Physiographic Zones for the Southland Region

Classification system validation and testing report

November 2016

Prepared by:

Ton Snelder

Brydon Hughes

Karen Wilson*

Katie Dey

*Landpro Ltd

For any information regarding this report please contact:

Ton Snelder

Phone: 03 377 3755
Email: ton@lwp.nz

LWP Ltd
PO Box 70
Lyttelton 8092
New Zealand

LWP Client Report Number:
Report Date: July 2016
LWP Project:

Quality Assurance

Version	Reviewed by	
1	Scott Larned (NIWA)	24 July 2016
2		12 August 2016
3		6 September 2016

Table of Contents

Figures

Tables

Executive Summary

Environment Southland have established a spatial framework for managing land use activities using nine non-contiguous physiographic zones. The physiographic zones have been mapped according to a conceptual model that relates biogeochemical and hydrological processes that determine potential water quality states to inherent characteristics of the Southland landscape. This report describes tests of the mapped physiographic zones that comprised three components:

- General 'omnibus' tests to assess whether the physiographic zones represent unique assemblages of drivers, environmental characteristics and observed water quality states;
- Tests of differences in individual water quality variables to determine the extent to which the physiographic zones explain spatial variation in magnitudes and temporal variability in water quality; and,
- Hypothesis testing to assess the validity of specific expectations for water quality variables that arise from the conceptual model that underpins the physiographic zones.

The results show physiographic zones strongly discriminate unique combinations of the drivers and characteristics, and that differences between zones were strongest where recharge mechanism differed. Variation in the characteristic magnitudes of assemblages of river water quality data were well explained by the physiographic zones. The characteristic magnitudes of assemblages of groundwater quality variables was less well explained by the physiographic zones. This may reflect better characterisation of river water quality due to more frequent sampling compared to groundwater. The better performance of the physiographic zones for river water quality may also be because river sites represent an integrated measure of the overall hydrological and biogeochemical characteristics of the physiographic zones occurring across the upstream catchment area, whereas groundwater sites represent a single point in what may be a rather heterogeneous system.

Tests showed that physiographic zones strongly discriminate between site differences in the magnitudes of individual water quality variables, particularly for river water quality. The physiographic zones also discriminated between site differences in temporal variation (overall variability and seasonal variation) of many individual water quality variables. However, the physiographic zones did not explain variation between sites in temporal variability of the groundwater quality variables. It is expected that the physiographic zone variants (sub-zones) will increase the explanation of overall variability and seasonality of water quality of the physiographic zones. However, the variants could not be tested in this study due to limitations in the availability of water quality data.

Tests of hypotheses concerning expected differences in the magnitude of individual water quality variables between, and within physiographic zones were largely consistent with expectations developed from the conceptual models for individual physiographic zones (Hughes *et al.,* 2016). Tests of hypotheses concerning the variability and seasonality of individual water quality variables were rarely inconsistent with the conceptual water quality risk framework (Hughes et al., 2016), but were often inconclusive due to a lack of data.

1 Introduction

The physiographic zones of Southland are a spatial framework for managing the effects of land use on freshwater quality. The framework assigns all parts of the Southland region to one of nine physiographic zones, each of which comprises a unique set of biogeochemical and hydrological controls over potential water quality state. The nine physiographic zones occur in an irregular patchwork of non-contiguous areas across the region that is determined by each location’s specific environmental characteristics including climate, geology and soils and proximity to rivers. The physiographic zones are based on a conceptual model that postulates that a range of biogeochemical and hydrological processes vary across the region and cause variation in water quality state (Rissmann *et al.*, 2016). The extent to which these variations are realised is largely dependent on land use.

Environment Southland have utilised the physiographic zones as a spatial framework for managing particular land use activities in order to maintain and improve water quality outcomes for the region in the Proposed Southland Water and Land Plan (Environment Southland, 2016). It is important that the physiographic zones of Southland are tested to assess the extent to which they discriminate variation in water quality and whether these differences are consistent with the underlying conceptual model, on which proposed policies are based.

This report does not evaluate the ability of key drivers to predict hydrochemistry and water quality states. The latter was undertaken by Rissmann *et al.* (2016). Rather, the primary focus is to assess the strength of the physiographic zones as a system of classification of inherent environmental variables and associated water quality outcomes (Hughes *et al.,* 2016) and spatial and temporal water quality risk (Hughes *et al*., 2016). It is outside of the scope of this report to test the accuracy of the delimiting boundaries of the mapped physiographic zones (Figure 1).

This study is concerned with whether, at a regional scale, the physiographic zones discriminate unique combinations of hydrochemical drivers, associated environmental characteristics, and of observed water quality. The testing and validation analyses therefore had three specific objectives:

1. to determine whether the physiographic zones discriminate spatial variation in unique assemblages of drivers, characteristics, and observed water quality states,
2. to determine whether the physiographic zones discriminate and explain the spatial distribution of the magnitudes and temporal behaviours of individual measures of water quality and hydrochemical indicators,
3. to determine whether differences in the magnitudes and temporal behaviours of specific individual water quality measures and hydrochemical indicators are consistent with the underlying conceptual basis for the physiographic zones.

2 Conceptual model and differences between the physiographic zones

The physiographic zones were developed using a conceptual model that distinguishes water quality risk according to variation in attenuation and dilution of contaminants along differing drainage pathways. When co-occurring at a location, particular combinations of hydrochemical drivers and hydrological factors are expected to result in different water quality states. The water quality state that is expected to be the most strongly explained by the physiographic zones are characteristic magnitudes (i.e. central tendency) associated with dissolved nutrients (Hughes *et al.,* 2016) and related hydrochemical indicators. However, the physiographic zones also account for spatial variation in dominant contaminant transport pathways, which are associated with between-zone variation in the temporal behaviour of water quality. It is expected there will be within-physiographic zone variation in temporal behaviour of water quality due to local differences in the dominant transport pathways. This within-unit spatial variability in the temporal behaviour of water quality constituents is partially addressed by physiographic zone variants, (i.e. sub-zones) described in Hughes *et al*., (2016).

The conceptual framework underlying the physiographic zones postulates that differences in water quality states are primarily associated with between-zone differences in contaminant transport pathways and attenuation processes (particularly dilution and reduction potential). These differences are anticipated to result in predictable differences in the magnitudes and temporal behaviours of water quality variables. Table 1 provides a summary of the contaminant transport pathways the associated expectations for characteristic magnitudes and temporal behaviours of different water quality variables and hydrochemical indicators by physiographic zone based on the characterisation of individual physiographic zones outlined in Hughes *et al*., (2016). The influence of land use on potential water quality states has been incorporated in a risk assessment (Hughes, *et al*, 2016).

Based on the conceptual model, a set of mapping rules were derived to delineate a spatial framework (i.e. produce a map) of the region comprising nine physiographic zones (Hughes, *et al*., 2016). The criteria for defining the physiographic zones were based on multiple mapping variables with each physiographic zone representing the occurrence of unique combinations of those variables (Hughes *et al.,* 2016). The physiographic zones are therefore expected to represent specific combinations of hydrological and biogeochemical drivers and associated environmental characteristics (e.g. soil reduction potential is associated with soil drainage characteristics). In addition, physiographic zones are expected to discriminate spatial and temporal patterns in water quality as described by many individual chemical, physical and microbiological variables (Rissmann *et al.,* 2016 and Hughes *et al.,* 2016).

Physiographic zone variants have been developed to spatially depict areas within individual zones where is increased water quality risk associated with particular drainage pathways when soils are saturated (Hughes *et al.,* 2016). Variants are expected to modify the potential for contaminant loss to surface water due to within zone variability in drainage mechanisms (specifically artificial drainage or overland flow). The variants have not been included in the analyses undertaken for this report due to insufficient water quality data.

Table 1. Summary of the physiographic zones and associated predicted water quality and hydrochemical characteristics. *See Figure 1 for the location of the physiographic zones.*

Physiographic zone	Dominant Contaminant Pathway(s)	Predicted water quality and hydrochemical characteristics
Alpine	Overland flow	▪ Very dilute water associated with precipitation containing low dissolved solute concentrations reflecting low marine aerosol load, orographic enhancement of rainfall, well flushed colluvium and low intensity land use ▪ Low concentrations of redox sensitive variables due to oxidising soils and geology ▪ Elevated pH in surface waters reflecting headwater degassing
Bedrock/Hill Country	Overland flow, lateral flow, artificial drainage	▪ Dilute waters containing low solute concentrations reflecting orographic rainfall enhancement and rainout of marine aerosol load (especially with increased distance from the coast) ▪ Although soils exhibit an elevated reduction potential the prominence of overland flow limits the potential for redox in the soil zone to influence water quality. Elevated concentrations of redox sensitive variables occur in surface waters during base flow conditions reflecting the greater contribution of reduced waters associated with soil zone drainage
Central Plains	Natural bypass flow, artificial drainage	▪ High marine aerosol load due to low elevation and proximity to coast ▪ Elevated pH due to the buffering capacity of young mafic parent materials with a few smaller areas influenced by the incorporation of limestone into the alluvial materials ▪ High alkalinity due to elevated soil zone pH (unlimited CO_2 production in the soil zone so alkalinity is pH limited) ▪ Elevated electrical conductivity reflecting both high marine aerosol load, elevated soil base saturation and pH due to parent material types (relatively young mafic and lesser carbonate materials) ▪ High nitrate concentrations in groundwater and surface water ▪ Elevated microbes in surface waters reflecting artificial drainage intersecting soils cracks (when soils are dry) and poor internal drainage (when soils are wet)
Gleyed	Artificial drainage	▪ Mixed marine aerosol loads due to range of geographical locations ▪ Elevated soil reduction potential and associated denitrification in the soil zone reflecting imperfect to poor soil profile drainage ▪ Strong soil zone signature in surface waters reflecting artificial drainage as dominant pathway for contaminant export. ▪ Episodic losses of dissolved and particulate contaminants via artificial drainage ▪ Low to moderate groundwater nitrate concentrations due to variable soil reduction potential ▪ Episodically elevated nitrate concentrations in surface water ▪ Dense subsoils and higher clay content increase the extent of ion exchange which removes many of the soil zone tracers in deep drainage

Physiographic zone	Dominant Contaminant Pathway(s)	Predicted water quality and hydrochemical characteristics
Lignite/Marine Terraces	Deep drainage, overland flow, artificial drainage	▪ Variable marine aerosol loads due to range of geographical locations ▪ Elevated concentrations of redox sensitive variables in groundwater reflecting elevated reduction potential in the saturated zone (although can be variable depending on proximity to carbonaceous sediments) ▪ Elevated dissolved reactive phosphorus (DRP) in groundwater ▪ Elevated alkalinity in groundwater compared to surface waters due to reduction processes ▪ Concentrations of redox sensitive parameters may vary seasonally according to soil moisture
Old Mataura	Deep drainage	▪ Low concentrations of chemically reduced species (i.e., dissolved iron and manganese) and dissolved organic carbon (DOC) due to oxidised (well drained) soil- and ground- waters ▪ Low sulphate concentrations in groundwater reflecting the advanced weathering of sediments and preferential removal by aluminium-oxides and hydroxides within the soil and/or unsaturated zones ▪ High nitrate concentrations in groundwater due to low reduction potential in the soil and saturated zone and recharge exclusively from land surface recharge (little riverine flushing) ▪ Low marine aerosols due to inland location ▪ Low alkalinity due to the lower pH of weathered soil and unsaturated zone materials
Oxidising	Deep drainage, overland flow, artificial drainage	▪ Mixed marine aerosol loads due to range of geographical locations ▪ Low concentrations of redox sensitive variables in groundwater due to low reduction potential in soils and underlying geology and a high proportion of land surface recharge (little riverine flushing)
Peat Wetlands	Artificial drainage, lateral flow	▪ Variable marine aerosol loads due to range of geographical locations but with high marine aerosol load in low altitude, coastal areas ▪ Waters are strongly acidic due to the high organic acid content in sediments and low pH buffering capacity. This also results in low alkalinity in surface waters ▪ Elevated concentrations of chemically reduced species (dissolved manganese and iron) in groundwater and surface water ▪ Very low nitrate concentration in ground- and surface waters due to high reduction potential in soils and underlying geology ▪ Dissolved calcium (Ca) concentrations are low due to the lack of mineral content in sediments and low soil base saturation ▪ Elevated Potassium (K) concentrations in groundwater due to the limited mineral material for regulation through ion exchange (making it very leachable) ▪ Elevated phosphorus (P) in surface and ground waters due to poor P-retention in soils and aquifers and enhanced P-mobility under reducing conditions. During high flow events, P concentrations increase markedly in developed

		catchments but remain largely unchanged in undeveloped areas.
Riverine	Deep drainage	▪ Dilute waters (low solute concentrations) reflecting the dominance of alpine derived waters ▪ Low to moderate nitrate reflecting dilution of local land surface recharge and pristine recharge areas ▪ Electrical conductivity is higher in groundwater than surface waters reflecting variable contributions from lowland land surface recharge

3 Data

3.1 Physiographic zones

Our validation and testing procedures were based on multiple variables that described hydrochemical drivers, environmental characteristics and water quality observations and used both univariate statistical procedures (tests based on a single response variable) and multivariate procedures (involving multiple response variables in a single test). All analyses took the mapped physiographic zones (Hughes *et al.,* 2016) as the fundamental input and the map of the physiographic zones was “taken as given” (Figure 1). Validation and testing data was independently derived for many locations across the region and assigned to a physiographic zone using a Geographic Information System (GIS).

3.2 Drivers and characteristics testing data

The first dataset is referred to as the drivers and characteristics (Figure 2). These data are independent of the spatial data used to define the physiographic zones (i.e. to define the boundaries of the zones). Each of the variables represented in these datasets were obtained from continuous spatial coverages that described variation across the entire Southland region, which were obtained from the GIS.

The individual spatial variables were separated into two categories. We refer to the first category as the drivers. The conceptual model developed by Rissmann *et al. (*2016) postulates that four key drivers (Section 2) control spatial variability of hydrochemistry and water quality across the Southland region. Spatial variation in the four key drivers are represented by the following variables:

- precipitation source;
- recharge mechanism;
- soil redox setting;
- geological (aquifer) redox setting;
- landform age (geomorphic surface age);
- surficial substrate (rock and sediment type) composition; and,
- sub-surface substrate composition.

All driver variables were categorical and are described by the Physiographics of Southland Part 1 report (Rissmann *et al.,* 2016; Technical Chapters 1, 2, 6 and 8).

We refer to the second category of variables as characteristics. The characteristics describe landscape features that are proximate outcomes of the drivers and describe aspects of the environmental setting of each physiographic zone that relate specifically to potential water quality risks associated with that zone (Hughes *et al.,* 2016). The characteristic variables are derived from the Physiographics of Southland Technical Sheets (Wilson *et al., in prep.*):

- elevation;
- slope;
- mean annual rainfall;
- drainage density;
- stream size;
- overland flow potential;
- soil type;
- soil profile drainage;
- soil base saturation;
- soil anion storage capacity;
- mean annual drainage season;
- artificial drainage;

- lateral drainage potential;
- depth to water table;
- aquifer permeability;
- active groundwater storage; and,
- deep drainage potential.

The characteristics data comprise continuous and categorical variables.

3.3 Water quality testing data

River water and groundwater have been sampled at sites distributed across Southland for compliance, state of environment monitoring (SOE) and scientific investigations for more than two decades. The samples have been analysed for a range of constituents some of which represent measures of quality (fitness for purpose) and some of which represent hydrochemical indicators of the provenance of water or processes that have or are occurring during the transit of water. Both types of variables are relevant to testing the physiographic zones. We refer to them collectively as “water quality variables” for brevity but note that the hydrochemical indicators are not associated with the water’s fitness for purpose.

The most comprehensive and long term datasets are associated with SOE monitoring. Monthly SOE monitoring of river water in Southland began in July 1995 with 32 sites, including 6 sites that belong to NIWA’s National River Water Quality Network (NRWQN). Since 1995, 40 additional sites have been added to the SOE river water quality network to improve coverage. The most comprehensive and long term datasets of groundwater quality are associated with GNS Science’s National Groundwater Monitoring Programme (NGMP) and SOE monitoring. The NGMP has been monitoring 7 sites quarterly since 1998. In 2000, Environment Southland established additional quarterly groundwater quality SOE monitoring at 49 sites. In 2001, a further 20 sites were added in order to improve the coverage of nitrate sampling. Over time, many of the nitrate-only sites have been incorporated into the SOE programme. A large number (1477) of groundwater sites have also been sampled on a bi-annual basis for a restricted number of water quality variables as part of Environment Southland’s resource consent compliance monitoring programme. A variety of physical, chemical and biological indicators of groundwater quality are measured at all sites quarterly (Wilson *et al.,* 2012).

Data were obtained and groomed for analysis in several steps. All available site information and water quality observations data, including date, were initially extracted from Environment Southland[1]’s water quality database. Unless otherwise stated, we made no distinction between data collected at regional council SOE sites, NRWQN and NGMP sites, or groundwater consent compliance sites and we refer to the sites collectively as the “river monitoring sites” of which there were 155 and “groundwater quality sites”, of which there were 1546.

Sites were assigned to physiographic zones to enable representative monitoring sites to be identified for each zone. Assignments were made by overlaying site locations on the mapped physiographic classification in the GIS. Groundwater sites were assigned to a single physiographic zone on the basis of spatial location. The 155 river sites were only assigned to a physiographic zone when >60% of the upstream catchment area was located in a single physiographic zone. Sites that did not meet this criterion were discarded. The exception to this rule was the selection of river sites for the Riverine physiographic zone. Limiting this zone to >60% catchment area resulted in a bias to more pristine headwater sites dominated by alpine runoff. As this zone is the ultimate receiving environment for all waters, monitoring sites on the main-stem rivers (i.e. Mataura, Oreti, Aparima and Waiau Rivers), that were assigned to the Riverine zone, were included. Inclusion of these sites ensured that the

---

[1] Environment Southland is the trading name of the Southland Regional Council.

Riverine zone was represented by sites with varying amounts of land surface recharge and alpine runoff without a strong bias towards one recharge mechanism.

For the retained sites, we obtained data describing the proportion of upstream catchment (river monitoring sites) and a 1km buffer zone for groundwater quality sites (i.e. a circular area of 1km diameter centred on each bore from which groundwater quality data was collected) that were occupied by high producing exotic grassland as defined by the New Zealand Land Cover Database (LCDB3). We refer to this explanatory variable for both river and groundwater sites as %Pasture. The LCDB3 differentiates New Zealand into 33 categories of land cover based on analysis of satellite imagery from 2008 (lris.scinfo.org.nz). The proportion of the catchment occupied by high producing exotic grassland (%Pastoral) has been shown to be a strong predictor of surface water quality in several studies (Larned *et al.*, 2016; McDowell *et al.*, 2012) and is used in the statistical analyses that follow to control for differences in levels of agricultural land use between sampling sites.

Cumulative probability plots were used to identify obvious erroneous values, which were removed from the data set. Sites with a bore depth of >50m were also removed from the dataset in order to ensure that the retained sites represented shallow, soil-influenced groundwater rather than groundwater contained in underlying tertiary sediments, which exhibit characteristic water quality reflecting the geological characteristics and associated redox status of these aquifers.

The next step involved filtering sites based on rules for inclusion. The inclusion rules involved limiting data to the recent past to reduce the impact of water quality trends, which are evident at many sites (Larned *et al.*, 2016). In addition, rules were defined to ensure the sites had sufficient samples to provide robust data in the analyses that follow. In particular, we required sites to have sufficient samples to provide reasonably precise measures of central tendency and to be amenable to temporal analyses. For the available data, long term sampling was limited to monthly temporal resolution at river sites. The groundwater data consisted of a mix of sampling frequencies with some quarterly data at SOE monitoring sites and a larger number of sites monitored twice a year, nominally during the wet and dry periods, for compliance purposes. Therefore, the requirements for recent and sufficient data were conflicting and our filtering rules represent a trade-off between recent data and replication.

After considering the trade-offs, river sites were restricted to those that had a been sampled in every month in at least two of the last 10 years (i.e. had been sampled 24 times in last 10 years with individual months being represented by sampling occasions in at least two years). This resulted in data for 43 river sites. Individual water quality variables were not necessarily sampled on each sampling occasion, resulting in differing numbers of samples for each variable. Figure 1 shows the location of the water quality sites used in this analysis.

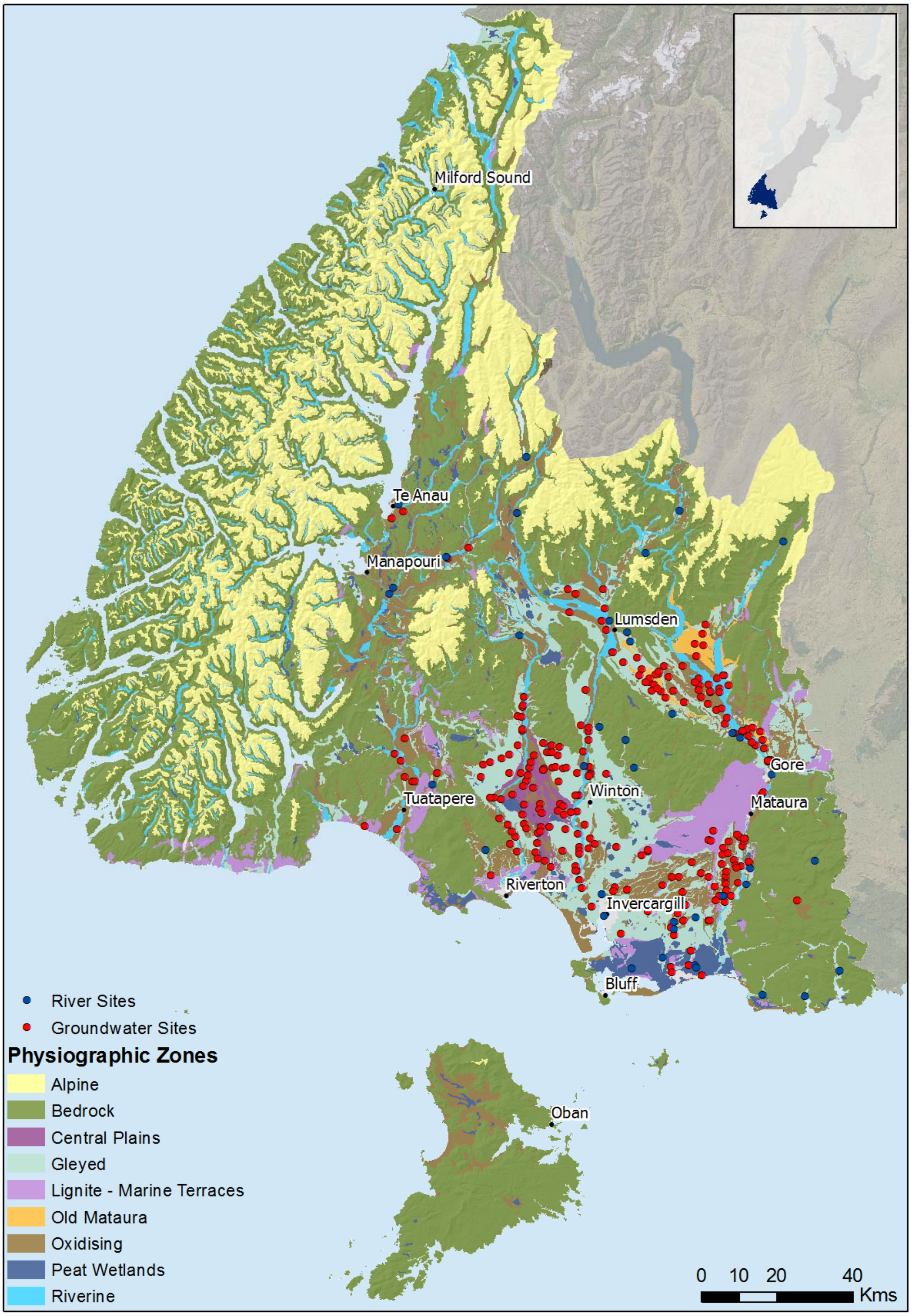


Figure 1. Map of the Southland physiographic zones and groundwater and river water quality sites used in the testing.

For groundwater we first assigned the sample occasions to one of two seasons: wet and dry. The wet season was nominally defined as the months from June to November (i.e. winter and spring) with the remaining months being the dry season (i.e. summer and autumn). We then imposed the rule that sites must have been sampled at least 4 times, with a minimum of twice in any one season within the last 10 years. This reduced the sites to 254, indicating that many sites in the original data had very few sample occasions. Of the remaining sites, many had uneven numbers of samples in the summer and winter periods. To avoid seasonal bias, we randomly sampled occasions in the over-represented season to obtain the same number of samples as the number of sampling occasions in the under-represented season.

Another consideration was the variation in the number of sample occasions that occurred between the water quality variables. A large suite of variables was sampled at most sites but some variables were sampled infrequently or never at some sites. In particular, some relevant hydrochemical variables including chloride, bromide, dissolved iron and manganese were infrequently analysed. For rivers, we further required that any variable was sampled on at least four occasions at a site to be included in that site’s variable suite. For groundwater, the only variable that was consistently sampled at all sites was nitrate-nitrogen. There was significant variability between sites in the number of times the other variables were included in the sample. This restricted the numbers of sites that were included in the analyses that follow and these restrictions are explained in the relevant methods sections below and the results.

River water quality was described by 20 variables that correspond to physical, chemical and microbiological conditions (Table 2). In this report, we use “river water quality” as a general term to refer to some or all of these 20 variables. Groundwater quality was described by 18 variables that correspond to physical, chemical and microbiological conditions (Table 3). In this report, we use “groundwater quality” as a general term to refer to some or all of the 18 variables. We use the term “water quality variables” to refer to the groundwater and river water quality datasets.

The water quality data contained censored values for several variables, for which the true values on some sampling occasions were too low or too high to be measured with precision. For very low values of a variable, the minimum acceptable precision corresponds to the “detection limit” for that variable; for very high values of a variable, the minimum acceptable precision corresponds to the “reporting limit” for that variable. Censored values had been replaced with substituted values in the data we received. The commonly used substitutions of 0.5×detection limit and 1.1×reporting limit are generally used. Our investigations indicated that these substitutions were not made for all observations with values below the detection limit and that this convention was not rigorously adhered to. We investigated the number of censored values for some variables whose nominal detection limits were known. These included DRP <0.05 mg L^{-1} until 2008 then <0.04 mg L^{-1}, NNN < 0.010 mg L^{-1} until 2008 then <0.002 mg L^{-1}, NO_2N < 0.002 mg L^{-1}, NH_4N < 0.01 mg L^{-1}, Br <0.05 mg L^{-1}, TSS <3 mg L^{-1}, Fe^{II} <0.02 mg L^{-1} and Mn^{II} < 0.001 mg L^{-1}. In the analyses that follow, the existence of censored values was of minor importance because they occur rarely at most sites. The analyses of magnitude are based on median site values and these are not affected by censored values unless more than 50% of samples were below the detection limit. We assessed the extent to which site medians were affected by censored values (i.e. medians values were below detection) and report this in the results section.

For each site and variable we calculated three test variables;

1. the median of the observations, representing the characteristic magnitude,
2. the coefficient of variation representing the overall variability,

3. a seasonality index representing the characteristic variation in magnitude associated with seasons.

The coefficient of variation was calculated as the standard deviation of the observations divided by the mean of the observations. The seasonality index was calculated as the coefficient of variation (i.e. standard deviation/mean) of the median values of samples in each season (i.e. the variability of the characteristic magnitudes for each season). For the river water quality data, we defined four seasons; summer (December to February), autumn (March to May), winter (June to August) and spring (September to November). That is, for each site and variable, calculate seasonal median values, then calculate the mean of the four medians and divide by their standard deviation. Because groundwater sites are sampled less frequently, we used the two (wet and dry) seasons only.

Finally, for the river water quality sites most samples (i.e. 75%) were associated with flow observations pertaining to the sample occasion. For some of the analyses that follow, we needed to be able to stratify the data according to the flow at the time of sampling. We stratified the flow state at sampling into three classes; ‘baseflow’, ‘highflow’ and ‘other’. We initially defined base flow and high flow thresholds to be the 25th and 75th percentile flows. However, we subsequently had to relax the high flow threshold to the 65th percentile flow to have a reasonable number of samples. These thresholds are therefore arbitrary but identify samples taken from reasonably contrasting river flow states. Because the majority of sites were not associated with flow recorders, we used modelled flow duration curves (FDC) to estimate the magnitude of the 25th and 65th percentile flows at all sites. We used the FDC model of (Booker and Snelder, 2012) to predict these flow percentiles for each sampling site.

Table 2. River water quality variables included in this study. *The means and ranges were calculated from site median values at the sites included in the analyses that follow.*

Variable	**Abbreviated name**	**Units**	**Mean (range) of site median values**
Total Alkalinity	TA	mg/L	24.24 (0.75 - 61)
Calcium	Ca	mg/L	7.41 (1.25 - 17.81)
Chloride	Cl	mg/L	15.38 (0.69 - 43.54)
Bromine	Br	mg/L	0.07 (0.03 - 0.19)
Electrical Conductivity	COND	uS/cm (at 25degC)	118 (29 - 249)
Iron	Fe^{II}	mg/L	0.24 (0.01 - 1.35)
Magnesium	Mg	mg/L	2.86 (0.75 - 5.3)
Manganese	Mn^{II}	mg/L	0.01 (0.0003 - 0.06)
Nitrite Nitrogen	NO_2N	mg/L	0.004 (0.0005 - 0.02)
Ammoniacal nitrogen	NH_4N	mg/L	0.02 (0.004 - 0.15)
Total Nitrogen	TN	mg/L	1.01 (0.06 - 4.3)
Dissolved Reactive Phosphorus	DRP	mg/L	0.01 (0.002 - 0.08)
Total Phosphorus	TP	mg/L	0.03 (0.002 - 0.14)
Nitrate Nitrite Nitrogen	NNN	mg/L	0.64 (0.003 - 2.7)
Kjeldahl Nitrogen	TKN	mg/L	0.29 (0.05 - 1.48)
Escherichia coli	ECOLI	CFU*/100 mL	281.1 (5 - 1600)
Sulfate	SO_4	mg/L	6.2 (0.75 - 23.24)
Potassium	K	mg/L	0.93 (0.17 - 2.98)
Total Suspended Solids	TSS	mg/L	4.23 (1.5 - 15)
Visual clarity	CLAR	m	1.35 (0.4 - 6.3)

* CFU = colony forming units.

***Table 3. Groundwater quality variables included in this study**. The means and ranges were calculated from site median values at the sites included in the analyses that follow.*

Variable	**Abbreviated name**	**Units**	**Mean (range)**
Total Alkalinity	TA	mg/L	50.7 (7 - 200)
Boron	B	mg/L	0.02 (0.01 - 0.08)
Bromine	Br	mg/L	0.09 (0.01 - 0.45)
Calcium	Ca	mg/L	20.2 (3.85 - 122)
Chloride	Cl	mg/L	23.68 (3 - 102.5)
Electrical Conductivity	COND	uS/cm (at 25degC)	250.38 (73 - 768)
Escherichia coli	ECOLI	MPN /100 mL	6.24 (0.5 - 441.5)
Iron	Fe^{II}	mg/L	0.72 (0 - 22)
Magnesium	Mg	mg/L	8.09 (1.64 - 27)
Manganese	Mn^{II}	X	0.11 (0 - 1.65)
Nitrate Nitrite Nitrogen	NNN	mg/L	4.98 (0 - 21)
Ammoniacal nitrogen	NH_4N	mg/L	0.04 (0 - 1.8)
Dissolved Reactive Phosphorus	DRP	mg/L	0.02 (0 - 0.26)
Potassium	K	mg/L	1.31 (0.42 - 13.7)
Silica	Si	mg/L	22.88 (10.1 - 48.02)
Sodium	Na	mg/L	17.56 (3.1 - 54)
Sulfate	SO_4	mg/L	10.74 (0.25 - 62.09)
pH	pH		6.57 (5.24 - 7.9)

* MPN = most probable number

4 Statistical tests

4.1 Preliminary data preparation

4.1.1 Selection of drivers and characteristics data

The drivers and characteristics data were sampled at the mid-point of approximately 100,000 grid cells distributed across a regular mesh that covered the entire Southland region. Each cell in the grid was described by a value for most of the driver and characteristic variables. Our analysis was based on taking a stratified random sample of 100 grid cells in each of the nine physiographic zones.

Some variables had data missing in some locations. Where our subset of driver or characteristic variables had more than 20% of data missing, the variable was removed from analysis. The variables removed from analysis mainly relate to soil properties, for which available data are spatially restricted. The drivers and characteristics used in the study are summarised in the Physiographics of Southland Technical Sheets User Guide (Wilson, *et al., in prep.*).

4.1.2 Preliminary inspection and filtering of water quality data

In order to determine whether the data was sufficiently representative to provide for robust results in the analyses that follow, we performed a number of preliminary steps in which we assessed the number of sample occasions for each water quality site. We inspected box and whisker plots of the site median values of the water quality variables and assessed the normality of the data distributions. We also inspected scatter plots of appropriately transformed values of the site median values against the %Pasture for evidence of relationships. For the groundwater quality data, we inspected scatter plots of appropriately transformed site median values against the reported bore depth for evidence of relationships. We also inspected plots of variation in the water quality variables by season at individual sites and also of variation between physiographic zones of an index of seasonality. The index of seasonality for each site was derived by first obtaining the median value of samples taken in each of four seasons (Summer – December to February, Autumn – March to May, Winter June to August, Spring September to November). For each site we calculated the coefficient of variation of the four seasonal medians and used this as a measure of the intensity of seasonal variability.

4.2 General approach

We carried out three sets of tests that were designed to respond to the three objectives of this study and were performed on different sets of test data (Figure 2). The testing procedures and data commenced with more general “omnibus” tests at step 1 and proceeded through to very specific tests using specific test variables at step 3 (Figure 2). The testing was based on the two types of test data: the drivers and characteristics data and the water quality (river and groundwater) observations (Figure 2).

We refer to the first set of tests as classification strength tests. Environmental classifications, such as the physiographic zones, are intended to discriminate spatial variation in multiple variables. These types of classifications are often tested using statistics that assess their ability to discriminate spatial variation in multiple variables simultaneously (Leathwick *et al.*, 2011; Snelder *et al.*, 2005). These tests can be broadly referred to as classification strength tests and are based on multivariate tests of discrimination (Hawkins *et al.*, 2000). We performed tests of classification strength using the drivers, characteristics and the water quality variables (Tables 2 and 3) as input data (Figure 2).

The second set of tests assessed the ability of the physiographic zones to discriminate spatial variation in individual water quality variables (Figure 2). These tests were performed only on the water quality observation data. The tests provide information about how consistent the magnitudes and temporal behaviour of the individual water quality variables were within particular zones and how these varied between the zones. We applied these tests to three sets of test variables that were obtained for each site by analysis of the individual observations of each variable (i.e. the median, coefficient of variation and the seasonality index).

The third set of analyses assessed specific expectations concerning differences in the magnitude and temporal behaviours (seasonality, variability and at differing flows) of specific water quality variables (Figure 2). The expectations were based on the conceptualisation of individual physiographic zones (Hughes *et al*., 2016) and assumptions regarding contaminant transport pathways and attenuation processes (Hughes *et al*., 2016). These tests were based on hypotheses concerning individual variables within and between particular physiographic zones and contaminant pathways, which could be posed based on an understanding of the conceptual framework that underlies the physiographic zones and associated water quality risk assessment. Physiographic zones were grouped according to these hypothesised differences and were then used to test whether the hypothesised differences were manifest as differences in the magnitude, variability and seasonality of specific water quality variables.

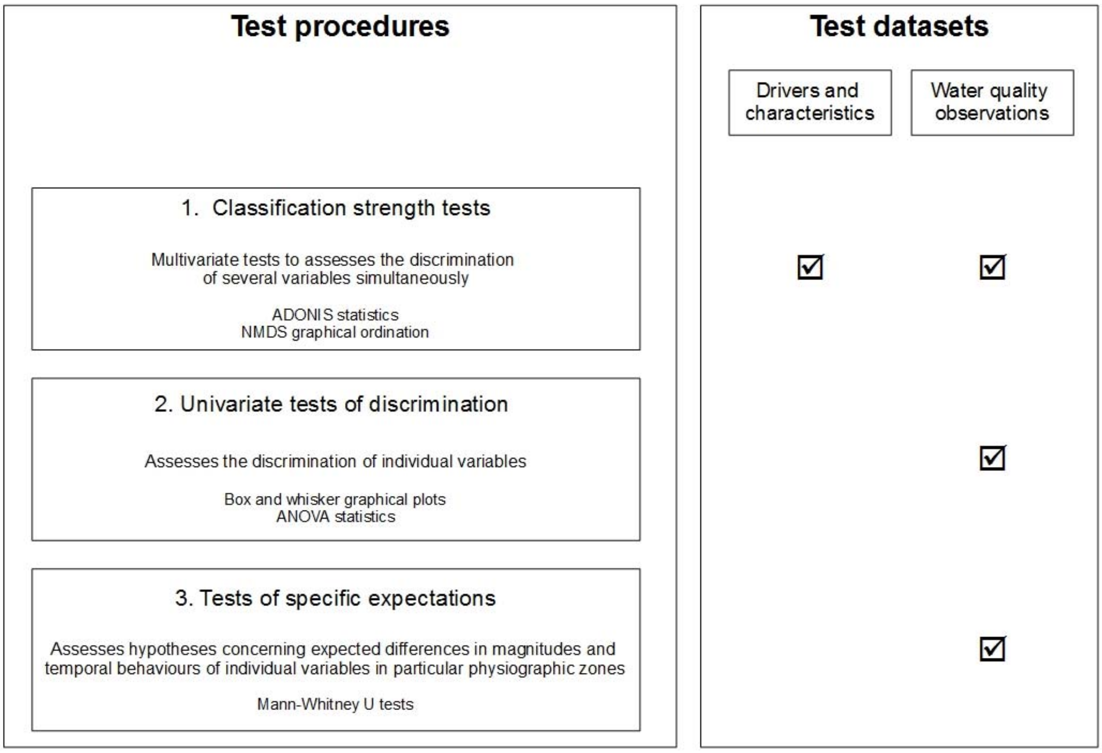


Figure 2. Schematic diagram of the test procedures and datasets.

4.3 Classification strength tests

4.3.1 Multivariate analysis of variance

We tested whether the physiographic zones discriminate unique combinations of drivers, characteristics and water quality variables using a global non-parametric analysis of variance using distance matrices (ADONIS; Anderson, 2001). ADONIS quantifies the extent to which the physiographic zones explained variation in multiple dependent variables (i.e. the variables represented by the drivers, characteristics and water quality data). The basis for this test is that if the physiographic zones do explain variation in multiple dependent variables, then groups of sites that are defined by physiographic zones should be isolated and compact in the multivariate space defined by the dependent variables.

ADONIS is a multivariate equivalent to analysis of variance (ANOVA). ANOVA compares the means of a single dependent variable between different groups and determines whether any of those means are significantly different from each other. ADONIS also compares groups but does this for more than one dependent variable at a time. The test can be understood as a comparison of the differences in the location of the centres of the groups (the group centroids) in the multivariate space defined by the dependent variables. In these analyses the groups were defined by the physiographic zones and the dependent variables were the variables in the driver, characteristics and water quality datasets.

ADONIS takes as its input a distance matrix that is computed from the dependent variables. For the drivers and characteristics data, distance matrices were based on the Gower distance measure, which is type of city block or Manhattan measure of distance between two points in a multivariate space (Legendre and Legendre, 1998). The Gower distance measure was used for two reasons. First, the Gower distance measure standardises the dependent variables to have a range between zero and one. This means that all variables have equal weight in the analysis. Second, the Gower distance can be used with categorical as well as continuous variables, which was important because both types of variables were represented in our test datasets. To calculate the Gower distance between two sites, first the distances between sites are calculated for each variable, taking the type of the variable into account, then the contributions are averaged. For a categorical variable the contribution to distance is zero if the sites have the same category and one if the categories are different.

For the water quality data, distance matrices were based on Gower distance measure, using the median site values of each variable for both the river and groundwater quality data. Calculation of distance measures requires that no variables have missing values. However, in our water quality datasets, some sites had never been sampled for some variables, resulting in missing values for that site-variable combination. If the values are missing at only a few sites, the missing values can be replaced by a measure of central tendency of that value so the distance measurement calculation is possible. This imputation of values has only minor effect on the outcome provided that there are not a large number of missing values. We replaced missing values with the median value for all other sites provided the number of missing values represented fewer than 20% of the sites. Where the number of missing values represented more than 20% of sites, we removed that variable before calculating the distance matrix.

ADONIS calculates a multivariate *F*-ratio based on the sum of squared distances for the within group pairs of cases (sites) compared to the sum of squared distances of all pairs of cases. This is equivalent to the sum of squared differences between cases and mean values, which is the basis of ANOVA, because the sum of squared distances divided by the number of cases is numerically equal to the sum of squared distances of cases to the centroids (Anderson, 2001).

Significance tests are performed by randomly permuting the raw data to estimate the distribution of the *F*-ratio under the null hypothesis of no differences in the location of the centres of the groups. ADONIS produces an R^2 value, as does ANOVA, which can be interpreted as the variation in distances that are explained by the grouping. Large R^2 values indicate that sites in a physiographic zone occupy a compact area of the multivariate space (i.e. that sites in the zone are similar with respect to their response variables) and that they are isolated in the multivariate space (i.e. that the zone is distinctive with respect to their response variables).

ADONIS is a global test, i.e. it indicates that at least two groups have differences in their locations in the multivariate space but does not indicate which groups are different. We followed up all significant global tests with post-hoc tests of differences by applying the ADONIS test to all pairs of physiographic zones. We adjusted the significance level for these tests to account for multiple comparisons using the false discovery rate (FDR) adjustment method (Benjamini and Hochberg, 1995).

Statistically significant results for both the global and post-hoc tests were interpreted as verifying the underlying conceptual basis for the physiographic zones and the decisions made regarding the definition of boundaries (i.e. mapping rules). The magnitude of the test statistics for the post-hoc tests were interpreted as reflecting the relative difference between the physiographic zones.

4.3.2 Graphical ordination

The goal of graphical ordination is to provide a simplified representation of a multivariate space defined by the response variables. In this study, we used graphical ordination to support and visualise the tests of classification strength (i.e. the ADONIS tests). It was expected that the physiographic zones would produce groups of sites in the multivariate space that were compact and isolated. The multivariate space has many dimensions and therefore cannot be represented graphically. Graphical ordination collapses (i.e. simplifies) the multivariate space to a lower number dimension (e.g. 2 to 4) so that the relationships between sites and groups can be shown graphically. Because graphical ordination is a simplification of the multivariate space it does not show all the detail of the actual relationships between sites and groups but strong differences between groups are highlighted.

We used a method of graphical ordination called non-metric multidimensional scaling (NMDS). NMDS operates on the distance matrices that were also used in the ADONIS tests described above. NMDS uses the rank orders of the distances between cases and uses an iterative procedure to find a low dimensional representation that preserves as closely as possible the original rank orders. The degree to which the original rank orders are distorted is represented by a “stress” value. The numerical value of stress can be viewed as a measure of how reliably the ordination represents the true relationships (dissimilarities) between the cases. NMDS plots with stress values equal to or below 0.05 indicate good fit, stress values less than 0.20 indicate an acceptable representation of the original dissimilarities and those with stress values above 0.30 are a poor representation of the original dissimilarities. We plotted the ordinations and coded the individual sites according to their physiographic zone assignment to aid interpretation.

4.4 Univariate tests of discrimination

For each of the individual water quality variables in the groundwater and river water quality datasets, we tested the extent to which the physiographic zones discriminated variation in the characteristic magnitudes (site median values), overall variability (coefficient of variation) and the seasonality (seasonality index) using graphical and analytical methods. Graphical representation of the distribution of each test variable were provided using box and whisker plots with sites grouped by their assigned physiographic zone.

4.4.1 Site median values

We used ANOVA tests to quantify the extent to which the physiographic zones discriminated differences in the characteristic magnitudes of the individual variables. Magnitude was represented by the site median values. ANOVA R^2 values indicate the proportion of the variation in site median values of each variable that was explained by physiographic zones. The significance values (*p*-values) for the individual tests were adjusted using the false discovery rate (Benjamini and Hochberg, 1995) to correct for multiple comparisons.

We also performed ANOVA on the site median values after controlling for the variation explained by the %Pasture variable. For each variable, we regressed the log (base 10) of the site median values against the %Pasture. The residuals of this regression represent an estimate of that variable in the absence of variation in high producing exotic grassland, and by association in the absence of variation in land use, between sites. We subjected these values (referred to hereafter as the regression residuals) to analysis by ANOVA to determine whether the physiographic zones discriminated the variation in the water quality variables after controlling for %Pasture.

4.4.2 Site coefficient of variation values

ANOVA was used to test the extent to which the physiographic zones discriminated differences in the overall variability of the individual water quality variables. Variability was represented by the site coefficient of variation values. ANOVA R^2 values indicate the proportion of the variation in test data that was explained by physiographic zones. The significance values (*p*-values) for the individual tests were adjusted using the false discovery rate (Benjamini and Hochberg, 1995) to correct for multiple comparisons.

4.4.3 Site seasonality index

ANOVA was used to test the extent to which the physiographic zones discriminated differences in the seasonality of the individual water quality variables. Seasonality was represented by the site seasonality index. ANOVA R^2 values indicate the proportion of the variation in test data that was explained by physiographic zones. The significance values (*p*-values) for the individual tests were adjusted using the false discovery rate (Benjamini and Hochberg, 1995) to correct for multiple comparisons.

4.5 Tests of specific expectations

4.5.1 Hypothesised water quality states by physiographic zone

Differences in the contaminant transport pathways and attenuation associated with the physiographic zones (Table 1) leads to specific expectations concerning the characteristic magnitudes and temporal behaviours of different water quality variables by physiographic zone. These expectations form the basis of a suite of hypotheses that can be tested using the water quality data (Table 2 and Table 3).

Hypotheses concerning the character of individual physiographic zones and specific differences between zones are summarised in Table 4. Hypotheses concerning differences in the magnitudes and temporal behaviours of different water quality variables associated with particular contaminant pathways and summarised in Table 5.

Table 4. Summary of hypothesised differences in the magnitudes and temporal behaviours of water quality variables in different physiographic zones

Physiographic zone	**Test**	**Predicted water quality states**	
		Surface water	**Groundwater**
Alpine	Compared to all other physiographic zones	Lower magnitude in Cl, NNN, TN, TKN, TA, Fe^{II}, Mn^{II}, COND Lower seasonality in COND, NNN, TN, TKN	
	Compared to other flow states within the physiographic zone	Higher magnitude TSS in high flow	
Bedrock/Hill Country	Compared to all other physiographic zones	Higher magnitude in Fe^{II}, Mg Higher variability in TSS, ECOLI, TP	Higher magnitude in Fe^{II}, Mn^{II} Higher variability in COND, NNN, Cl Higher seasonality in COND, NNN, Cl
	Compared to other flow states within the physiographic zone	Higher magnitude TSS, ECOLI, TP in high flow	
	Comparison of river water and groundwater within the physiographic zone	Higher magnitude in Mg	Higher magnitude in COND
Central Plains	Compared to all other physiographic zones		Higher magnitude in TA, pH, Mg, COND, NNN, Ca, Higher seasonality in COND, NNN
Gleyed	Compared to all other physiographic zones	Higher magnitude in Fe^{II}, Mn^{II}, Mg, K, SO_4, NH_4N, COND and lower magnitude in NNN, TN, TKN Higher variability in TSS, ECOLI, TP Higher seasonality in COND, NNN, TN, TKN	Lower magnitude in K, NNN, TA Lower seasonality in NNN
	Compared to other flow states within the physiographic zone	Higher magnitude TSS, ECOLI, TP in high flow Lower magnitude NNN, TN, TKN in baseflow	
	Comparison of river water and groundwater within the physiographic zone	Higher magnitude in K, SO_4, Mg	

Physiographic zone	Test	Predicted water quality states	
		Surface water	Groundwater
Lignite/Marine Terraces	Compared to all other physiographic zones		Lower magnitude in NNN and higher median Fe^{II}, DRP Higher seasonality in NNN
	Compared to groundwater in the Gleyed zone		Higher magnitude in NNN
Old Mataura	Compared to all other physiographic zones		Higher magnitude in NNN and lower magnitude in COND, SO_4, TA, Fe^{II}, Mn^{II}, Cl Higher seasonality in COND, NNN
Oxidising	Compared to all other physiographic zones	Lower magnitude in TA and higher magnitude in K, SO_4 Higher magnitude in NNN, TN, TKN in baseflow	Lower median TA, Fe^{II}, Mn^{II} and higher median NNN
	Compared to high flow in all other physiographic zones	Lower magnitude in ECOLI, TSS, TP in high flow	
	Comparison of river water and groundwater within the physiographic zone	Higher magnitude in Fe^{II}, Mn^{II}	
Peat Wetlands	Compared to all other physiographic zones	Lower magnitude in TA and higher magnitude in DRP, TP, Cl Lower magnitude NNN, TN, TKN in baseflow	Lower magnitude in pH, NNN, Mn^{II}, Ca, Mg and higher magnitude in Cl, K
	Compared to other flows within the physiographic zone	Higher magnitude ECOLI, TSS, TP in high flow	
	Comparison of river water and groundwater within the physiographic zone		Lower magnitude in DRP and higher magnitude in K
Riverine	Compared to all other physiographic zones	Lower magnitude in COND, SO_4, Cl, K Higher seasonality in NNN, TN, TKN, TSS Higher variability in ECOLI, TSS, TP	Lower magnitude in COND, SO_4, Fe^{II} Higher seasonality in NNN
	Comparison of river water and groundwater within the physiographic zone		Higher magnitude in COND

Table 5. Summary of hypothesised differences in the magnitudes and temporal behaviours of water quality variables in different pathway categories

Pathway category	**Test**	**Predicted water quality states**	
		Surface water	**Groundwater**
Overland flow	Compared to all other pathways	Higher magnitude in TSS Higher magnitude TSS in high flow Lower magnitude COND, NNN, TN, TKN in baseflow and higher magnitude Fe^{II} in baseflow Higher variability in DRP, Fe^{II}, ECOLI, TSS, NH_4N and lower in variability NNN, TN, TKN Higher seasonality in DRP, Fe^{II}, ECOLI, TSS, NH_4N and lower seasonality in NNN, TN, TKN	
Artificial drainage	Compared to all other pathways	Higher magnitude in NH_4N Higher magnitude NNN, ECOLI in high flow Higher variability in NNN, DRP, ECOLI Higher seasonality in NNN, DRP, ECOLI	
Deep drainage (oxidising)	Compared to all other pathways	Higher magnitude NNN, TN, TKN in baseflow and lower magnitude Fe^{II}, SO_4, Mg in baseflow	Higher magnitude in NNN and lower magnitude in Fe^{II}, COND Higher variability in NNN Higher seasonality in NNN
Deep drainage (reducing)	Compared to all other pathways	Lower magnitude NNN, TN, TKN in baseflow and higher magnitude Fe^{II}, SO_4 in baseflow	

4.5.2 Testing the hypotheses

We tested the hypotheses posed in Table 4 and Table 5 by converting them into to quantitative tests in the following ways:

1. Where it was hypothesised that a physiographic zone or pathway category would exhibit the highest or lowest magnitude of a given variable, a test for differences was performed that compared the median values of that variable for sites in the zone (or pathway category) to the median values observed at all other sites (i.e. inter- zone comparison).

2. Where it was hypothesised that a physiographic zone or pathway category would exhibit the highest or lowest variability (overall or seasonal) of a given variable, a test for differences was performed that compared the coefficients of variation or seasonal indices for sites in the zone (or pathway category) to the same values at all other sites.

3. Where it was hypothesised that differences occur between groundwater and surface water quality within a zone, a test for differences was performed that compared the median values of that variable in groundwater and surface water for sites in the unit (i.e. intra- zone comparison);

4. Where it was hypothesised that a physiographic zone or pathway category would have elevated or low values of a given variable associated with particular flow states (base flow or high flow), we first stratified all site samples by either the Q25 or the Q65. For each site, we then calculated the median of values for strata defined by flows either lower than Q25 or higher than Q65 (depending on the hypothesis). We then computed the ratio of the stratified samples for all sites. We then performed a test for differences that compared the ratios for sites in the zone (or pathway category) to the same values at all sites in other units (inter- zone) or at other flows within the zone (intra- zone).

Our tests for difference were all performed using the Mann-Whitney U test. This test is a non-parametric test of difference between two groups. Although this test is less statistically powerful than a t-test, it has the advantage that it is free of distributional assumptions. We took this precaution because of the difficulty involved in assessing assumptions of normality in all of the many tests outline above.

For each hypothesis we first tested if there were differences in the two groups of values. We adjusted the significance level for these tests to account for multiple comparisons using the False Discovery Rate (FDR) adjustment method (Benjamini and Hochberg, 1995). If the adjusted *p*-value for the test was <0.05, we then assessed whether the difference between the two groups was as predicted by the hypothesis. If this was found to be true we categorised the hypothesis as consistent. Hypotheses were categorised as inconsistent if the adjusted *p*-value was <0.05 and the difference between the two groups was the opposite of that predicted by the hypothesis. Hypotheses were categorised as inconclusive when p > 0.05, irrespective of the difference between the groups.

5 Results

5.1 Discrimination of drivers

The number of grid cells representing physiographic zones was not even, but all zones were represented by at least 480 cells. Less than 2% of the approximately 100,000 grid cells had missing values for any driver variable and these were replaced by the median of all non-missing values. A balanced subset was taken for the classification strength tests by randomly selecting 100 cells in each zone.

The ADONIS test produced a global R^2 value of 0.6, indicating that 60% of the total variance of the driver data was explained by the physiographic zones (*p*-value <0.01). It is noted that achieving significance is to be expected because of the large number of cases that were used in the test.

The R^2 values for pairwise ADONIS tests ranged from approximately 0.1 to 0.8 (Figure 3). All pairs were significant with p<0.01 (after correction for multiple comparisons using the FDR adjustment).

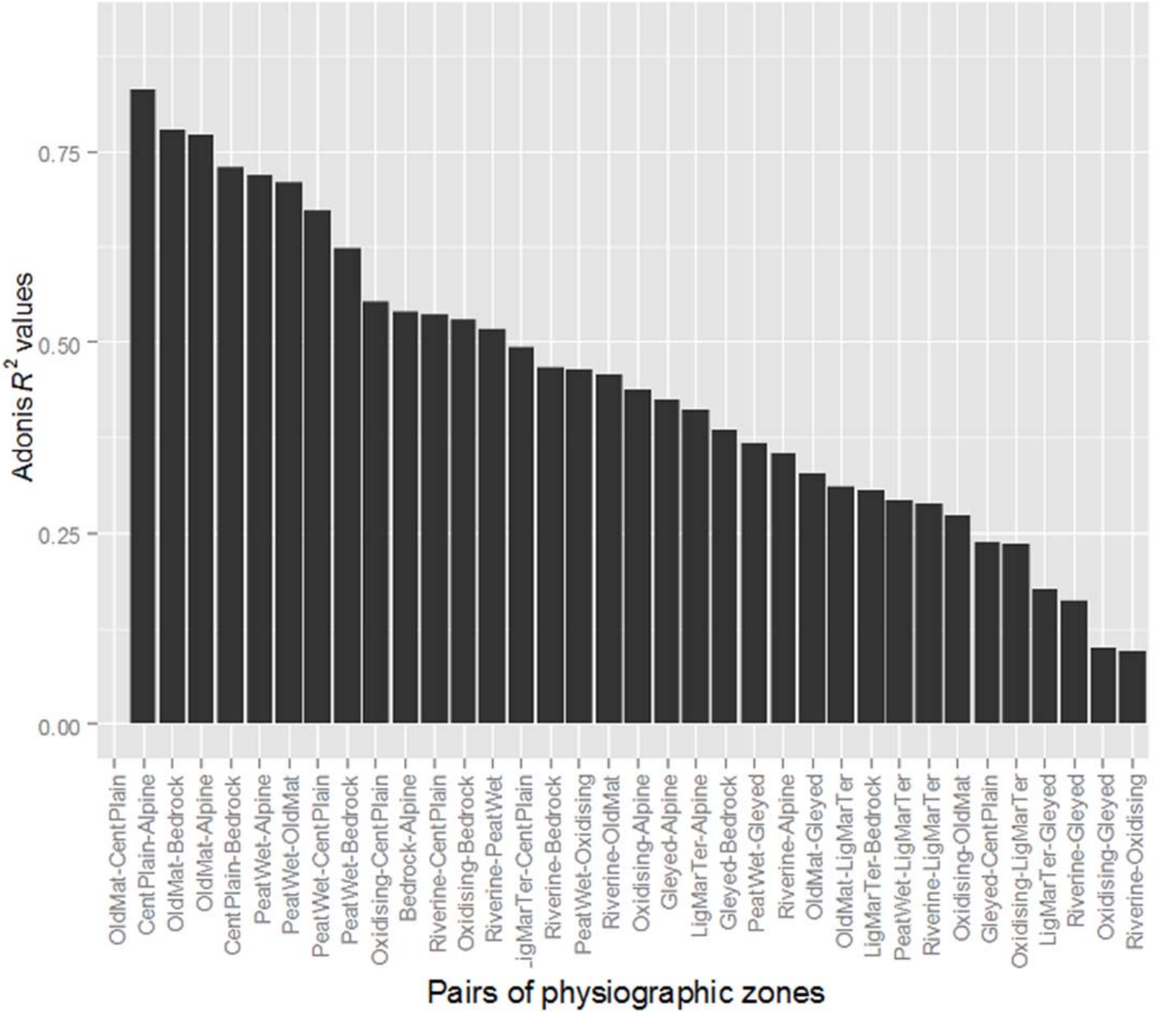


Figure 3. Results of ADONIS tests (R^2 values) of all pairs of physiographic zones based on the driver variables*. All tests were significant (p < 0.01) after correction for multiple comparisons using the FDR adjustment.*

The exact stress value for the three dimensional NMDS depended on the data subset but results from 10 trials were in the range 0.09-0.14. These stress values indicate that the ordination provides a reasonable representation of the original driver-space. Cells belonging to specific physiographic zones were clustered in specific locations on at least one of NMDS plots (Figure 4). This indicates that the physiographic zones occupied relatively isolated and compact parts of the multivariate space defined by the driver variables. For example, all zones, with the exception of Lignite Marine Terraces, occupied isolated positions of the gradient that was defined by the 1st dimension with Alpine, Bedrock and Riverine situated on the left side of this gradient and Peat Wetlands on the opposite end (Figure 4). The zones were less well discriminated on the other dimensions but isolation of some zones was evident, for example Central Plains, Bedrock and Old Mataura were reasonably isolated to parts of the 2nd dimension (Figure 4).

Some zones were not strongly isolated on any dimension but were compact (i.e. cells in the zone occupied a very restricted sector on the plots). This is particularly apparent considering that all units are represented in these plots by 100 cells. For example, the Central Plains zone appeared on the plots as only three points (Figure 4). This means that these zones are very homogeneous with respect to the variables representing the drivers.

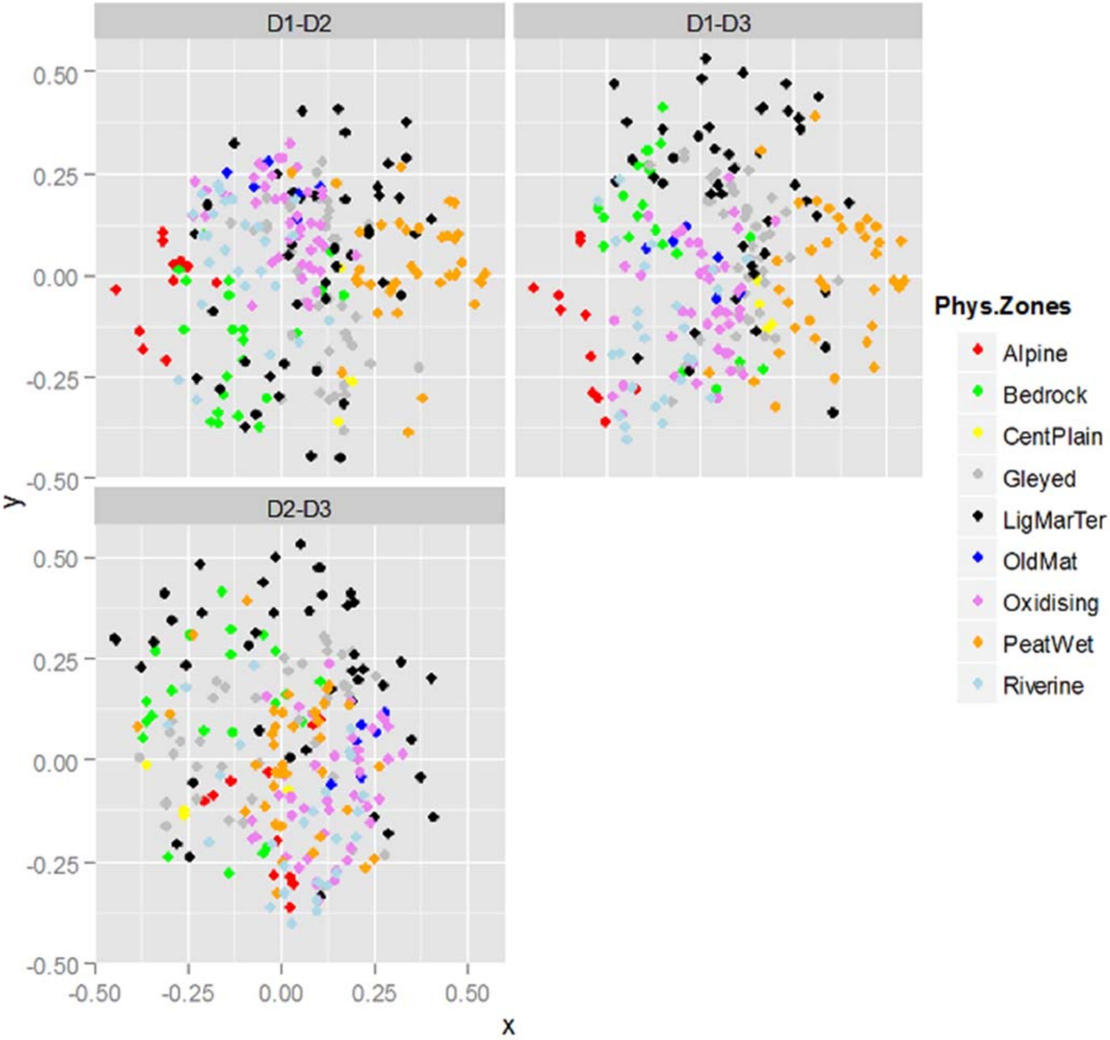


Figure 4. NMDS plot based on the driver data. *The plotted points represent 100 individual grid cells, which have been coloured according to the physiographic zones they are located within. The number of points appearing on the plot for some zones are fewer than 100, indicating that all driver variables took the same or very similar values for many cells. The first dimension named each of the panel titles refers to the x-axis and second to the y-axis.*

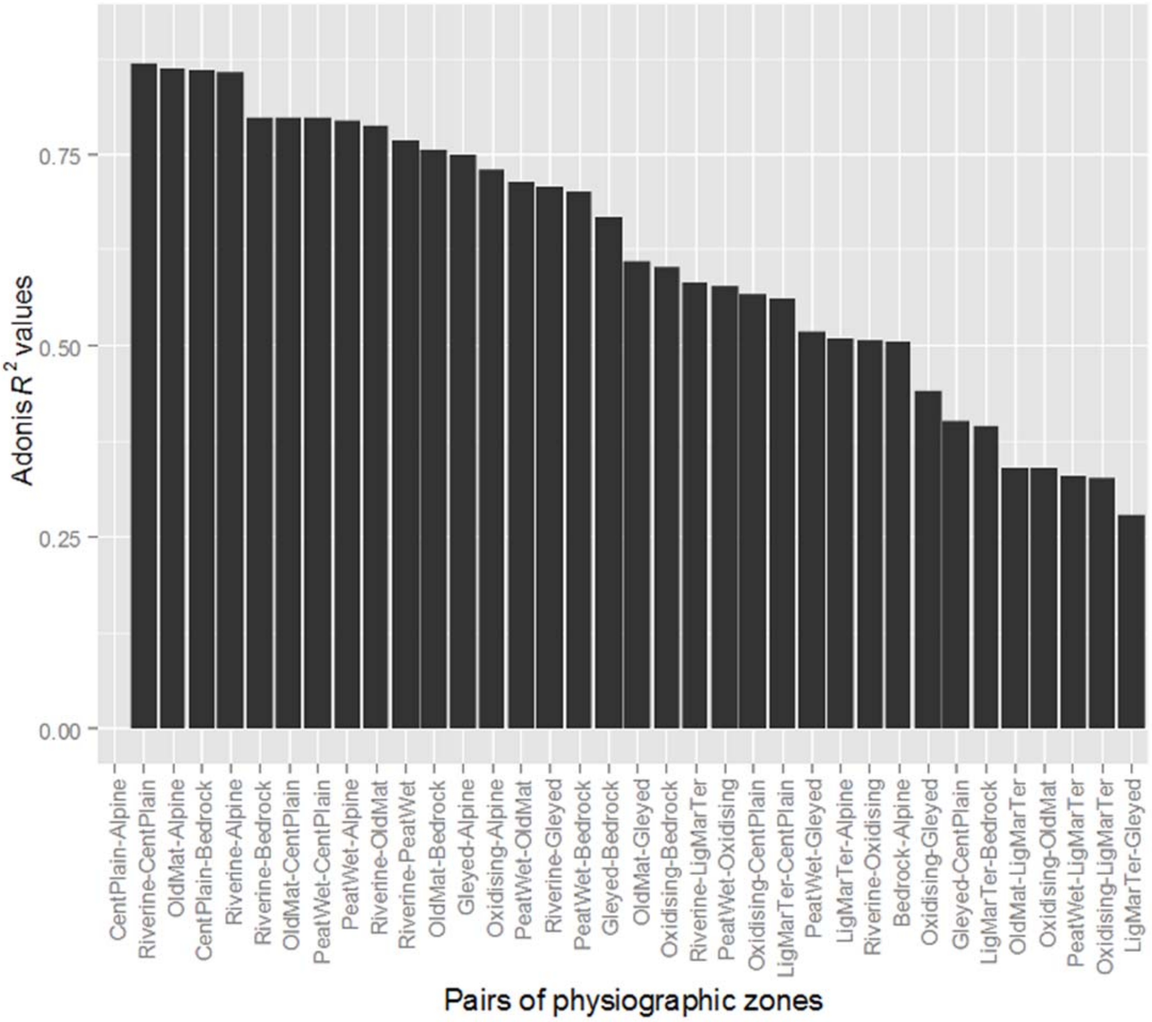
5.2 Discrimination of characteristics

The number of grid cells representing the physiographic zones was not even, but all zones were represented by at least 480 cells. Many cells had missing data for at least one variable and five variables that had more than 20% missing data were removed from the analysis; profile drainage, base saturation, anion storage capacity, mean annual drainage season, water table depth class. This left 12 variables for which missing values within some cells were replaced with the median of all non-missing values. A balanced subset was taken for the classification strength tests by randomly selecting 100 cells in each zones.

The non-parametric multivariate analysis of variance (ADONIS) produced a global R^2 value of 0.75, indicating that 75% of the total variance of the characteristics data was explained by the physiographic zones (p-value <0.01). It is noted that achieving significance is to be expected because of the large number of cases that were used in the test.

The R^2 values for pairwise ADONIS tests ranged from approximately 0.2 to 0.9 (Figure 5). All pairs were significant with $p<0.01$ (after correction for multiple comparisons using the FDR adjustment).

The exact stress value for the three dimensional NMDS depended on the data subset but results from 10 trials were in the range of 0.10-0.13. This indicates that the NMDS provides a fair representation of the original multivariate space represented by the characteristics data. Plots of the 1st, 2nd and 3rd NMDS dimensions indicated that the physiographic zones occupied relatively isolated and compact parts of the characteristics-space (Figure 6). For example, the location of some physiographic zones were distributed along the 1st dimension, in particular Alpine, Bedrock Hill County, Peat Wetlands and Oxidising and others on the 2nd dimension (e.g., Riverine, Old Mataura and Gleyed) (Figure 6).



***Results of ADONIS tests (R^2 values) of all pairs of physiographic zones based the characteristics variables**. All tests were significant ($p < 0.01$) after correction for multiple comparisons using the FDR adjustment.*

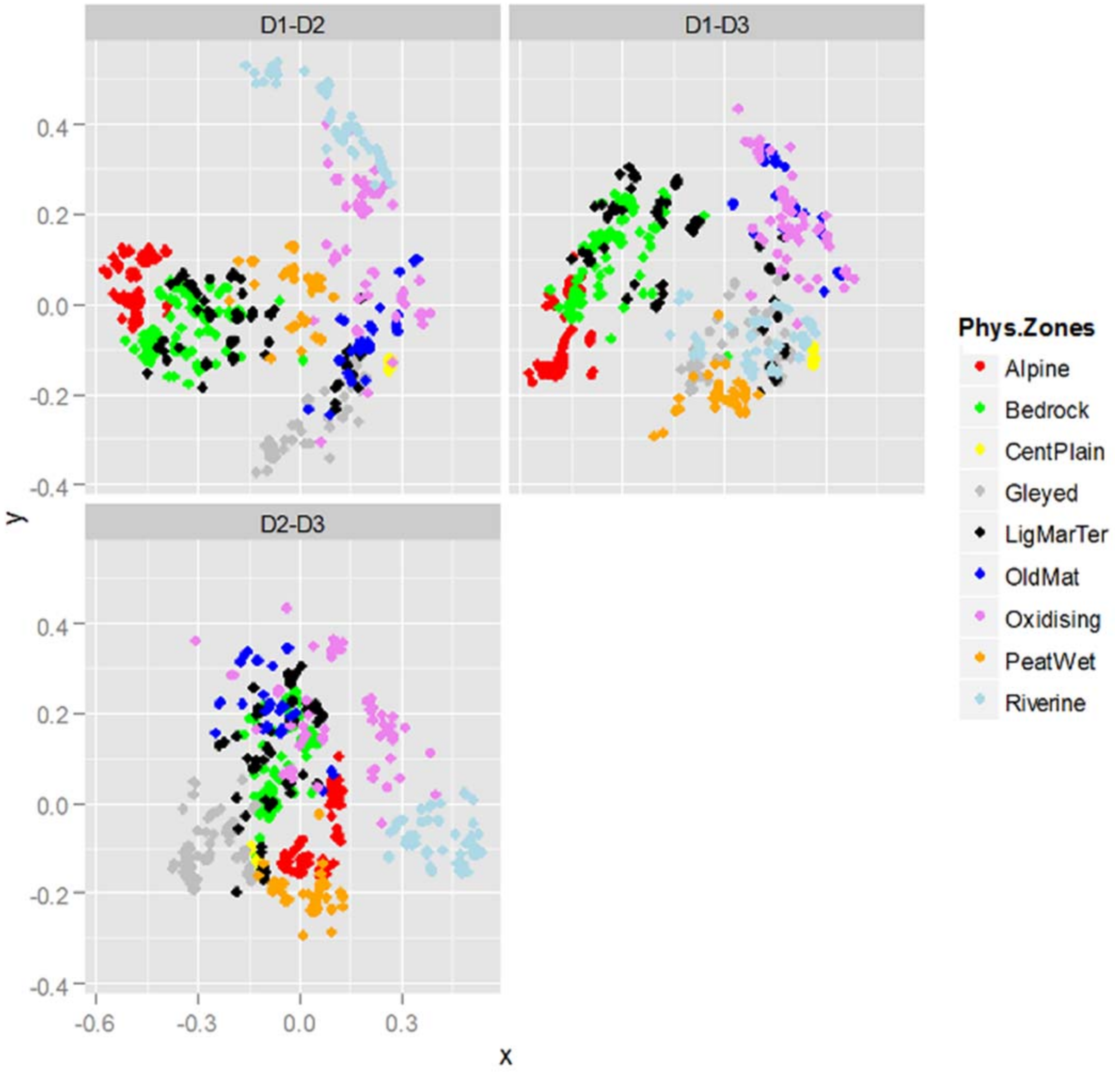


Figure 6. NMDS plot based on the characteristics data. *The plotted points represent 100 individual grid cells, which have been coloured according to the physiographic zones they are located within. The number of points appearing on the plot for some zones are fewer than 100, indicating that all driver variables took the same or very similar values for many cells. The first dimension named each of the panel tittles refers to the x-axis and second to the y-axis.*

5.3 Discrimination of river water quality

5.3.1 Preliminary steps

The 43 river monitoring sites that were retained for analysis were unevenly spread over physiographic zones (Figure 7). There were no sites in the Gleyed, Lignite-Marine and Central Plains zones, only one site in the Oxidising zone, three sites in the Peat Wetlands and Alpine zones, but at least six sites in the other three physiographic zones (Figure 7).

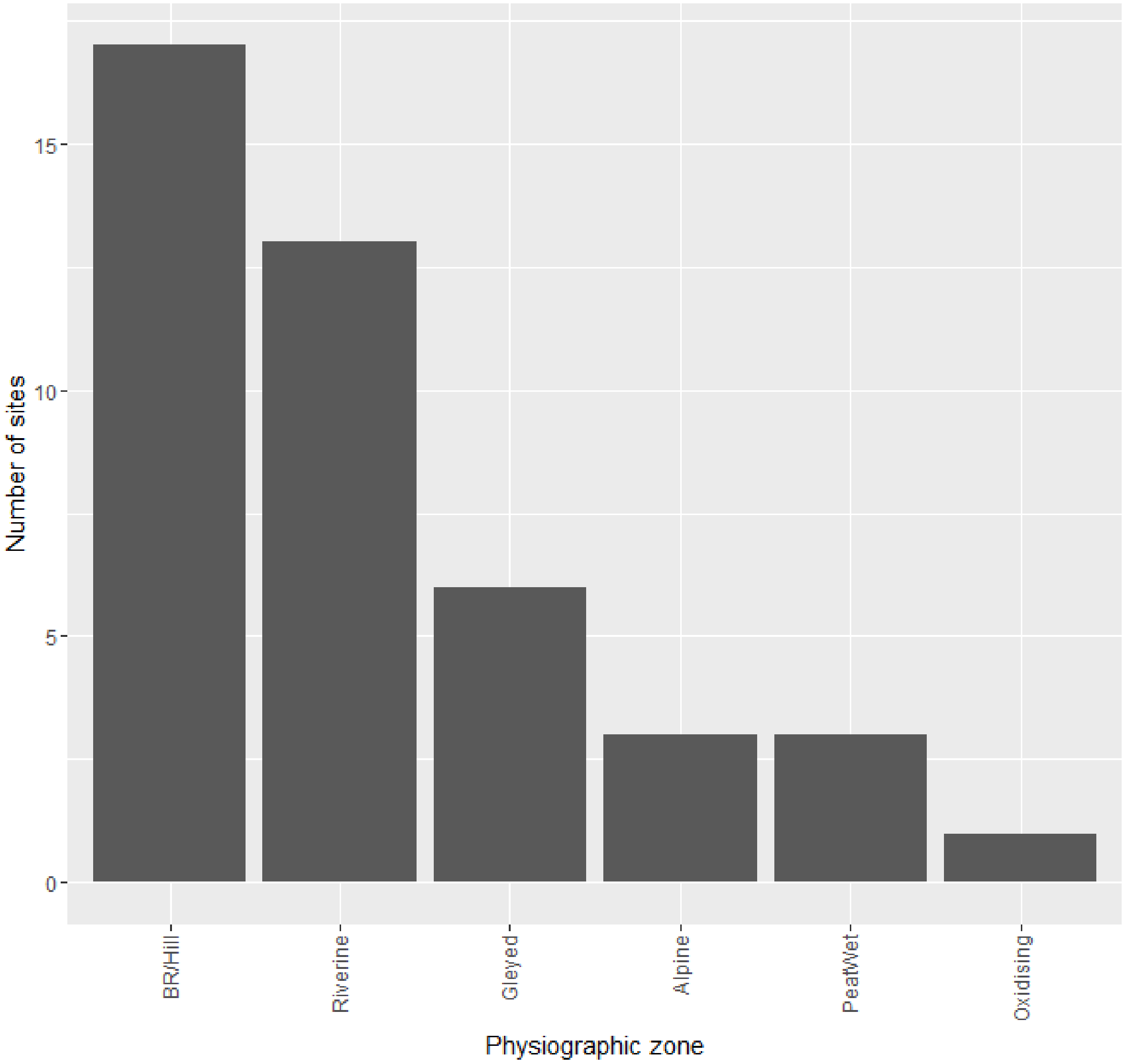


Figure 7. Distribution of river water quality sites over physiographic zones.

At all sites, all variables were sampled on four or more occasions, with the exception of CLAR, which was never sampled at 12 sites. A number of variables including COND, TN, TP, NNN and TKN were sampled on 30 or more occasions at all sites. A further subset of variables including Cl, Br, NH_4N, DRP, ECOLI and VSS were sampled on more than 11 occasions at all sites.

A large proportion of sites had median values less than the established detection limits for NH_4N and Br (Figure 8). Fewer sites had median values less than detection for NO_2N, DRP, Fe^{II} and Mn^{II} and no sites had medians less than detection for NNN and TSS (Figure 8).

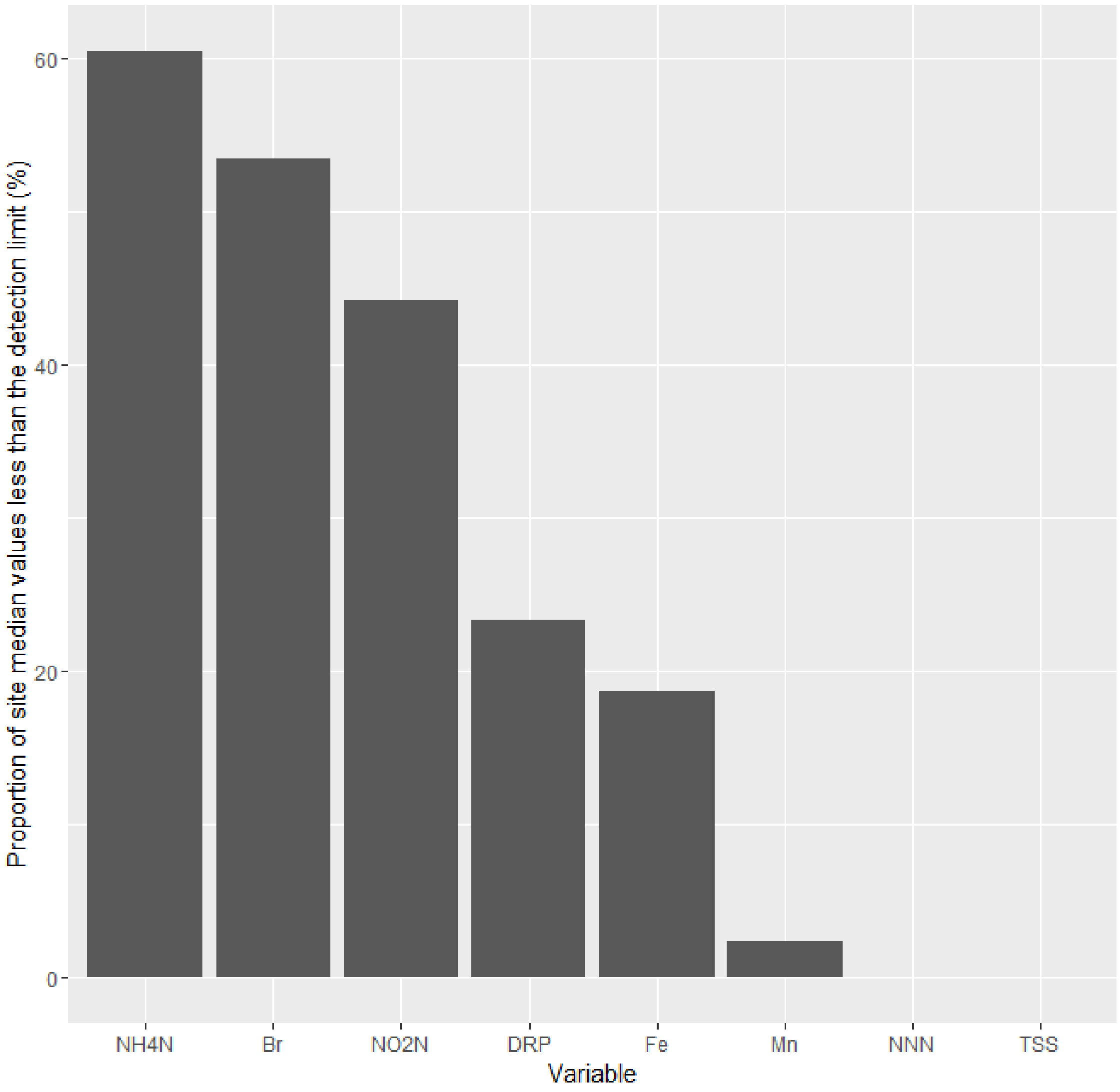


Figure 8. Proportion of site median values less than the detection limit for eight variables.

All 20 variables had reasonably strong relationships with %Pasture (Figure 9 and Figure 10). Linear regressions using %Pasture explained between 14% and 68% of the variation in the $\log_{10}$ transformed site median values (blue lines shown in Figure 9). The residuals from these models were obtained and used in subsequent analyses.

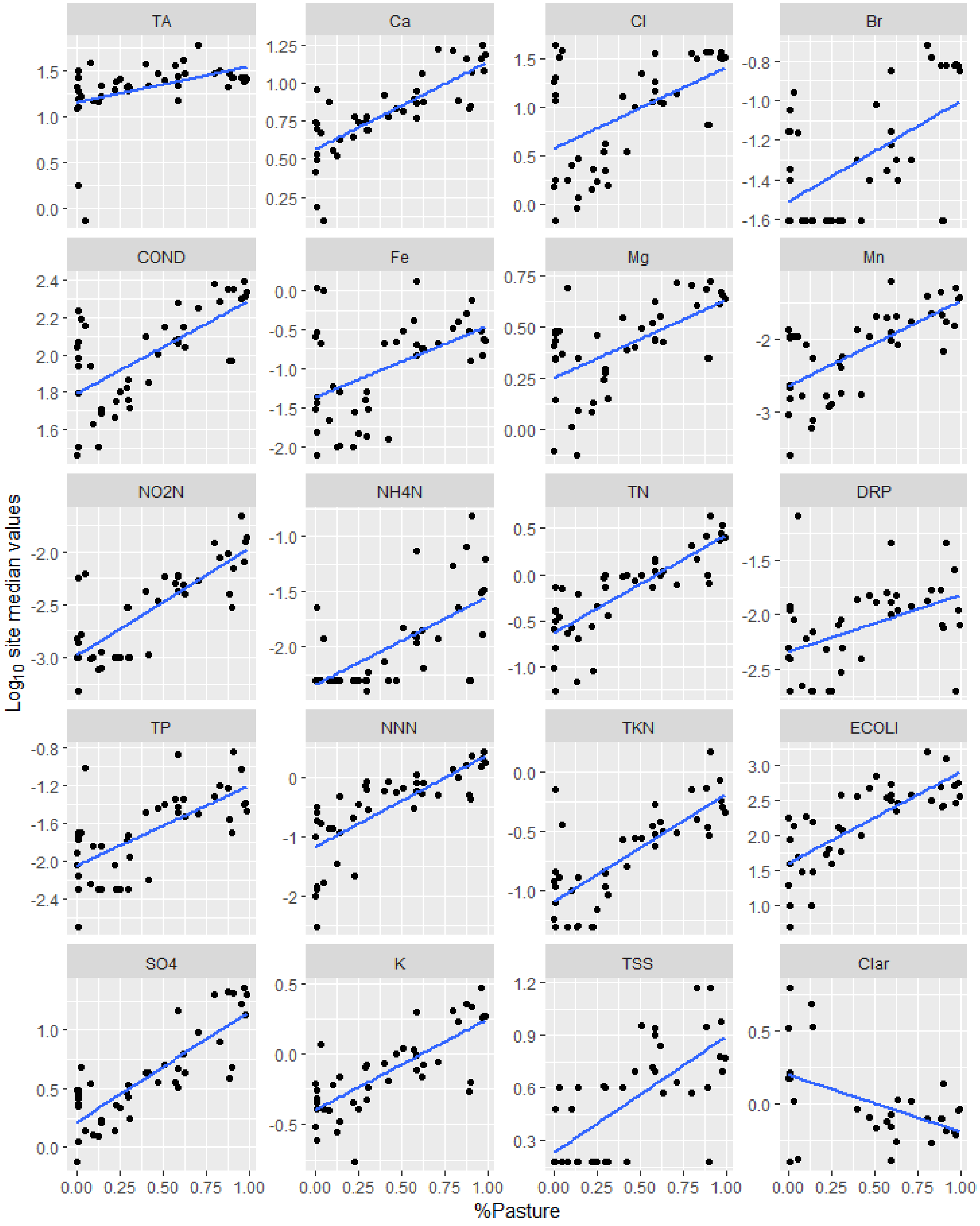


Figure 9. Scatterplots showing relationships of site median river water quality with %Pasture. *The blue lines represent linear regressions from which the residuals were obtained.*

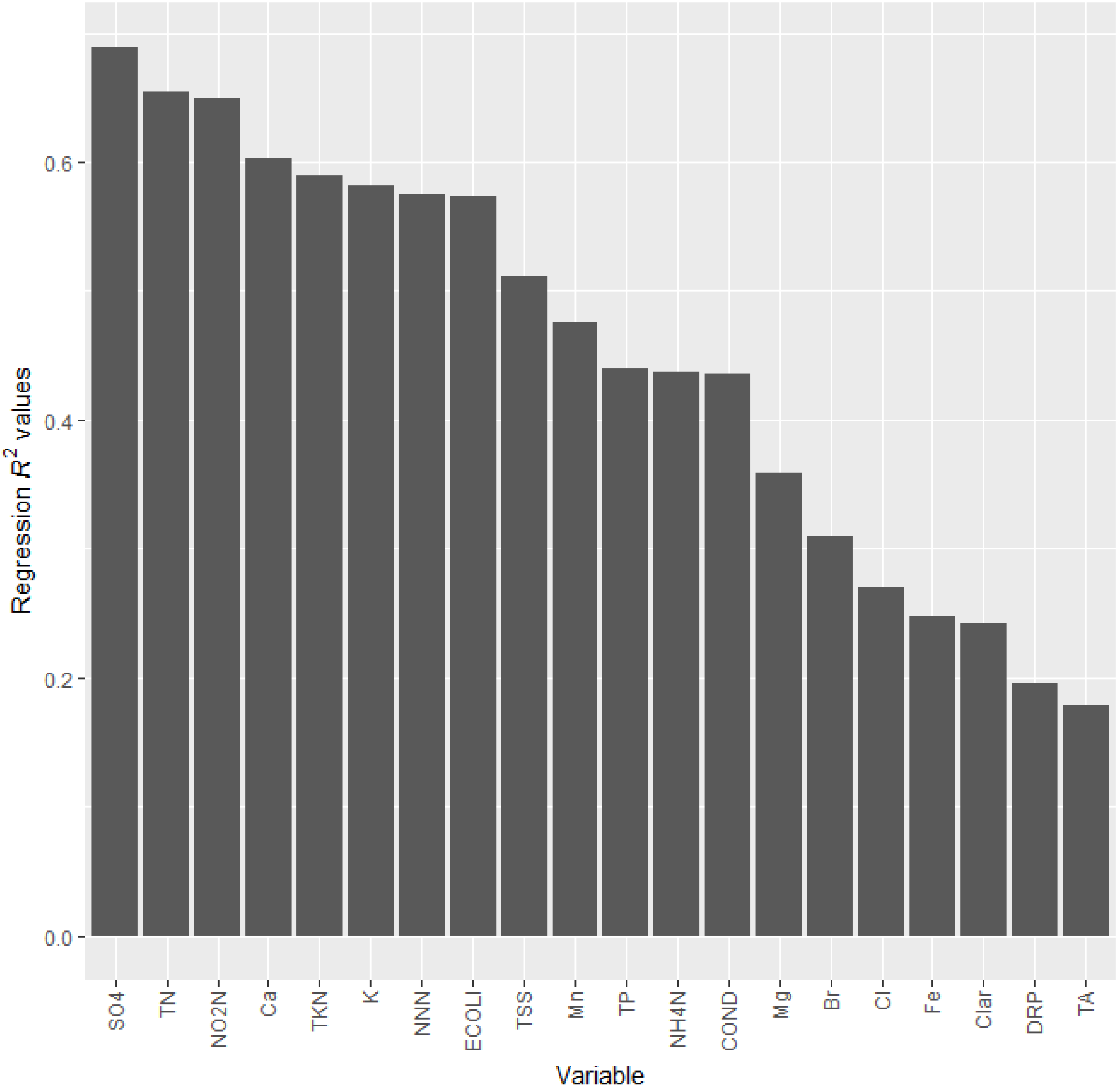


Figure 10. Coefficients of determination (R^2) values for regression of site median water quality values against %Pasture.

5.3.2 Strength of classification of river water quality by physiographic zones

The classification strength test was performed on the complete set of site median values of all variables. The missing CLAR values at 12 sites were replaced with the median of that variable (i.e. median of all non-missing sites). The global ADONIS R^2 value was 0.69, indicating that the physiographic zones explained 69% of the multivariate variation in river water quality ($p < 0.001$).

The pairwise ADONIS R^2 values indicated that many pairs of physiographic zones differed significantly with respect to their combinations of river water quality characteristics. The R^2 values exceeded 0.1 for all but one contrast (Figure 11). The adjusted *p*-values were < 0.05 for nine contrasts: Riverine-Gleyed, Gleyed-Alpine, Riverine-Peat Wetlands, Gleyed-Bedrock/Hill Country, Riverine-Alpine, Peat Wetlands-Gleyed, Bedrock/Hill Country-Alpine, Peat Wetlands-Bedrock/Hill Country, Riverine-Bedrock/Hill Country. Six contrasts did not achieve statistical significance, which was due in part to low sample size for several zones (Figure 11).

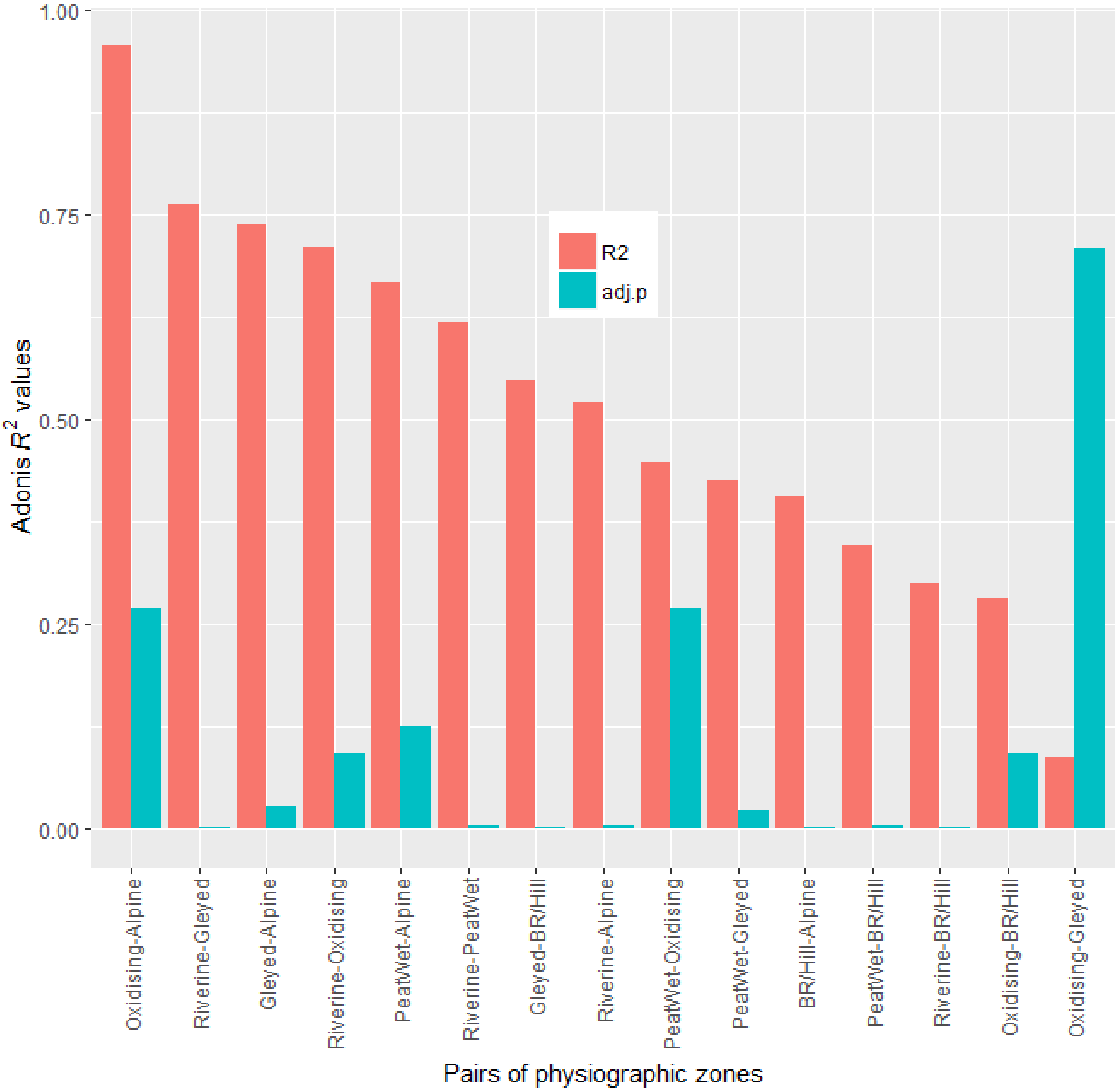


***Figure 11. Results of ADONIS tests (R^2 values) for all pairs of physiographic zones based the water quality variables**. The adjusted p-values (using the false discovery rate to correct for multiple comparisons) are <0.05 for nine pairs.*

The physiographic zones were strongly discriminated on the NMDS plots. The stress of this 3-dimensional NMDS plot was 0.04, which indicates that the ordination provides a good representation of the original space. The Alpine zone was particularly compact and isolated on the ordination diagram, indicating that this zone has very distinctive combinations of water quality variables and small within zone variation. The Riverine zone was also compact and adjacent to the Alpine zone, which it could be expected to be most similar to. The Peat Wetland zone, was not as compact but these sites were very isolated from all other sites.

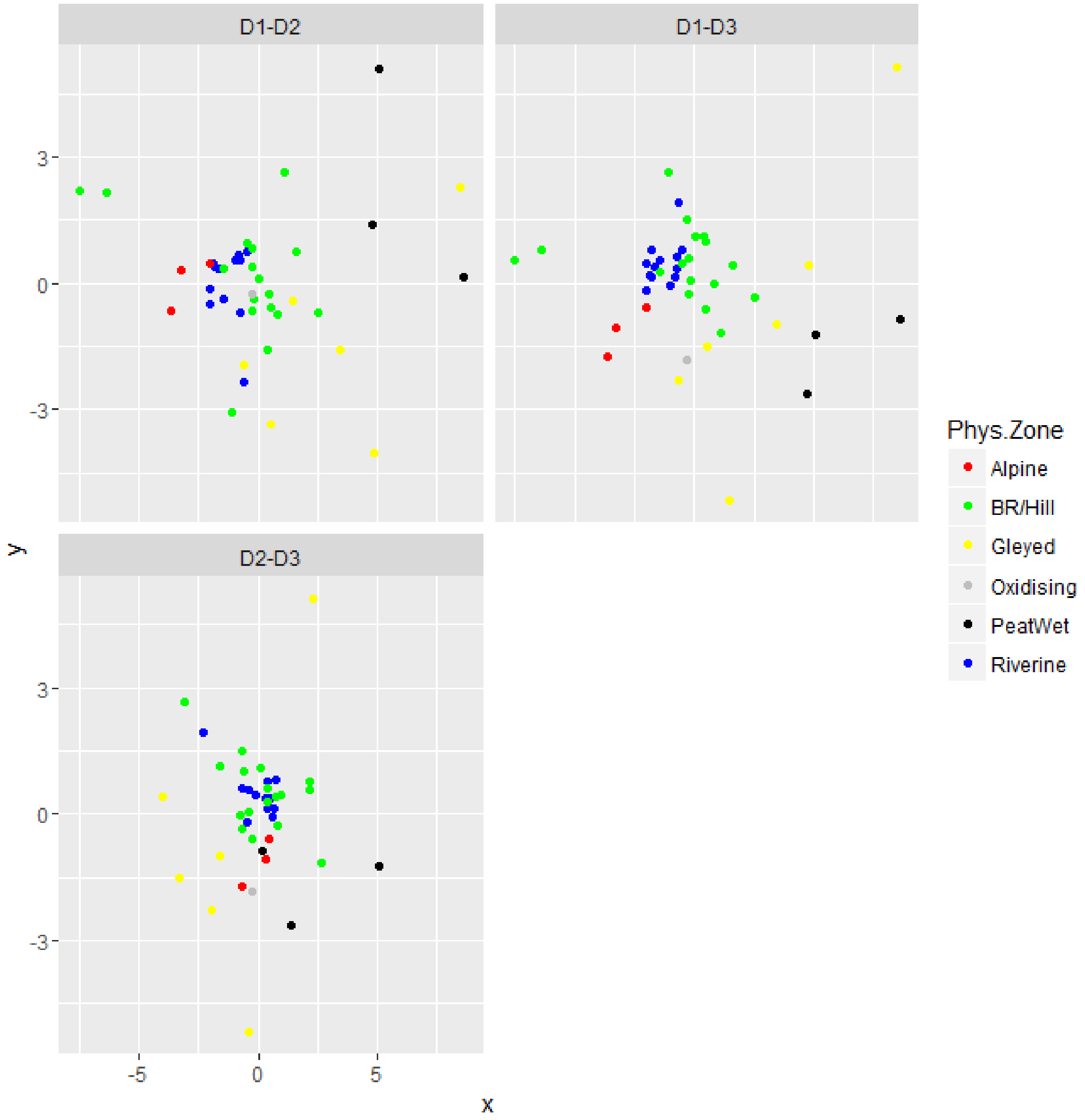


Figure 12. NMDS plot based on the river water quality data*. The plotted points represent individual sites, which have been coloured according to their physiographic zones.*

The classification strength test was repeated after replacing the site median values for each water quality variable with the residuals of the regressions of site median values against %Pasture. The missing residuals for CLAR at 12 sites were replaced with the median residual calculated over the other sites.

The global ADONIS R^2 value was 0.41, indicating that the physiographic zones explained 41% of the multivariate variation in the regression residuals (p<0.001). The pairwise ADONIS R^2 values indicated that all pairs of physiographic zones differed with respect to their combinations of water quality characteristics (R^2 >0.1 for all but two contrasts) and that the level of difference varied between contrasts (Figure 13). Statistical significance (adjusted p-values) was achieved for only five contrasts: Riverine-Peat Wetland, Riverine-Alpine, Riverine-Gleyed, Peat Wetland-Bedrock/Hill, Gleyed- Bedrock/Hill (Figure 13). All other adjusted *p*-values were >0.05.

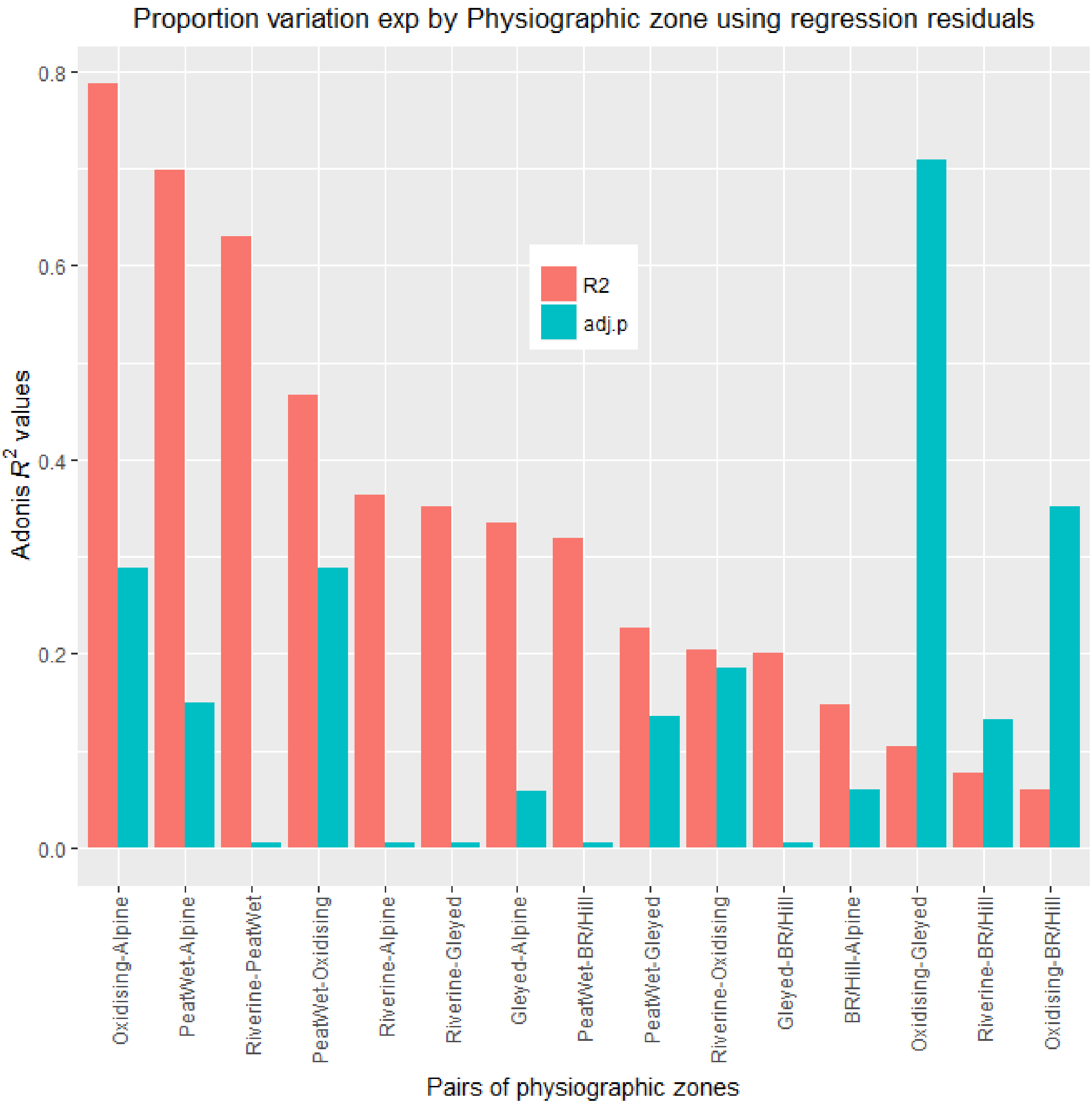


Figure 13. Results of ADONIS tests (R^2 values) of all pairs of physiographic zones based the residual of water quality variables regressed against %Pasture*. The adjusted p-values (using the false discovery rate to correct for multiple comparisons) are <0.05 for the contrasts for which the blue bar is only just visible.*

The physiographic zones were strongly discriminated on NMDS plots based on the regression residuals (Figure 14). The stress of this 3-dimensional NMDS plot was 0.07, which indicates that the ordination provides a reasonable representation of the original space. The Alpine zone was generally isolated and compact. The Riverine zone was also very compact but not as isolated as the Alpine zone. The Peat Wetlands zone was less compact, but was very isolated.

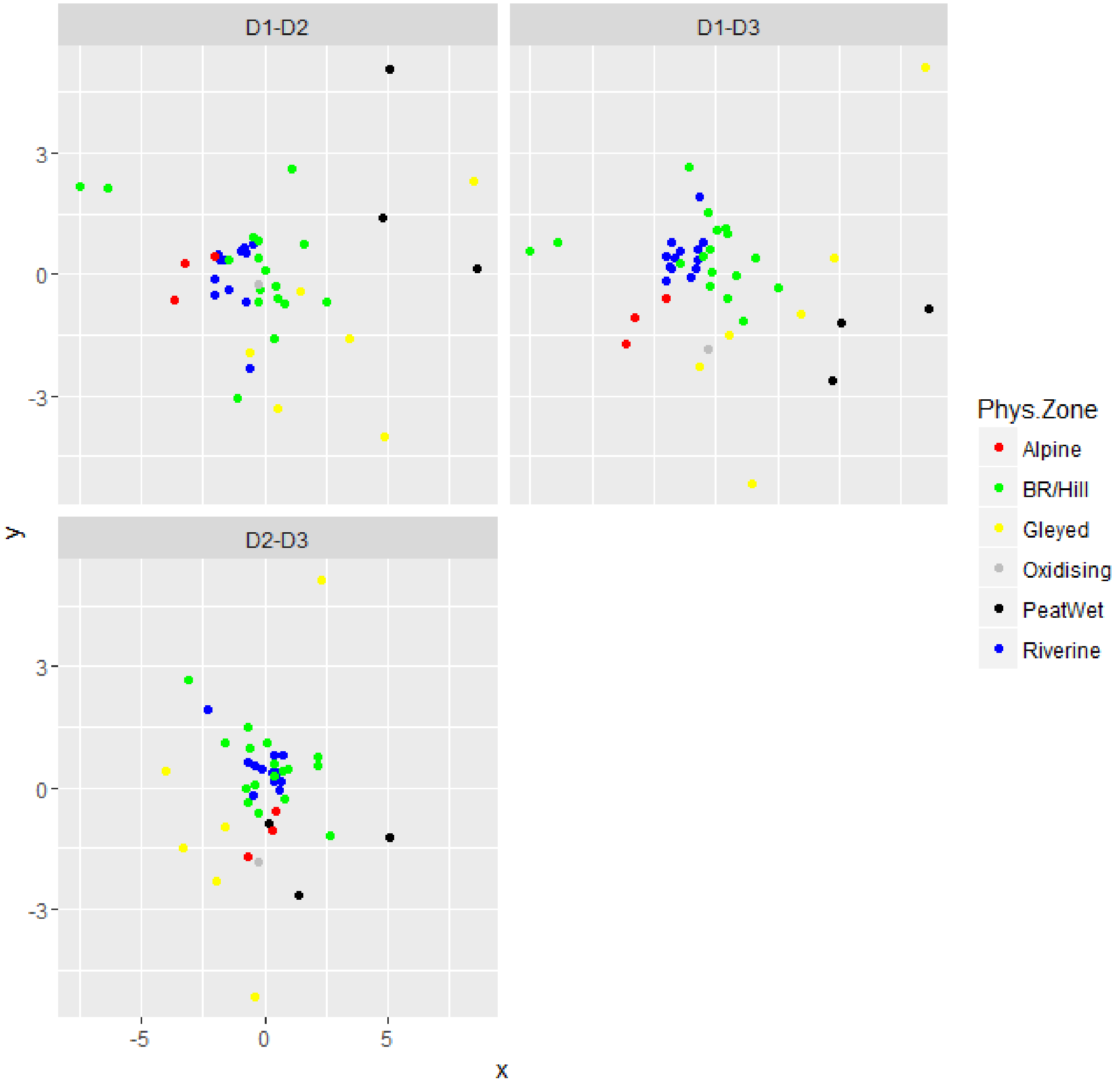


Figure 14. NMDS plot based the residuals of the site median values of the water quality variables regressed against %Pasture*. The plotted points represent individual sites, which have been coloured according to their physiographic zones.*

5.3.3 Discrimination of magnitudes of water quality variables by physiographic zones

A box and whisker plot (Figure 15) indicated that physiographic zones explain variation in the characteristic magnitudes of variables (site median values) to varying degrees. The plot indicates that median values for some variables exhibited large differences among physiographic zones (e.g., Ca, TN, COND and NNN) but others were less well discriminated (e.g., ECOLI, TKN and TP). Variables associated with redox (e.g. Fe^{II}, NNN), precipitation source (Cl) and substrate composition (e.g. Ca, TA) are generally well discriminated.

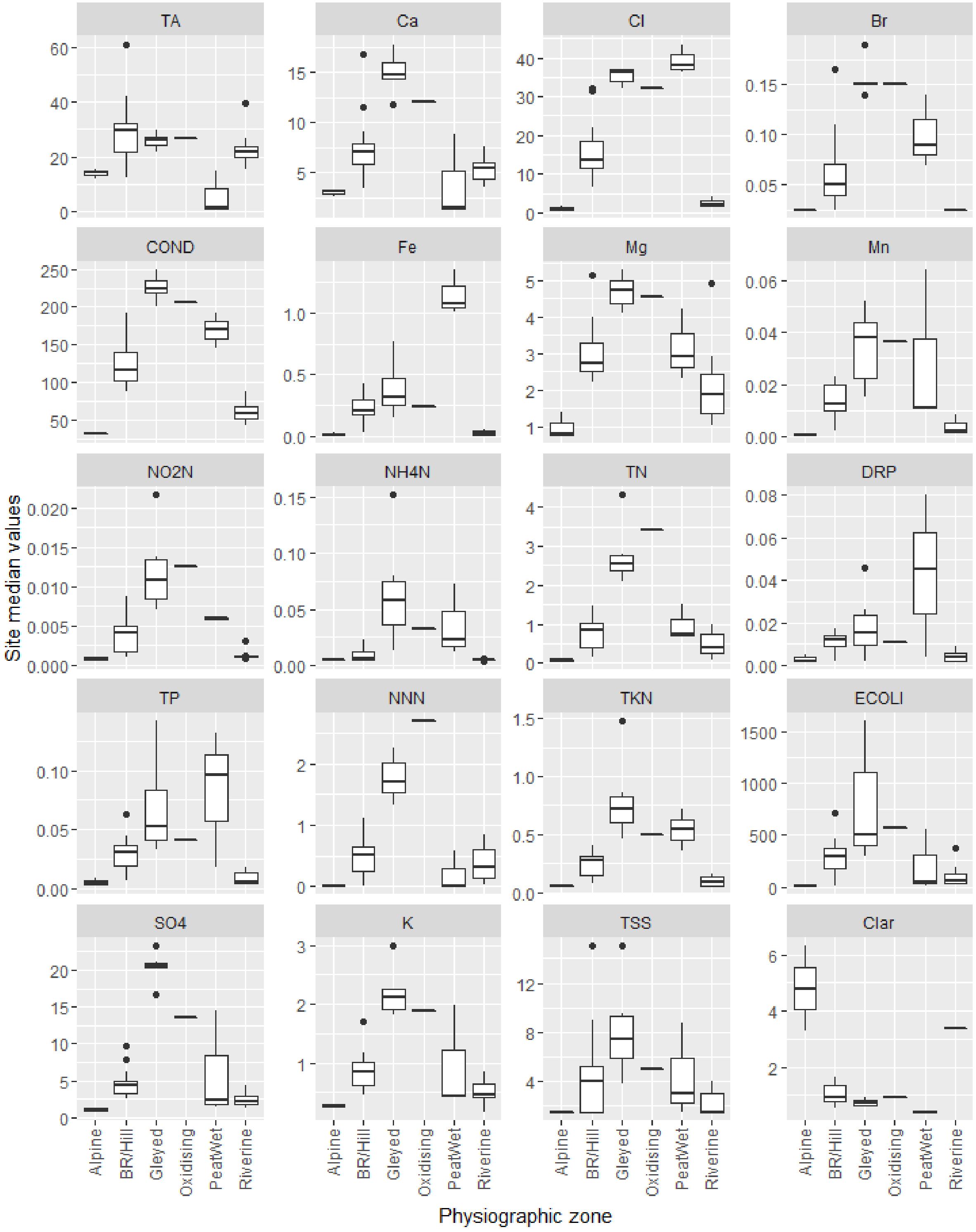


Figure 15. Box and whisker plots showing the distribution of site median values of the river water quality variables by physiographic zone. *The top and bottom lines of the rectangle represent the 3rd and 1st quartiles and the central line represents the median. The whiskers extend to the 3rd and 1st quartiles plus and minus 1.5 times the interquartile range. The solid dots represent data beyond 1.5 times the interquartile range.*

ANOVA tests performed on the log (base 10) transformed site median values shown in Figure 16 indicated the physiographic zones explained differing levels of the variation in the individual water quality values. All results were highly significant (Figure 16).

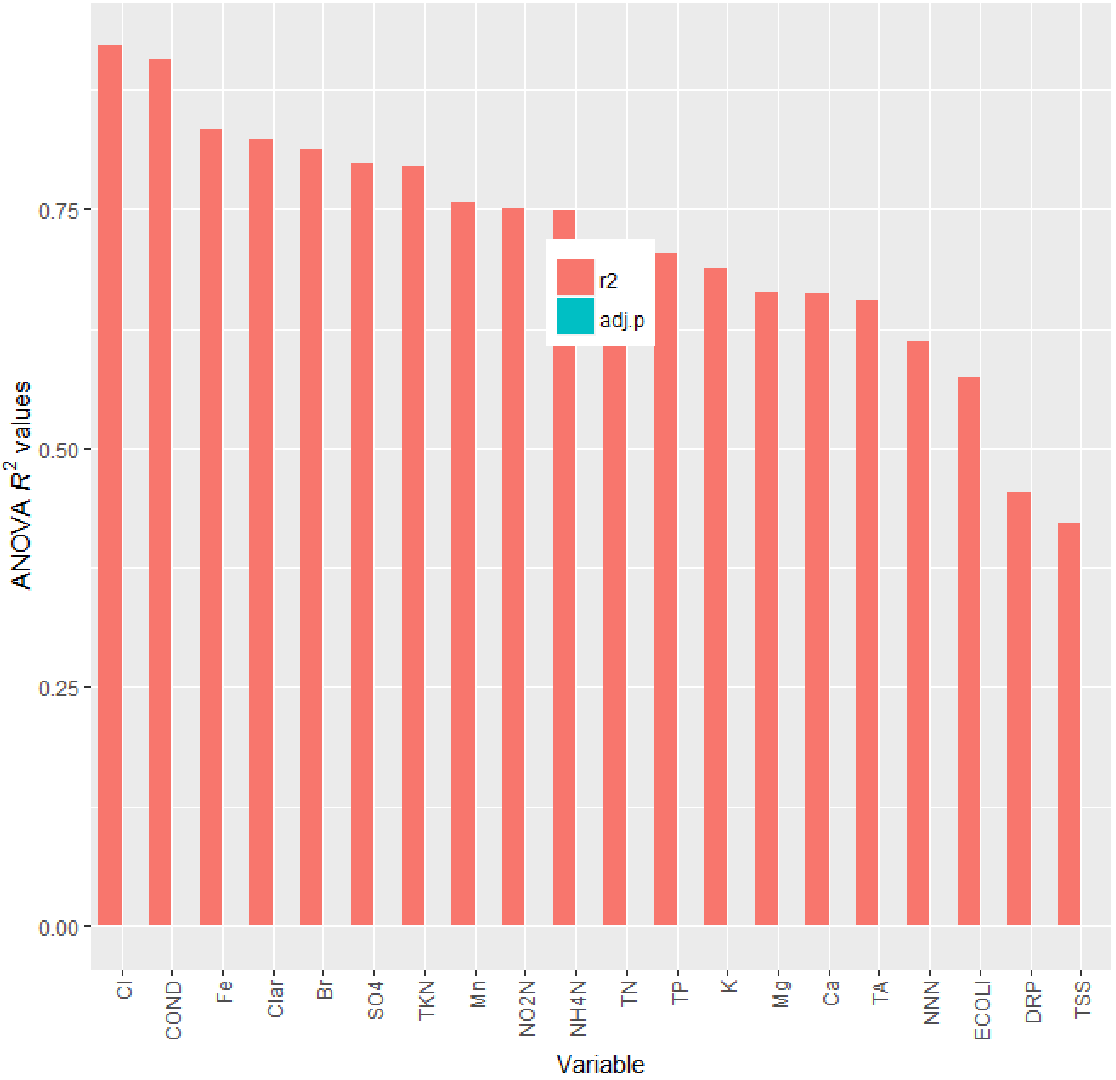


Figure 16. Results of ANOVA tests performed on site median values of river water quality variables. *The site median values were log (base 10) transformed to normalise their distributions. All results were highly significant so the adjusted p-values (using the false discovery rate to correct for multiple comparisons) are too small to be visible on the graph (p-values all <0.01).*

Boxplots indicated that some of the variation in the residuals of the regression of each variable versus %Pasture (the ‘regression residuals’) was explained by physiographic zone Figure 17). The regression residual associated with variables that reflect recharge mechanism (e.g., COND) and associated soil zone processes (e.g., K, SO_4, TA) were more strongly discriminated by physiographic zone than other variables (Figure 17). The regression residuals representing the dissolved solute variables (e.g., Cl, Mn^{II} and Br) were lowest in the Alpine and Riverine zones. The regression residuals for variables that reflect redox (e.g. DRP, Fe^{II}, Mn^{II}) were also discriminated by the physiographic zones, with zones that are expected to be reducing (i.e. Peat Wetlands, Gleyed and Bedrock/Hill Country) having the highest values. It is noted that while NNN did not achieve statistical significance, total forms of nitrogen (e.g. TKN, TN) were significantly discriminated by physiographic zone and the relative differences in values of these variables between zones were consistent with expected differences in redox.

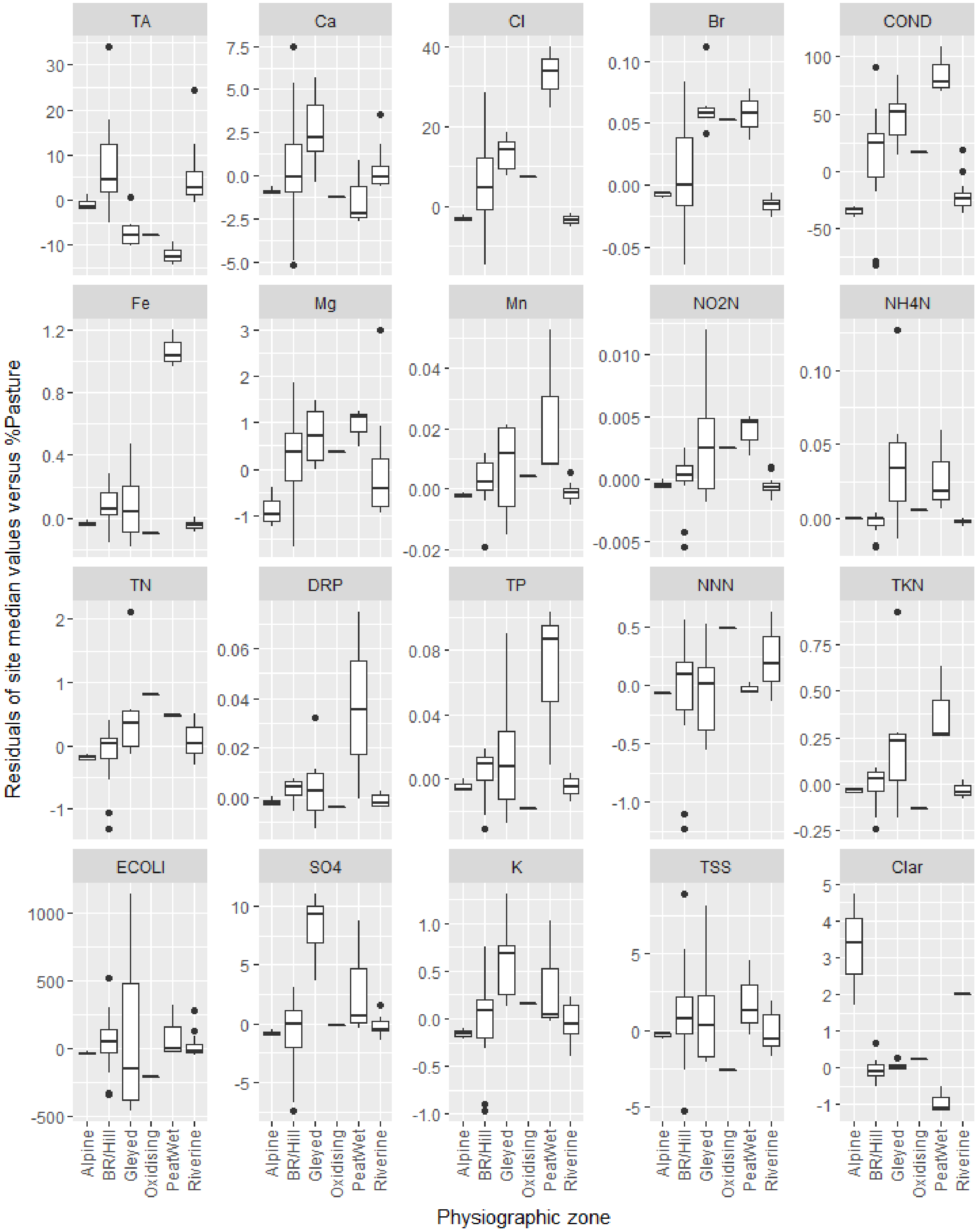


Figure 17. Box and whisker plots showing the distribution of regression residuals by physiographic zone. *The top and bottom lines of the rectangle represent the 3rd and 1st quartiles and the central line represents the median. The whiskers extend to the 3rd and 1st*

quartiles plus and minus 1.5 times the interquartile range. The solid dots represent data beyond 1.5 times the interquartile range.

ANOVA tests performed on the data shown in Figure 17 indicated the physiographic zones explained differing levels of the variation in the regression residuals (Figure 18). Most variables had significant p values ($p<0.05$) but the adjusted *p*-values for Mg, Ca, NNN, TSS, and ECOLI were >0.05 (Figure 18).

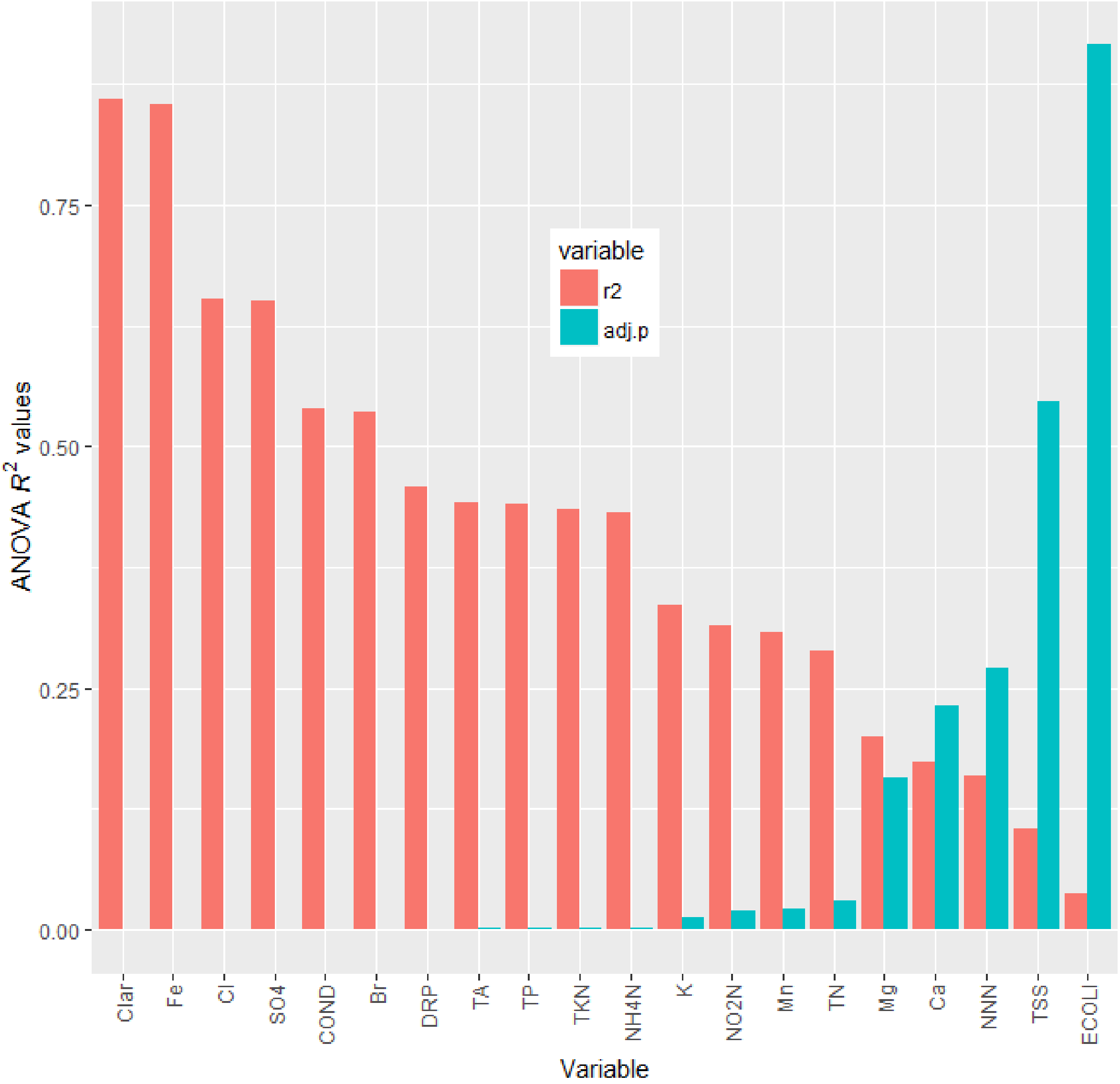


Figure 18. Results of ANOVA tests performed on the residuals of regressions of water quality variables versus %Pasture. *The adjusted p-values (using the false discovery rate to correct for multiple comparisons) are < 0.05 for the variable left of Mg.*

5.3.4 Discrimination of general variability by physiographic zone

A box and whisker plot (Figure 19) indicated that variation of the individual water quality variables (site median values) were discriminated to varying degrees by physiographic zones. Generally, those zones which exhibited the lowest magnitudes for individual variables (Figure 14) exhibit the highest corresponding coefficients of variation in Figure 19 (e.g., COND, TN, Br, SO_4 in the Alpine zone and CLAR, TA, Ca in Peat Wetlands). We note that variables that reflect soil zone signatures (e.g. K) exhibit greater variability in zones dominated by land surface recharge.

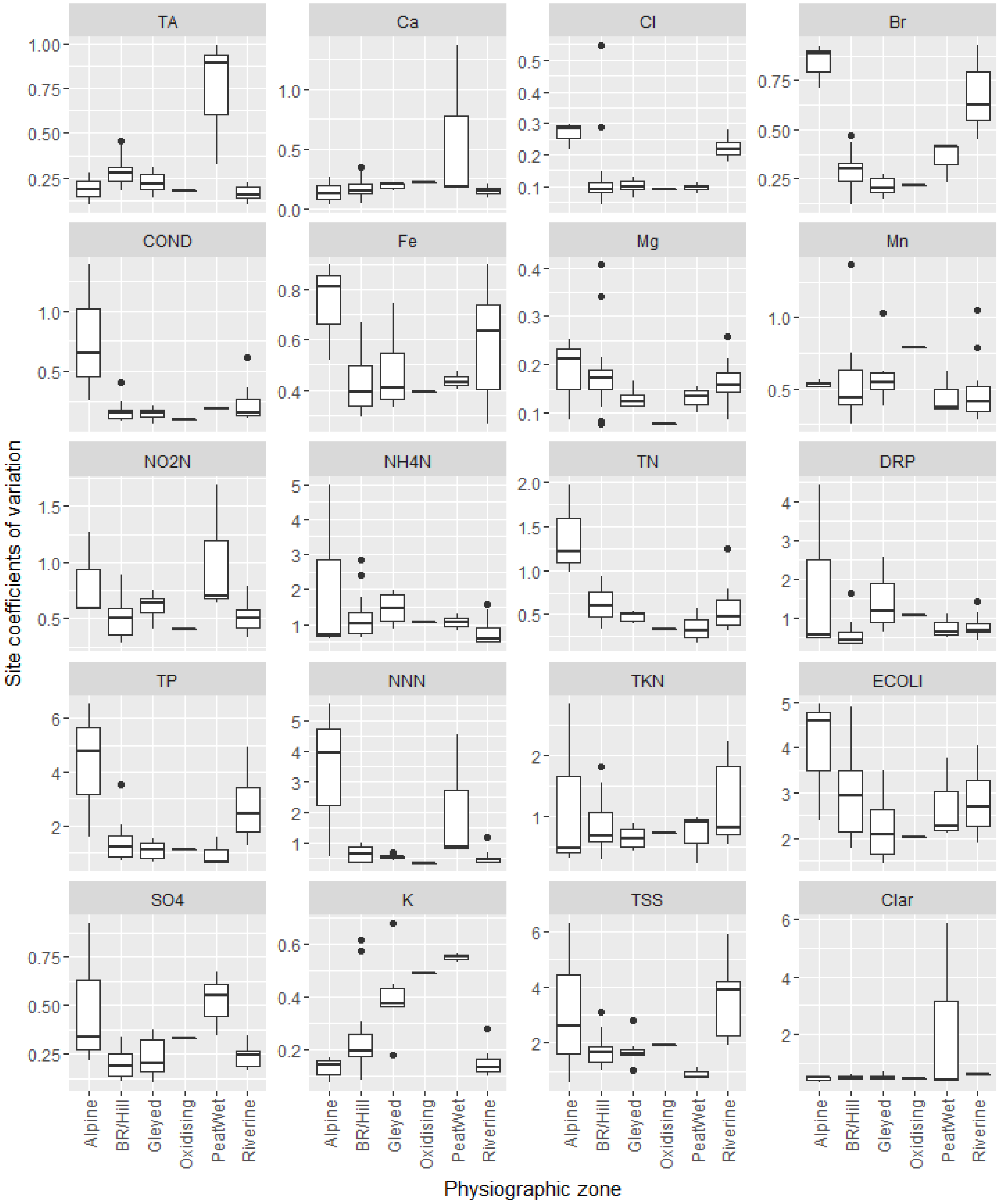


***Figure 19. Box and whisker plots showing the distribution of site coefficients of variation of the river water quality variables by physiographic zone.** The top and bottom lines of the rectangle represent the 3rd and 1st quartiles and the central line represents the median. The whiskers extend to the 3rd and 1st quartiles plus and minus 1.5 times the interquartile range. The solid dots represent data beyond 1.5 times the interquartile range.*

ANOVA tests performed on the site coefficient of variation values shown in Figure 19 indicated the physiographic zones explained differing levels of the variation in the individual water quality values. Fourteen results were significant (Figure 20).

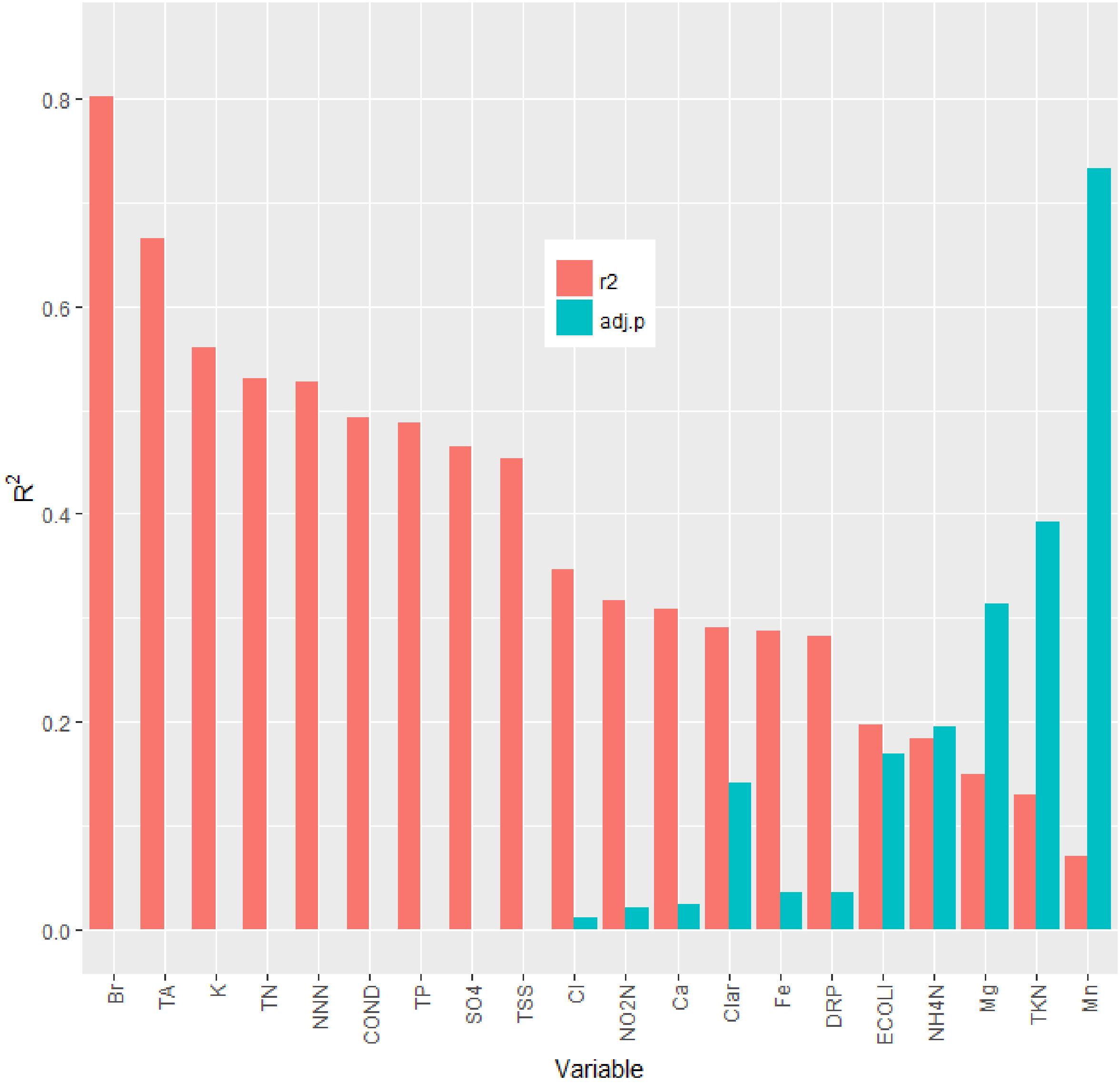


Figure 20. Results of ANOVA tests performed on the site coefficient of variation values of the individual water quality variables. *The adjusted p-values (using the false discovery rate to correct for multiple comparisons) are <0.05 for all variables left of ECOLI but excluding CLAR.*

5.3.5 Discrimination of variation in seasonality by physiographic zone

The water quality variables generally exhibited seasonal variation but the degree of seasonality varied considerably between sites (Figure 21). The variation of site seasonal indices was discriminated to some extent by physiographic zones (Figure 21).

The observation that physiographic zone explains some of the variation of site seasonal indices was confirmed by ANOVA (Figure 22). The physiographic zones explained significant differences in seasonality index for the following water quality variables: Br, TA, NH_4N, K, TSS, Mn^{II}, TKN, TP, Cl, SO_4, DRP, COND, ECOLI and TN (adjusted p-values <0.05). Figure 19 shows seasonality was well differentiated for Br and Cl. This was most pronounced in the Alpine, Riverine and Bedrock/Hill Country zones. Across other zones, seasonality was greatest in SO_4, K, TSS, and TA. Redox sensitive variables (e.g. Fe^{II}, NNN) exhibited limited seasonality.

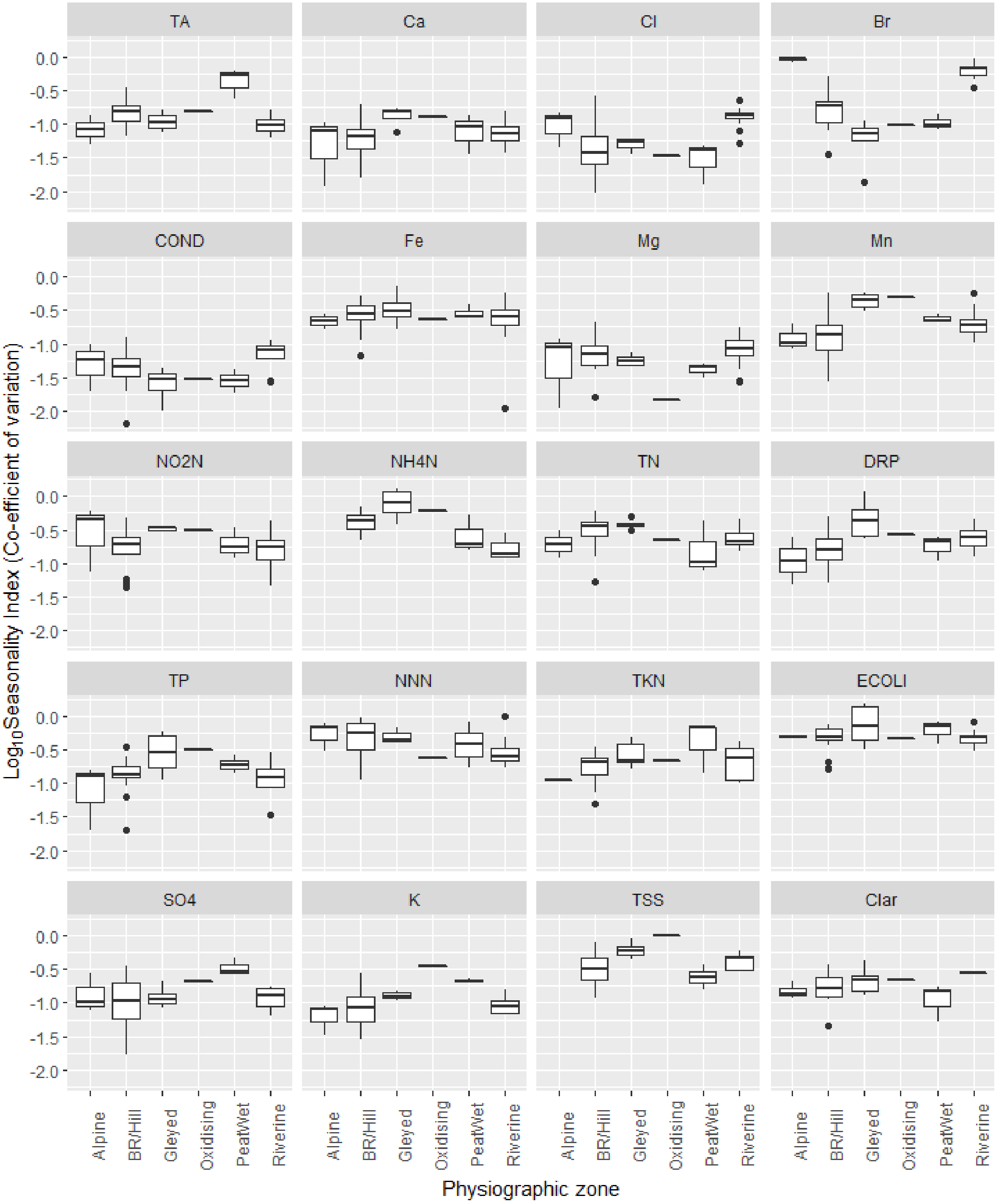


Figure 21. Box and whisker plots showing the distributions of site seasonality index values of the river water quality variables by physiographic zone. *The top and bottom*

lines of the rectangle represent the 3rd and 1st quartiles and the central line represents the median. The whiskers extend to the 3rd and 1st quartiles plus and minus 1.5 times the interquartile range. The solid dots represent data beyond 1.5 times the interquartile range.

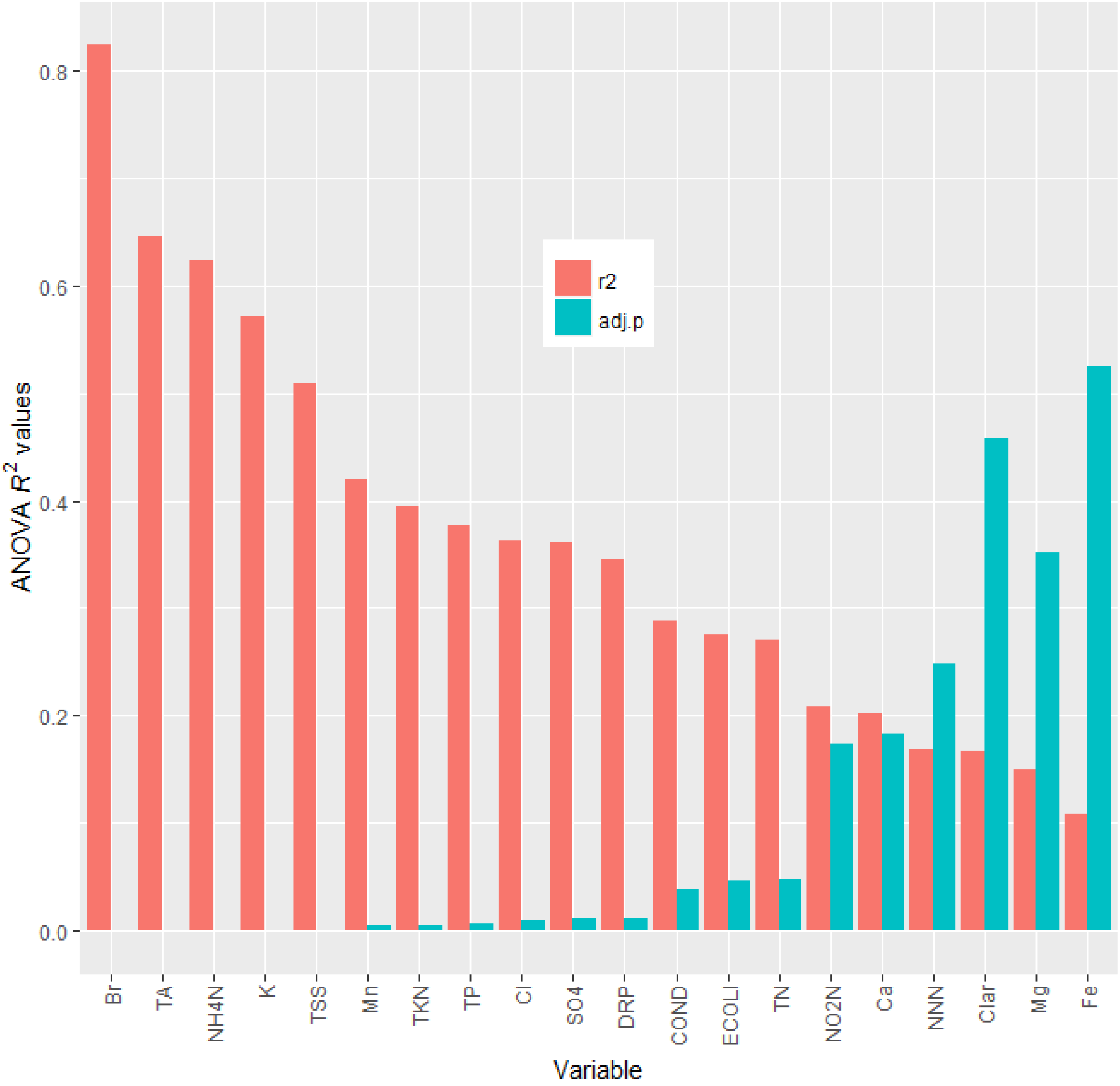


***Figure 22. Results of ANOVA tests performed on site seasonality indices for all river water quality variables.** All results right of TN did not achieve statistical significance (adjusted p-values >0.05).*

5.4 Discrimination of groundwater quality

5.4.1 Preliminary steps

The 254 sites that were retained for analysis were unevenly spread over physiographic zones (Figure 23). There were no sites representing the Alpine zone, less than five sites in the Bedrock/Hill Country and Peat Wetlands zones but at least 10 sites in all other physiographic zones (Figure 23).

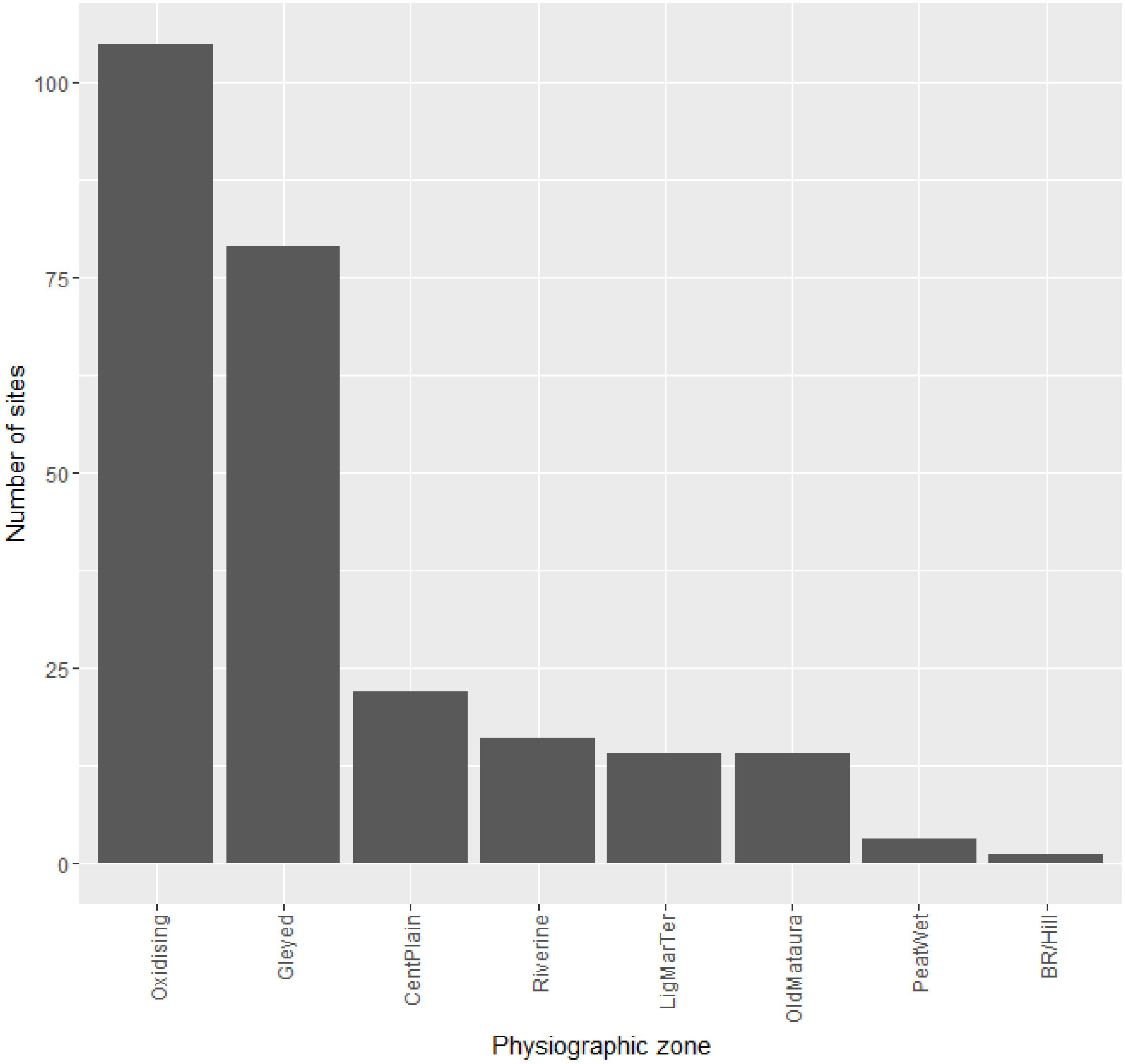


Figure 23. Distribution of groundwater quality sites over physiographic zones.

The number of sites that had data for each of the groundwater variables ranged from all 254 sites for COND and NNN to less than 100 sites for B and Si (Figure 24).

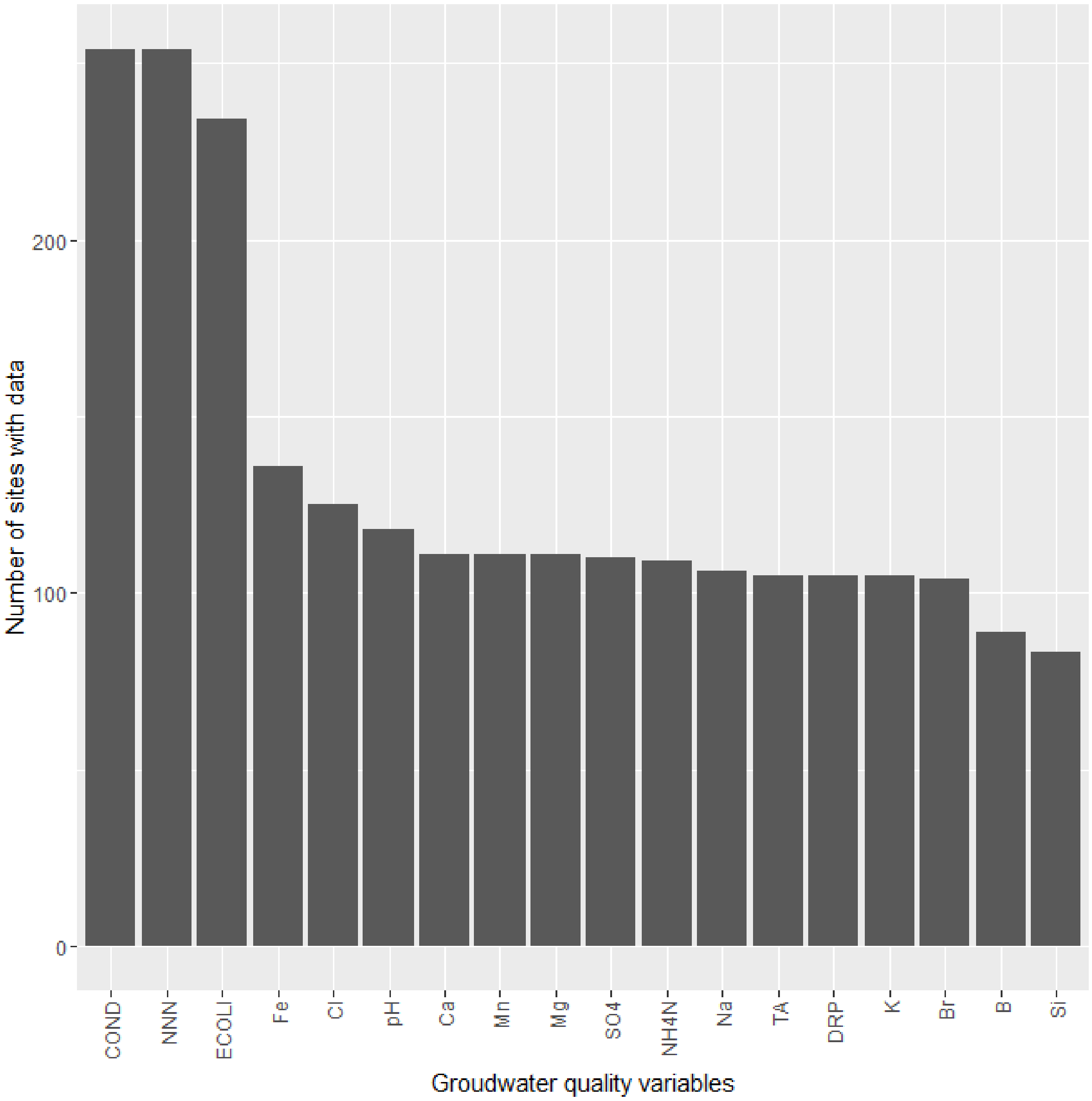


Figure 24. Numbers of sites with data for each of the groundwater quality variables.

A small proportion of sites had median values less than the established detection limits for DRP, NNN, Fe^{II} and Mn^{II} (Figure 25). No sites had median values less than detection for Br (Figure 25).

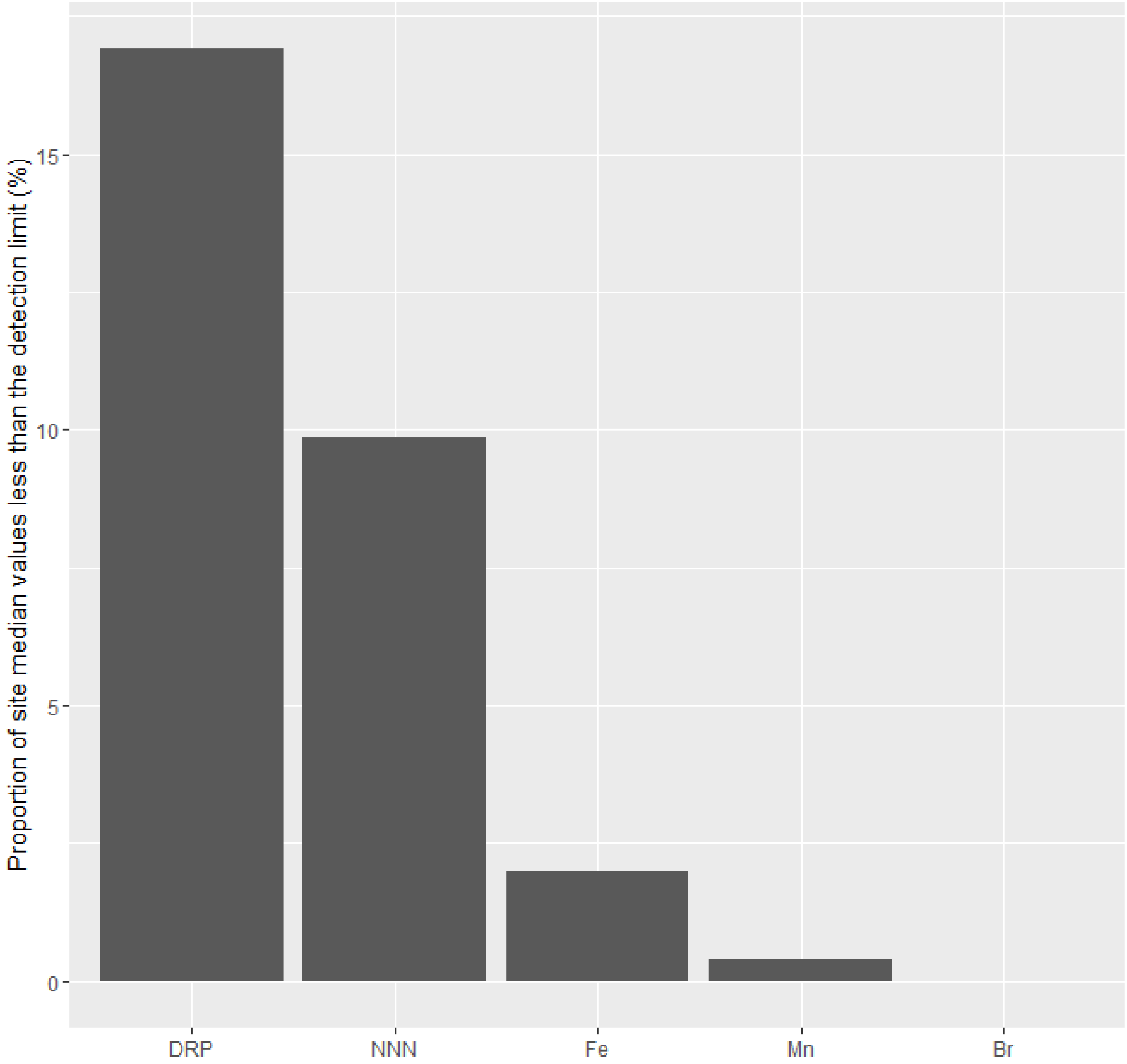


Figure 25. Proportion of site median values less than the detection limit for five variables.

Samples were taken from a range of depths across all physiographic zones (Figure 26). Inspection of scatter plots of site median values of the water quality variables versus bore depth indicated that there were not strong relationships (Figure 27). The only significant relationships ($p<0.05$) were for NNN, pH, SO_4 and TA, which had R^2 values of 0.07, 0.3, 0.10 and 0.15 respectively. We concluded that there was little justification for stratifying the data by bore depth, especially given that this would reduce the sample sizes in the analyses that follow. For the analyses that follow, all data was retained and sites were not stratified. The majority (>83%) of bores were less than 20 metres deep.

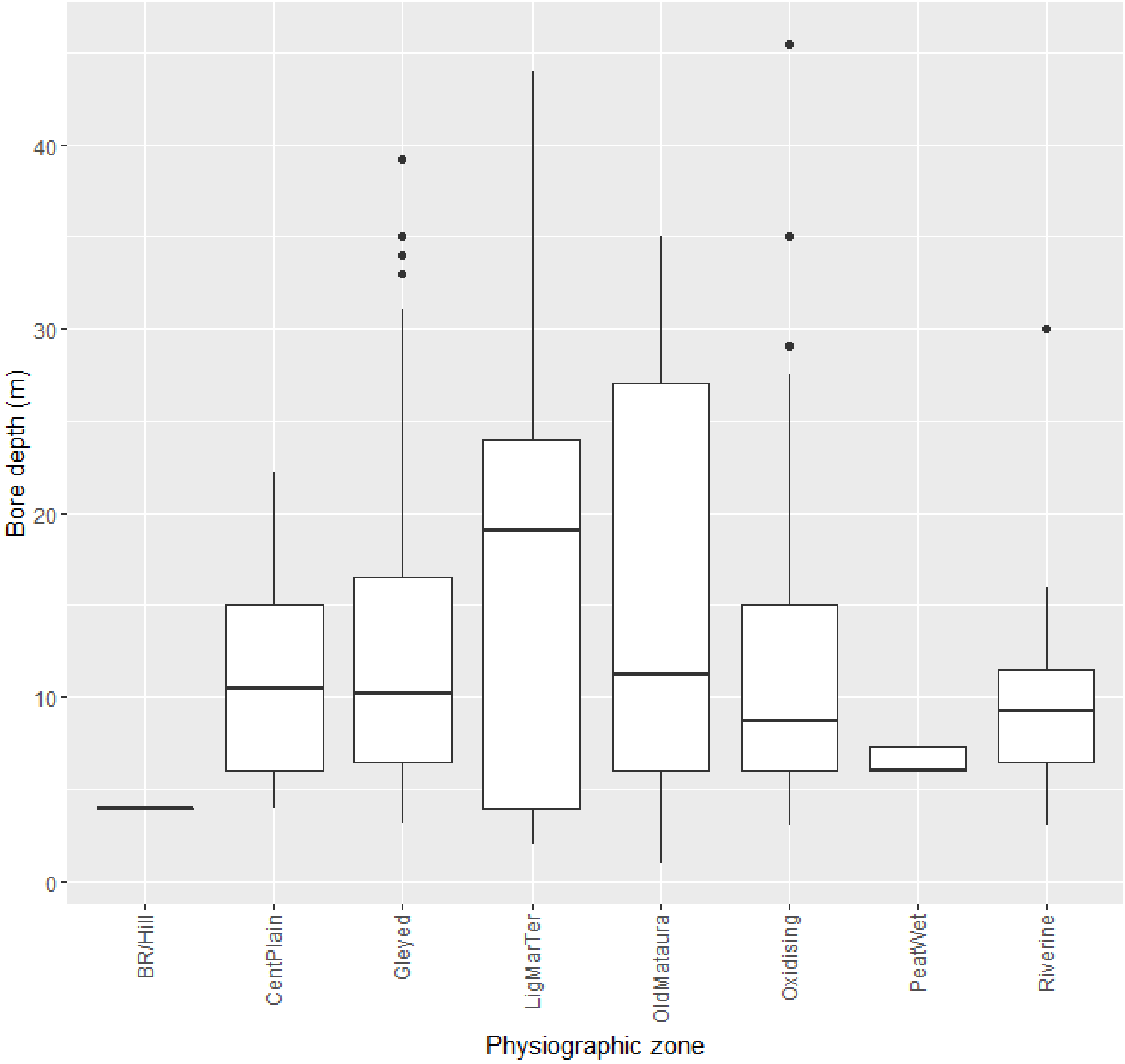


Figure 26. Distribution of bore depths from which groundwater samples were taken. *The top and bottom lines of the rectangle represent the 3rd and 1st quartiles and the central line represents the median. The whiskers extend to the 3rd and 1st quartiles plus and minus 1.5 times the interquartile range. The solid dots represent data beyond 1.5 times the interquartile range.*

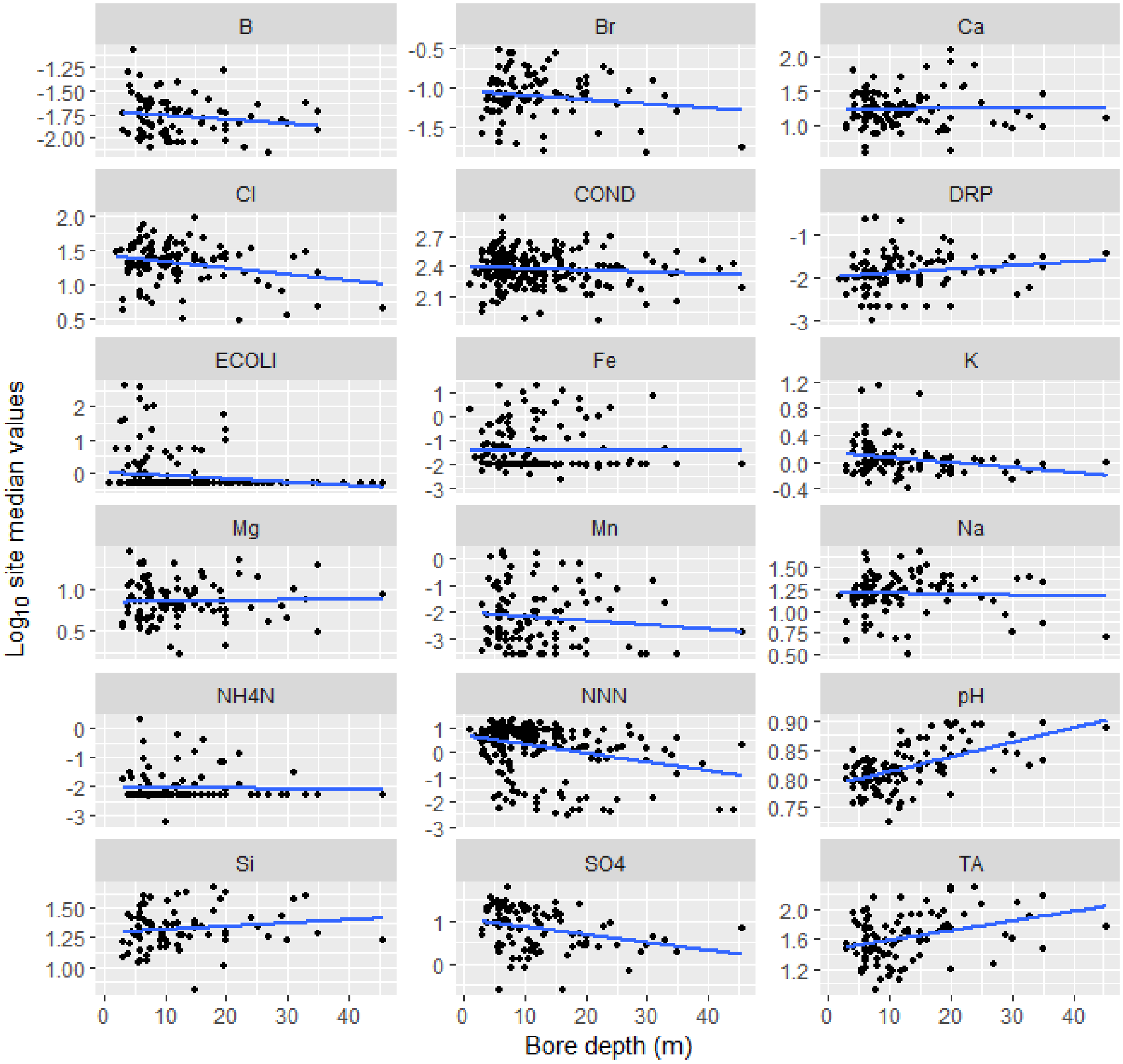


Figure 27. Relationships between site median values of the water quality variables versus bore depth. *The blue lines represent fitted linear relationships, which were only significant ($p < 0.05$) for NNN, pH, SO4 and TA.*

Relationships between the groundwater quality variables and the proportion of land in the 1 km diameter buffer zone surrounding the site that was occupied by pastoral land cover (%Pasture; Figure 28) were weak. There were only five variables for which the linear regression of log$_{10}$ site medians versus %Pasture were significant (Br, Fe^{II}, NNN, NH$_4$H and DRP). The highest R^2 values for these significant models were 0.24 and 0.18 for NH$_4$N and Fe and the remaining models had R^2 values lower than 0.06. We did not, therefore, control for %Pasture in the analyses that follow.

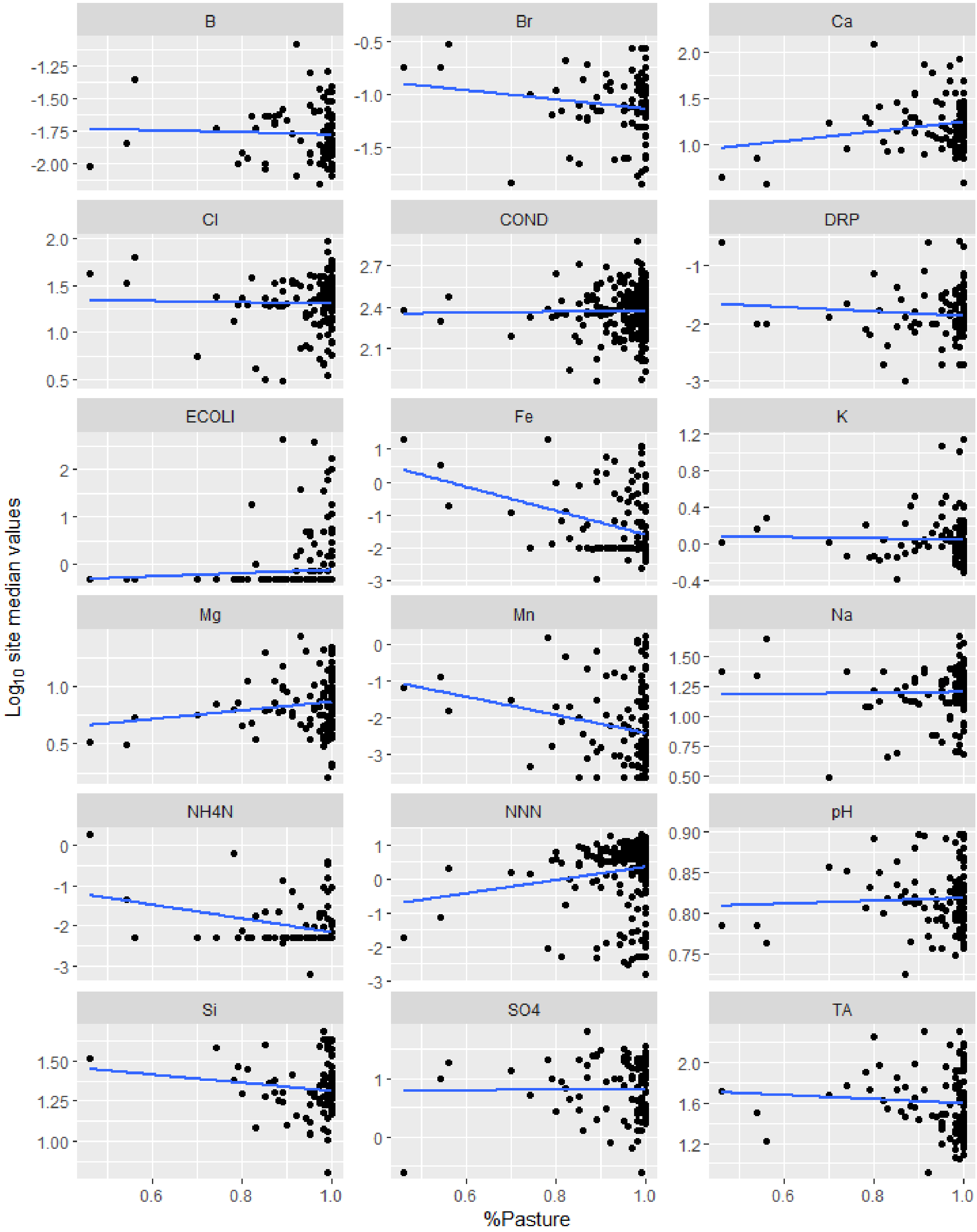


Figure 28. Relationship between the site median values of the groundwater quality variables and the proportion of land surrounding the site that was occupied by pastoral land cover (%Pasture). *The blue lines represent fitted linear relationships.*

5.4.2 Strength of classification of groundwater quality by physiographic zones

Of the 18 groundwater quality variables, only Cl, COND, ECOLI, Fe^{II} and NNN had missing values at fewer than 20% of the sites. We therefore restricted our test of classification strength to these variables and replaced missing values for each variable with the median values. This restricted the number of sites that were used in the classification strength test to 149.

The global ADONIS R^2 value was 0.15 ($p<0.001$), indicating that the physiographic zones explained 15% of the multivariate variation in groundwater quality. The pairwise ADONIS R^2 values varied between 0.51 and close to zero (Figure 29). Most of the between- zone comparisons had low R^2 values, for example only 9 pairs had R^2 values greater than 0.15. After adjustment for multiple comparisons, 13 of the contrasts achieved statistical significance (adjusted p-value <0.05; Figure 29).

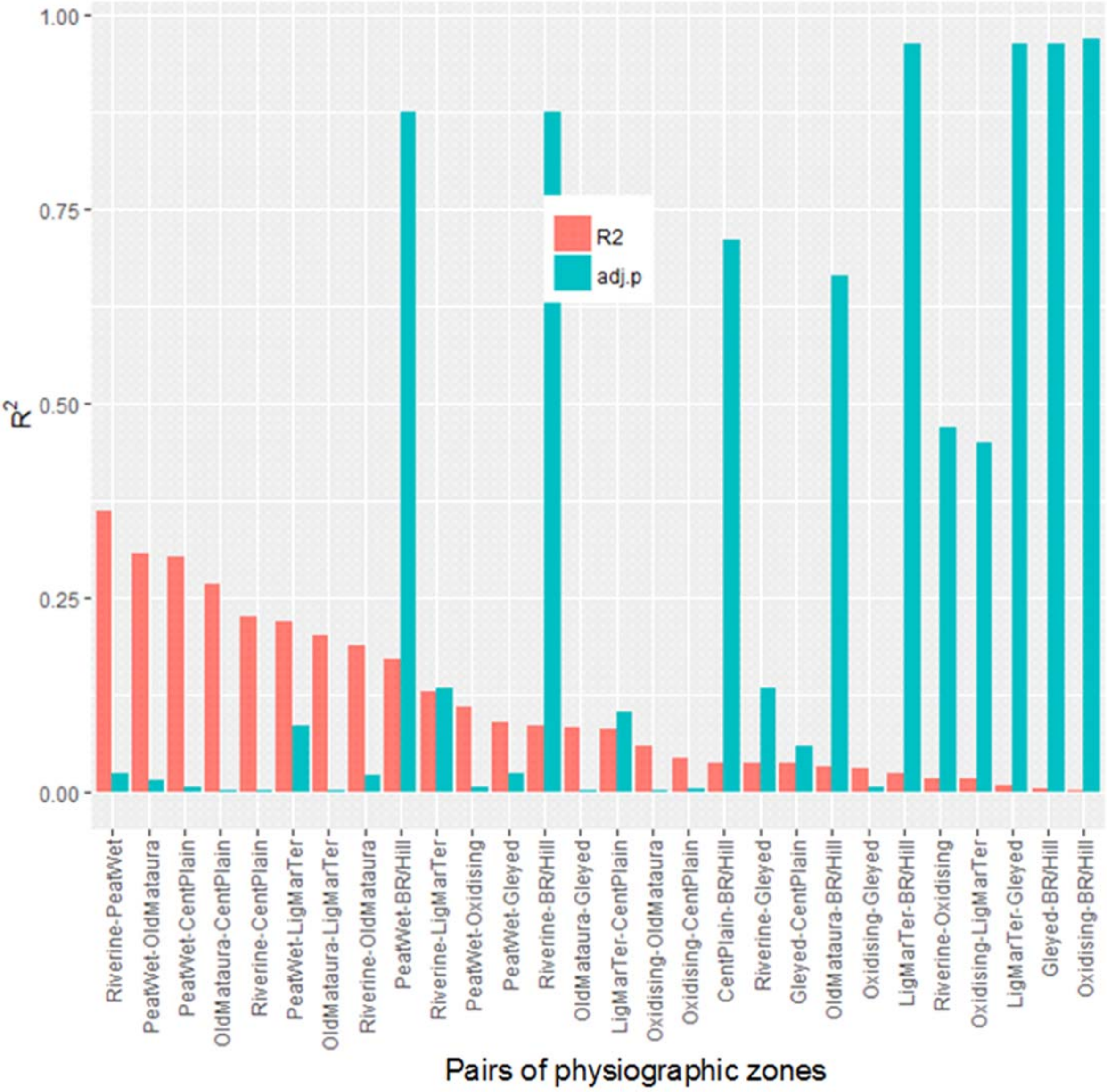


***Figure 29. Results of ADONIS tests (R^2 values) for all pairs of physiographic zones based the groundwater quality variables**. The adjusted p-values (using the false discovery rate to correct for multiple comparisons) were <0.05 for 13 contrasts.*

The NMDS had a stress value of 0.04, which indicates that the ordination provides a good representation of the original high dimensional variable-space (Figure 30). The plot of the

NMDS was consistent with the poor ADONIS performance with the physiographic zones poorly discriminated on any of the ordination dimensions (Figure 30).

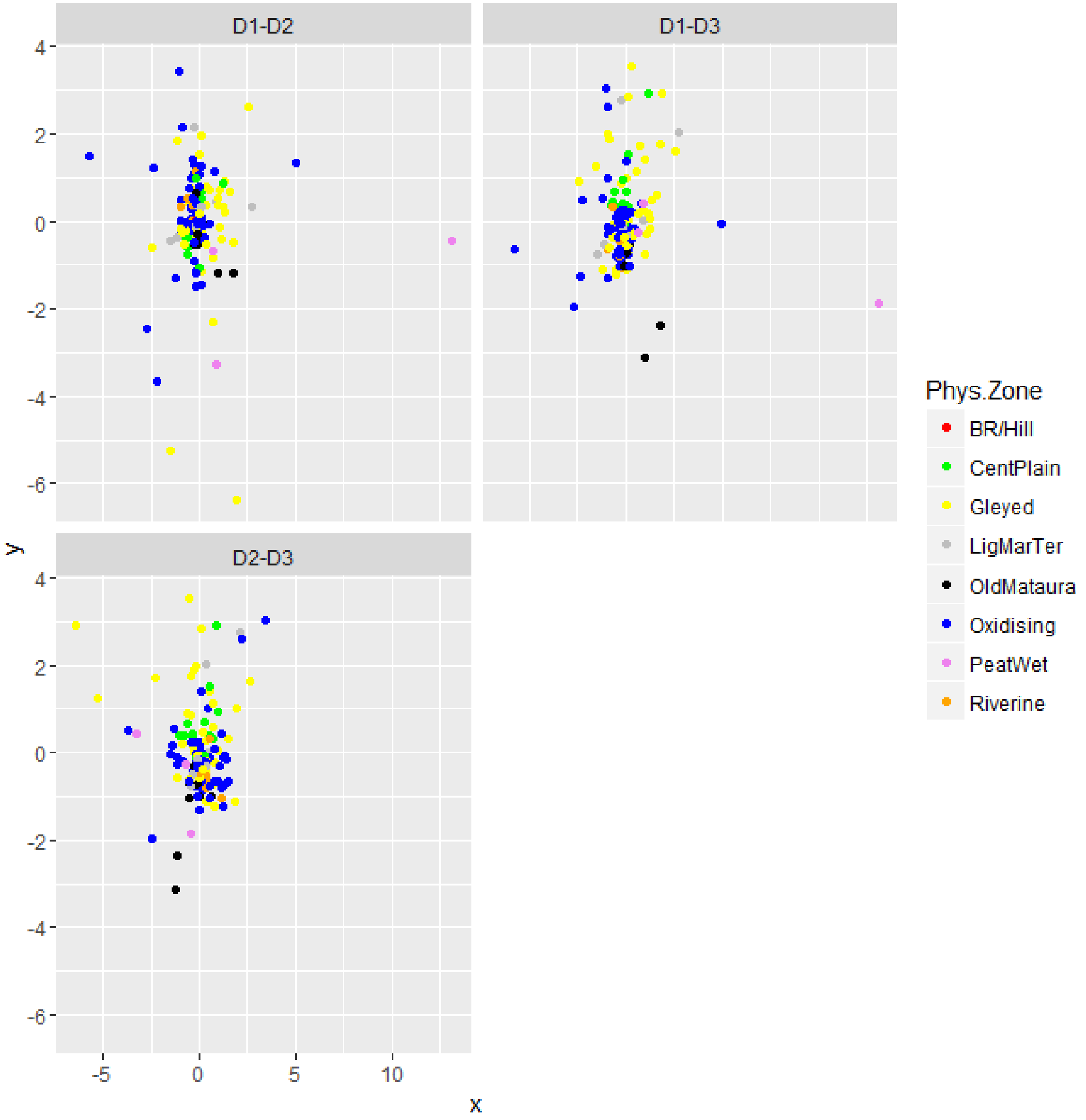


***Figure 30. NMDS plot based on the groundwater quality data**. The plotted points represent individual sites, which have been coloured according to the physiographic zones they are assigned to.*

The points of a simple ordination based on only the two most consistently sampled values measured at all 254 sites were also not well discriminated by the physiographic zones (Figure 31). The Riverine zone occupied the lower left quadrant of the plot and the Old Mataura zone predominantly occupied the upper left quadrant. However, there was considerable overlap between all physiographic zones. This variability likely reflects heterogeneity in hydrogeological settings and land use associated with the capture zones of individual sample sites.

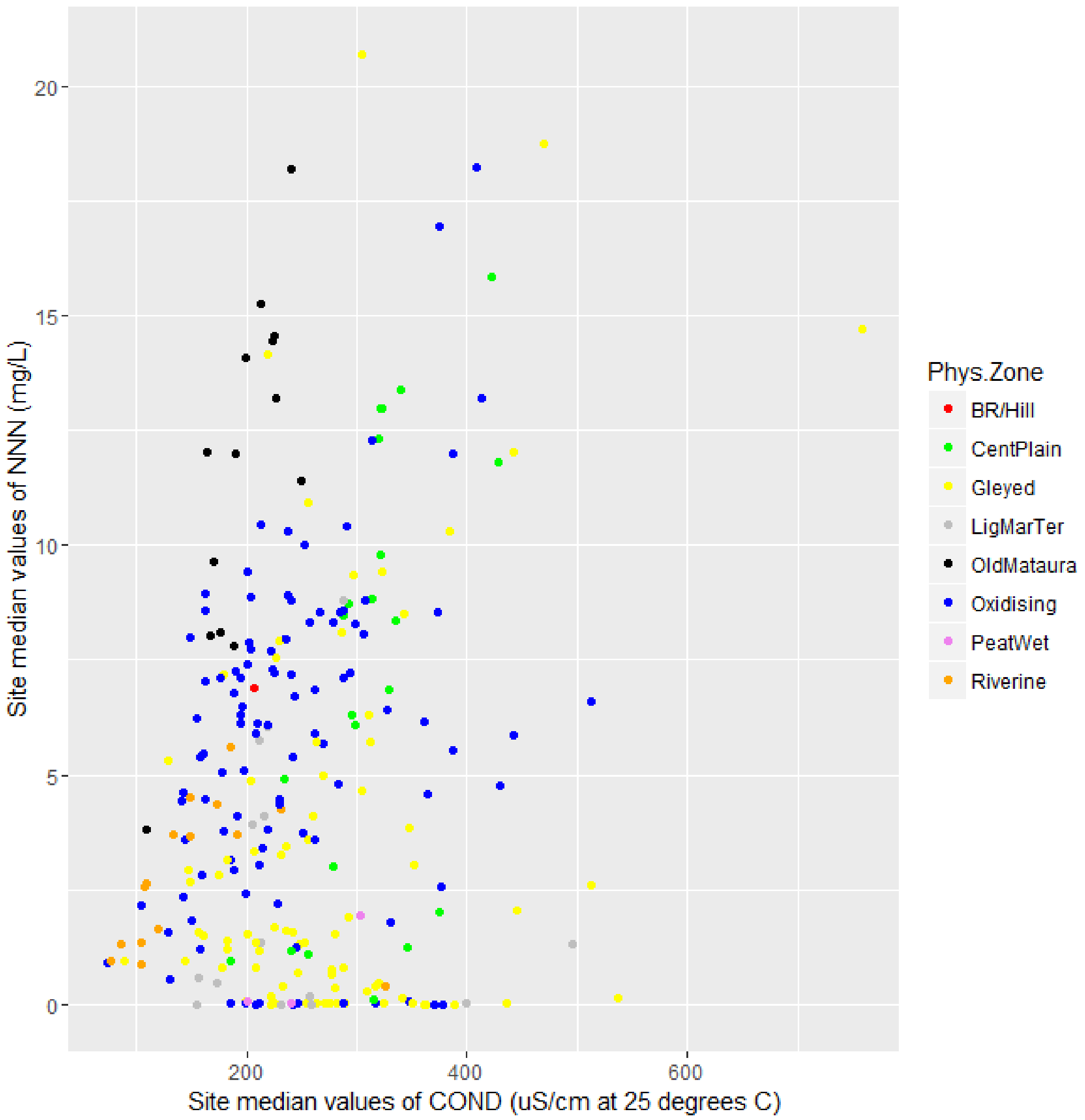


Figure 31. Scatterplot of the site median values of NNN versus COND. *The plotted points represent individual sites, which have been coloured according to the physiographic zones they are assigned to.*

5.4.3 Discrimination of characteristic magnitudes of the individual groundwater quality variables by physiographic zones

Box and whisker plots (Figure 32) indicated that physiographic zones explain variation in the magnitudes of variables (site median values) to varying degrees. The plots indicate that some variables were more strongly discriminated than others. For example, median values of NH_4N, NNN, Na, SO_4, Cl and COND differed considerably between physiographic zones, whereas variation in ECOLI, pH, K, DRP, and Mg was poorly discriminated by zone (Figure 32).

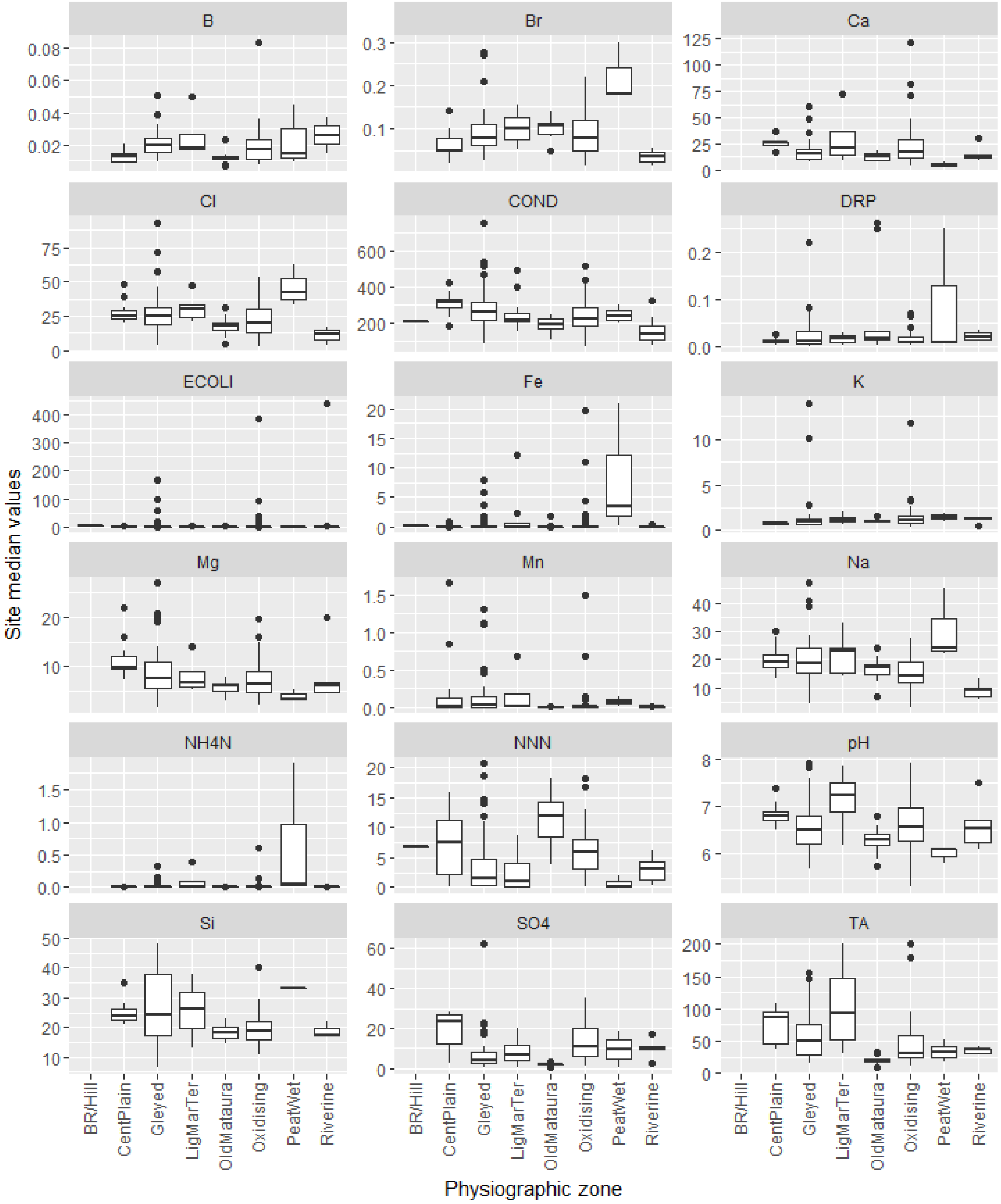


Figure 32. Box and whisker plots showing the distribution of site median values of the groundwater quality variables by physiographic zone. *The top and bottom lines of the rectangle represent the 3rd and 1st quartiles and the central line represents the median. The whiskers extend to the 3rd and 1st quartiles plus and minus 1.5 times the interquartile range. The solid dots represent data beyond 1.5 times the interquartile range.*

ANOVA tests indicated the physiographic zones explained differing levels of the variation in the $\log_{10}$ transformed site median values of the groundwater variables (Figure 33). Results were significant (adjusted *p*-values <0.05) for 14 of the groundwater quality variables (Figure 33) but R^2 values were generally less than 0.3.

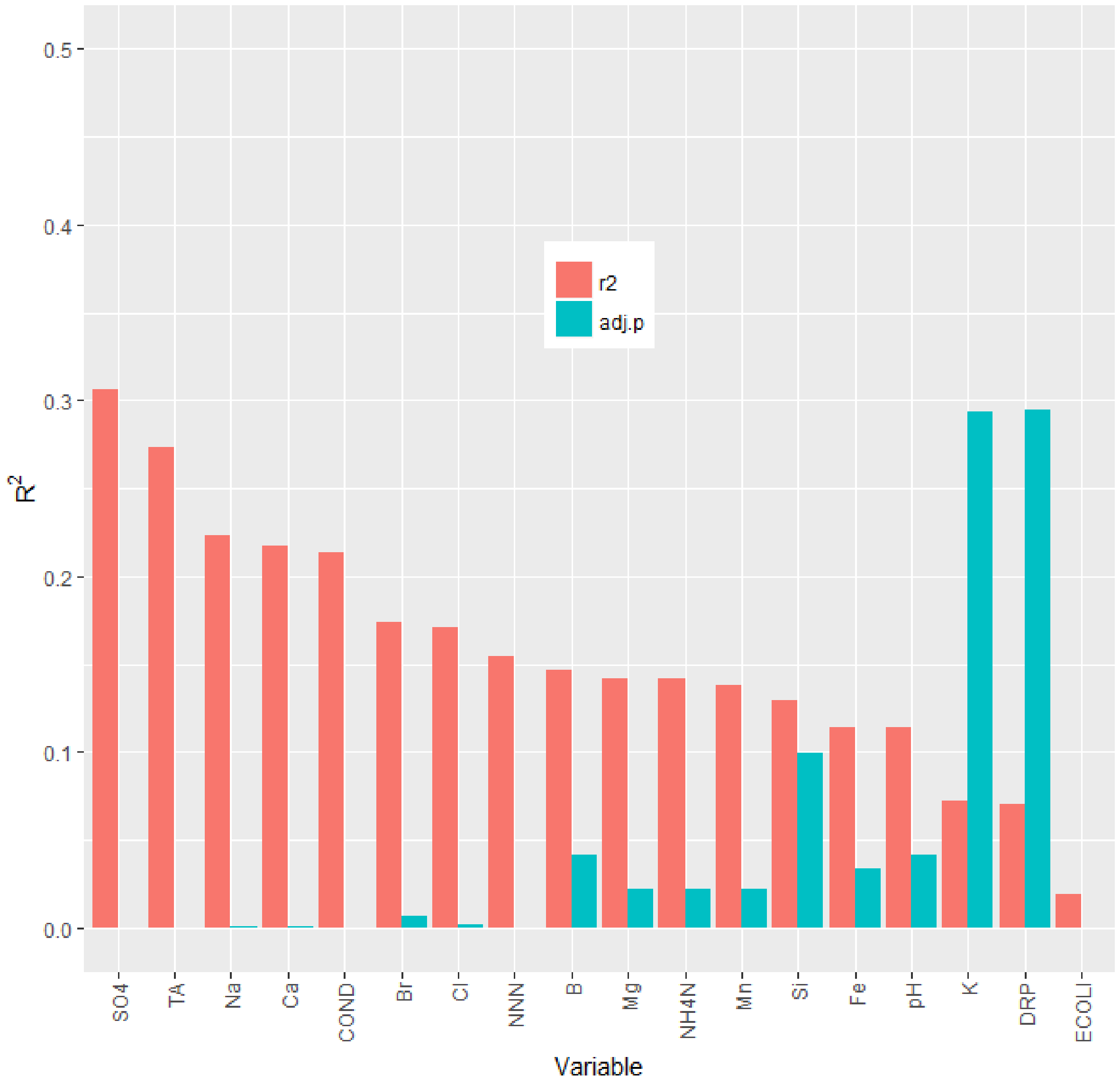


Figure 33. Results of ANOVA tests performed on all groundwater quality variables. *Four results did not achieve statistical significance (Si, K, DRP and ECOLI) (adjusted p-values >0.05).*

5.4.4 Discrimination of general variability by physiographic zone

Box and whisker plots (Figure 34) indicated that variability (coefficients of variation) in individual water quality variables were poorly discriminated by physiographic zones. Only SO_4 exhibited a statistically significant (adjusted *p*-value <0.05) result with a significantly higher median value in the Riverine zone.

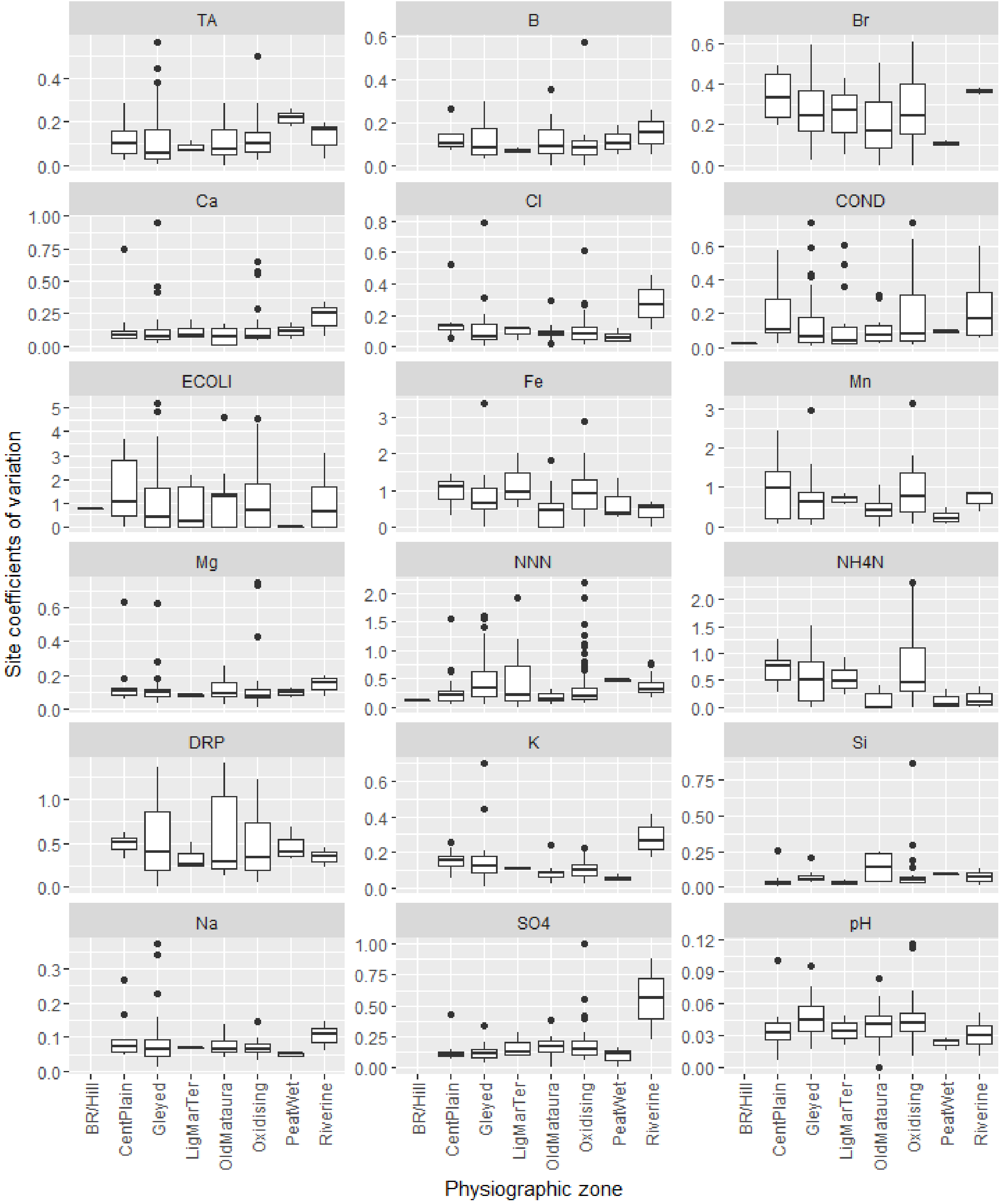


Figure 34. Box and whisker plots showing the distribution of site coefficients of variation of the groundwater quality variables by physiographic zone. *The top and bottom lines of the rectangle represent the 3rd and 1st quartiles and the central line represents the median. The whiskers extend to the 3rd and 1st quartiles plus and minus 1.5 times the interquartile range. The solid dots represent data beyond 1.5 times the interquartile range.*

ANOVA tests performed on the site coefficient of variation values shown in Figure 34 indicated the physiographic zones explained only small amounts of variation for most of the individual water quality values. Only one result was significant (Figure 35).

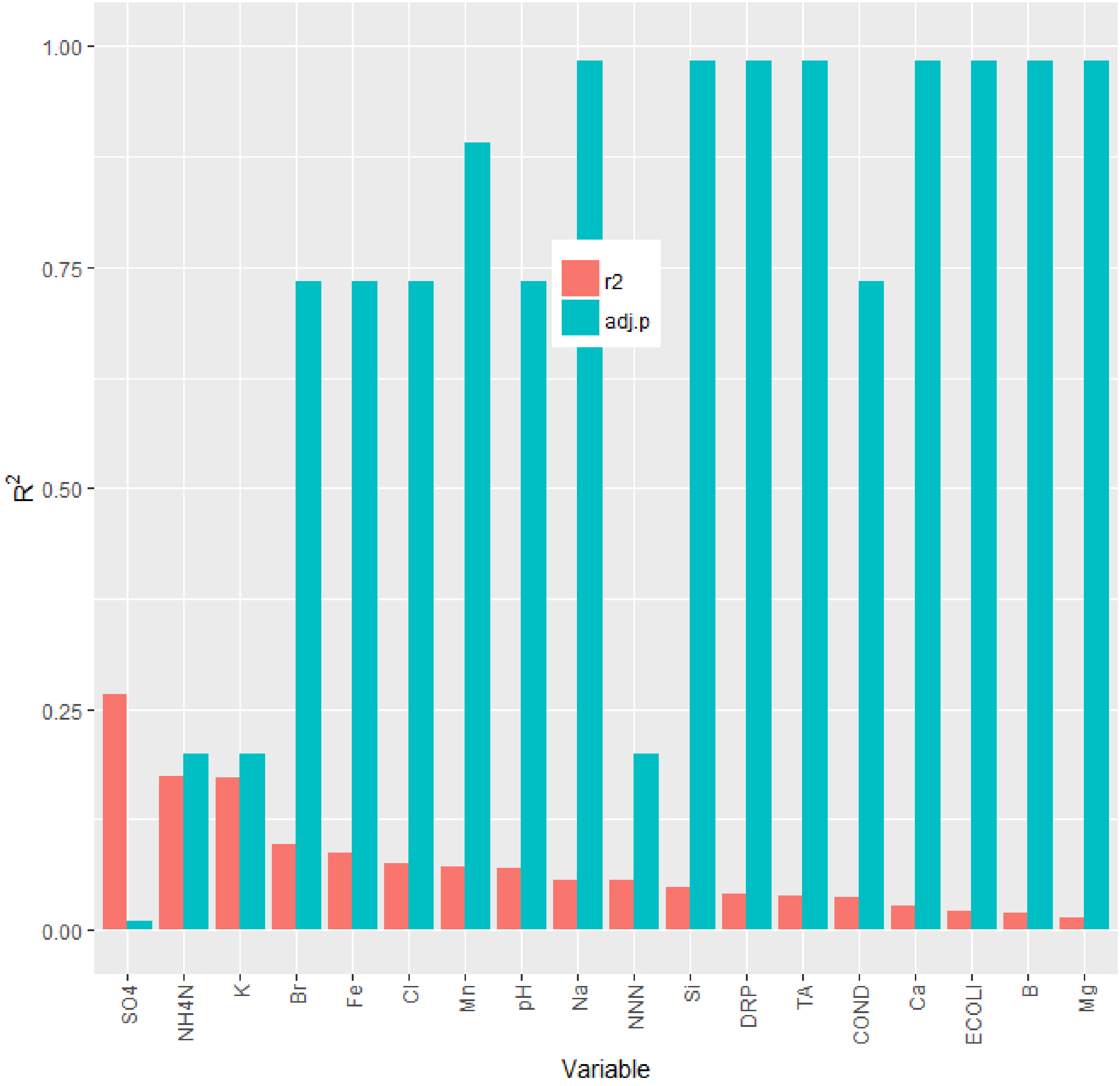


Figure 35. Results of ANOVA tests performed on the site coefficient of variation values of the individual water quality variables. *The only variable that was significantly explained was SO4.*

5.4.5 Discrimination of variation in seasonality by physiographic zone

Figure 36 indicates that variation in site seasonal indices of the individual groundwater quality variables were poorly discriminated by the physiographic zones. The plots indicate that although sites varied appreciably in the degree of seasonality (seasonality indices ranging from zero to more than 1; Figure 36), there was little systematic variation in the indices between physiographic zones. These observations were confirmed by ANOVA conducted on the same data (Figure 37). No variable had significant differences in seasonality index when the *p*-values were adjusted for multiple comparisons.

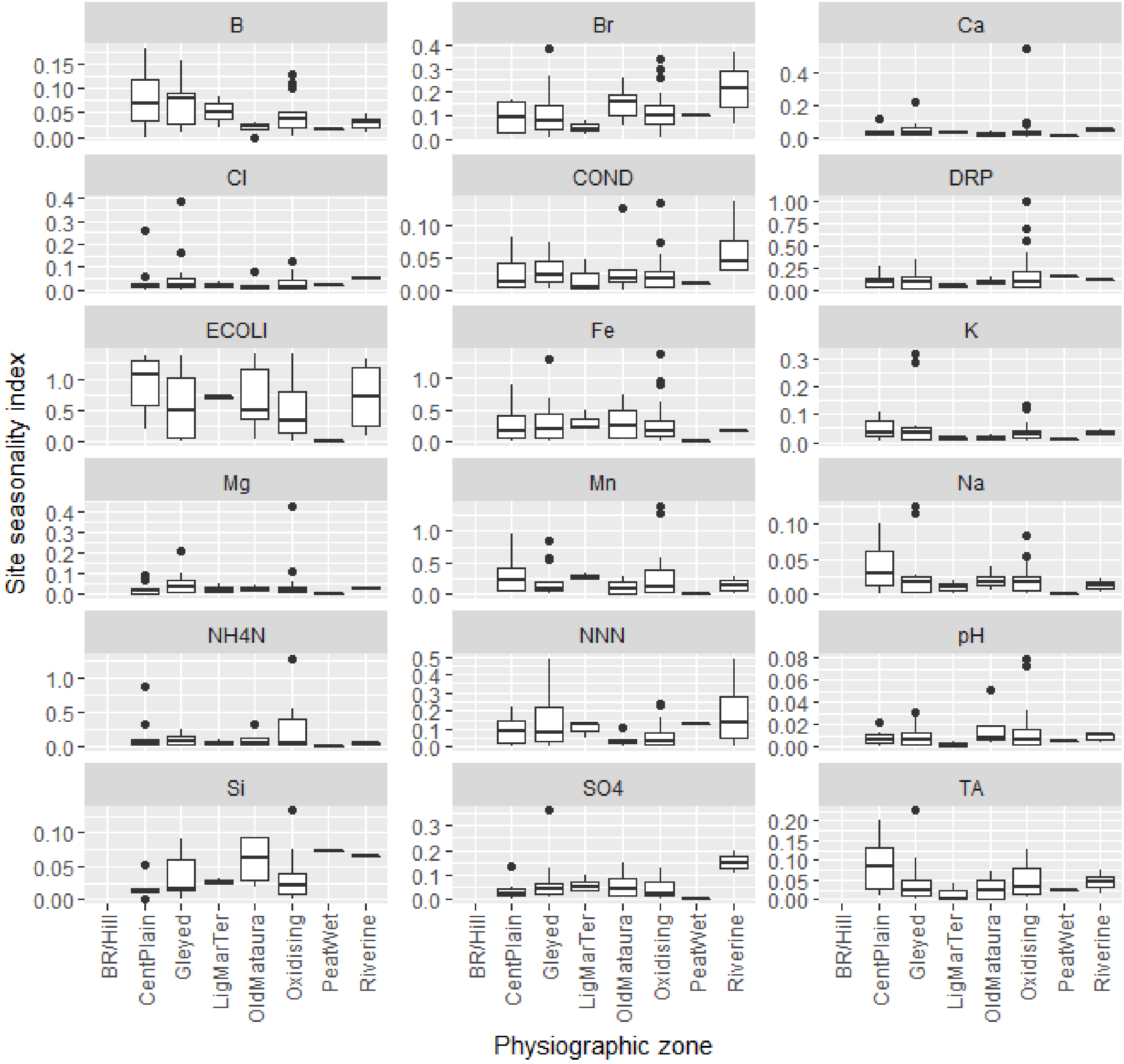


Figure 36. Distribution of site seasonality index values of the groundwater quality variables by physiographic zone. *The top and bottom lines of the rectangle represent the 3rd and 1st quartiles and the central line represents the median. The whiskers extend to the 3rd and 1st quartiles plus and minus 1.5 times the interquartile range. The solid dots represent data beyond 1.5 times the interquartile range.*

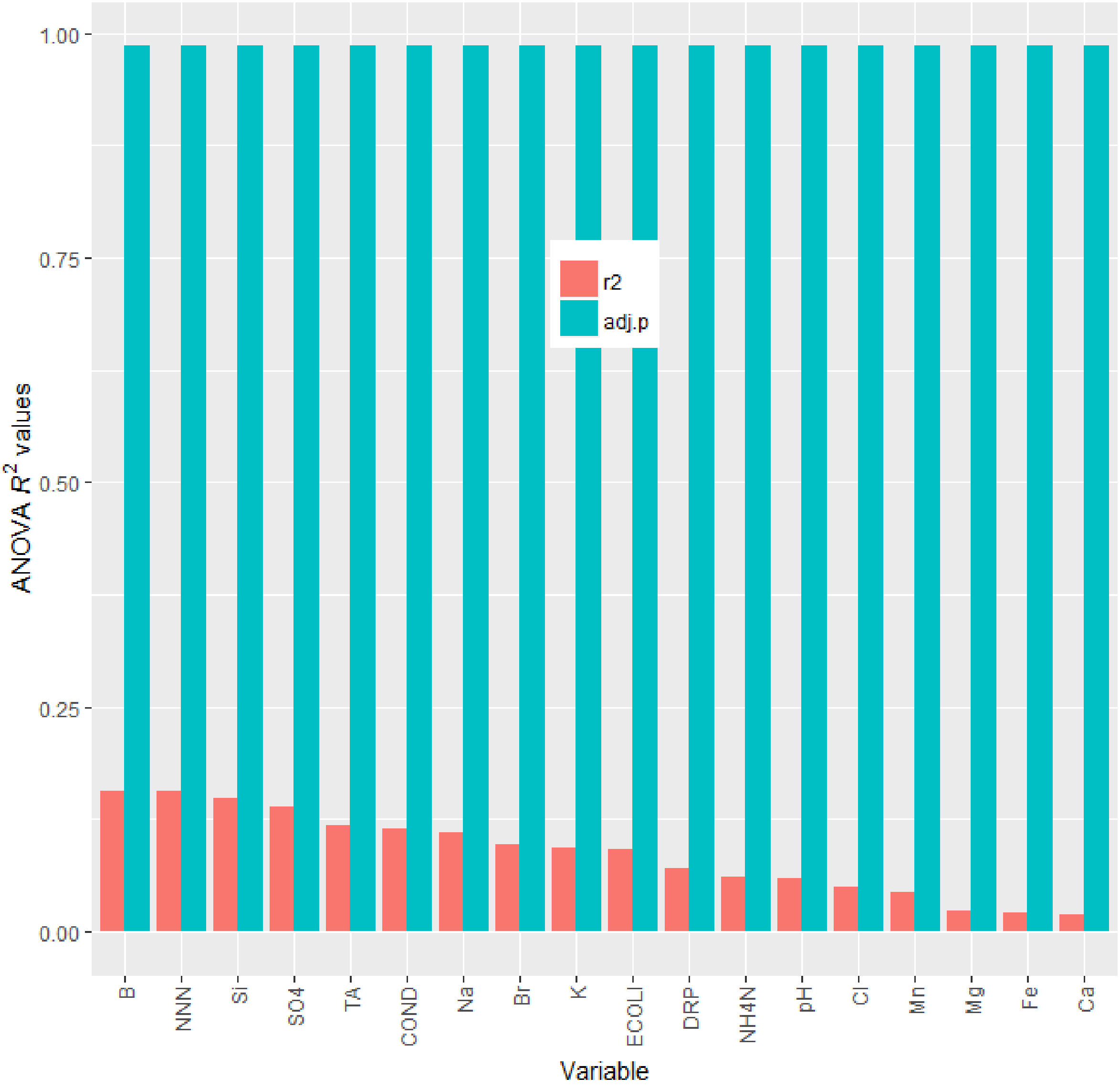


Figure 37. Results of ANOVA tests performed on site seasonality indices for all ground water quality variables. *No results were significant (p-values adjusted using the false discovery rate to correct for multiple comparisons).*

5.5 Hypotheses

Tests were performed on the hypothesised water quality and hydrochemical outcomes outlined in Table 4 and Table 5 using the Mann-Whitney U-test. Results of individual hypothesis tests are provided in Appendix 1.

5.5.1 Hypothesis testing by physiographic zone

Our tests of the physiographic zone hypotheses were consistent with expectations for 65% (79/122) of the tests, comprising 65% (45/69) of the river water, 61% (27/44) of the groundwater and 78% (7/9) of the river compared with groundwater hypotheses. Only 2% of the tests (comprising 3 groundwater tests) indicated that the data were inconsistent with the postulated hypotheses. The remaining 33% of tests were inconclusive (i.e. *p*-values > 0.05).

Results of water quality hypothesis testing at the physiographic zone level are summarised in Table 6 and Table 7. Table 6 indicates that the majority of differences in the magnitude of individual water quality variables and hydrochemical indicators both between and within (i.e. between different flow states or between groundwater and river water) physiographic zones were consistent with expectations derived from the individual physiographic zone conceptualisations (Hughes *et al*., 2016). However, the majority of results were inconclusive for the hypotheses concerning the variability and seasonality of individual water quality variables and hydrochemical indicators (Table 6). Table 7 shows a majority of hypothesis were consistent with observed water quality across all physiographic zones (ranging from 40% in Lignite/Marine Terraces to 80% in Peat Wetlands). A very small minority of tests indicated that the data were inconsistent with the hypotheses.

5.5.2 Hypothesis testing by contaminant pathway

Our tests of the pathways hypotheses were consistent with expectations for 36% (4/11) of the groundwater tests and 37% (16/43) of the surface water tests. Only 4% of the tests indicated that the data were inconsistent with the postulated hypotheses (0% and 18% of the river water and groundwater tests respectively). The remaining 32 tests were inconclusive (i.e. *p*-values > 0.05).

Results of water quality hypothesis testing at a contaminant pathway level are summarised in **Error! Reference source not found.** and **Error! Reference source not found.**. Overall, the testing confirmed water quality data were consistent with 41% ($p<0.05$) of the postulated contaminant pathway hypotheses derived from Hughes *et al*., (2016). The results indicate that the observed water quality data were reasonably consistent with the postulated hypotheses concerning magnitudes. However, the majority of pathway test results were inconclusive for the hypotheses concerning the variability and seasonality of individual water quality variables and hydrochemical indicators (**Error! Reference source not found.**). A very small minority of tests indicated that the data were inconsistent with the hypotheses.

Table 6. Summary of hypothesis testing results of all water quality metrics. *Testing was undertaken between physiographic zones (inter-class) and within physiographic zones (intra-class) as described in Section 4.4. A complete set of results are provided in Appendix 1.*

Metric	Test	Groundwater (GW)/ river water (RW)	Hypotheses tested	Hypotheses consistent	Hypotheses inconsistent	Hypotheses inconclusive
Magnitude	Inter-class	RW	34	25	3	6
		GW	34	28	0	6
	Intra-class	RW	13	12	0	1
		RW/GW	9	7	0	2
Seasonality	Inter-class	RW	13	2	0	11
		GW	8	0	0	8
Variability	Inter-class	RW	9	3	0	6
		GW	2	2	0	0

Table 7. Summary of hypothesis testing results by physiographic zones. *Testing was undertaken using groundwater and surface water quality datasets, as described in Section 4.4. A complete set of results are provided in Appendix 1.*

Physiographic zone	Groundwater				River water				Groundwater-River water comparison			
	Hypotheses tested	Hypotheses consistent	Hypotheses inconsistent	Hypotheses inconclusive	Hypotheses tested	Hypotheses consistent	Hypotheses inconsistent	Hypotheses inconclusive	Hypotheses tested	Hypotheses consistent	Hypotheses inconsistent	Hypotheses inconclusive
Alpine	0				14	10	0	4	0	0	0	0
Bedrock/Hill Country	3	2	0	1	7	5	0	2	1	1	0	0
Central Plains	8	6	0	3	0	0	0	0	0	0	0	0
Gleyed	4	1	0	3	23	13	0	10	3	2	0	1
Lignite/Marine Terraces	5	2	0	3	0	0	0	0	0	0	0	0
Old Mataura	9	6	1	2	0	0	0	0	0	0	0	0
Oxidising	4	3	1	0	3	2	0	1	0	0	0	0
Peat Wetlands	7	6	0	1	11	9	0	2	2	1	0	1
Riverine	4	1	1	2	11	6	0	5	1	1	0	0

***Table 8. Summary of hypothesis testing results of all water quality metrics for contaminant pathways**. Testing was undertaken by grouping physiographic zones into their dominant contaminant pathways, as described in Section 4.4. A complete set of results are provided in Appendix 1.*

Metric	Test	Groundwater (GW)/ river water (RW)	Hypotheses tested	Hypotheses consistent	Hypotheses inconsistent	Hypotheses inconclusive
Magnitude	Inter-class	RW	21	10	0	11
		GW	6	4	2	0
Seasonality	Inter-class	RW	11	1	0	10
		GW	2	0	0	2
Variability	Inter-class	RW	11	5	0	6
		GW	3	0	0	3

***Table 9. Summary of hypothesis testing results of all contaminant pathways.** Testing was undertaken by grouping physiographic zones into their dominant contaminant pathways, as described in Section 4.4.* A complete set of results are provided in Appendix 1.

Contaminant Pathway	Groundwater				River water			
	Hypotheses tested	Hypotheses consistent	Hypotheses inconsistent	Hypotheses inconclusive	Hypotheses tested	Hypotheses consistent	Hypotheses inconsistent	Hypotheses inconclusive
Overland flow	0			0	23	7	0	16
Artificial drainage	0			0	9	6	0	3
Deep drainage - oxidising	5	2	1	2	6	1	0	5
Deep drainage - reducing	6	2	1	3	5	2	0	3

6 Discussion

The available data enabled us to perform tests that corresponded to the three objectives of this study. However, as is very often the case with tests of environmental classifications, we were limited by available data. Partly these limitations arose because we implemented a series of data filtering rules to ensure our site level data were representative of characteristic magnitudes and variability. We consider that these rules make the results we have achieved robust but the rules represent trade-offs between numbers of sites and robustness that required subjective judgements. We have been transparent about the choices we made so that our testing procedure is repeatable.

It is noted additional analyses could compare the performance of the physiographic zones with performance associated with similar environmental classifications such as River Environment Classification (Snelder and Biggs, 2002) or Land Environments of New Zealand (Leathwick *et al.*, 2003). Comparing our results with similar test results based on more general environmental classification systems would provide a way to assess the relevance and usefulness of the conceptual model that underlies the physiographic zones. However, this type of testing was beyond the scope of this report.

6.1 Drivers and characteristics

Classification strength tests and NMDS plots showed the physiographic zones strongly discriminate unique combinations of the drivers and characteristics. Pairwise testing indicated the differences between drivers were more distinctive between zones that have differing recharge mechanisms. This is not surprising given the focus on recharge mechanism in the conceptual model and in the delineation of the zones. Physiographic zones which were less distinctive tended to be delineated on the basis of geology (e.g. Lignite/Marine Terraces), or where the redox setting is mixed (e.g. Gleyed).

Pairwise testing indicated that the most distinctive physiographic zones tend to be those with distinct soils (e.g. Peat Wetlands) or topographic settings (e.g. Alpine, Bedrock/Hill Country). This is attributable in part to the multiple variables in the characteristics data that represent topographic setting (i.e. slope, elevation and mean annual rainfall) and soils (soil type, soil profile drainage, soil base saturation, soil anion storage capacity). This means the multivariate space defined by the characteristics data was strongly associated with gradients in topography and soils. Physiographic zones that occupy distant parts of these gradients will be interpreted in both the classification strength tests and NMDS as strongly differing.

6.2 Water quality observations

6.2.1 Water quality data

We found that for the majority of sites and variables, censored values (where detection limits could be established) represented fewer than 50% of the data. This means that estimates of most site median values were unaffected by censoring, and therefore that tests of magnitude were unlikely to have been affected. Censored values may have affected our tests of temporal variation (i.e. overall variability and seasonal variation in magnitude). However, because all censored values were set to half the detection limit, their existence only has the effect of homogenising the data. The effect of this on our tests of temporal variation would have been to reduce the number of significant results, rather than introducing any bias.

The inclusion of several correlated variables (particularly nitrogen species) will have weighted certain gradients in our tests of classification strength (i.e. the ADONIS and NMDA multivariate analyses). The inclusion of correlated variables will have influenced the results by more heavily weighting the gradients associated with the correlated variables. Although it is possible to reduce the effect of weighting gradients in multivariate analyses, we did not do this as decisions to do so are inherently subjective. We consider the classification strength tests are descriptive and indicative and that the subsequent univariate analyses clarify which individual variables are most strongly discriminated.

We could have used more flexible models (e.g. non parametric or non-linear) to control for %Pasture and bore depth. We used relatively simple linear models because these appeared to be reasonable for most variables (Figure 9) and it avoided introducing further complexity into the analyses. We consider that its very unlikely that our conclusion would be altered by use of more sophisticated methods to control for %Pasture and bore depth.

We found that %Pasture correlated with most of the river water quality variables but several of these are probably not causative relationships. Some correlations may reflect the spatial distribution of pasture and not the effect of pasture on water quality *per se*. For example, correlations between %Pasture and Na and Cl are most likely to be associated with the generally closer proximity of highly agricultural catchments in coastal as opposed to inland areas. We reported results for all variables after controlling for %Pasture to avoid making subjective decisions about excluding variables that we did not think were mechanistically related.

6.2.2 Overall results

Overall, our results indicate the physiographic zones discriminate both indicators of water provenance and processes (i.e. hydrochemistry) and water quality (i.e. variables indicating water's fitness for purpose). Our results indicate that the framework discriminates the characteristic magnitudes of most water quality variables and their characteristic temporal variation (for river water). The results provide strong support for the conceptual model underlying the physiographic zones and the procedure used to map the zone boundaries.

Quantitative measures of discrimination (e.g. R^2 values) were not as high as reported for some analyses that were associated with the Part 1 report (Rissmann *et al.*, 2016) and some R^2 values in our tests were not very high, although tests were often significant. These two observations are to be expected for several reasons the most important of which is that the physiographic zones represent a simplification of the more detailed conceptual model, which itself is simplification of reality. An important simplification made by the physiographic zones is associated with simplifying the underlying conceptual model into a small number of categories. A small number of mapped categories is necessary to provide a relatively simple, understandable and clear (i.e. mapped boundaries) framework for management purposes. There are trade-off judgements inherent in choosing the number of physiographic zones – choosing more would increase discrimination but would decrease the simplicity and usability of the framework as a basis for regulation. In addition, the physiographic zones are effectively subdividing and categorising multiple gradients associated with both biogeochemical and hydrological processes. The complexity of these gradients means that there is necessarily a degree of within category variation and potential for overlap between categories, at least for some variables. This reduces the discrimination by the physiographic zones of any specific variable, but produces a compromise that performs reasonably well over all variables while being relatively simple.

In all tests, surface water quality results were much stronger than groundwater results. There are two possible reasons for this. First, there were generally more samples, collected at a higher frequency, at the river sites than groundwater sites. The median values and seasonality indices may therefore have been more accurate descriptors of river water quality than the corresponding groundwater measures. Second, surface water samples represent an integrated measure of the biogeochemical drivers and environmental characteristics of the physiographic zones on water quality because they are derived from entire catchments of known extent whereas groundwater systems exhibit much greater heterogeneity. Groundwater samples represent the water quality state from a potentially extensive but in this study, undetermined capture zones which incorporates greater temporal variability (associated with lag-times and depth). Further, wells may or may be screened at discrete intervals or drilled selectively to intersect particular water bearing layers and depths that position intakes below the zone of maximum water table variation (to stop the well running dry) and therefore may not be representative of shallow, soil-influenced groundwater. Groundwater sites cannot be assumed to be entirely representative of the physiographic zone in which the sample site is situated, due primarily to vertical stratification of aquifer systems and associated heterogeneity which is not well defined even though this is the assumption made in our analysis. The latter are key issues in assessing the representativeness of groundwater.

6.2.3 River water

There was insufficient data to include the Central Plains, Lignite/Marine Terraces and Old Mataura zones in our river water quality tests and only one site represented the Oxidising zones. Despite these limitations, the results of the classification strength tests indicate that the physiographic zones strongly discriminate unique combinations of river water quality and hydrochemical variables. We interpret the good performance of the physiographic zones in these tests as strong support for the underlying conceptual model. Furthermore, we consider that because the conceptual model (Rissmann *et al*, 2016) was systematically applied to the conceptualisation of individual physiographic zones (Hughes *et al*., 2016), it is likely that inclusion of the absent zones would not change our conclusions, were data available.

Pairwise testing of physiographic zones showed most zones differed with respect to their combinations of water quality characteristics however the level of difference varied between individual pairs. Generally, the physiographic zones explained a greater amount of variability between zones exhibiting contrasting redox settings. We interpret this as strong support for aspects of the underlying physiographic zone conceptualisations that are concerned with transformations of contaminants in groundwater by redox processes.

Analysis of variation in the magnitudes of individual water quality variables indicated that some were strongly discriminated by the physiographic zones (e.g. Ca, TN, COND, NNN) while others were less well discriminated (e.g. ECOLI, TKN, TP). Variables associated with precipitation source (e.g. Cl), soil zone processes (e.g. SO_4, K), substrate composition (e.g. Ca, TA) and redox sensitive parameters (e.g. Fe^{II}, Mn^{II}, NNN) were generally well discriminated. Analysis that controlled for %Pasture showed the physiographic zones explained statistically significant variation for all individual variables except for Mg, Ca, NNN, TSS and ECOLI. We interpret this as strong support for the underlying conceptual model that postulates that sources and transformations of water borne constituents vary between physiographic zones due to variation in natural drivers that are independent of land use.

Explanation of variation in DRP showed little change between the raw data and data that was controlled for %Pasture. We interpret this as a reflection of the strong influence redox setting and soil characteristics (Anion Storage Capacity) have over DRP, which are independent of land use.

We found that the physiographic zones did not discriminate differences in temporal behaviour (overall variability and seasonal magnitude variation) of the river water quality variables as well as their characteristic magnitudes. We consider this result is because of temporal variability in contaminant pathways within individual physiographic zones. However, the temporal behaviour of some variables was strongly discriminated (e.g. R^2 values > 40%; Figure 20 and 22). We consider that some variation in the temporal behaviour of variables is attributable to the specific environmental characteristics that vary significantly within individual physiographic zones. This variability is recognised by the delineation of physiographic zone variants for the water quality risk assessment outlined in the Physiographics of Southland Part 3 report (Hughes *et al.,* 2016). However, as noted in Section 3.1, water quality data was not available at sufficient spatial (for river water) or temporal (for groundwater and river water) resolution to adequately test water quality behaviours at a variant scale.

Our tests treated all variability as having the same importance by expressing both the overall and seasonal variation as proportions of the long term characteristic site magnitude. This means that sometimes sites with high variability have absolute concentrations that are very low and the variations have little significance in terms of management. The variability tests should be interpreted as a reflection of processes occurring within physiographic zones rather than having direct significance to management. We left all measures of variation (i.e. overall variability and seasonal indices) in the analysis to avoid making subjective judgements concerning including or not including sites on the basis that variation was occurring at low absolute magnitudes.

6.2.4 Groundwater

There was insufficient data to analyse the Alpine zone and the Bedrock/Hill Country and Peat Wetlands zones had few sites (<5). As a result, it is noted two of the five recharge mechanism types; alpine river recharge and bedrock river recharge, were not well represented in the groundwater data.

The classification strength tests and NMDS plots showed the physiographic zones only weakly discriminate unique combinations of groundwater quality (ADONIS global R^2 = 0.15). These tests suggest that redox setting is associated with the largest differences between zones with pairs of zones with the greatest expected differences in redox setting exhibiting the pairwise R^2 values (e.g., Peat Wetland and Riverine).

Analysis of individual parameters indicated some variables were more strongly discriminated by the physiographic zones than others. For example, NH_4N, NNN, Na, SO_4, Cl and COND were more strongly discriminated than ECOLI, pH, K, DRP, Mg. In addition, our results indicate that some zones are distinctive due to the characteristic magnitude of specific variables and these are often consistent with the underlying conceptual model. For instance, Peat Wetlands was clearly differentiated from the other zones with respect to Fe^{II}, Na, Br, Cl and Na. We interpret this as a reflection of this zone’s predominately coastal precipitation source and strongly reducing redox setting. Old Mataura and Riverine were also distinctive with respect to individual water quality variables TA, NNN, pH for the former and Cl, Br, COND in the case of the latter. Again, these differences are consistent with the underlying physiographic zone conceptualisations.

The tests performed on site coefficient of variation and seasonality indicated that variability in groundwater quality is poorly discriminated by the physiographic zones. These results may reflect reduced temporal variability and greater heterogeneity in groundwater systems compared to river waters.

6.3 Hypotheses

6.3.1 Physiographic zones

The hypothesis testing demonstrated that, in most cases, predicted differences in the relative magnitude of individual water quality variables were consistent with observations of groundwater and surface water quality. We interpret these results as supporting the steady-state aspects of the conceptual model that underlies the physiographic zones.

Hypotheses regarding variation in water quality with flow and between groundwater and river water within individual physiographic zones were largely consistent with the water quality observations. A very small minority of tests indicated that the data were inconsistent with the hypotheses. These results support aspects of the physiographic zone conceptualisations that are concerned with transport and transformation processes. However, tests of the hypotheses associated with the variability and seasonality of the individual water quality variables were largely inconclusive. This result may be attributable to two factors. First, we were limited by the size of our dataset. Where the size of the differences between test groups was small our tests lacked the statistical power to achieve significance. Second, we expect that the size of the differences in variability and seasonality will be greater at the level of the physiographic zone variants. The variants are specifically concerned with depiction of the spatial and temporal variability in contaminant pathways within individual physiographic zones (Hughes *et al.,* 2016). As noted in Section 3.1, water quality data was not available at sufficient spatial (for surface water) or temporal (for groundwater and surface water) resolution to adequately test temporal behaviour at a variant scale. Future monitoring could be targeted to collecting data to test the variants.

6.3.2 Contaminant pathways

Results of hypothesis testing for contaminant pathways were similar to those for those made for individual physiographic zones. Differences in the relative magnitude of individual water quality and hydrochemical variables were largely consistent with differences associated with variation in dominant contaminant pathways. A very small minority of tests indicated that the data were inconsistent with the hypotheses. However, tests of the hypotheses associated with the variability and seasonality of the individual water quality variables were largely inconclusive.

We consider that a significant proportion of the inconclusive results are attributable to data limitations and spatial variation in contaminant pathways within individual physiographic zones, which is associated with temporal hydrological variability. This temporal variability has been addressed by delineation of physiographic zone ‘variants’ for the water quality risk assessment (Hughes *et al.*, 2016). However, insufficient water quality data was available to undertake testing at a variant scale within this report.

7 Conclusions

Overall, the physiographic zones perform well at discriminating steady-state (differences in magnitude) water quality states. However, results were weaker for tests that assessed the discrimination of temporal behaviour of water quality variables by the physiographic zones. The water quality risk assessment (Hughes *et al.,* 2016) incorporates variants to address temporal variability in contaminant pathways within physiographic zones. However, variants have not been included in this report because there is insufficient water quality data to test spatial variability at the variant scale.

Acknowledgements

We thank Stephen Aldridge and Dean Petrichevich of Environment Southland for provision of data and assistance with GIS tasks and Scott Larned (NIWA) for reviewing this report.

References

Anderson, M.J., 2001. A New Method for Non-Parametric Multivariate Analysis of Variance. Austral Ecology 26:32–46.

Benjamini, Y. and Y. Hochberg, 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological):289–300.

Booker, D.J. and T.H. Snelder, 2012. Comparing Methods for Estimating Flow Duration Curves at Ungauged Sites. Journal of Hydrology.

Environment Southland, 2016. Proposed Southland Water and Land Plan. Environment Southland, Invercargill.

Hawkins, C.P., R.H. Norris, J. Gerritsen, R.M. Hughes, S.K. Jackson, R.K. Johnson, and R.J. Stevenson, 2000. Evaluation of the Use of Landscape Classifications for the Prediction of Freshwater Biota: Synthesis and Recommendations. Journal of the North American Benthological Society 19:541–556.

Hughes, B., Wilson, K., Rissmann, C., and 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, Invercargill. Publication number 2016/11. 109p.

Larned, S.T., T.H. Snelder, M. Unwin, and G.B. McBride, 2016. Water Quality in New Zealand Rivers: Current State and Trends. NZJMFR.

Leathwick, J.R., J.M. Overton, and M. McLeod, 2003. An Environmental Domain Analysis of New Zealand, and Its Application to Biodiversity Conservation. Conservation Biology 17:1612–1623.

Leathwick, J., T. Snelder, W. Chadderton, J. Elith, K. Julian, and S. Ferrier, 2011. Use of Generalised Dissimilarity Modelling to Improve the Biological Discrimination of River and Stream Classifications. Freshwater Biology 56:21–38.

Legendre, P. and L. Legendre, 1998. Numerical Ecology. Elsevier, Amsterdam, The Netherlands.

McDowell, R., T. Snelder, D. Booker, N. Cox, and R. Wilcock, 2012. Establishment of Reference or Baseline Conditions of Chemical Indicators in New Zealand Streams and Rivers Relative to Present Conditions. Marine and Freshwater Research. http://www.publish.csiro.au/view/journals/dsp_journals_pip_abstract_scholar1.cfm?nid=126&pip=MF12153. Accessed 12 Jun 2013.

Rissmann, C., Rodway, E., Beyer, M., Hodgetts, J., Snelder, T., Pearson, L., Killick, M., Marapara, T.R., Akbaripasand, 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., 2016. Physiographics of Southland Part 1: Delineation of key drivers of regional hydrochemistry and water quality. Environment Southland publication number 2016/3, Invercargill. 146p.

Snelder, T.H. and B.J.F. Biggs, 2002. Multi-Scale River Environment Classification for Water Resources Management. Journal of the American Water Resources Association 38:1225–1240.

Snelder, T.H., R. Woods, and B.J.F. Biggs, 2005. Improved Eco-Hydrological Classification of Rivers. River Research and Applications 21:609–628.

Wilson, K., Meijer, K., Larkin, G., Hicks, A., 2012. Water quality methodology for Southland Water 2010 report. Environment Southland, Invercargill. Publication number 2012-05. 107p.

Wilson, K., Hughes, B., Rissmann, C., Rodway, E., Hodges, J., 2016 *in prep.* Physiographics of Southland Technical Sheets.

Wilson, K., Hughes, B., Rissmann, C., Rodway, E., Hodges, J., 2016 *in prep.* Physiographics of Southland Technical Sheets User Guide.

Appendix 1 - Hypotheses Testing Results

Physiographic zones

Class 1	Variable	Metric	Class 2	Expectation Class 1	Adjusted P Value
Alpine (rw)	Cl	Median	All other zones (rw)	-	2.02E-59
Alpine (rw)	COND	Median	All other zones (rw)	-	4.4E-196
Alpine (rw)	NNN	Median	All other zones (rw)	-	7.8E-162
Alpine (rw)	TN	Median	All other zones (rw)	-	1.3E-186
Alpine (rw)	TKN	Median	All other zones (rw)	-	4.22E-55
Alpine (rw)	TA	Median	All other zones (rw)	-	7.25E-37
Alpine (rw)	Fe^{II}	Median	All other zones (rw)	-	5.12E-44
Alpine (rw)	Mn^{II}	Median	All other zones (rw)	-	2.23E-57
Alpine (rw)	COND	Seasonality	All other zones (rw)	-	0.743852
Alpine (rw)	NNN	Seasonality	All other zones (rw)	-	0.943217
Alpine (rw)	TN	Seasonality	All other zones (rw)	-	0.06767
Alpine (rw)	TKN	Seasonality	All other zones (rw)	-	0.010735
Alpine (rw)	Cl	Seasonality	All other zones (rw)	-	0.943217
Alpine (rw) - high flow	TSS	Median	Baseflow & mid flow (Alpine)	+	1.10E-06
Bedrock/Hill Country (rw) - high flow	TSS	Median	Baseflow & mid flow (Bedrock/Hill Country)	+	6.84E-06
Bedrock/Hill Country (rw) – high flow	*E.coli*	Median	Baseflow & mid flow (Bedrock/Hill Country)	+	1.73E-09
Bedrock/Hill Country (rw) – high flow	TP	Median	Baseflow & mid flow (Bedrock/Hill Country)	+	1.44E-07
Bedrock/Hill Country (rw)	TSS	Variability	All other zones (rw)	+	1
Bedrock/Hill Country (rw)	*E.coli*	Variability	All other zones (rw)	+	1.55E-11
Bedrock/Hill Country (rw)	TP	Variability	All other zones (rw)	+	1
Bedrock/Hill Country (rw)	Fe^{II}	Median	All other zones (rw)	+	7.23E-29
Bedrock/Hill Country (gw)	Fe^{II}	Median	All other zones (gw)	+	0.200757
Bedrock/Hill Country (gw)	COND	Variability	All other zones (gw)	-	3.40E-05
Bedrock/Hill Country (gw)	NNN	Variability	All other zones (gw)	-	0.000185
Bedrock/Hill Country (gw)	COND	Median	Bedrock/Hill Country (rw)	+	1.84E-06
Central Plains (gw)	TA	Median	All other zones (gw)	+	4.14E-22
Central Plains (gw)	pH	Median	All other zones (gw)	+	3.47E-09
Central Plains (gw)	COND	Median	All other zones (gw)	+	4.94E-62
Central Plains (gw)	NNN	Median	All other zones (gw)	+	2.16E-21
Central Plains (gw)	Mg	Median	All other zones (gw)	+	2.64E-57
Central Plains (gw)	Ca	Median	All other zones (gw)	+	3.67E-68
Central Plains (gw)	COND	Seasonality	All other zones (gw)	+	0.692623
Central Plains (gw)	NNN	Seasonality	All other zones (gw)	+	0.310985
Gleyed (rw) - high flow	TSS	Median	Baseflow % mid flow (Gleyed)	+	1.05E-08
Gleyed (rw) - high flow	*E.coli*	Median	Baseflow & mid flow (Gleyed)	+	0.30949
Gleyed (rw) - high flow	TP	Median	Baseflow & mid flow (Gleyed)	+	1.64E-09

Class 1	Variable	Metric	Class 2	Expectation Class 1	Adjusted P Value
Gleyed (rw)	TSS	Variability	All other zones (rw)	+	1
Gleyed (rw)	*E.coli*	Variability	All other zones (rw)	+	1
Gleyed (rw)	TP	Variability	All other zones (rw)	+	1
Gleyed (rw)	Fe^{II}	Median	All other zones (rw)	+	1.42E-29
Gleyed (rw)	Mn^{II}	Median	All other zones (rw)	+	2.44E-43
Gleyed (rw)	NH_4N	Median	All other zones (rw)	+	6.7E-158
Gleyed (rw)	NNN	Median	All other zones (rw)	-	1
Gleyed (rw)	TN	Median	All other zones (rw)	-	1
Gleyed (rw)	TKN	Median	All other zones (rw)	-	1
Gleyed (rw)	COND	Median	All other zones (rw)	+	0
Gleyed (rw)	COND	Seasonality	All other zones (rw)	+	1
Gleyed (rw)	NNN	Seasonality	All other zones (rw)	+	0.264006
Gleyed (rw)	TN	Seasonality	All other zones (rw)	+	0.02635
Gleyed (rw)	TKN	Seasonality	All other zones (rw)	+	0.193437
Gleyed (rw)	Mg	Median	All other zones (rw)	+	1.69E-60
Gleyed (rw)	K	Median	All other zones (rw)	+	6.57E-77
Gleyed (rw)	SO_4	Median	All other zones (rw)	+	1.6E-83
Gleyed (rw)	Mg	Median	Gleyed (gw)	+	1
Gleyed (rw)	K	Median	Gleyed (gw)	+	7.32E-49
Gleyed (rw)	SO_4	Median	Gleyed (gw)	+	5.06E-36
Gleyed (rw) - baseflow	NNN	Median	Mid flow & high flow (Gleyed)	-	9.14E-16
Gleyed (rw) - baseflow	TN	Median	Mid flow & high flow (Gleyed)	-	4.09E-16
Gleyed (rw) - baseflow	TKN	Median	Mid flow & high flow (Gleyed)	-	2.46E-14
Gleyed (gw)	K	Median	All other zones (gw)	-	0.055546
Gleyed (gw)	NNN	Median	All other zones (gw)	-	1.65E-83
Gleyed (gw)	NNN	Seasonality	All other zones (gw)	-	1
Gleyed (gw)	TA	Median	All other zones (gw)	-	1
Lignite/Marine Terraces (gw)	NNN	Median	All other zones (gw)	-	1.65E-15
Lignite/Marine Terraces (gw)	NNN	Seasonality	All other zones (gw)	-	0.683033
Lignite/Marine Terraces (gw)	Fe^{II}	Median	All other zones (gw)	+	0.992755
Lignite/Marine Terraces (gw)	DRP	Median	All other zones (gw)	+	4.86E-07
Lignite/Marine Terraces (gw)	COND	Seasonality	NNN in Gleyed (gw)	+	0.862714
Old Mataura (gw)	NNN	Median	All other zones (gw)	+	7.93E-65
Old Mataura (gw)	COND	Median	All other zones (gw)	-	8.33E-29
Old Mataura (gw)	Cl	Median	All other zones (gw)	-	8.38E-31
Old Mataura (gw)	SO_4	Median	All other zones (gw)	-	1.69E-60
Old Mataura (gw)	TA	Median	All other zones (gw)	-	1.15E-35
Old Mataura (gw)	COND	Seasonality	All other zones (gw)	+	0.290247
Old Mataura (gw)	NNN	Seasonality	All other zones (gw)	+	0.862714
Old Mataura (gw)	Fe^{II}	Median	All other zones (gw)	-	5.66E-05
Old Mataura (gw)	Mn^{II}	Median	All other zones (gw)	-	5.13E-22

Class 1	Variable	Metric	Class 2	Expectation Class 1	Adjusted P Value
Oxidising (rw)	TA	Median	All other zones (rw)	-	1
Oxidising (rw)	K	Median	All other zones (rw)	+	5.72E-16
Oxidising (rw)	SO_4	Median	All other zones (rw)	+	2.73E-16
Oxidising (rw)	Fe^{II}	Median	Oxidising (gw)	+	2.79E-25
Oxidising (rw)	Mn^{II}	Median	Oxidising (gw)	+	0.84E-20
Oxidising (gw)	TA	Median	All other zones (gw)	-	1.73E-09
Oxidising (gw)	Fe^{II}	Median	All other zones (gw)	-	0.030356
Oxidising (gw)	Mn^{II}	Median	All other zones (gw)	-	1.85E-06
Oxidising (gw)	NNN	Median	All other zones (gw)	+	3.02E-21
Peat Wetlands (rw) - high flow	*E.coli*	Median	Baseflow & mid flow (Peat Wetlands)	+	1.69E-29
Peat Wetlands (rw) - high flow	TSS	Median	Baseflow & mid flow (Peat Wetlands)	+	1.63E-15
Peat Wetlands (rw) - high flow	TP	Median	Baseflow & mid flow (Peat Wetlands)	+	4.72E-20
Peat Wetlands (rw) - baseflow	NNN	Median	All other zones (rw) - baseflow	-	0.001095
Peat Wetlands (rw) - baseflow	TN	Median	All other zones (rw) - baseflow	-	1
Peat Wetlands (rw) - baseflow	TKN	Median	All other zones (rw) - baseflow	-	1
Peat Wetlands (rw)	TA	Median	All other zones (rw)	-	2.09E-38
Peat Wetlands (rw)	Cl	Median	All other zones (rw)	+	1.43E-82
Peat Wetlands (rw)	DRP	Median	All other zones (rw)	+	1.3E-116
Peat Wetlands (rw)	TP	Median	All other zones (rw)	+	3.7E-105
Peat Wetlands (gw)	NNN	Median	All other zones (gw)	-	1.03E-11
Peat Wetlands (gw)	K	Median	All other zones (gw)	+	3.32E-07
Peat Wetlands (gw)	Mn^{II}	Median	All other zones (gw)	-	1
Peat Wetlands (gw)	Ca	Median	All other zones (gw)	-	3.96E-17
Peat Wetlands (gw)	Mg	Median	All other zones (gw)	-	9.32E-10
Peat Wetlands (gw)	Cl	Median	All other zones (gw)	+	2.7E-15
Peat Wetlands (gw)	DRP	Median	Peat Wetlands (rw)	-	0.612392
Peat Wetlands (gw)	K	Median	Peat Wetlands (rw)	+	0.002611
Riverine (rw)	COND	Median	All other zones (rw)	-	0
Riverine (rw)	SO_4	Median	All other zones (rw)	-	1.76E-65
Riverine (rw)	Cl	Median	All other zones (rw)	-	8.5E-138
Riverine (rw)	K	Median	All other zones (rw)	-	2.32E-51
Riverine (rw)	NNN	Seasonality	All other zones (rw)	+	1
Riverine (rw)	TN	Seasonality	All other zones (rw)	+	1
Riverine (rw)	TKN	Seasonality	All other zones (rw)	+	0.455152
Riverine (rw)	TSS	Seasonality	All other zones (rw)	+	1
Riverine (rw)	*E.coli*	Variability	All other zones (rw)	+	0.329078
Riverine (rw)	TSS	Variability	All other zones (rw)	+	0
Riverine (rw)	TP	Variability	All other zones (rw)	+	0
Riverine (gw)	COND	Median	All other zones (rw)	-	8.61E-33
Riverine (gw)	COND	Median	Riverine (rw)	+	3.38E-72

Class 1	Variable	Metric	Class 2	Expectation Class 1	Adjusted P Value
Riverine (gw)	NNN	Seasonality	All other zones (gw)	+	0.30949
Riverine (gw)	SO_4	Median	All other zones (gw)	-	0.209885
Riverine (gw)	Fe^{II}	Median	All other zones (gw)	-	0.000885

Contaminant Pathways

Class 1	Variable	Metric	Class 2	Expectation Class 1	Adjusted P Value
Overland flow (rw)	NNN	Seasonality	All other pathways (rw)	-	1
Overland flow (rw)	TN	Seasonality	All other pathways (rw)	-	1
Overland flow (rw)	TKN	Seasonality	All other pathways (rw)	-	0.058665
Overland flow (rw)	DRP	Seasonality	All other pathways (rw)	+	1
Overland flow (rw)	Fe^{II}	Seasonality	All other pathways (rw)	+	1
Overland flow (rw)	*E.coli*	Seasonality	All other pathways (rw)	+	1
Overland flow (rw)	TSS	Seasonality	All other pathways (rw)	+	1
Overland flow (rw)	NH_4N	Seasonality	All other pathways (rw)	+	1
Overland flow (rw)	NNN	Variability	All other pathways (rw)	-	1
Overland flow (rw)	TN	Variability	All other pathways (rw)	-	1
Overland flow (rw)	TKN	Variability	All other pathways (rw)	-	2.36E-43
Overland flow (rw)	DRP	Variability	All other pathways (rw)	+	1
Overland flow (rw)	Fe^{II}	Variability	All other pathways (rw)	+	1
Overland flow (rw)	*E.coli*	Variability	All other pathways (rw)	+	2.27E-62
Overland flow (rw)	TSS	Variability	All other pathways (rw)	+	1
Overland flow (rw)	NH_4N	Variability	All other pathways (rw)	+	4.06E-29
Overland flow (rw)	TSS	Median	All other pathways (rw)	+	0.955538
Overland flow (rw)- high flow	TSS	Median	All other pathways (rw)	+	0.31136
Overland flow (rw) - baseflow	COND	Median	All other pathways (rw)	-	1.97E-12
Overland flow (rw) - baseflow	NNN	Median	All other pathways (rw)	-	2.85E-08
Overland flow (rw) - baseflow	TN	Median	All other pathways (rw)	-	1.20E-11
Overland flow (rw) - baseflow	TKN	Median	All other pathways (rw)	-	0.00148
Overland flow (rw) - baseflow	Fe^{II}	Median	All other pathways (rw)	+	1
Artificial drainage (rw)	NNN	Seasonality	All other pathways (rw)	+	0.49016
Artificial drainage (rw)	DRP	Seasonality	All other pathways (rw)	+	0.021521
Artificial drainage (rw)	*E.coli*	Seasonality	All other pathways (rw)	+	0.16868
Artificial drainage (rw)	NNN	Variability	All other pathways (rw)	+	5.29E-70
Artificial drainage (rw)	DRP	Variability	All other pathways (rw)	+	5.4E-110
Artificial drainage (rw)	*E.coli*	Variability	All other pathways (rw)	+	1
Artificial drainage (rw) – high flow	NNN	Median	All other pathways (rw)	+	4.06E-53
Artificial drainage (rw) – high flow	*E.coli*	Median	All other pathways (rw)	+	2.94E-21
Artificial drainage (rw)	NH_4N	Median	All other pathways (rw)	+	3.4E-241
Deep drainage (oxidising) (rw) – baseflow	NNN	Median	All other pathways (rw)	+	1

Class 1	Variable	Metric	Class 2	Expectation Class 1	Adjusted P Value
Deep drainage (oxidising) (rw) – baseflow	TN	Median	All other pathways (rw)	+	1
Deep drainage (oxidising) (rw) – baseflow	TKN	Median	All other pathways (rw)	+	1
Deep drainage (oxidising) (rw) – baseflow	Fe^{II}	Median	All other pathways (rw)	-	0.00015
Deep drainage (oxidising) (rw) – baseflow	SO_4	Median	All other pathways (rw)	-	0.23307
Deep drainage (oxidising) (rw) – baseflow	Mg	Median	All other pathways (rw)	-	0.24090
Deep drainage (oxidising) (gw)	NNN	Median	All other pathways (gw)	+	2.34E-59
Deep drainage (oxidising) (gw)	NNN	Seasonality	All other pathways (gw)	+	1
Deep drainage (oxidising) (gw)	NNN	Variability	All other pathways (gw)	+	1
Deep drainage (oxidising) (gw)	Fe^{II}	Median	All other pathways (gw)	-	2.85E-08
Deep drainage (oxidising) (gw)	COND	Median	All other pathways (gw)	-	4.08E-51
Deep drainage (reducing) (gw)	NNN	Median	All other pathways (gw)	-	2.8E-120
Deep drainage (reducing) (gw)	NNN	Seasonality	All other pathways (gw)	-	1
Deep drainage (reducing) (gw)	NNN	Variability	All other pathways (gw)	-	1
Deep drainage (reducing) (gw)	Fe^{II}	Median	All other pathways (gw)	+	2.30E-12
Deep drainage (reducing) (gw)	COND	Median	All other pathways (gw)	+	5.68E-07
Deep drainage (reducing) (gw)	Fe^{II}	Variability	All other pathways (gw)	+	1
Deep drainage (reducing) (rw) - baseflow	NNN	Median	All other pathways (rw)	-	1
Deep drainage (reducing) (rw) - baseflow	TN	Median	All other pathways (rw)	-	1
Deep drainage (reducing) (rw) - baseflow	TKN	Median	All other pathways (rw)	-	1
Deep drainage (reducing) (rw) - baseflow	Fe^{II}	Median	All other pathways (rw)	+	5.65E-10
Deep drainage (reducing) (rw) - baseflow	SO_4	Median	All other pathways (rw)	+	1.06E-09

(rw) = river water

(gw) = groundwater