### 2.2.2. Geospatial datasets

To best define the effective landscape properties of the province, attributes were selected from several national and regional GIS datasets. This included: topography, elevation, and altitude from an 8 m digital elevation model (DEM; [Land Information New Zealand, 2012](#)); geological attributes from the 1:250,000 geological map series (QMAP) covering Southland ([Turnbull and Allibone, 2003](#)), and the New Zealand Land Resource Inventory (NZLRI; 1:50,000; [Newsome et al., 2008](#); [Lynn et al., 2009](#)); soil attributes from Topoclimate South (1:50,000; [Topoclimate South, 2001](#)) and Fundamental Soils Layer (1:50,000; [Landcare Research, 2010](#)), land cover from the Land Cover Database (LCDBv4.1; [Landcare Research, 2015](#)), and; land use from the Southland Land Use Map ([Pearson and Couldrey, 2016](#)). Regarding hydrological input data: riverlines, stream Strahler order, and catchment areas (capture zones) were sourced from the River Environment Classification (REC; [Snelder and Biggs, 2002](#)); precipitation volume from the National Climate Database ([Macara, 2013](#));  $\delta^{18}\text{O}\text{-H}_2\text{O}$  of precipitation from the national isoscape model of [Baisden et al. \(2016\)](#), and; unconfined aquifer type (e.g. lowland, terrace, riparian) and extent from Environment Southland (unpublished). Land use intensity was derived by combining the Land Use Capability classification of [Lynn et al. \(2009\)](#) with the Land Cover Database ([Fig. 1c](#)).

### 2.2.3. Long-term water quality monitoring sites

A subset of 93 surface water sites, comprising 7028 repeat measures for the period 2013–2018, were used for evaluating the performance of PAG maps to estimate spatial variation in hydrochemical measures and for subsequent development of models for estimating spatial variation in both surface water hydrochemistry and water quality ([Section 3.3](#); [Fig. 2](#)). Hierarchical cluster analysis (HCA) was applied to the log transformed and z-scored median data for each site (following the recommendations of [Güler et al., 2002](#)) in order to subdivide waters into distinct hydrochemical groups for hypothesis development and model evaluation. HCA assigned the surface water sites into 12 clusters on the basis of 22 measures (pH,  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{SiO}_2$ , total alkalinity,  $\text{Cl}^-$ , I, F, B,  $\text{SO}_4^{2-}$ , DO,  $\text{Mn}^{2+}$ ,  $\text{Fe}^{2+}$ , TON, TKN, TP, DRP,  $\delta^{18}\text{O}\text{-H}_2\text{O}$ ,  $\delta^2\text{H}\text{-H}_2\text{O}$ , and  $\delta^{13}\text{C}\text{-DIC}$ ).

### 2.3. Conceptual framework and general methodology

A brief summary of the dominant processes responsible for the spatial variation in surface water hydrochemistry and their associated PAGs for Southland is summarised in [Table 2](#) and [Appendix A](#). In depth analysis and discussion of PAGs is available in a series of technical reports ([Beyer et al., 2016a, 2016b](#); [Rissmann et al., 2016a, 2016b, 2016c, 2016d, 2016e, 2016f](#); [Rodway et al., 2016](#); [Rissmann et al., 2018b](#); [Pearson et al., 2018](#)).

The drivers of spatial variation in hydrochemistry across a catchment or a region are explored with a view to building a series of controlling factor hypotheses for each dominant process which are then tested against long-term water quality monitoring hydrochemical data ([Fig. 3](#); [Rissmann et al., 2016a](#); [Rissmann et al., 2018b, 2018c](#)).

Conceptually, the PoAM framework recognises that hydrochemical and water quality measures are driven by more than one PAG occurring across multiple domains (e.g. atmospheric, soil, shallow aquifer domains) and scales ([Rissmann et al., 2016a](#); [Rissmann et al., 2018b](#)). Therefore, overall compositional variation between surface water monitoring sites, within a catchment or across a region, is the sum of multiple PAGs of varying steepness ([Rissmann et al., 2016a](#); [Rissmann et al., 2018b](#)).

#### 2.3.1. Step 1: Spatial hydrochemical exploration for hypothesis development

Hypothesis development includes specification of the relative sensitivity (importance) and magnitude (direction) of the response of a hydrochemical signature with respect to one or more effective landscape attribute gradients, and is supported by local knowledge of those gradients, relevant literature, and/or through the inclusion of existing controlling factor classifications (e.g. [Snelder and Biggs, 2002](#); [Rissmann, 2011](#); [Killick et al., 2015](#); [Beyer et al., 2016a, 2016b](#); [Rodway et al., 2016](#); [Rissmann et al., 2016a, 2016b, 2016c, 2016d, 2016e, 2016f](#); [Rissmann et al., 2018a, 2018b](#)).

In practice, hydrochemical and water quality measures are imported into GIS and provide the reference point for association with pre-existing spatial classifications of soil, geology, and hydrological representations of

**Table 2**  
The dominant processes and process-attribute gradients for estimating spatial variation in hydrochemistry and water quality.

Process	Role	Controlling factor	Process attribute gradient	Scale	Data scale
Hydrological	The deposition of marine aerosols (wet and dry) and the stable isotopes of water under the local climatic setting, prior to redistribution by the hydrological network.	Topographic controls of elevation and distance from coast over $\delta^2\text{H-H}_2\text{O}$ and $\delta^2\text{H-H}_2\text{O}$ signatures and marine aerosol ( $\text{Na}^+$ , $\text{Cl}^-$ , $\text{Br}^-$ , $\text{B}$ , $\text{Mg}^{+2}$ , $\text{SO}_4^{2-}$ ) loading.	Atmospheric (ATM)	Macro (1000–10,000 km <sup>2</sup> )	Regional-scale
			Recharge domain (RCD)	Meso to Macro (1–1000 km <sup>2</sup> )	Regional-scale higher order stream network and aquifers
	The transport and mixing of solutes and particulates by water through the surface water and shallow groundwater network (where present).	Topographic domains (e.g. alpine, hill, lowland) and connectivity associated with distinct hydrological tracer signatures in surface waters and groundwaters. Soil series scale hydrological pathways occurring within recharge domains.	Overland flow (OLF) Deep drainage (DD) Lateral drainage (LAT) Natural soil bypass (BP) Artificial drainage (ART) Vector (riverlines)	Micro (0.01–10 km <sup>2</sup> )	Soil polygon for 1:50,000 maps
Redox	Low temperature (<35 °C), microbially mediated succession of terminal electron species ('redox') in unsaturated zone (soil) and saturated zone (aquifer) materials.	Property and paddock flow paths associated small scale drainage basins (<10 ha) and associated low order ephemeral, intermittent, and perennial streams. Soil redox potential associated with soil drainage class and redox indicators (mottling and gleying). Lithological-scale electron donor abundance of subsurface geology.	Soil reduction potential (SRP)	Micro (0.01–10 km <sup>2</sup> )	Soil polygon for 1:50,000 maps
			Geological reduction potential (GRP)	Meso to Macro (10–10,000 km <sup>2</sup> )	Geological polygon for 1:50,000–1:250,000 maps Soil series polygon for 1:50,000 soil maps
Weathering	The Acid Neutralisation Capacity (ANC) of unsaturated zone and saturated zone materials (unconfined aquifers).	Lewis base concentration of soil. Lewis base concentration of geological materials that host aquifers.	Soil acid neutralisation capacity (SANC)	Micro (0.01–10 km <sup>2</sup> )	Soil series polygon for 1:50,000 soil maps
			Geological acid neutralisation capacity (GANC)	Meso to Macro (10–10,000 km <sup>2</sup> )	1:50,000–1:250,000 geological maps

river and aquifer networks. Multivariate methods, such as principal component analysis (PCA), are useful for refining the relative sensitivity of landscape attributes with regards to the spatial variation in hydrochemical signatures (Güler et al., 2002; Rodway et al., 2016; Beyer et al., 2016a; Rissmann et al., 2018a, 2018b, 2018c).

### 2.3.2. Step 2: Process-attribute gradient map construction

The controlling landscape attributes identified during Step 1 are extracted from one or more existing GIS datasets and combined in an attempt to replicate, as accurately as possible, gradients for each dominant process (Fig. 3; Table 3). If an important controlling factor is not represented by existing classifications, e.g. the tendency of clay-rich smectitic soils to crack in response to soil moisture deficit, it may be necessary to generate a representation by extracting and combining relevant controlling factor information (Beyer et al., 2016a).

Historical water quality measures, which often occur at higher spatial densities, are used to guide the editing of geospatial layers so that they represent as faithfully as possible actual PAG (Beyer et al., 2016a, 2016b; Rissmann et al., 2016b, 2016c, 2016d, 2016e, 2016f; Rissmann et al., 2018a, 2018b, 2018c; Fig. 2). For example, historical measures of electrical conductivity and  $\text{NO}_3^-$  from unconfined aquifers may be used to refine the *effective* hydrological boundary, i.e. solute gradient, that exists between a riparian aquifer recharged by an alpine stream and an adjacent lowland aquifer recharged by local precipitation (Beyer et al., 2016a).

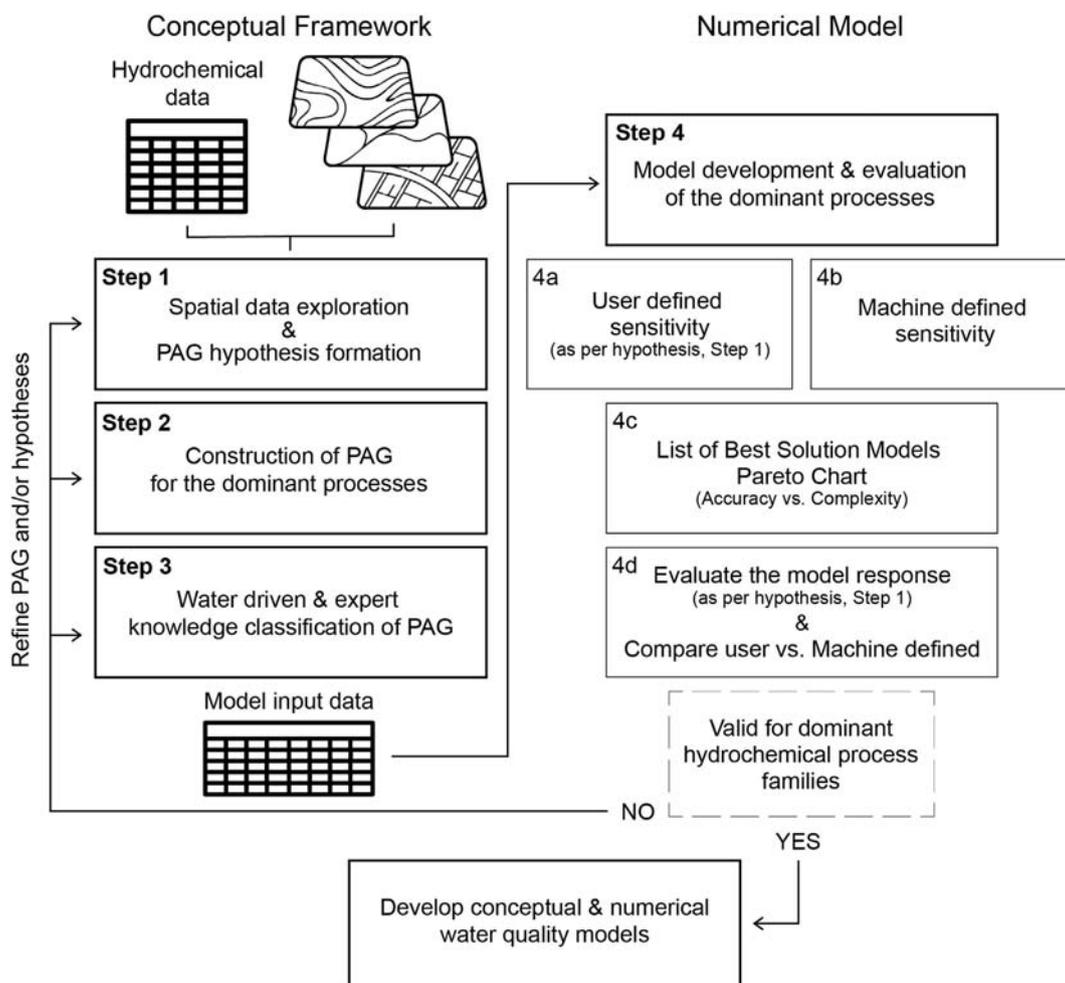
### 2.3.3. Step 3: Process-attribute gradient classification

Grouping of landscape attributes into *effective* classes is data driven, using hydrochemical and/or a large number of low-resolution water quality measures (e.g.  $\text{NO}_3^-$  and conductivity; Beyer et al., 2016a, 2016b; Rissmann et al., 2018a, 2018b, 2018c). Hierarchical cluster analysis (HCA) combined with post hoc significance testing of resultant classes is used for grouping the landscape attributes according to hydrochemical and/or water quality measures (Rodway et al., 2016; Beyer et al., 2016a, 2016b; Rissmann et al., 2018b, 2018c).

Given the objective is to represent only the effective attributes of the landscape, the native complexity of existing classifications of landscape attributes is often reduced (Rissmann, 2011; Rissmann et al., 2016a, 2016b, 2016c, 2016d, 2016e, 2016f, 2016g; Rissmann et al., 2018b, 2018c). For example, a regional geological survey containing 100 or more individual rock types may be reduced to 3 or 4 main classes according to the relative abundance of bioavailable electron donors observed to drive redox succession in shallow aquifer systems (Beyer et al., 2016b; Rissmann et al., 2018b, 2018c). Despite the reduction in the complexity of native classifications, a lack of hydrochemical or water quality data may necessitate a subjective, albeit expert, hydrochemical grouping of attributes (Rissmann et al., 2018b).

Resultant classes are subsequently ranked according to process signals, and numeric scores assigned. For example, HCA derived clusters defining macro-scale recharge domains are ranked in terms of most to least dilute and lowest to highest recharge altitude according to the concentration of  $\text{Na}^+$ ,  $\text{Cl}^-$ ,  $\text{Br}^-$ , and the stable isotopes of water (Beyer et al., 2016a; Rissmann et al., 2018b, 2018c); meso-scale soil and rock acid neutralising capacity are ranked from least to most neutralising according to pH and alkalinity (Rissmann et al., 2018b); and macro-scale geological reduction potential of unconfined aquifer systems ranked from most oxidising to most reducing according to the concentration of redox sensitive species (Beyer et al., 2016b; Rissmann et al., 2018b, 2018c).

The input data for model development and evaluation is compiled from the PAG classification in GIS. Capture zones are generated for the long-term surface water monitoring sites and combined with each PAG, before being exported in a tabular format (Rissmann et al., 2018b, 2018c). The tabulated PAG scores are proportionally weighted by area for each capture zone and linked with median hydrochemical and water quality measures for each monitoring site. Weighted PAG scores and median hydrochemical data for each site are subsequently log transformed and Z-scored prior to correlation analysis (Appendix C) and model development and evaluation (Appendix D).



**Fig. 3.** A flow chart summary of key steps in process-attribute mapping (PoAM). The conceptual model provides a spatial representation of the key processes and contaminant transport pathways whilst the numerical models provide an estimate of steady-state median surface water concentration.

#### 2.3.4. Step 4: Model development and evaluation

The model development and evaluation process assess if the classed PAGs respond as hypothesised (Step 1), and whether they are sufficiently representative of actual hydrological, redox, and weathering gradients (Rissmann et al., 2018b). Tabulated data is imported into the machine intelligence software package *Eureqa* (version 1.24.0; Schmidt and Lipson, 2015) which utilises symbolic regression (SR) as the genetic programming algorithm, Pareto front optimisation for model selection, provides measures of predictor sensitivity and magnitude, and cross validates by splitting the data via a disjunctive approach where the training and validation data sets are kept separate (e.g. 90% training: 10% validation with a random shuffle before splitting the data; Schmidt and Lipson, 2009, 2015; Dubčáková, 2011).

Modelling proceeds by the user identifying what are considered the most sensitive PAGs for explaining spatial variation in each dominant process (i.e., user defined in Step 1; Fig. 3). The SR algorithm then produces numeric expressions by randomly combining mathematical building blocks, such as algebraic operators, analytical functions, constant variables, and state variables (Khu et al., 2001; Schmidt and Lipson, 2009). The SR algorithms retain those equations that best model the PAG relationships and abandons unpromising solutions (Khu et al., 2001; Schmidt and Lipson, 2009). Importantly, if a potential causative PAG offers little explanatory power, relative to others, it is automatically excluded during the evolutionary process. After numerous iterations, the SR algorithm returns a set of models (i.e. explicit

mathematical functions) that explain the spatial variation in hydrochemical processes as a function of one or more PAGs. A subsequent model run is generated whereby all PAGs are selected as an input and the SR algorithm determines which to retain (i.e. machine defined). Model outputs include numeric scores of the sensitivity, magnitude of response, and form the basis for testing the hypotheses formulated in Step 1, as well as for evaluating the adequacy with which PAGs are replicated. The sensitivity of a given PAG is defined according to its relative impact over a hydrochemical process signal. A positive magnitude occurs when increases in a PAG score leads to increases in the target variable, and vice-versa for a negative magnitude.

Model complexity and accuracy is assessed by the goodness of fit ( $R^2$ ), correlation coefficient ( $r$ ), root mean squared error (RMSE) and mean absolute error (MAE) methods (Schmidt and Lipson, 2009, 2015; Dubčáková, 2011). Model selection is guided by the trade-off between  $R^2$  and model simplicity, where the  $R^2$  and MAE are analogues to the Akaike information criteria (Aho et al., 2014).

Only when models are consistent with the hypotheses for the dominant processes (Step 1) and are accurate (i.e. a cross-validated  $R^2$  of  $\geq 0.75$ ), are the mapped PAGs considered fit to be combined with a LUI gradient for the estimation of spatial variation in steady-state surface water quality. If the combination of PAG maps with land use intensity provides a reasonable estimate of spatial variation in water quality the model is then considered appropriate for estimation of steady-state surface water quality across data poor areas (Rissmann et al., 2018b).

**Table 3**  
A summary of the various geospatial and hydrochemical (sample type, analyte, GIS data) used to define each process attribute gradient, the dominant process they are associated with, the scale at which the process attribute gradient is applied, and an example application.

Process	PAG	Sample types	Hydrochemical analytes	GIS dataset(s) and resolution	PAG scale	Example application
Hydrological	ATM	Precipitation, soil water, ground and surface waters	Na <sup>+</sup> , Cl <sup>-</sup> , δ <sup>18</sup> O-H <sub>2</sub> O, δ <sup>2</sup> H-H <sub>2</sub> O	8 m DEM, δ <sup>18</sup> O-H <sub>2</sub> O precipitation isoscape (4 km <sup>2</sup> pixel)	Macro (1000–10,000 km <sup>2</sup> )	Rodway et al., 2016; Rissmann et al., 2018a, 2018b.
	RCD	Soil waters, ground and surface waters	Na <sup>+</sup> , Cl <sup>-</sup> , Br <sup>-</sup> , δ <sup>18</sup> O-H <sub>2</sub> O, δ <sup>2</sup> H-H <sub>2</sub> O, δ <sup>13</sup> C-DIC, EC, NO <sub>3</sub> <sup>-</sup>	Soil surveys (1:50,000), Aquifer type and extent (1:50,000)	Meso to macro (1–1000 km <sup>2</sup> )	Rissmann et al., 2018a, 2018b
	OLF	OLF runoff, surface waters	Clarity, TSS	Soil surveys (1:50,000), 8 m DEM	Micro to meso (0.01–1 km <sup>2</sup> )	Pearson, 2015b
	DD	Ground and surface waters	Clarity, TSS	Soil surveys (1:50,000)	Micro to meso (0.01–1 km <sup>2</sup> )	Rissmann et al., 2018a, 2018b
	LAT	Soil water, ground and surface waters	Clarity, TSS	Soil surveys, 8 m DEM	Micro to meso (0.01–1 km <sup>2</sup> )	Rissmann et al., 2018a, 2018b
	ART	Soil water	Clarity, TSS, DO, NO <sub>3</sub> <sup>-</sup> , Mn <sup>2+</sup> , Fe <sup>2+</sup> , SO <sub>4</sub> <sup>2-</sup> , DOC	Soil surveys, 8 m DEM, Land Cover (1 Ha)	Micro to meso (0.01–1 km <sup>2</sup> )	Pearson, 2015a
	BP	Ground water, soil artificial drainage	DO, NO <sub>3</sub> <sup>-</sup> , Mn <sup>2+</sup> , Fe <sup>2+</sup> , SO <sub>4</sub> <sup>2-</sup> , DOC	Soil surveys (1:50,000)	Micro to meso (0.01–1 km <sup>2</sup> )	Beyer et al., 2016a; Rissmann et al., 2018a, 2018b
Redox	SRP	Soil water, ground and surface waters	DO, NO <sub>3</sub> <sup>-</sup> , Mn <sup>2+</sup> , Fe <sup>2+</sup> , SO <sub>4</sub> <sup>2-</sup> , DOC	Soil surveys (1:50,000); soil chemistry profile points.	Micro to meso (0.01–1 km <sup>2</sup> )	Killick et al., 2015; Beyer et al., 2016a, 2016b; Rissmann et al., 2018a, 2018b.
	GRP	Ground and surface waters	DO, NO <sub>3</sub> <sup>-</sup> , Mn <sup>2+</sup> , Fe <sup>2+</sup> , SO <sub>4</sub> <sup>2-</sup> , DOC	Geological surveys (1:50,000–1:250,000)	Meso to macro (10–1000 km <sup>2</sup> )	Rissmann, 2011, Beyer et al., 2016a, 2016b; Rissmann et al., 2018a, 2018b.
Weathering	SANC	Soil water, ground and surface waters	pH, Ca <sup>2+</sup> , DIC, bicarbonate alkalinity, carbonate alkalinity, δ <sup>13</sup> C-DIC	Soil surveys (1:50,000); soil chemistry (profile points)	Micro to meso (0.01–1 km <sup>2</sup> )	Rissmann et al., 2016a, 2018a.
	GANC	Ground and surface waters	pH, Ca <sup>2+</sup> , DIC, bicarbonate alkalinity, carbonate alkalinity, δ <sup>13</sup> C-DIC	Geological surveys (1:50,000–1:250,000); soil chemistry (profile points)	Meso to macro (10–1000 km <sup>2</sup> )	Rissmann et al., 2016a, 2018a.

ATM = atmospheric loading; RCD = recharge domain; OLF = overland flow; DD = deep drainage (vertical soil profile drainage); LAT = lateral drainage (horizon permeable drainage); ART = artificial drainage (subsurface mole-pipe and open ditch drainage); BP = bypass (soil moisture deficit induced cracking of clay-rich soil and bypass of the soil matrix); SRP = soil reduction potential; GRP = geological reduction potential; SANC = soil acid neutralisation capacity; GANC = geological acid neutralisation capacity.

### 3. Results

#### 3.1. Main process hypotheses (step 1)

The hypothesised drivers of spatial variation in Southland hydrochemistry are summarised in Table 4 with additional context provided in Appendix B. These hypotheses are specific to the regional scale and vary when applied at catchment and/or subcatchment scales (Rissmann et al., 2016a; Rissmann et al., 2018b).

#### 3.2. Process-attribute gradient maps (steps 2–3)

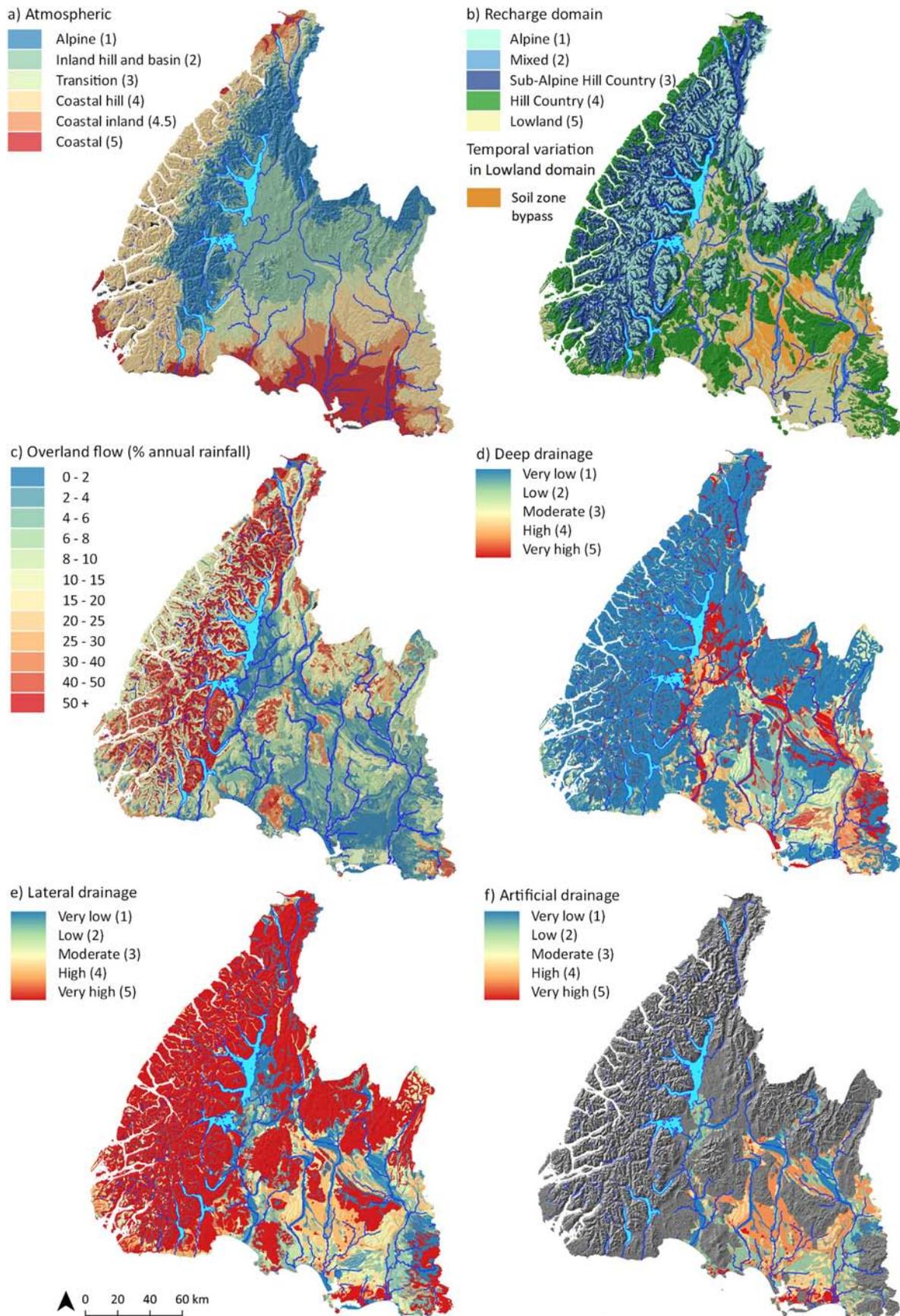
Eleven PAGs, defining the dominant processes governing hydrochemistry were produced for the Southland province (Figs. 4–6;

Tables 2–4, Appendix B). Of the 11 PAGs, 7 are associated with the hydrological drivers of spatial variation in hydrochemistry (Fig. 4). Of the 7 hydrological PAGs, the atmospheric (ATM) and recharge domain and hydrological connectivity (RCD) layers are macro in scale (Table 3). The remaining 5 hydrological PAGs are all associated with the soil zone and represent meso-scale hydrological gradients in percent precipitation occurring as overland flow (OLF), lateral drainage (LAT), shrink-swell mediated soil zone bypass (BP), deep drainage through the soil profile (DD), and artificial drainage of the soil profile (ART) (Table 3; Pearson, 2015a, 2015b; Pearson et al., 2018). The redox and weathering processes are comprised of separate soil and geological PAGs (Figs. 5 and 6; Beyer et al., 2016b). The soil PAG overlies the geological PAG, except where bedrock outcrops (Rissmann et al., 2016a). Across the lowland areas of the Southland province, the geological

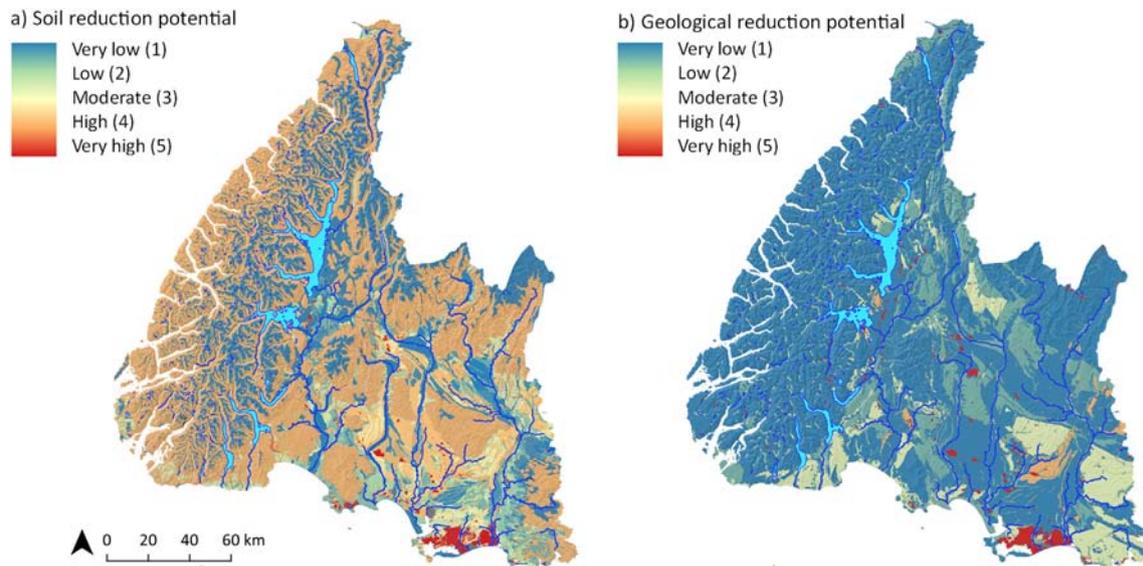
**Table 4**  
A summary of the main hypotheses for each dominant process used in Step 1, Southland province, New Zealand.

Dominant process	Narrative	Sensitivity of PAG	Magnitude
Hydrological	Spatial variation in the indicators of water source and hydrological connectivity, i.e. the conservative hydrological tracers Na <sup>+</sup> , Cl <sup>-</sup> , Br <sup>-</sup> , δ <sup>18</sup> O-H <sub>2</sub> O, δ <sup>2</sup> H-H <sub>2</sub> O, and δ <sup>13</sup> C-DIC will be explained best by combination of the macro-scale atmospheric (ATM) and recharge domain (RCD) process-attribute gradients.	Mostly ATM > RCD	Na <sup>+</sup> , Cl <sup>-</sup> , Br <sup>-</sup> , and the stable isotopes of water will show a positive magnitude, and δ <sup>13</sup> C-DIC a negative magnitude, across both macro-scale hydrological process-attribute gradients.
Redox	Spatial variation in redox indicators, i.e. DO, Mn <sup>2+</sup> , Fe <sup>2+</sup> , and DOC will be explained best by the combination of macro-scale hydrological (ATM, RCD), meso-scale hydrological soil bypass (BP), and meso-scale soil and geological redox potential (SRP, GRP) process-attribute gradients.	Hydrological > Redox	Mn <sup>2+</sup> , Fe <sup>2+</sup> , and DOC will show a positive magnitude, and DO a negative magnitude, across the macro-scale hydrological and meso-scale redox process-attribute gradients; Mn <sup>2+</sup> , Fe <sup>2+</sup> , and DOC will show a negative, and DO a positive, magnitude across the meso-scale BP process-attribute gradient.
Weathering	Spatial variation in the indicators of weathering, i.e. pH and total alkalinity will be explained best by combination of the macro-scale hydrological (ATM, RCD) and meso-scale weathering (SANC, GANC) process-attribute gradients.	Weathering ≥ Hydrological	Total alkalinity will show a positive magnitude across the meso-scale weathering and macro-scale hydrological process-attribute gradients; pH will show a positive magnitude across the weathering process-attribute gradients but a negative magnitude across the macro-scale hydrological process-attribute gradients.
Overall hydrochemical composition	Spatial variation in HCA defined cluster membership will be explained best by combining hydrological (ATM, RCD, BP), redox (SRP, GRP), and weathering (SANC, GANC) process-attribute gradients.	Hydrological > Redox = Weathering	Not applicable.

PAG = Process-attribute gradient; ATM = atmospheric loading; RCD = recharge domain; SRP = soil reduction potential; GRP = geological reduction potential; SANC = soil acid neutralisation capacity; GANC = geological acid neutralisation capacity; BP = bypass (soil moisture deficit induced cracking of clay-rich soil and bypass of the soil matrix).



**Fig. 4.** Hydrological process-attribute gradient maps of a) atmospheric (ATM), b) recharge domain and hydrological connectivity (RCD) including soil zone bypass (BP), c) overland flow (OLF), d) deep drainage (DD), e) lateral drainage (LAT) and, f) artificial drainage (ART). The base map is an 8 m DEM of the Southland province, New Zealand. Large lakes (light blue) and major rivers (dark blue) are shown for reference. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 5.** The process-attribute gradient maps for the redox potential of a) soil zone and b) geological substrate. The base map is an 8 m DEM of the Southland province, New Zealand. Large lakes (light blue) and major rivers (dark blue) are shown for reference. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

PAGs represent the upper portion of the unconfined aquifer system (Rissmann et al., 2016a).

### 3.3. Model development and evaluation (step 4)

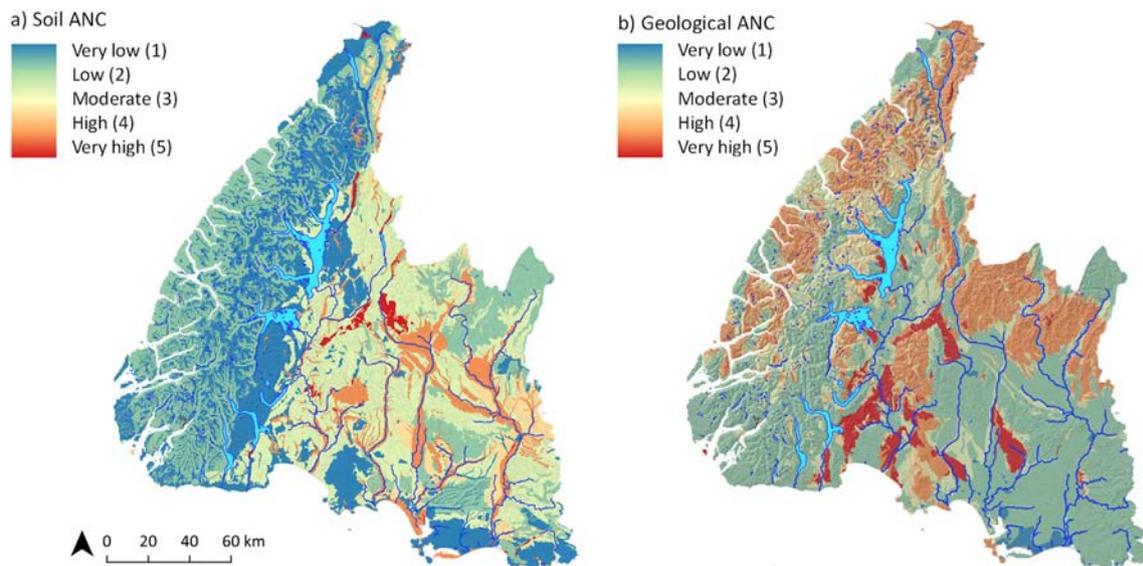
#### 3.3.1. Hydrochemistry

The performance of PoAM to replicate the effective hydrological, redox, and weathering gradients of the Southland province is presented in Table 5. User defined gradients for the dominant processes all responded as hypothesised (Step 1; Table 4, Appendix D); and with the exception of dissolved oxygen (DO,  $R^2 = 0.56$ ), cross-validated  $R^2$  values of 0.75 to 0.95 indicated good to very good representation of macro and meso-scale hydrological gradients, and meso-scale redox and weathering gradients (Table 5).

Machine defined runs retained a greater number of PAGs (predictors), maximum of 6 of the 11 PAGs, and overall produced slightly less

complex and more accurate models ( $R^2 = 0.79$  to 0.96 excluding DO; Table 5). The main difference between the user and the machine defined models was the inclusion and retention of the meso scale hydrological, redox and weathering PAGs often in place of, or in addition to the macro-scale hydrological PAGs. Notably, many of the meso-scale PAGs are significantly correlated with macro-scale ATM and RCD PAGs due to the regional setting (Appendix C).

Accounting for spatial correlation between macro- and meso-scale PAGs, the pattern of predictor retention by the machine defined runs were also consistent with the hypotheses proposed in Step 1 for Southland (Table 4). Specifically, the hydrological PAGs were collectively the most sensitive predictors retained by the machine defined runs for the hydrological processes; hydrological PAGs followed by the redox (soil reduction potential, geological reduction potential and indirectly artificial drainage, deep drainage, and soil zone bypass) PAGs were the most sensitive predictors retained for the redox processes; hydrological



**Fig. 6.** The process-attribute gradient maps for the acid neutralisation capacity of the a) soil zone and b) geological substrate. The base map is an 8 m DEM of the Southland province, New Zealand. Large lakes (light blue) and major rivers (dark blue) are shown for reference. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

**Table 5**

User defined and machine defined hydrochemical model and performance for dominant processes, Southland province, New Zealand.

User function	User defined model	R <sup>2</sup>	Complexity	Machine function	Machine defined model	R <sup>2</sup>	Complexity	
Hydrology	Cl <sup>-</sup> = f(ATM, RCD)	Cl <sup>-</sup> = 0.09 + 0.58 * RCD + 0.58 * ATM + 0.85 * ATM <sup>2</sup> + 0.71 * ATM <sup>3</sup> - 0.26 * RCD * ATM - 0.31 * RCD <sup>2</sup> - 0.648 * RCD * ATM <sup>2</sup>	0.94	43	Cl <sup>-</sup> = f(ATM, RCD, BP, SRP, GRP)	Cl <sup>-</sup> = 0.7788 * ATM + 0.4589 * RCD - 0.06051 - 0.1009 * SRP <sup>2</sup> - 0.159 * RCD * GRP * BP - 0.3949 * RCD * ATM * BP	0.96	31
	Na <sup>+</sup> = f(ATM, RCD)	Na <sup>+</sup> = 0.07 + 0.67 * RCD + 0.27 * ATM + 0.56 * ATM <sup>3</sup> + 0.35 * ATM <sup>2</sup> + 0.048 * RCD <sup>3</sup> - 0.51 * RCD * ATM - 0.66 * RCD * ATM <sup>2</sup>	0.95	45	Na <sup>+</sup> = f(ATM, RCD, SRP)	Na <sup>+</sup> = RCD + 0.1951 * ATM + 0.4502 * ATM <sup>3</sup> + 0.082 * SRP * RCD <sup>2</sup> - 0.08205 - 0.5496 * RCD * ATM <sup>2</sup>	0.96	31
	Br = f(ATM, RCD)	Br = 0.26 + ATM + 0.38 * RCD + ATM <sup>2</sup> - 0.18 * ATM <sup>5</sup> - 0.53 * ATM <sup>4</sup>	0.75	33	Br = f(ATM, SRP, BP, GANC, LUI)	Br = 0.1185 + 0.8819 * ATM + 0.1509 * LUI * SRP - 0.2469 * SRP - 0.2824 * GANC * BP - 0.4477 * ATM * BP	0.91	27
	δ <sup>18</sup> O-H <sub>2</sub> O = f(ATM, RCD)	δ <sup>18</sup> O-H <sub>2</sub> O = 0.17 + 0.64 * ATM <sup>2</sup> + 0.62 * ATM <sup>3</sup> + 5.8 * ATM * cos(0.49 * RCD) - RCD <sup>2</sup> - 5.3 * ATM	0.93	35	δ <sup>18</sup> O-H <sub>2</sub> O = f(ATM, RCD, GRP, BP, SANC, DD)	δ <sup>18</sup> O-H <sub>2</sub> O = 0.1255 + 0.4749 * ATM + 0.2946 * RCD + 0.1877 * BP + 0.1604 * ATM <sup>3</sup> - 0.4146 * RCD * BP - 0.1773 * LUI * RCD * ATM	0.96	33
	δ <sup>2</sup> H-H <sub>2</sub> O = f(ATM, RCD)	δ <sup>2</sup> H-H <sub>2</sub> O = 0.26 + 0.5 * ATM + 2.7 * sin(RCD) + 0.5 * ATM <sup>3</sup> + 0.5 * ATM <sup>2</sup> - 2.8 * RCD - 1.1 * RCD <sup>2</sup> - 0.27 * RCD * ATM <sup>2</sup>	0.94	41	δ <sup>2</sup> H-H <sub>2</sub> O = f(ATM, RCD, LUI, BP, OLF)	δ <sup>2</sup> H-H <sub>2</sub> O = 0.05265 + 0.3761 * RCD + 0.3254 * ATM + 0.3213 * ATM <sup>3</sup> + 0.15 * LUI * OLF * BP - 0.1673 * LUI * RCD * ATM	0.96	37
	δ <sup>13</sup> C-DIC = f(ATM, RCD)	δ <sup>13</sup> C-DIC = 0.16 + 2.9 * RCD * ATM <sup>2</sup> + 1.15 * RCD * ATM <sup>3</sup> - ATM * sin(0.16 + 2.9 * ATM) - 0.27 * ATM - 1.85 * RCD	0.79	38	δ <sup>13</sup> C-DIC = f(LUI, SANC, ART, BP, SRP, DD)	δ <sup>13</sup> C-DIC = 0.1849 * ART <sup>2</sup> + 0.4848 * SANC * DD <sup>2</sup> - 0.6779 * LUI - 0.1484 * LUI * DD - 0.2849 * BP * SRP <sup>2</sup>	0.83	31
Redox	DO = f(SRP, GRP, ATM, RCD, BP)	DO = 0.1952 * RCD * BP + 1.166 * GRP * sin(1.294 * BP * SRP <sup>2</sup> ) + 0.2484 * sin(0.7313 + 5.059 * GRP + 2.627 * RCD) - sin(GRP * BP <sup>3</sup> )	0.56	46	DO = f(DD, SRP, ART, BP, SANC)	DO = 0.151 * LUI * GANC + 0.2229 * ATM * SRP * OLF * ART + 0.7206 * ART * SANC <sup>2</sup> * sin(BP) - 0.8206 * SRP * GRP * sin(BP)	0.67	39
	Mn <sup>2+</sup> = f(SRP, GRP, ATM, RCD, BP)	Mn <sup>2+</sup> = 0.41 + 1.3 * ATM + 0.5 * GRP * BP + 0.36 * ATM <sup>2</sup> - 0.42 * ATM * BP - 0.17 * SRP <sup>2</sup>	0.87	29	Mn <sup>2+</sup> = f(ART, ATM, GRP, LUI, BP)	Mn <sup>2+</sup> = 0.1843 + 0.6967 * ART + 0.4423 * ATM + 0.08515 * GRP + 0.007284 / GRP + 0.2722 * LUI * GRP - 0.1366 * LUI <sup>2</sup> - 0.4796 * LUI * ART * BP	0.86	38
	Fe <sup>2+</sup> = f(SRP, GRP, ATM, RCD, BP)	Fe <sup>2+</sup> = 0.31 + 0.54 * ATM + 0.54 * ATM * SRP + 0.50 * RCD * SRP + 0.16 * SRP <sup>2</sup> + 0.51 * SRP * ATM <sup>2</sup> - 0.27 * RCD * BP <sup>2</sup>	0.83	45	Fe <sup>2+</sup> = f(ATM, GRP, ART, SANC, LUI, BP)	Fe <sup>2+</sup> = 0.2799 * GRP + 0.1212 * ART + 0.3753 * LUI * GRP + 0.1103 * SANC <sup>2</sup> + sin(0.4566 * ATM) - 0.2136 * SANC * GRP - 0.225 * ART * BP	0.84	38
	DOC = f(SRP, GRP, ATM, RCD)	DOC = 0.21 * RCD + 0.07 * SRP + 0.10 * RCD * SRP + 0.12 * ATM * GRP <sup>2</sup> + 0.24 / (2.5 - 3.0 * ATM) - 0.04	0.94	32	DOC = f(ART, SANC, BP, ATM)	DOC = 0.2317 * RCD + 0.04685 * OLF - 0.3357 - 0.1576 * OLF * GRP <sup>2</sup> - 0.3028 * SANC * ATM <sup>2</sup> * exp(ATM)	0.92	31
	pH = f(SANC, GANC, RCD, ATM)	pH = 0.068 + 0.24 * GANC + 0.25 * SANC * exp(ATM) + GANC * sin(0.2642 * RCD) + 0.02 * factorial(0.46 * GANC - 0.54 - RCD)	0.79	50	pH = f(GRP, SANC, OLF, BP, ART, ATM)	pH = 0.3372 * ATM * OLF * GRP <sup>2</sup> + sin(0.2367 + 0.2923 * SANC + 0.3331 * ART * BP + 0.3189 * ATM * SANC - 0.06645 * GRP <sup>4</sup> )	0.79	40
	Weathering	Total Alkalinity = f(SANC, GANC, RCD, ATM)	Total Alkalinity = 0.29 + 0.17 * SANC + 0.10 * RCD * SANC - sin(0.10 * GANC <sup>2</sup> ) - 0.31 * RCD * ATM - 0.17 * ATM <sup>2</sup> - 0.18 * ATM <sup>3</sup>	0.81	40	Total Alkalinity = f(BP, ATM, RCD, GANC, ART)	Total Alkalinity = 0.08035 + 0.1719 * BP + 0.03631 / BP + 0.4292 * RCD * GANC + 0.3947 * ATM * GANC * BP + 0.3267 * BP * ART <sup>2</sup> + 0.0004306 / (0.01053 + 0.1719 * BP)	0.90

For detailed performance results see Appendix D. HCA = Hierarchical Cluster Analysis membership; ATM = atmospheric loading; RCD = recharge domain; SANC = soil acid neutralisation capacity; GANC = geological acid neutralisation capacity; SRP = soil reduction potential; GRP = geological reduction potential; OLF = overland flow; DD = deep drainage (vertical soil profile drainage); LAT = lateral drainage (horizon permeable drainage); ART = artificial drainage (subsurface mole-pipe and open ditch drainage); BP = bypass (soil moisture deficit induced cracking of clay-rich soil and bypass of the soil matrix).

followed by macro-scale weathering PAGs and their spatially correlated proxies were the most sensitive predictors retained for the weathering processes.

The performance of the hydrological, redox, and weathering PAGs, to estimate spatial variability in hydrochemical facies, as defined by the 12 surface water HCA clusters, was also strong with a cross-validated R<sup>2</sup> of 0.90 for both the user and machine defined model runs. In both model runs, PAGs representing all three dominant processes, were retained as significant predictors reflecting the collective influence of each process over surface water hydrochemical composition.

Although not tabulated, machine defined runs of conductivity (μS/cm), major cations, SiO<sub>2</sub>, I, F, B, DIC, HCO<sub>3</sub><sup>-</sup> alkalinity, and SO<sub>4</sub><sup>2-</sup> exhibited responses consistent with the process knowledge developed in Step 1 and achieved cross-validated R<sup>2</sup> values of ≥0.85. Although model responses were consistent with process knowledge, the lower accuracy of DO (R<sup>2</sup> = ≤0.67) was probably due to re-equilibration of soil and

aquifer drainage with atmospheric oxygen within the stream channel, a factor that is not accounted for by the PAGs.

### 3.3.2. Water quality

As PAGs were observed to adequately represent regional hydrochemical process gradients, land use intensity (LUI) was subsequently incorporated and the ability to estimate spatial variation in water quality assessed. User defined functions responded in a manner consistent with process knowledge and produced cross-validated R<sup>2</sup> values of 0.65 to 0.91, indicative of a moderate to very good estimation of spatial variation in steady-state water quality for Southland (Table 6). For example, total nitrogen and total oxidised nitrogen increased across the macro-scale ATM and RCD PAGs and decreased across the meso-scale SRP and GRP PAGs (Appendix D).

Overall, machine defined runs once again produced less complex and more accurate water quality models than user defined models (R<sup>2</sup> 0.72 to 0.92; Table 6). Machine defined models for particulate species,

*E. coli* ( $R^2 = 0.72$ ) and total suspended sediment ( $R^2 = 0.73$ ), were significantly less complex and more accurate than user defined models ( $R^2$  of 0.65 and 0.68, respectively). Varying degrees of spatial correlation between macro and meso and meso scale PAG and LUI was observed and was readily interpretable (Appendix C). For example, LUI exhibits a relatively strong positive correlation ( $r = 0.76$  and  $0.63$ ) with the macro-scale RCD and ATM PAGs, respectively reflecting the controls of topography and climate over the regional LUI gradient. Spatial correlation also appears to be the main difference between the user and the machine defined models of water quality, with the inclusion and retention of finer scale PAGs (e.g. BP, ART) and in some instances LUI in place of larger scale redox, hydrological and weathering PAGs.

Land use intensity was retained as an important estimator for all water quality measures except for total phosphorus and *E. coli*. In both instances, ART appears to be a more sensitive predictor of high risk areas for these water quality measures due to an inherent spatial correlation with areas of intensive land use, drainage modification, and a greater surficial runoff risk (% OLF).

## 4. Discussion

### 4.1. PoAM performance

Model evaluation indicates that PoAM was effective at replicating the actual process-attribute gradients that govern the hydrochemical evolution of surface water across the Southland province. A random forest evaluation of the performance of the macro-scale hydrological and meso-scale redox and weathering PAGs for Southland drew a similar conclusion (Snelder, 2016). The observation that spatial variation in hydrochemical measures can be reliably estimated as a function of landscape attributes is consistent with national and international studies (Johnson et al., 1997; Hale et al., 2004; Snelder and Biggs, 2002; Snelder et al., 2004; King et al., 2005; Dow et al., 2006; Becker et al., 2014). When spatial correlation was accounted for, there was little difference in the response and performance of user defined versus machine defined models of the dominant processes controlling hydrochemical evolution of surface water in

Southland - supporting the rigour of the process level knowledge and indicating that the generated PAGs reasonably replicated actual gradients.

The PoAM approach, when combined with a LUI gradient, also effectively explained spatial variation in steady-state total nitrogen, total oxidised nitrogen, total phosphorus, and dissolved reactive phosphorus across the 93 surface water sites for Southland. Overall, machine defined models of water quality provided less complex and more accurate estimates of spatial variation in water quality. Both user and machine defined models of *E. coli* ( $R^2 = 0.72$ ) and total suspended sediment ( $R^2 = 0.73$ ) were the least accurate water quality indices. The lower accuracy of models for the particulate species *E. coli* and total suspended sediment may reflect: (i) failure to adequately represent important PAGs; (ii) use of LUI gradients that are too coarse, and/or; (iii) the need to normalise highly positively skewed water quality metrics by flow prior to model evaluation and development. A recent application of PoAM within the Waituna Lagoon catchment (c. 22,000 ha), Southland, noted poor representation of enterprise type (e.g. agricultural versus conservation estate) at scales <300 ha, was a key source of estimation error for particulate and particulate bound species (Pearson et al., 2018; Rissmann et al., 2018b). Therefore, it is likely that improvements in the spatial resolution of landscape attributes, LUI, and/or consideration of different statistical measures are required in order to provide more accurate estimates of all water quality measures, especially particulate and particulate bound species.

In terms of model performance, bias is associated with the subset of 93 surface water monitoring sites, that for pragmatic reasons, are associated with higher order streams ( $\geq 3$ ) and larger drainage basin areas. Specifically, 50% of the 93 capture zones are  $\geq 70$  km<sup>2</sup>, with only eight <5 km<sup>2</sup> (500 ha). This is relevant because most farms across Southland range between 300 and 500 ha (Moran et al., 2017). Average farm size globally is even smaller (Samberg et al., 2016). As such, the ability of the PoAM approach to estimate water quality at property scales is currently unknown. This is important given that error within water quality models tends to increase as drainage basin size decreases, reflecting increasing sensitivity to the resolution, and hence accuracy, of landscape attributes (Troy et al., 2008; Matott et al., 2009; Moriasi et al., 2015).

**Table 6**  
User defined and machine defined water quality model and performance for dominant processes, Southland province, New Zealand.

	User Functions	User Defined Model	R <sup>2</sup>	Complexity	Machine Functions	Machine Defined Model	R <sup>2</sup>	Complexity
<b>Nitrogen</b>	TN = f(SRP, GRP, RCD, ATM, BP, LUI)	$TN = 0.15 + 0.15*ATM + 0.37*ATM^2 + 0.15*ATM^3 + \exp(0.63*LUI) - 1.22*\cos(0.68*LUI^2*ATM^2*BP^3)$	0.91	48	TN = f(LUI, BP, OLF, ATM)	$TN = 0.058 + 0.60*LUI + 0.087*BP + LUI*GRP*LAT*BP - LAT*\sin(0.60*LUI^3*BP^2) - 0.28*OLF - 0.22*LUI*OLF$	0.92	42
	TON = f(LUI, SRP, GRP, ATM, RCD, BP)	$TON = 0.75*LUI + 0.28*LUI^2 + 1.30*LUI*RCD*BP + LUI*BP*\sin(0.75*LUI*BP^2) - 1.53*LUI*\sin(0.75*LUI*BP^2)$	0.85	47	TON = f(RCD, LUI, BP)	$TON = 0.68*LUI + RCD*BP + 0.24*LUI^2 + 1.12*BP*RCD^2 + 0.16*BP^2*\sin(0.74*LUI*BP^2)$	0.85	38
<b>Phosphorus</b>	TP = f(LUI, SRP, GRP, RCD, ATM, DD, ART, BP)	$TP = \sin(0.60 + 0.70*ART - ATM^2 - 0.42*ATM - 0.54*ART*ATM^2) - 0.60*LUI^2*\sin(0.82*SRP*BP)$	0.72	39	TP = f(ATM, ART, GRP)	$TP = 0.6836*ART*\cos(ATM) + \cos(ATM) + \tan(0.2321*GRP^2) - 0.9595$	0.84	52
	DRP = f(LUI, SRP, GRP, RCD, ATM, ART, BP)	$DRP = 0.10*LUI*GRP + 0.10*SRP*BP + \sin(0.36*LUI + 0.16*GRP + 0.10*SRP + 0.36*LUI*BP + 0.17*GRP*BP^2) - 0.10 - 0.15*RCD$	0.81	84	DRP = f(GANC, GRP, BP, LUI, SRP)	$DRP = 0.063*factorial(GANC) + 0.067*LUI*GANC^2 + BP/(0.93*factorial(GANC) + SRP*\cos(BP) + factorial(GRP)) - 0.36$	0.81	41
<b>Sediment</b>	TSS = f(LUI, RCD, ATM, ART, DD, BP, OLF)	$TSS = (0.93 + 0.47*ART + 0.56*RCD*ATM - ATM^2 - 0.18*DD^2)/factorial(BP + ART*BP - 0.31 - BP*DD^2)$	0.65	38	TSS = f(ATM, DD, BP, GRP, SRP, RCD, LUI, ART)	$TSS = ATM*\cos(SRP) + 2.311*\cos(SRP)*\cos(DD)*\cos(BP)*\cos(1.392 + RCD*GRP*DD) - 0.2832*LUI*ART$	0.73	40
<b>Microbes</b>	<i>E. coli</i> = f(LUI, RCD, ATM, ART, DD, BP, OLF)	$E. coli = (0.20 + 0.70*LUI + 0.70*ATM*ART - 0.70*RCD*BP)/\cos(0.24*ART*BP^2) - 0.75*ATM^2 - 0.17*RCD*ATM*ART$	0.68	73	<i>E. coli</i> = f(ATM, ART, BP, LAT, GRP)	$E. coli = 1.428*\sin(0.3939 + 0.3079*ATM + 0.3079*GRP*ART - 0.3079*ATM*LAT) - 0.1168*BP - 0.2441*ART^2 - 0.5312*ATM^2$	0.72	38

For detailed performance results see Appendix D. HCA = Hierarchical Cluster Analysis membership; ATM = atmospheric loading; RCD = recharge domain; SANC = soil acid neutralisation capacity; GANC = geological acid neutralisation capacity; SRP = soil reduction potential; GRP = geological reduction potential; OLF = overland flow; DD = deep drainage (vertical soil profile drainage); LAT = lateral drainage (horizon permeable drainage); ART = artificial drainage (subsurface mole-pipe and open ditch drainage); BP = bypass (soil moisture deficit induced cracking of clay-rich soil and bypass of the soil matrix).

The loss of accuracy at small scales is particularly relevant given most geospatial layers used in the case study were only as fine as 1:50,000 scale, with finer scales often considered more appropriate for assessing controls at property scales.

Options for extending the relevance of PoAM towards property and paddock (<10 ha) scales include the use of passive (e.g. airborne gamma ray spectroscopy) and remotely sensed datasets (e.g. MODIS: Moderate Resolution Imaging Spectroradiometer satellite) that offer high resolution constraints over effective landscape properties (Beamish, 2014; Wilford et al., 2016; Rissmann et al., 2018d). That higher resolution and more accurate depiction of landscape attributes can be used to improve the representation of PAGs is hardly surprising given that the same dominant processes also drive hydrochemical and water quality variation at small scales (Moldan and Černý, 1994; Clark and Fritz, 1997; James and Roulet, 2006; Inamdar, 2011; Tratnyek et al., 2012).

Overall, the validity and observed performance of the PoAM approach is supported by the 'dominant process concept' proposed by Grayson and Bloschl (2000), namely:

- (i) that the response of environmental systems is often well explained by the representation of a small number of dominant processes, and;
- (ii) that "a logical way to identify the dominant processes governing a system is by evaluating the sensitivity of the system to each of the individual processes (believed to have influence) through a (high-order) multi-variable sensitivity analysis and selecting those variables that are found to have a 'noticeably significant influence" (Sivakumar, 2004, 2008).

#### 4.2. Water quality and hydrochemical data needs

Collection of hydrochemical data for PoAM can be expensive, with variable data-richness between catchments and provinces. To overcome differences in data-richness, a top-down hydrochemical controlling factor approach can be utilised for data-poor areas, and a stronger hybrid approach, PoAM, that incorporates hydrochemical measures for areas with a greater number of monitoring points and associated repeat measures. The role of different landscape attributes over many of the key hydrochemical and water quality measures is well established, and as such, can be used to produce a reasonable proxy of likely PAGs, without the need for large hydrochemical or water quality datasets. This allows for targeted water sampling, minimising the number of samples that are required for model refinement and validation. Further, as the approach is looking at process signals, historical water quality data is considered a valuable source of process level information.

For example, application of the method to the Northland Region of New Zealand resulted in comparable performance to that reported for Southland for many of the same water quality measures (Rissmann et al., 2018c, unpublished data). The latter lends support to the observation that a strong understanding of landscape controls over the dominant processes governing water quality outcomes can be used to produce a reasonable proxy of likely PAGs, without the need for large hydrochemical or water quality datasets.

#### 4.3. The influence of lagged contaminant transport

Water quality and compositional measures are by their nature a by-product of the inherent properties of the landscape, which also includes lags, so that any potential limitation is a factor of the currency and representativeness (i.e. adequate sampling of flow range) of existing water monitoring datasets, and the adequacy with which hydrological connectivity has been mapped. The PoAM approach assesses hydrological connectivity in terms of the sensitivity, response, and accuracy with which the conservative hydrological tracers, volumetric (i.e.  $\delta^{18}\text{O}\text{-H}_2\text{O}$ ,  $\delta^2\text{H}\text{-H}_2\text{O}$ ), and solute (e.g.  $\text{Na}^+$ ,  $\text{Cl}^-$ , Br) were estimated across long-

term surface water monitoring sites. The model proposes that the ability to accurately estimate spatial variation in hydrological tracers, especially solute tracers, is unambiguous evidence that hydrological connectivity, and hence lags, have been sufficiently represented for dissolved species. Any lags associated with stored particulate or particulate bound species is not represented by the hydrological model.

## 5. Conclusion

Process-attribute mapping utilises process signals within water to identify effective landscape attributes and to guide the classification of the landscape via a hybrid controlling factor approach. Due to a focus on mapping dominant process gradients (e.g. redox, hydrological), opposed to a singular reaction (e.g. denitrification) or component process (e.g. overland flow), PoAM accounts for multiple different water quality measures and is not limited by reliance upon poorly constrained coefficients or 'decay constants' to represent transport and attenuation. The combination of PAGs with symbolic regression to produce explicit mathematical models that describe spatial variation in hydrochemistry and water quality, offers greater transparency over both model response and the underlying drivers of spatial variability for multiple water quality measures.

The strong performance of PoAM to estimate a wide range of hydrochemical and water quality measures highlights the importance of landscape attributes over spatial variation in water composition. PAGs provide a visual-spatial representation of the 'effective' attributes of the landscape that mediate land use pressure and resulting water quality in distinct, and yet predictable ways. As such, they can be used in conjunction with land use activity to communicate, at a dominant process and landscape level, 'how' and 'why' water quality varies spatially - a feature that is highly valued by land users and resource managers (Duncan, 2016; Hughes et al., 2016; Pearson et al., 2018; Rissmann et al., 2018a).

Uncertainties associated with PoAM reflect the resolution of existing geospatial layers typically created for purposes other than water quality and that of the surface water monitoring network used for model evaluation. Both are particularly limiting for application at farm scales, although the utilisation of higher resolution, passive or remotely sensed data, offers an opportunity to overcome this limitation via the provision of higher resolution maps of effective landscape attributes. The PoAM approach does not currently include a transient component, although this is a current focus of research along with the generation of higher resolution depictions of effective landscape attributes.

Process-attribute mapping reveals the value of understanding the relationship between the dominant processes controlling spatial variation in the hydrochemical evolution of water and in conjunction with land use, water quality. As such, PoAM is considered to provide a more realistic, transparent, accurate and accessible picture of the processes governing spatial variability in hydrochemistry and water quality to land users and resource managers.

## Acknowledgments

This work was supported in part by a research grant from Our Land and Water, National Science Challenge, New Zealand, under the Physiographic Environments of New Zealand project. Hydrochemical and water quality data is supplied by Environment Southland (Southland Regional Council). The early conceptual hydrochemical understanding that underpins this work occurred under the umbrella of the 'Physiographics of Southland' science project, Environment Southland. We acknowledge Dr. Mike Friedel, Lincoln Agritech and Dr. Shailesh Singh, National Institute for Water and Atmosphere, New Zealand for review and commentary over the numerical modelling approach taken here. We thank the Waterways Centre for Freshwater Management and affiliated staff at both Lincoln University and the University of Canterbury for their support and interest in this work. Finally, we thank three unnamed reviewers for their comments which greatly improved the clarity and rigour of this work.

## Appendix A. Conceptual framework

The PoAM framework recognises that a single landscape attribute may drive spatial variation in more than one PAG as is seen in the landscape attribute 'soil drainage class,' which may be common to both soil zone redox potential (Klefoth et al., 2014; Killick et al., 2015; Clough et al., 2017; Rissmann et al., 2018b) and the hydrological pathway taken by water across the land surface (McDowell et al., 2005; Lin, 2012; Pearson, 2015a, 2015b; Rissmann et al., 2018b). Therefore, processes may be spatially linked at the attribute level, although the relative steepness in process signals may not be correlated across the same scale due to fundamental differences in process mechanics (e.g. microbially mediated redox succession versus saturation excess overland flow). Other attributes may be spatially independent (e.g. the electron donor abundance of geological units and topographic relief).

Spatial correlation also exists between inherent landscape attributes (e.g. slope and soil drainage class), and anthropogenic gradients such as land cover and land use (Harmsworth, 1996; Lynn et al., 2009) and drainage modification (Pearson, 2015a). The challenges of discriminating between 'natural' versus 'anthropogenic' influences over water quality measures are overcome in PoAM by combining PAGs from both inherent landscape attributes (i.e. hydrology, redox and weathering) and anthropogenically influenced attributes (i.e. land use intensity and artificial drainage).

Solute gradients are considered the most appropriate basis for identifying and mapping the landscape attributes most critical to water quality (Moldan and Černý, 1994; Clark and Fritz, 1997; Beyer et al., 2016a; Rissmann et al., 2018b, 2018c). Given a small volume of water with a high solute load (e.g.  $\text{NO}_3^-$  or  $\text{Cl}^-$ ) may rapidly overwhelm the solute concentration of very dilute (e.g. alpine sourced) waters, and yet result in little change in the signature of volumetric tracers (i.e.,  $\delta^{18}\text{O}\text{-H}_2\text{O}$  and  $\delta^2\text{H}\text{-H}_2\text{O}$ ).

## Appendix B. Process-attribute gradient descriptions and response

The atmospheric PAG depicts regional gradients in marine aerosol load and the  $\delta^{18}\text{O}\text{-H}_2\text{O}/\delta^2\text{H}\text{-H}_2\text{O}$  prior to redistribution by the hydrological network (Baisden et al., 2016; Rodway et al., 2016). Marine aerosol load decays exponentially with distance from the coast and altitude and  $\delta^{18}\text{O}\text{-H}_2\text{O}$  becomes progressively more negative (Rodway et al., 2016; Fig. 4a). Atmospheric gradients are steepest where elevation increases rapidly over a short lateral distance (e.g. west coast of Fiordland). The recharge domain PAG depicts the redistribution and temporary storage of water by the hydrological network (Beyer et al., 2016a; Rissmann et al., 2018a). This includes hydrological tracer distinct source water domains, between and within domain hydrological connectivity, surface and shallow aquifer connectivity and associated dilution potential (Fig. 4b; Beyer et al., 2016a).

In areas of intensive land use, overland flow is often associated with a disproportionately large contaminant load to surface waterways, especially sediment, phosphorus and microbes (Smith and Monaghan, 2003; Curran Cournane et al., 2011; Goldsmith and Ryder, 2013; Orchiston et al., 2013; Rissmann et al., 2018b). Across Southland however, percent precipitation is most prevalent pathway across alpine and subalpine recharge domains (Pearson, 2015b; Rissmann et al., 2016a). These areas are mostly characterised by natural state land cover and as such are considered 'source limited' and exhibit little, if any, evidence of anthropogenic impact (Fig. 4c and 1). Percent OLF occurring within developed hill country or lowland recharge domains, although constituting a smaller area, poses the greatest risk to water quality due to higher land use intensity and low dilution potential (Hughes et al., 2016; Pearson et al., 2018; Rissmann et al., 2018b).

Deep drainage, BP, LAT and ART constitute different pathways water takes through the soil zone (Pearson, 2015a, 2015b; Beyer et al., 2016b). Deep drainage is highest across areas of gravely silt loam soils with moderate to rapid permeability where imperfect drainage, textural contrasts, pan formation or shallow bedrock is largely absent (Fig. 4d and e; Pearson, 2015a, 2015b; Rissmann et al., 2016c, 2016d, 2016e). In agricultural areas, imperfectly drained soils (low deep drainage) and high lateral flow (horizon permeable flow) are most often artificially drained providing an artificial conduit for contaminants to bypass the soil matrix (Fig. 4f; Pearson, 2015a; Rissmann et al., 2018b). In areas of high LUI, BP mediated by shrink-swell clays (Fig. 4b) and high ART result in higher concentrations of contaminants entering both shallow aquifers and the surface water network (Houlbrooke and Monaghan, 2009; Beyer et al., 2016a; Hughes et al., 2016; Rissmann et al., 2018b).

The redox and weathering processes are comprised of separate soil and geological PAGs (Figs. 5 and 6; Beyer et al., 2016b). The soil PAG overlies the geological PAG, except where bedrock outcrops (Rissmann et al., 2016a). Across the lowland areas of the Southland province, the geological PAGs represent the upper portion of the unconfined aquifer system (Rissmann et al., 2016a). Oxidic recharge waters are associated with areas of well drained soils (deep drainage) and alluvial gravel aquifers, whilst recharge from imperfectly drained and organic soils is reducing (Killick et al., 2015; Beyer et al., 2016b; Beyer and Rissmann, 2016). Permeable, unconfined aquifer systems composed of Quaternary gravels and/or fractured rock with low organic carbon content (electron donor abundance) tend to be strongly oxidising, whilst poorly permeable and/or carbon rich aquifers tend to be reducing (Rissmann, 2011; Beyer et al., 2016b; Rissmann et al., 2018b). Hill country streams have a mixed oxidic-anoxic (see Jürgen et al., 2009) redox signature due to thin, often organic carbon rich soils, that overlie poorly permeable bedrock. Here seasonal saturation at the contact with poorly permeable bedrock results in reducing soil waters (Fig. 5a; Rissmann et al., 2016c, 2016d). Where reducing, clay-rich soils crack in response to soil moisture deficit aquifer recharge bypasses (BP) the reducing soil matrix transporting oxidised waters to the underlying aquifer (Beyer et al., 2016b; Hughes et al., 2016; Rissmann et al., 2016g).

The ANC of soil (SANC) and geological (GANC) materials reflect the abundance of Lewis bases and the degree of weathering of the substrate (Fig. 6; Rissmann et al., 2016a; Rissmann et al., 2018b). Within a recharge domain, alkalinity, pH and base cation concentrations increase as the ANC of soil or underlying geology increases (Rissmann et al., 2016b, 2016c, 2016d, 2016e, 2016f; Rissmann et al., 2018b).

## Appendix C. Process-attribute gradient and land use intensity correlation matrix

**Table C.1**

Pearson correlation matrix using pairwise significance test ( $\alpha = 0.05$ ) for 93 long-term monitoring sites from Southland province, New Zealand.

		LUI	RCD	ATM	SANC	GANC	SRP	GRP	OLF	DD	LAT	ART	BP
LUI	Pearson r	1	0.7546	0.6335	0.2254	-0.1970	-0.3586	-0.3761	-0.8023	0.6965	-0.6639	0.6355	0.4248
	p (2-tailed)	-	<0.0001	<0.0001	0.0298	0.0584	0.0004	0.0002	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
RCD	Pearson r	0.7546	1	0.8942	-0.1374	-0.4774	0.0525	0.1718	-0.7420	0.4607	-0.4137	0.7679	0.3145
	p (2-tailed)	<0.0001	-	<0.0001	0.1891	<0.0001	0.6169	0.0996	<0.0001	<0.0001	<0.0001	<0.0001	0.0021
ATM	Pearson r	0.6335	0.8942	1	-0.3681	-0.6122	0.0985	0.2564	-0.7051	0.3244	-0.3036	0.7333	0.1606
	p (2-tailed)	<0.0001	<0.0001	-	0.0003	<0.0001	0.3477	0.0131	<0.0001	0.0015	0.0031	<0.0001	0.1241

Table C.1 (continued)

		LUI	RCD	ATM	SANC	GANC	SRP	GRP	OLF	DD	LAT	ART	BP
SANC	Pearson r	0.2254	-0.1374	-0.3681	1	0.4449	-0.2307	-0.4356	0.0212	0.2492	-0.2278	-0.2316	0.3465
	p (2-tailed)	0.0298	0.1891	0.0003	-	<0.0001	0.0261	<0.0001	0.8399	0.0160	0.0281	0.0255	0.0007
GANC	Pearson r	-0.1970	-0.4774	-0.6122	0.4449	1	-0.1197	-0.5045	0.5108	-0.0302	0.0337	-0.4776	0.1471
	p (2-tailed)	0.0584	<0.0001	<0.0001	<0.0001	-	0.2531	<0.0001	<0.0001	0.7736	0.7481	<0.0001	0.1595
SRP	Pearson r	-0.3586	0.0525	0.0985	-0.2307	-0.1197	1	0.6514	0.1733	-0.7846	0.8189	0.0767	0.0864
	p (2-tailed)	0.0004	0.6169	0.3477	0.0261	0.2531	-	<0.0001	0.0966	<0.0001	<0.0001	0.4648	0.4102
GRP	Pearson r	-0.3761	0.1718	0.2564	-0.4356	-0.5045	0.6514	1	-0.0510	-0.5340	0.5554	0.2126	-0.1763
	p (2-tailed)	0.0002	0.0996	0.0131	<0.0001	<0.0001	<0.0001	-	0.6272	<0.0001	<0.0001	0.0407	0.0910
OLF	Pearson r	-0.8023	-0.7420	-0.7051	0.0212	0.5108	0.1733	-0.0510	1	-0.4435	0.4100	-0.7556	-0.3031
	p (2-tailed)	<0.0001	<0.0001	<0.0001	0.8399	<0.0001	0.0966	0.6272	-	<0.0001	<0.0001	<0.0001	0.0031
DD	Pearson r	0.6965	0.4607	0.3244	0.2492	-0.0302	-0.7846	-0.5340	-0.4435	1	-0.9874	0.1665	0.0561
	p (2-tailed)	<0.0001	<0.0001	0.0015	0.0160	0.7736	<0.0001	<0.0001	<0.0001	-	<0.0001	0.1106	0.5935
LAT	Pearson r	-0.6639	-0.4137	-0.3036	-0.2278	0.0337	0.8189	0.5554	0.4100	-0.9874	1	-0.1444	-0.0293
	p (2-tailed)	<0.0001	<0.0001	0.0031	0.0281	0.7481	<0.0001	<0.0001	<0.0001	<0.0001	-	0.1672	0.7802
ART	Pearson r	0.6355	0.7679	0.7333	-0.2316	-0.4776	0.0767	0.2126	-0.7556	0.1665	-0.1444	1	0.4023
	p (2-tailed)	<0.0001	<0.0001	<0.0001	0.0255	<0.0001	0.4648	0.0407	<0.0001	0.1106	0.1672	-	0.0001
BP	Pearson r	0.4248	0.3145	0.1606	0.3465	0.1471	0.0864	-0.1763	-0.3031	0.0561	-0.0293	0.4023	1
	p (2-tailed)	<0.0001	0.0021	0.1241	0.0007	0.1595	0.4102	0.0910	0.0031	0.5935	0.7802	0.0001	-

Correlation is significant at the 0.05 level (2-tailed). LUI = land use intensity; ATM = atmospheric loading; RCD = recharge domain; SANC = soil acid neutralisation capacity; GANC = geological acid neutralisation capacity; SRP = soil reduction potential; GRP = geological reduction potential; OLF = overland flow; DD = deep drainage (vertical soil profile drainage); LAT = lateral drainage (horizon permeable drainage); ART = artificial drainage (subsurface mole-pipe and open ditch drainage); BP = bypass (soil moisture deficit induced cracking of clay-rich soil and bypass of the soil matrix).

Appendix D. Hydrochemical and water quality response and performance

Table D.1

User defined hydrochemical model performance for dominant processes, Southland province, New Zealand.

Dominant processes	User defined functions	PAG	Sensitivity	Positive (%)	Positive mag.	Negative (%)	Negative mag.	R <sup>2</sup>	r	ME	MSE	MAE	Coeff.	Complexity		
Hydrology	Cl = f(ATM, RCD)	ATM	1.33	100%	1.33	0%	0.00	0.94	0.97	0.83	0.05	0.15	8	43		
		RCD	0.45	100%	0.45	0%	0.00									
	Na = f(ATM, RCD)	ATM	2.06	100%	2.06	0%	0.00	0.95	0.97	0.93	0.05	0.15	8	45		
		RCD	0.38	100%	0.38	0%	0.00									
	Br = f(ATM, RCD)	RCD	0.75	100%	0.75	0%	0.00	0.75	0.87	1.62	0.27	0.34	4	33		
		ATM	0.65	100%	0.65	0%	0.00									
	δ <sup>18</sup> O-H <sub>2</sub> O = f(ATM, RCD)	ATM	1.34	100%	1.34	0%	0.00	0.93	0.97	0.70	0.07	0.20	6	35		
		RCD	0.31	100%	0.31	0%	0.00									
	δ <sup>2</sup> H-H <sub>2</sub> O = f(ATM, RCD)	ATM	1.20	99%	1.22	1%	0.19	0.94	0.97	0.81	0.06	0.19	7	41		
		RCD	0.13	100%	0.13	0%	0.00									
	δ <sup>13</sup> C-DIC = f(ATM, RCD)	ATM	2.35	33%	1.45	67%	2.78	0.79	0.89	1.39	0.19	0.29	7	38		
		RCD	0.45	0%	0.00	100%	0.45									
DO = f(SRP, GRP, ATM, RCD, BP)	BP	1.02	35%	0.88	65%	1.10	0.56	0.75	1.82	0.33	0.37	7	46			
	GRP	0.65	53%	0.59	47%	0.72										
Redox	Mn <sup>2+</sup> = f(SRP, GRP, ATM, RCD, BP)	SRP	0.17	48%	0.14	52%	0.20									
		RCD	0.13	64%	0.14	36%	0.11									
	Fe <sup>2+</sup> = f(SRP, GRP, ATM, RCD, BP)	ATM	Not retained by model													
		ATM	1.00	81%	1.17	19%	0.30	0.87	0.94	0.72	0.08	0.21	6	29		
	DOC = f(SRP, GRP, ATM, RCD)	SRP	0.43	81%	0.31	19%	0.96									
		GRP	0.24	100%	0.24	0%	0.00									
	pH = f(SANC, GANC, RCD, ATM)	BP	0.22	100%	0.22	0%	0.00									
		RCD	Not retained by model													
	Total alkalinity = f(SANC, GANC, RCD, ATM)	ATM	1.04	100%	1.04	0%	0.00	0.83	0.91	2.07	0.13	0.21	8	45		
		SRP	0.13	100%	0.13	0%	0.00									
	Weathering	HCA# = f(SRP, ATM, GRP, SANC, GANC)	BP	0.10	34%	0.01	66%	0.14								
			RCD	0.04	0%	0.00	100%	0.04								
pH = f(SANC, GANC, RCD, ATM)		GRP	Not retained by model													
		ATM	6.83	100%	6.83	0%	0.00	0.94	0.97	0.64	0.03	0.12	8	32		
Total alkalinity = f(SANC, GANC, RCD, ATM)		RCD	0.43	100%	0.43	0%	0.00									
		GRP	0.08	56%	0.06	44%	0.17									
Total alkalinity = f(SANC, GANC, RCD, ATM)		SRP	0.01	100%	0.01	0%	0.00									
		GANC	0.31	46%	0.34	54%	0.27	0.79	0.89	0.60	0.04	0.15	9	50		
Total alkalinity = f(SANC, GANC, RCD, ATM)		RCD	0.18	100%	0.18	0%	0.00									
		SANC	0.17	100%	0.51	0%	0.00									
Total alkalinity = f(SANC, GANC, RCD, ATM)		ATM	<0.001	0%	0.00	100%	<0.001									
		RCD	0.62	100%	0.62	0%	0.00	0.81	0.90	1.87	0.14	0.26	7	40		
Total alkalinity = f(SANC, GANC, RCD, ATM)	GANC	0.38	46%	0.32	54%	0.43										
	SANC	0.25	100%	0.25	0%	0.00										
Total alkalinity = f(SANC, GANC, RCD, ATM)	ATM	Not retained by model														
	RCD	0.62	100%	0.62	0%	0.00	0.90	0.95	0.97	0.11	0.21	5	37			
Total alkalinity = f(SANC, GANC, RCD, ATM)	ATM	0.32	100%	0.32	0%	0.00										
	GANC	0.20	0%	0.00	100%	0.00										
Total alkalinity = f(SANC, GANC, RCD, ATM)	GRP	0.05	0%	0.00	100%	0.05										
	SRP	Not retained by model														

(continued on next page)



Table D.2 (continued)

Process family	Water quality measure	PAG retained by model	Sensitivity (%)	Positive (%)	Positive magnitude	Negative (%)	Negative magnitude	R <sup>2</sup>	r	ME	MSE	MAE	Coefficients	Complexity
Hydrochemical cluster membership	Total alkalinity	BP	2.056	74%	1.887	26%	2.540	0.90	0.95	0.94	0.07	0.16	9	41
		ATM	0.057	100%	0.057	0%	0.000							
		RCD	0.056	100%	0.056	0%	0.000							
		GANC	0.052	100%	0.052	0%	0.000							
		ART	0.032	48%	0.037	52%	0.028							
	Cluster membership	ATM	0.418	100%	0.418	0%	0.000	0.90	0.95	1.04	0.10	0.19	6	34
		ART	0.325	91%	0.340	9%	0.163							
		LUI	0.256	99%	0.259	1%	0.254							
		BP	0.025	100%	0.025	0%	0.000							
		GANC	0.023	100%	0.023	0%	0.000							
	GRP	0.007	100%	0.007	0%	0.000								

where R<sup>2</sup> = cross-validated goodness of fit; r = correlation coefficient; ME = maximum error; MSE = mean squared error; MAE = mean absolute error; Coeff. = the number of coefficients retained by the model; Complexity = the complexity of the model; ATM = atmospheric loading; RCD = recharge domain; SANC = soil acid neutralisation capacity; GANC = geological acid neutralisation capacity; SRP = soil reduction potential; GRP = geological reduction potential; OLF = overland flow; DD = deep drainage (vertical soil profile drainage); LAT = lateral drainage (horizon permeable drainage); ART = artificial drainage (subsurface mole-pipe and open ditch drainage); BP = bypass (soil moisture deficit induced cracking of clay-rich soil and bypass of the soil matrix).

Table D.3

Machine defined water quality model performance, Southland province, New Zealand.

Water Quality Measure	PAG Retained by Model	Sensitivity	Positive (%)	Positive Magnitude	Negative (%)	Negative Magnitude	R <sup>2</sup>	r	ME	MSE	MAE	Coeff.	Complexity
Total Nitrogen (TN)	DD	4.95	100%	4.95	0%	0.00	0.89	0.95	1.05	0.09	0.22	7	39
	BP	4.25	100%	4.25	0%	0.00							
	OLF	0.72	0%	0.00	100%	0.72							
	LUI	0.43	100%	0.43	0%	0.00							
Total Oxidisable Nitrogen (TON)	LAT	0.50	0%	0.00	100%	0.50	0.87	0.94	1.44	0.12	0.21	6	31
	BP	0.46	100%	0.46	100%	0.00							
	LUI	0.32	100%	0.32	100%	0.00							
Total Phosphorus (TP)	ATM	1.08	46%	1.17	54%	1.00	0.84	0.92	1.43	0.10	0.21	7	52
	ART	0.69	100%	0.69	0%	0.00							
	GRP	0.49	24%	1.10	76%	0.30							
Dissolved Reactive Phosphorus (DRP)	GANC	1.43	36%	1.66	64%	1.30	0.81	0.90	1.10	0.04	0.13	4	41
	GRP	1.39	14%	5.42	86%	0.74							
	BP	0.80	100%	0.80	0%	0.00							
	LUI	0.18	100%	0.18	0%	0.00							
<i>E. coli</i> (MPN)	SRP	0.18	100%	0.18	0%	0.00	0.72	0.85	1.74	0.20	0.31	8	38
	ATM	1.38	60%	1.74	40%	0.86							
	ART	0.67	49%	0.72	51%	0.62							
	BP	0.15	0%	0.00	100%	0.15							
	LAT	0.13	97%	0.14	3%	0.01							
Total Suspended Sediment (TSS)	GRP	0.07	0%	0.00	100%	0.07	0.73	0.86	1.53	0.20	0.31	3	40
	ATM	0.86	100%	0.86	0%	0.00							
	DD	0.25	43%	0.27	57%	0.24							
	BP	0.23	63%	0.20	37%	0.30							
	GRP	0.19	4%	0.06	96%	0.19							
	SRP	0.17	56%	0.17	44%	0.16							
	RCD	0.08	9%	0.05	91%	0.08							
	LUI	0.04	100%	0.04	0%	0.00							
	ART	0.001	100%	0.001	0%	0.00							

Where R<sup>2</sup> = cross-validated goodness of fit; r = correlation coefficient; ME = maximum error; MSE = mean squared error; MAE = mean absolute error; Coeff. = the number of coefficients retained by the model; Complexity = the complexity of the model; LUI = land use intensity; ATM = atmospheric loading; RCD = recharge domain; SANC = soil acid neutralisation capacity; GANC = geological acid neutralisation capacity; SRP = soil reduction potential; GRP = geological reduction potential; OLF = overland flow; DD = deep drainage (vertical soil profile drainage); LAT = lateral drainage (horizon permeable drainage); ART = artificial drainage (subsurface mole-pipe and open ditch drainage); BP = bypass (soil moisture deficit induced cracking of clay-rich soil and bypass of the soil matrix).

## References

- Aho, K., Derryberry, D., Peterson, T., 2014. Model selection for ecologists: the worldviews of AIC and BIC. *Ecology* 95, 631–636. <https://doi.org/10.1890/13-1452.1>.
- Baisden, W.T., Keller, E.D., Van Hale, R., Frew, R.D., Wassenaar, L.L., 2016. Precipitation isoscapes for New Zealand: enhanced temporal detail using precipitation-weighted daily climatology. *Isot. Environ. Health Stud.* 52 (4–5), 343–352. <https://doi.org/10.1080/10256016.2016.1153472>.
- Beamish, D., 2014. Peat Mapping Associations of Airborne Radiometric Survey Data. , pp. 521–539 <https://doi.org/10.3390/rs6010521>.
- Becker, J.C., Rodibaugh, K.J., Labay, B.J., Bonner, T.H., Zhang, Y., Nowlin, W.H., 2014. Physiographic gradients determine nutrient concentrations more than land use in a Gulf Slope (USA) river system. *Freshw. Sci.* 33 (3), 731–744. <https://doi.org/10.1086/676635>.
- Beven, K., Germann, P., 2013. Macropores and water flow in soils revisited. *Water Resour. Res.* 49 (6), 3071–3092. <https://doi.org/10.1002/wrcr.20156>.
- Beyer, M., Rissmann, C., 2016. Tile Drain and Soil Water Sample Assessment: Relationships Between Soil Denitrification Potential and Redox Processes. GNS Science, Lower Hutt, N.Z <https://doi.org/10.21420/G2PP4Q> GNS Science report 2016/60 ii, 16 p.
- Beyer, M., Rissmann, C., Rodway, E., Marapara, T.R., Hodgetts, J., 2016a. Technical Chapter 2: Discrimination of Recharge Mechanism and Vulnerability to Bypass Flow. Environment Southland, Invercargill, New Zealand Technical Report. No: 2016/3.
- Beyer, M., Rissmann, C., Rodway, E., Killick, M., Pearson, L., 2016b. Technical Chapter 6: Influence of Soil and Geological Composition over Redox Conditions for Southland Groundwater and Surface Waters. Environment Southland, Invercargill, New Zealand Technical Report. No: 2016/3.
- Bracken, L.J., Wainwright, J., Ali, G.A., Tetzlaff, D., Smith, M.W., Reaney, S.M., Roy, A.G., 2013. Concepts of hydrological connectivity: research approaches, pathways and future agendas. *Earth Sci. Rev.* <https://doi.org/10.1016/j.earscirev.2013.02.001>.
- Burbury, L., 2012. Analysis of Groundwater Level Data: Waituna Lagoon. Envirolink Advice Grant ESRC152 (Report No 1008-2-R1).
- Clark, I., Fritz, P., 1997. *Environmental Isotopes in Hydrogeology*. 2nd edition. CRC Press 1-56670-249-6.
- Clough, T.J., Lanigan, G.J., De Klein, C.A.M., Samad, M.S., Morales, S.E., Rex, D., Bakken, L.R., Johns, C., Condrion, L.M., Grant, J., Richards, K.G., 2017. Influence of soil moisture on

- codenitrification fluxes from a urea-affected pasture soil. *Sci. Rep.* <https://doi.org/10.1038/s41598-017-02278-y>.
- Curran Courmane, F., McDowell, R., Littlejohn, R., Condron, L., 2011. Effects of cattle, sheep and deer grazing on soil physical quality and losses of phosphorus and suspended sediment losses in surface runoff. *Agric. Ecosyst. Environ.* 140 (1), 264–272. <https://doi.org/10.1016/j.agee.2010.12.013>.
- Daughney, C., Rissmann, C., Friedel, M., Morgenstern, U., Hodson, R., van der Raaij, R., Rodway, E., Martindale, H., Pearson, L., Townsend, D., Kees, L., Moreau, M., Millar, R., Horton, T., 2015. Hydrochemistry of the Southland Region, GNS Science Report 2015/24.
- Doctor, D.H., Kendall, C., Sebestyen, S.D., Shanley, J.B., Ohte, N., Boyer, E.W., 2008. Carbon isotope fractionation of dissolved inorganic carbon (DIC) due to outgassing of carbon dioxide from a headwater stream. *Hydrol. Process.* <https://doi.org/10.1002/hyp.6833>.
- Dow, C.L., Arscott, D.B., Newbold, J.D., 2006. Relating major ions and nutrients to watershed conditions across a mixed-use, water-supply watershed. *J. N. Am. Benthol. Soc.* 25 (4), 887–911. [https://doi.org/10.1899/0887-3593\(2006\)025\[0887:RMIANT\]2.0.CO;2](https://doi.org/10.1899/0887-3593(2006)025[0887:RMIANT]2.0.CO;2).
- Dubčáková, R., 2011. Eureka: software review. *Genet. Program Evolvable Mach.* 12 (2), 173–178.
- Duncan, R., 2016. Ways of knowing – out-of-sync or incompatible? Framing water quality and farmers' encounters with science in the regulation of non-point source pollution in the Canterbury region of New Zealand. *Environ. Sci. Pol.* <https://doi.org/10.1016/j.envsci.2015.10.004>.
- Dymond, J.R., 2010. Soil erosion in New Zealand is a net sink of CO<sub>2</sub>. *Earth Surf. Process. Landf.* 35, 1763–1772 ([doi.org/10.1002/esp.2014](https://doi.org/10.1002/esp.2014)).
- Giller, P.S., Malmqvist, B., 1998. The biology of streams and rivers. *Biol. Habitats* <https://doi.org/10.1145/2789211>.
- Goldsmith, R., Ryder, G., 2013. Factors Affecting Contaminant Loss in Overland Flow. Ryder Consulting Limited, Dunedin, New Zealand Technical Review for Environment Southland.
- Gray, D.P., Harding, J.S., Elberling, B., Horton, T., Clough, T.J., Winterbourn, M.J., 2011. Carbon cycling in floodplain ecosystems: out-gassing and photosynthesis transmit soil δ<sup>13</sup>C gradient through stream food webs. *Ecosystems* <https://doi.org/10.1007/s10021-011-9430-1>.
- Grayson, R., Blochl, G., 2000. Summary of pattern comparison and concluding remarks. Spatial Patterns in Catchment Hydrology: Observations and Modelling. Cambridge University Press, Cambridge, UK.
- Güler, C., Thyne, G.D., McCray, J.E., Turner, A.K., 2002. Evaluation of graphical and multivariate statistical methods for classification of water chemistry data. *Hydrogeol. J.* 10 (4), 455–474. <https://doi.org/10.1007/s10040-002-0196-6>.
- Hale, S.S., Paul, J.F., Heltshe, J.F., Hale, S.S., Paul, J.F., Heltshe, J.F., 2004. Watershed landscape indicators of estuarine benthic condition. *Estuaries* 27 (2), 283–295. <https://doi.org/10.1007/BF02803385>.
- Harmsworth, G.R., 1996. Land Use Capability Classification of the Northland Region. A Report to Accompany the Second Edition (1:50 000) New Zealand Land Resource Inventory Worksheets. Landcare Research Science Series 9/Manaaki Whenua Press, Lincoln (269 p).
- Houlbrooke, D.J., Monaghan, R.M., 2009. The Influence of Soil Drainage Characteristics on Contaminant Leakage Risk Associated With the Land Application of Farm Dairy Effluent (AgResearch report prepared for Environment Southland).
- Hughes, B., 2001. In: Rosen, M.R., White, P.A., et al. (Eds.), Southland. New Zealand Hydrological Society, Wellington, N.Z. ISBN: 0-473-07816-3 498p.
- Hughes, B., Wilson, K., Rissmann, C., Rodway, E., 2016. Physiographics of Southland: development and application of a classification system for managing land use effects on water quality in Southland. Environment Southland Technical Report No. 2016/11 (Invercargill, New Zealand).
- Hume, T.M., Snelder, T., Weatherhead, M., Liefing, R., 2007. A controlling factor approach to estuary classification. *Ocean Coast. Manag.* <https://doi.org/10.1016/j.ocecoaman.2007.05.009>.
- Inamdar, S., 2011. In: Levina, D., Carlyle-Moses, D., Tanaka, T., et al. (Eds.), The use of geochemical mixing models to derive runoff sources and hydrologic flow paths. vol. 216. Springer, Netherlands. [https://doi.org/10.1007/978-94-007-1363-5\\_8](https://doi.org/10.1007/978-94-007-1363-5_8).
- James, A.L., Roulet, N.T., 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 (8). <https://doi.org/10.1029/2005WR004419>.
- Jamieson, R., Gordon, R., Sharples, K.E., Stratton, G., Madani, A., 2002. Movement and persistence of fecal bacteria in agricultural soils and subsurface drainage water: A review. *Can. Biosyst. Eng.* 44, 1.1–1.9.
- Johnson, L., Richards, C., Host, G., Arthur, J., 1997. Landscape influences on water chemistry in Midwestern stream ecosystems. *Freshw. Biol.* 37 (1), 193–208. <https://doi.org/10.1046/j.1365-2427.1997.d01-539.x>.
- Jürgen, B.C., McMahon, P.B., Chapelle, F.H., Eberts, S.M., 2009. An Excel® Workbook for Identifying Redox Processes in Ground Water.
- Kendall, C., McDonnell, J.J. (Eds.), 1998. Isotope tracers in catchment hydrology. Amsterdam, New York: Elsevier, p. 839 (p. 804–816).
- Khu, S.T., Liong, S.Y., Babovic, V., Madsen, H., Muttill, N., 2001. Genetic programming and its application in real-time runoff forecasting. *J. Am. Water Resour. Assoc.* <https://doi.org/10.1111/j.1752-1688.2001.tb00980.x>.
- Killick, M., Stenger, R., Rissmann, C., 2015. Estimating Soil Zone Denitrification for Southland. Technical Report. Environment Southland, Invercargill, New Zealand.
- King, R.S., Baker, M.E., Whigham, D.F., Weller, D.E., Jordan, T.E., Kazyak, P.F., Hurd, M.K., 2005. Spatial considerations for linking watershed land cover to ecological indicators in streams. *Ecol. Appl.* 15 (1), 137–153. <https://doi.org/10.1890/04-0481>.
- Klefoth, R.R., Clough, T.J., Oenema, O., Van Groenigen, J.W., 2014. Soil bulk density and moisture content influence relative gas diffusivity and the reduction of nitrogen-15 nitrous oxide. *Vadose Zone J.* <https://doi.org/10.2136/vzj2014.07.0089>.
- Krantz, D.E., Powars, D.S., 2000. Hydrogeologic setting and potential for denitrification in ground water, Coastal Plain of Southern Maryland. U.S. Geological Survey Water-Resources Investigations Report. 00–4051 19 p. <https://pubs.usgs.gov/wri/2000/4051/>.
- Kratzer, E.B., Jackson, J.K., Arscott, D.B., Aufdenkampe, A.K., Dow, C.L., Kaplan, L.A., Newbold, J.D., Sweeney, B.W., 2006. Macroinvertebrate distribution in relation to land use and water chemistry in New York City. *J. N. Am. Benthol. Soc.* 954–976 <https://doi.org/10.1899/0887-3593>.
- Kurtzman, D., Baram, S., Dahan, O., 2016. Soil-aquifer phenomena affecting groundwater under vertisols: a review. *Hydrol. Earth Syst. Sci.* <https://doi.org/10.5194/hess-20-1-2016>.
- Land Information New Zealand, 2012. NZ 8m Digital Elevation Model. [raster]. <https://data.linz.govt.nz/layer/51768-nz-8m-digital-elevation-model-2012/>.
- Landcare Research, 2010. Fundamental soil layer New Zealand soil classification. [vector polygon]. <https://iris.scinfo.org.nz/layer/48079-fsl-new-zealand-soil-classification/>.
- Landcare Research, 2015. Land Cover Database Version 4.1 Mainland New Zealand. [vector polygon]. <https://iris.scinfo.org.nz/layer/48423-lcdb-v41-land-cover-database-version-41-mainland-new-zealand/>.
- Langmuir, D., 1997. Aqueous Environmental Geochemistry. N.J. Prentice Hall, Upper Saddle River (c1997).
- Ledgard, G., 2013. Land Use Change in the Southland Region. Technical Report No. 2013–13/Environment Southland, Invercargill, New Zealand.
- Lin, H. (Ed.), 2012. Hydrogeology: Synergistic Integration of Soil Science and Hydrology. Academic Press. ISBN: 978-0-12-386941-8.
- Lydersen, E., Larssen, T., Fjeld, E., 2004. The influence of total organic carbon (TOC) on the relationship between acid neutralizing capacity (ANC) and fish status in Norwegian lakes. *Sci. Total Environ.* 326 (1–3), 63–69. <https://doi.org/10.1016/j.scitotenv.2003.12.005>.
- Lynn, I.H., Manderson, A.K., Page, M.J., Harmsworth, G.R., Eyles, G.O., Douglas, G.B., Mackay, A.D., Newsome, P.J.F., 2009. Land Use Capability Survey Handbook – A New Zealand Handbook for the Classification of Land. AgResearch Hamilton; Manaaki Whenua Lincoln; GNS Science, Lower Hutt, New Zealand.
- Macara, G.R., 2013. The Climate and Weather of Southland. 2nd edition. NIWA Science and Technology Series, No.631173-0382.
- Matott, L.S., Babendreier, J.E., Purucker, S.T., 2009. Evaluating uncertainty in integrated environmental models: a review of concepts and tools. *Water Resour. Res.* <https://doi.org/10.1029/2008WR007301>.
- McDowell, R.W., Monaghan, R.M., Wheeler, D., 2005. Modelling phosphorus losses from pastoral farming systems in New Zealand. *N. Z. J. Agric. Res.* 48 (1), 131–141.
- McMahon, P., Chapelle, F., 2008. Redox processes and water quality of selected principal aquifer systems. *Groundwater* 46 (2), 259–271.
- Ministry for the Environment, 2006. A National Protocol for State of the Environment Groundwater Sampling in New Zealand. Ministry for the Environment, Wellington, New Zealand 0-478-30112-X.
- Moldan, B., Černý, J.V., 1994. Biogeochemistry of Small Catchments: A Tool for Environmental Research. John Wiley and Sons Ltd, Chichester, England, UK.
- Moran, E., Pearson, L., Coudrey, M., Eyre, K., 2017. The Southland Economic Project: Agriculture and Forestry. Technical Report. Publication no. 2017-02. Environment Southland, Invercargill (340 p).
- Moreau, M., Hodson, R., 2015. Trends in Southland's water quality: a comparison between regional and national sites for groundwater and rivers. GNS Science Consultancy Report 2014/61 (110p).
- Moriasi, D.N., Gitau, M.W., Pai, N., and Daggupati, P. (2015). Hydrologic and water quality models: performance measures and evaluation criteria. (2015). *Trans. ASABE*, 58(6), 1763–1785. <https://doi.org/10.13031/trans.58.10715>.
- National Research Council. (2007). Models in Environmental Regulatory Decision Making. Washington, DC: The National Academies Press. <https://doi.org/10.17226/11972>.
- Newsome, P.F.J., Wilde, R.H., Willoughby, E.J., 2008. Land Resource Information System Spatial Data Layers. Landcare Research, Palmerston North, New Zealand.
- Orchiston, T.S., Monaghan, R., Laursen, S., 2013. Reducing overland flow and sediment losses from winter forage crop paddocks grazed by dairy cows. In: Currie, L.D., Christensen, C.L. (Eds.), Accurate and Efficient Use of Nutrients on Farms. Occasional Report No. 26. Fertilizer and Lime Research Centre, Massey University, Palmerston North, New Zealand <http://frc.massey.ac.nz/publications.html>. (7 pages).
- Parliamentary Commissioner for the Environment, 2018. Overseer and Regulatory Oversight: Models, Uncertainty and Cleaning up our Waterways. Wellington, New Zealand.
- Pearson, L., 2015a. Artificial Subsurface Drainage in Southland. Environment Southland, Invercargill, New Zealand Technical Report No. 2015-07.
- Pearson, L., 2015b. Overland Flow Risk in Southland. Environment Southland, Invercargill, New Zealand Technical Report No. 2015-06.
- Pearson, L., Coudrey, M., 2016. Methodology for GIS-Based Land Use Maps for Southland. Environment Southland, Invercargill, New Zealand Technical Report No. 2016-10.
- Pearson, L., Rissmann, C., Lindsay, J., 2018. Waituna Catchment: Physiographic Risk Assessment. Land and Water Science Report 2018/02. Prepared for Living Water (46p).
- Rissmann, C. (2011). Regional Mapping of Groundwater Denitrification Potential and Aquifer Sensitivity, 2011–12, 38p. <https://doi.org/10.13140/2.1.3412.5606>.
- Rissmann, C., Leybourne, M., Benn, C., Christenson, B., 2015. The origin of solutes within the groundwaters of a high Andean aquifer. *Chem. Geol.* <https://doi.org/10.1016/j.chemgeo.2014.11.029>.
- Rissmann, C., Rodway, E., Beyer, M., Hodgetts, J., Pearson, L., Killick, M., Marapara, T.R., Akbaripasad, A., Hodson, R., Dare, J., Millar, R., Ellis, T., Lawton, M., Ward, N., Hughes, B., Wilson, K., McMeeking, J., Horton, T., May, D., Kees, L., 2016a. Physiographics of Southland Part 1: Delineation of Key Drivers of Regional Hydrochemistry and Water Quality. Environment Southland Technical Report No. 2016/3 (Invercargill, New Zealand).
- Rissmann, C., Pearson, L., Killick, M., Rodway, E., Beyer, M., Marapara, T.R., Hodgetts, J., 2016b. Technical Chapter 3: Soil Chemistry. Environment Southland, Invercargill, New Zealand Technical Report. No: 2016/3.

- Rissmann, C., Rodway, E., Beyer, M., Pearson, L., Killick, M., Dare, J., 2016c. Technical Chapter 4: Relationship Between Soil Chemistry and Soil Water Chemistry. Environment Southland, Invercargill, New Zealand Technical Report. No: 2016/3.
- Rissmann, C., Rodway, E., Beyer, M., Marapara, T.R., Hodgetts, J., Pearson, L., Killick, M., 2016d. Technical Chapter 5: Comparison of Soil Water With Surface Water and Groundwater Chemistry. Environment Southland, Invercargill, New Zealand Technical Report. No: 2016/3.
- Rissmann, C., Rodway, E., Hughes, B., Beyer, M., Marapara, T.R., Hodgetts, J., 2016e. Technical Chapter 7: Spatial Controls Over Major Ion Facies of Southland's Ground and Surface Waters. Environment Southland, Invercargill, New Zealand Technical Report. No: 2016/3.
- Rissmann, C., Rodway, E., Hughes, B., Marapara, T.R., Hodgetts, J., 2016f. Technical Chapter 8: Geomorphic Surface Age and Substrate Compositional Influences Over Regional Hydrochemistry. Environment Southland, Invercargill, New Zealand Technical Report. No: 2016/3.
- Rissmann, C., Beyer, M., Rodway, E., Hodson, R., Ward, N., Ellis, T., Hodgetts, J., Pearson, L., Killick, M., Marapara, T., Akbaripasand, A., Dare, J., Kees, L., Snelder, T., Moriarty, E., Almond, P., Webster-Brown, J., 2016g. Shrink-swell (cracking) soils as important conduits of ground and surface water microbial contamination? Proceedings, New Zealand Freshwater Sciences Conference, Invercargill, New Zealand
- Rissmann, C., Pearson, L., Lindsay, J., Beyer, M., Marapara, T., Badenhop, A., Martin, A., 2018a. Integrated landscape mapping of water quality controls for farm planning – applying a high-resolution physiographic approach to the Waituna Catchment, Southland. In: Currie, L.D., Christensen, C.L. (Eds.), Farm environmental Planning – Science, Policy and Practice. Fertilizer and Lime Research Centre, Massey University, Palmerston North, New Zealand <http://frc.massey.ac.nz/publications.html>. Occasional Report No. 31. (19 pages).
- Rissmann, C., Pearson, L., Lindsay, J., Marapara, M., and Badenhop, A. (2018b). Waituna Catchment: Technical Information and Physiographic Application. Land and Water Science Report 2018/01. (Prepared for Living Water. 134p).
- Rissmann, C., Pearson, L., Lindsay, J., Couldrey, M., Lovett, A., 2018c. Application of physiographic science to the northland region: preliminary hydrological and redox process-attribute layers. Land and Water Science Report 2018/11, p. 88.
- Rissmann, C., Pearson, L., Lindsay, J., Couldrey, M., 2018d. Sediment Process-Attribute Layer for Northland. Land and Water Science Report 2018/35 p. 61.
- Rodway, E., Rissmann, C., Beyer, M., Marapara, T.R., Hodgetts, J. (2016). Technical Chapter 1: Precipitation. (Technical Report. No: 2016/3. Invercargill, New Zealand).
- Samberg, L.H., Gerber, J.S., Ramankutty, N., Herrero, M., West, P.C., 2016. Subnational distribution of average farm size and smallholder contributions to global food production. *Environ. Res. Lett.* 11 (12), 124010.
- Schmidt, M., and Lipson, H. (2009). Symbolic regression of implicit equations. *Genetic Programming Theory and Practice*. [https://doi.org/doi:https://doi.org/10.1007/978-1-4419-1626-6\\_5](https://doi.org/doi:https://doi.org/10.1007/978-1-4419-1626-6_5).
- Schmidt, M., Lipson, H., 2015. *Eureqa* (Version 1.24.0).
- Scott, D.T., Baisden, W.T., Davies-Colley, R., Gomez, B., Hicks, D.M., Page, M.J., Preston, N.J., Trustrum, N.A., Tate, K.R., Woods, R.A., 2006. Localised erosion affects national carbon budget. *Geophys. Res. Lett.* 33, L01402. <https://doi.org/10.1029/2005GL024644>.
- Sivakumar, B., 2004. Dominant processes concept in hydrology: moving forward. *Hydrol. Process.* <https://doi.org/10.1002/hyp.5606>.
- Sivakumar, B., 2008. Dominant processes concept, model simplification and classification framework in catchment hydrology. *Stoch. Env. Res. Risk A.* 22 (6), 737–748. <https://doi.org/10.1007/s00477-007-0183-5>.
- Smith, L.C., Monaghan, R.M., 2003. Nitrogen and phosphorus losses in overland flow from a cattle-grazed pasture in Southland. *N. Z. J. Agric. Res.* 46 (3), 225–237.
- Snelder, T., 2016. Performance testing of the physiographics of Southland using random forest (RF) models. Water Land People Report Prepared for Environment Southland.
- Snelder, T.H., Biggs, B.J.F., 2002. Multiscale river environment classification for water resources management. *J. Am. Water Resour. Assoc.* <https://doi.org/10.1111/j.1752-1688.2002.tb04344.x>.
- Snelder, T.H., Cattaneo, F., Suren, A.M., Biggs, B.J.F., 2004. Is the river environment classification an improved landscape-scale classification of rivers? *J. N. Am. Benthol. Soc.* [https://doi.org/10.1899/0887-3593\(2004\)0230580:ITRECA2.0.CO;2](https://doi.org/10.1899/0887-3593(2004)0230580:ITRECA2.0.CO;2).
- Stevens, L., Robertson, B., 2012. New River Estuary 2012 Broad Scale Habitat Mapping. Wriggle Coastal Management Report Prepared for Environment Southland (29p).
- Thomas, A.L., Dambrine, E., King, D., Party, J.P., Probst, A., 1999. A spatial study of the relationships between streamwater acidity and geology, soils and relief (Vosges, north-eastern France). *J. Hydrol.* 217 (1–2), 35–45. [https://doi.org/10.1016/S0022-1694\(99\)00014-1](https://doi.org/10.1016/S0022-1694(99)00014-1).
- Topoclimate South, 2001. Topoclimate South Soil Mapping Project. Electronic files held by Environment Southland. Crops for Southland, Invercargill, New Zealand.
- Tratnyek, P.G., Grundl, T.J., Haderlein, S.B. (Eds.), 2012. *Aquatic Redox Chemistry*. American Chemical Society Symposium Series 1071. Oxford University Press, p. 20.
- Troy, T.J., Wood, E.F., Sheffels, J., 2008. An efficient calibration method for continental-scale land surface modeling. *Water Resour. Invest.* <https://doi.org/10.1029/2007WR006513>.
- Tsakiris, G., Alexakis, D., 2012. Water quality models: an overview. *Eur. Water* 37, 33–46.
- Turnbull, I.M. and Allibone, A.H. (2003). Geology of the Murihiku Area. Institute of Geological and Nuclear Sciences Limited. 1:250 000 geological map 20. 1 sheet and 74 p. Lower Hutt, New Zealand.
- Wilford, J., Caritat, P. De, Bui, E., 2016. Applied geochemistry predictive geochemical mapping using environmental correlation. *Appl. Geochem.* 66, 275–288. <https://doi.org/10.1016/j.apgeochem.2015.08.012>.
- Wright, R.F., 1988. Influence of acid rain on weathering rates. In: Lerman, A., Meybeck, M. (Eds.), *Physical and Chemical Weathering in Geochemical Cycles*. Kluwer Academic Publishers, Oslo, Norway, pp. 181–196.