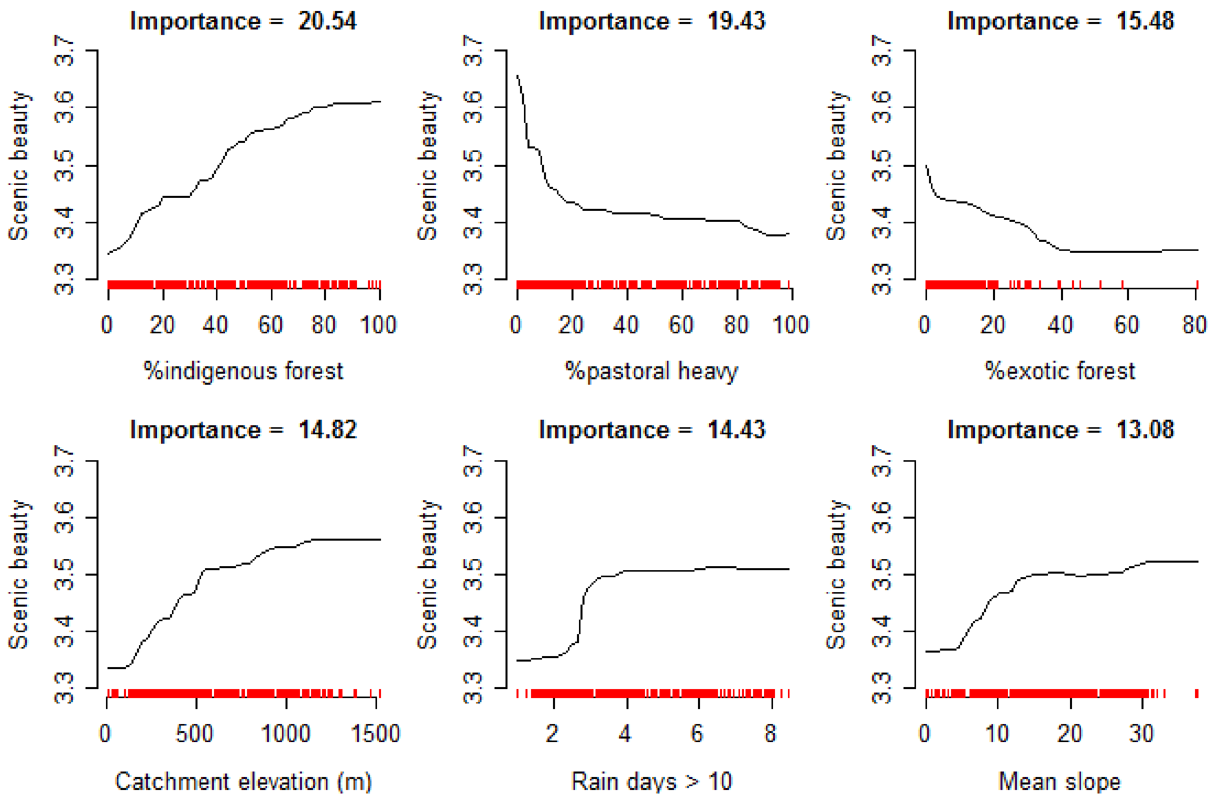
REC-based analyses of Fish & Game New Zealand angler survey data: an exploratory study

Prepared for Fish & Game New Zealand

October 2012



Authors/Contributors:
M J Unwin

For any information regarding this report please contact:
Martin Unwin

+64-3-343 7885
martin.unwin@niwa.co.nz

National Institute of Water & Atmospheric Research Ltd
10 Kyle Street
Riccarton
Christchurch 8011
PO Box 8602, Riccarton
Christchurch 8440
New Zealand

Phone +64-3-348 8987
Fax +64-3-348 5548

NIWA Client Report No:	CHC2012-124
Report date:	October 2012
NIWA Project:	FGC11504

Contents

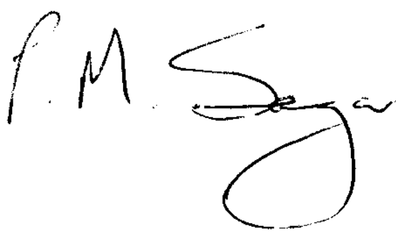
Tables

Figures

Reviewed by

Phil Jellyman

Approved for release by



Paul Sagar

Summary

Fish & Game New Zealand (FGNZ) holds extensive databases on usage of New Zealand rivers by anglers fishing for acclimatised trout and salmon. These are based on angler sample surveys at 6-7 year intervals, allowing trends in usage to be monitored over decadal time scales.

The River Environment Classification (REC) is a detailed representation of New Zealand's river network, derived from a digital topographical model overlaid with Geographic Information System (GIS) databases on climate, geology, and land cover. This report describes an exploratory study which uses the REC to objectively classify and characterise river fisheries by type, and to model angler usage and other fishery attributes derived from the angler survey data in terms of REC-based variables.

REC-based classification appears to group rivers into classes that are meaningful and relevant to FGNZ management goals, and provide considerable insight into long-term trends in angler usage. The most useful classifications appear to be those that apply over spatial sales of between about 50 and 200 km, and are based on either climate, source of flow, or a combination of these two factors.

Total annual usage by anglers fishing for trout declined by 13% between 1994/95 and 2007/08. Most of this decline has been associated with lowland rivers in areas of low to intermediate rainfall. By contrast, usage of rivers in wetter and/or higher areas has either stayed the same or increased. Overseas visitors show a moderate preference for rivers draining mountain and hill catchments rather than lowland catchments.

Predictive models using Random Forests, a multivariate modelling technique based on classification and regression trees, achieved mixed success. Angling usage was strongly related to predictors associated with catchment or river size, particularly catchment area, fishable length, and flow. The best models explained 49-50% of the observed variance in total annual usage, and are consistent with a steady increase in usage up to a threshold spatial scale of about 50 km (in length) or 2,500 km^2 (in catchment area).

An older (1979/81) FGNZ data set provides index scores for eight qualitative attributes which characterise individual rivers, such as scenic beauty and ease of access. Predictive models for these scores explained 43-51% of observed variance in scenic beauty, and for feelings of peace and solitude, attributes for which a link to catchment variables such as land cover are intuitively reasonable. Scores for both attributes declined rapidly as heavy pastoral landcover intensified, and scenic beauty was also strongly positively correlated with the extent of indigenous forest cover. Scores for angler expectations of catching a large fish were moderately well modelled, and increased with increasing catchment elevation.

A parallel set of models based on predicted water quality in New Zealand rivers, derived from another REC-based model, was generally consistent with these results, but offered few new insights. The main exception was expectation of catching a large fish, for which the water quality-based model hints at a threshold decline in expected fish size as conductivity increases. Such results should be treated with caution, but could potentially serve as a basis for developing hypotheses on which to base further studies.

The REC appears to add considerable value to FGNZ's existing angler usage data sets. Suggestions for developing this work further include revising and strengthening the core data set associated with rivers managed by FGNZ, exploring ways to further exploit these linkages when developing the next round of FGNZ angler surveys, and streamlining survey techniques so as to minimise recording and coding errors which confound attempts to match the survey data to the REC.

1 Introduction

Fish & Game New Zealand (FGNZ) manages angling for acclimatised fish species in all fresh waters except Lake Taupo and its inflowing tributaries (McDowall 1994). Under the 1990 Conservation Law Reform Act, FGNZ is tasked with monitoring "… sports fish and game populations…" and the "… success rate and degree of satisfaction of users of the sports fish and game resource…", while also being required to "…maintain and improve the sports fish and game resource". To fulfil this role, FGNZ requires reliable and up to date information on angler use of the freshwater fisheries resource.

Since the mid-1990s, FGNZ has used random sample surveys of fishing licence holders to estimate annual angling usage for all significant freshwater sports fisheries within the 12 FGNZ Regions. Three such surveys have been completed to date, using essentially the same methodology, for the 1994/95, 2001/02, and 2008/08 angling seasons (Unwin 2009a, Unwin & Brown 1998, Unwin & Image 2003). FGNZ intends to repeat these surveys every 6-7 years, to compile a long-term database providing up-to-date estimates of angling usage, and allowing trends in usage to be monitored over decadal time scales.

River fisheries span a broad range of angling activity, ranging from highly accessible and heavily used fisheries close to population centres (e.g., Mataura, Rakaia, Motueka), to remote headwater fisheries in pristine wilderness environments (e.g., Caples, Greenstone, Sabine). To help characterise this diversity, river fisheries are frequently grouped into one of three broad categories, defined as recreational (or lowland) fisheries, scenic (or backcountry) fisheries, and wilderness (or headwater) fisheries (Teirney & Richardson 1992, Teirney *et al.* 1982, Unwin & Brown 1998). This classification scheme has close parallels to the Recreational Opportunity Spectrum (ROS; Stankey & Wood 1982), but remains largely subjective and is qualitative rather than quantitative.

Considerable recent progress has been made towards developing objective schemes for characterising river environments, based on catchment values defined and measured over a broad range of spatial scales. One such scheme is the River Environment Classification (REC), first developed in 2002 (Snelder & Biggs 2002). The REC model assumes that physical regimes in rivers are controlled by six independent landscape components which form a descending hierarchy when ordered by spatial scale, including climate (10^3 - 10^5 km^2), topography (100-1000 km^2), land cover (10-100 km^2), and stream network position (1-10 km^2).

To date, attempts to link the REC to FGNZ's angling databases have been limited to creating maps characterising local and regional variation in angling effort (Unwin 2009a, 2009c). Recognising the potential for a more quantitative analysis of the merged data sets, FGNZ invited NIWA to undertake an exploratory study to identify relationships between metrics of angler usage and the REC, in anticipation of providing a more objective basis for characterising river fisheries in ways that would help to advance FGNZ management goals. These analyses are the subject of this report.

2 Data Sources

2.1 Angling data

2.1.1 Source data

Data on angling usage were obtained from two primary sources: the pooled database for the 1994/95, 2001/02, and 2007/08 National Angler Surveys (Unwin 2009a and references therein), and the archival database from the 1979/81 National Angling Survey conducted by the former

Ministry of Agriculture and Fisheries (Teirney *et al.* 1982). Both data sets were compiled into a Microsoft Access™ database, along with lookup tables to facilitate cross-matching.

The combined data set encompasses a suite of metrics which are relevant to FGNZ management goals, and are potentially related to REC variables. The 1994/95 and subsequent surveys measure annual usage in angler-days, i.e., one angler fishing a river on one day irrespective of the number of hours fished. The 1979/81 survey provides estimates of eight qualitative attributes relating to angler motivations for fishing a particular river, such as ease of access, scenic beauty, and catch rate (Teirney & Richardson 1992).

A total of 12 angler usage metrics were developed, comprising two from the 1994/95 – 2007/08 surveys, and eight from the 1979/81 surveys (Table 1). These were as follows.

Table 1: Angling usage metrics considered in this study.

Data source	Metric / attribute	Abbreviation	units
1994/95 – 2007/08 surveys	Total annual usage (all angling)	days total	angler-days
	Total annual usage (trout angling)	days trout	angler-days
	Total annual usage (overseas angler) [1]	days os	angler-days
	% usage, overseas anglers [1]	%os	%
1979/81 survey	Overall importance	overall importance	1-5 scale
	Close to home	close to home	1-5 scale
	Area of fishable water	area fishable	1-5 scale
	Ease of access	ease of access	1-5 scale
	Scenic beauty	scenic beauty	1-5 scale
	Feelings of peace & solitude	peace & solitude	1-5 scale
	Catch rate	catch rate	1-5 scale
	Size of fish [2]	size of fish	1-5 scale

[1] metrics from 2007/08 survey only

[2] "Size of fish" was not clearly defined in the 1979/81 survey. For the purposes of this study I have taken it to mean size of fish actually caught, but anecdotal evidence from some respondents to the original survey suggest that it may have also been interpreted as "expectation of catching a large fish", particular by individuals who fished a river without success.

1994/95 – 2007/08 surveys

Total annual usage (days.total): total usage in angler-days by New Zealand-resident anglers, irrespective of whether this was influenced by salmon angling. Overseas anglers were excluded as data on their fishing activity was available only for 2007/08.

Total annual usage for trout (days.trout): total usage in angler-days by New Zealand-resident anglers fishing for trout. The survey data do not differentiate between trout and salmon angling, so regional FGNZ staff were asked to estimate (to the nearest 10%) the proportion of total annual effort they considered likely to have been devoted to salmon on rivers which sustain recognised fisheries for both species. The resulting figures, for seven rivers in the Canterbury and Otago regions, were: Waiau River 50%; Hurunui River 50%; Waimakariri River 90%; Rakaia River 80%; Rangitata River 75%; Waitaki River 40%; Clutha River 10%. These proportions were then used to estimate the total usage for each river by trout anglers.

Total annual usage by overseas anglers (days.os): total usage in angler-days by overseas visitors. These data were available only for the 2007/08 survey.

Percentage of annual usage by overseas anglers (pos): *days.os* as a percentage of *days.trout*, on the assumption that overseas visitors fish for trout rather than salmon. These data were available only for the 2007/08 survey.

1979/81 survey

The 1979/81 survey invited respondents to identify whether each of seven listed attributes for each river fishery was an important factor influencing their decision to fish there. Respondents used a 5-point scale from 1 (least important) to 5 (most important). These attributes were: close to home, area of fishable water, ease of access, scenic beauty, feelings of peace and solitude, good catch rate, and size of fish. Respondents were also asked to provide a single rating, on the same 1-5 scale, for the overall importance of each river. These scores were used to derive average ratings for each river and attribute, interpreted as a continuous variable on a 1-5 numeric scale.

The 1979/81 metrics are now over 30 years old, and are potentially dated. Despite their age, however, these results are generally consistent with a more recent data set collected by FGNZ in 2008 (Unwin 2009b), giving some confidence that they give a consistent and stable characterisation of each river. This seems plausible given that attributes such as area of scenic beauty are likely to be associated with landscape values which change only slowly with time, and hence that a river considered scenically attractive in 1980 is likely to remain so today. FGNZ expects to run a full national survey updating the 1979/81 data set within the next 12-18 months, but – until these results are available – the 1979/81 data represents the best available information. Usage of these data is also fully consistent with the exploratory nature of the present study.

2.2 The River Environment Classification (REC)

2.2.1 The REC river network

The REC is a representation of New Zealand's river network derived from a Digital Elevation Model (DEM). The network has a spatial resolution of 50 m, and comprises ~570 000 unique river segments defined by upstream and downstream confluences with tributaries. Each segment (or NZReach[1]) is uniquely defined by its upstream and downstream node, with each node marking the junction of two segments. Segments are scaled to be commensurate with the 1:50,000 series of topographic maps, and are stored in a Geographic Information System (GIS) database as a set of (generally curved) polylines. Mean segment length is 740 m.

Each segment is associated with its own local watershed, allowing the catchment draining to each node to be characterised by accumulating attributes (also derived from the DEM) for all upstream segments. The network is linked to a GIS database describing the climate, topography, geology, land cover and hydrology of New Zealand, including layers for segment-specific catchment characteristics derived from catchment averaged values of each variable. Topographical and network attributes associated with each segment include centroid coordinates; stream order[2]; segment length; elevation (maximum, minimum, mean); catchment area (for the individual segment and for the entire upstream catchment); modelled mean flow (Woods *et al.* 2006); and the distance from the sea (obtained by summing the length of all downstream segments).

[1] The terms "segment' and "NZReach" are essentially synonymous, and are treated as such throughout this report.
[2] A measure of catchment complexity, defined as 1 for the uppermost segments in a given network, and incrementing by one whenever two segments of equal order meet at a node. The most complex catchments in New Zealand (e.g., the Clutha River below Cromwell, and the Waitaki River below Lake Benmore) are order 8. Stream order provides a natural measure for grouping rivers by size, independently of mean flow, and is used frequently throughout this report.

2.2.2 Predictor variables

Additional predictor variables potentially related to angling usage metrics were obtained from three other sources. Measures of catchment land cover were derived from Version 2 of the New Zealand Land Cover Database (LCDB2; Snelder *et al.* 2005). The LCDB2 recognises 43 distinct land cover descriptors, but many of these are sparsely used and of little relevance to rivers fished by anglers. These were pooled into seven groups: *bare* (bare ground); *exotic_forest* (plantations, deciduous forests, shelter belts); *indigenous_forest* (including contiguous sub-alpine and alpine vegetation); *pastoral_heavy* (cropland, vineyards, orchards, high producing exotic grassland); *pastoral_light* (low producing or depleted grassland, tussock); *scrub* (including fern, manuka, kanuka, gorse/broom, matagouri, and exotic shrubs); and *urban* (including dumps and mines).

The second set of predictors was obtained from the Freshwater Environments of New Zealand classification (FWENZ; Wild *et al.* 2005). These variables represent catchment geology and climate, and are weighted by mean annual catchment runoff rather than by catchment area to ensure that parts of the catchment with higher runoff are appropriately related to segment-level runoff characteristics.

The pooled LCDB2 and FWENZ data sets include several pairs of variables which are highly correlated ($|r| \geq 0.9$), and are therefore likely to be surrogates for other each other. I used a subset of 28 of these variables, which have previously been used to predict water quality in New Zealand rivers (Unwin *et al.* 2010), to compile a modified list (Appendix 1) containing 26 of the variables used for the water quality models together with two new variables (fishable length, downstream distance to sea) which I considered were potentially relevant to freshwater angling. The third set of predictor variables, derived from the water quality model referred to above, was based on predicted values for eight water quality analytes at all 570,000 river segments in the REC (Table 2, Appendix 1).

2.2.3 REC catchment categories

The various REC, LCDB2, and FWENZ variables associated with the REC provide continuous measures, and are well suited to numerical modelling. However, for many purposes, it is also convenient to condense these into a series of discrete, categorical variables structured to match the underlying hierarchical character of the REC. Four of the six available levels (climate, source of flow, landcover, network position) were used in this study; readers are referred to Snelder & Biggs (2002) for full details.

Table 2: Water quality variables used as predictor variables for modelling angling usage and related metrics. Values for each river are based on modelled data for each NZReach (Unwin *et al.* 2010), averaged over the fishable length.

Analyte type	Abbreviation	Description	Units
Physical	CLAR	Black disc visibility (clarity)	m
	COND	Electrical conductivity	µS/cm
Nutrients	NH4N	Ammoniacal nitrogen	mg/l
	NO3N	Oxidised nitrogen	mg/l
	TN	Total nitrogen	mg/l
	DRP	Dissolved reactive phosphorus	mg/l
	TP	Total phosphorus	mg/l
Bacteria count	ECOLI	*Escherichia coli*	/100 mL

The REC climate class is defined in terms of mean annual temperature (C = cool, W = warm) and precipitation (D = dry, W = wet, X = extremely wet), so as to yield six discrete classes (CD, CW, CX, WD, WW, WX). Preliminary inspection of the FGNZ angler survey databases confirmed that 77% (115) of the 149 angling rivers in the three warm climate classes were in the WW class, so these were collapsed into a single class to yield four classification levels henceforth denoted CD, CW, CX, W.

The next level in the REC hierarchy, source of flow, is defined using a rainfall weighted measure of catchment elevation. For example, catchments in which over 50% of annual rainfall occurs at elevations exceeding 1000 m are classified as mountain (M). The source of flow class also includes a measure of the proportion of the catchment draining lakes, yielding a four-level classification comprising M (mountain[3]); H (hill); L (lowland); and Lk (lake).

The fourth REC class (after geology, which was not considered in this study) indexes the predominant land cover (landcover) in the upstream catchment based on the LCDB2. The original LCDB (LCDB1) recognised 17 landcover categories, subsequently increased to 43 in the LCDB2. For REC purposes these were condensed into nine classes, dominated by pasture (P), indigenous forest (IF), and tussock (T); less common classes include bare ground (B), scrub (S), and exotic forest (EF). For the purposes of this study these were further condensed into just two classes, corresponding to natural landcover (indigneous forest, tussock, bare, scrub; class N), and modified landcover (pasture, exotic forest, urban; class M).

The fifth REC class, network position[4], is defined solely by stream order. The REC recognises three levels, defined as low, middle and high order, corresponding to stream orders 1-2, 3-4, and 5-8. For the purposes of this study, stream order was regrouped into five classes, representing orders 2-3, 4, 5, 6, and 7-8.

It is important to recognise that the climate, source of flow, and landcover classes for each NZReach are based on the upstream catchment rather than the local conditions at each segment. For example, the lower ~50 km of the Ashburton River is assigned to climate class CW (cold wet) and source of flow class H (hill), reflecting its headwater origins in the Canterbury foothills rather than its course over the more arid and low-lying Canterbury Plains.

An important technicality underlying the REC class structure is that, at each level, classes are defined by the concatenation of all higher level classes rather than just by the single descriptor associated with that level. For example, the second level (source of flow) is nested within the first level (climate), so that source of flow class H (hill) actually comprises six classes (e.g., CD/H, CW/H, CX/H etc.) representing all possible climate × source of flow classes. Since the REC was developed it has become common practice to refer to each level of the hierarchy as if it were a single entity, and I frequently do so in this report. However, this interpretation is potentially over-simplistic, and readers seeking a deeper understanding of the REC methodology are referred to Snelder & Biggs (2002).

2.3 Data matching

2.3.1 Angling rivers

The REC is based solely on satellite-derived data, and does not currently provide a natural way to link individual segments with named waterways. Because angling usage data are defined solely by river name, there is no direct way to match rivers as identified by anglers (e.g., Mataura River) with a specific subset of REC segments.

[3] Subsequent revisions to the REC have further subdivided the M class so as to add class G (glacial mountain). This distinction is ignored in the present study.
[4] The last REC class, valley landform, is essentially a measure of local stream gradient and applies only at segment scale. It is therefore not relevant to this study, which is concerned with spatial scales representative of a whole river.

To address this problem, all rivers in the angling usage data set were associated with three REC segments, defining their upstream and downstream extent, and the uppermost point at which the river was considered by regional FGNZ staff to sustain a viable fishery. Approximate coordinates for each segment were obtained from the NZMS260 1:50,000 map series, and then used to identify the closest NZReach for which the stream order matched that of the river in question. Downstream coordinates were defined either by the river mouth (for rivers flowing into the sea or a lake), or the confluence with another named river (for individual tributaries). Upstream coordinates were either a lake outlet (where appropriate), or the highest point in the headwaters (generally of order two or three) where the coalescing stream network first developed a clearly defined mainstem. These were co-located with the centroid of the nearest REC segment, and traced downstream from the uppermost segment to identify and name all segments between the two endpoints. Within each catchment, these analyses were performed working upstream from the river mouth in order of increasing tributary altitude to ensure that segments in streams joining an already named stream were named appropriately (Unwin 2009b).

For the 2001/02 and 2007/08 surveys, 25 river fisheries for which regional FGNZ managers sought more detailed information on usage patterns were subdivided into reaches, normally corresponding to well-defined geographical boundaries such as confluences with major tributaries (Unwin 2009a, Table 3). To obtain corresponding estimates for the 1994/95 survey, the estimated usage for each river was partitioned across reaches based on their average usage for the 2001/02 and 2007/08 surveys. With one exception, each reach was then treated as a separate river for subsequent data matching. The exception was the lower Waitaki River below Kurow, where the five angler survey reaches reflected the information needs of a proposed hydroelectric development (Project Aqua) rather than catchment characteristics.

2.3.2 Predictor variables

REC, LCDB2, and FWENZ predictor variables for each river were based on the values associated with the segment identified as marking their downstream limit. Fishable length was generally derived by summing lengths for all segments between the downstream NZReach and the upstream angling limit, but for very small streams (e.g., Boundary Creek, which drains into Lake Wanaka via a steep side valley), the fishable length was recorded as 0.5 km. Water quality variables were averaged over all segments within the fishable length.

2.4 Data analysis

2.4.1 Data sets

The merged angler usage/REC data set was used to compile three working data sets, each suited to a particular analysis. To compare annual usage trends from 1994/95 to 2007/08, trout angler usage (*days trout*) was tallied by year and REC class for combinations of classes (such as climate and source of flow) likely to represent meaningful groupings of river types. These analysis did not include usage by overseas anglers, who were surveyed only in 2007/08. This data set included 831 rivers (counting one river for each section of a subdivided river) for which at least one usage estimate was available over the period covered by the surveys.

The second data set was based on data for the most recent survey (2007/08), and was used to model total angling usage by New Zealand anglers (*days total* and *days trout*; Table 1). Both variables were approximately log-normally distributed (c.f. Unwin & Deans 2003), so were log transformed before fitting models. Overseas angler usage (days os, %os) was characterised by a high proportion (60%) of rivers with zero overseas angler usage and was poorly suited to the modelling approach used in this study (see Section 2.4.2), so analyses of these data were limited to simple cross-tabulation. The full data set consisted of usage estimates for 632 rivers, representing a total of 666,700 angler-days for *days total*, and 516,500 angler-days for *days trout* (77.0% of river angling in 2007/08).

The third data set was essentially the same as the second, with the four 2007/08 angling usage measures replaced by the eight attribute-related measures derived from the 1979/81 survey (Table 1). Rivers with fewer than ten responses were discarded, to minimise the risk of including spurious values based on poorly-estimated means for each attribute. The 1979/81 survey treated trout and salmon fishing in the same river as separate fisheries, so I applied an additional filter to limit the data to trout angling. The final data set included 355 rivers with a total usage (in 2007/08) of 483,900 angler days, i.e., 93.7% of the usage represented in the second data set.

Treating mean attribute scores as responses to be modelled in terms of catchment characteristics is not necessarily meaningful in all cases. Five attributes – overall importance, scenic beauty, peace & solitude, catch rate, and size of fish – can legitimately be expected to vary in ways that are related to, and perhaps defined by, factors such as river flow, altitude, and instream water quality. For the remaining three attributes – ease of access, area fishable, and distance from home – the relationship to catchment-scale measures is less clear, and all three could equally well be interpreted as potential *a priori* predictors of variables such as total usage and overall importance. For example, the *close to home* attribute for each river could be interpreted as a surrogate for mean travel distance, which is strongly related to total usage (Unwin & Deans 2003).

2.4.2 Random Forest models

Models for each usage metric as a function of selected subsets of predictors were estimated using Random Forest (RF; Breiman 2001, Cutler *et al.* 2007), a type of multivariate regression model which uses classification and regression trees[5] to identify the relationships between each predictor variable and the response variable, and to rank each predictor in order of importance. Fitting a RF involves constructing a large number of individual trees (hence "forest"), each of which uses a random subset of the available predictors to predict a random subset of the available cases, and evaluating the effect of each predictor by comparing the performance of trees which include it with those that do not. Readers are referred to Unwin *et al*. (2010) and references therein for a more detailed description of RF models, but for the purposes of this report their key features are:

1. they are free from assumptions about the underlying distributions of the predictor variables; and
2. they can automatically accommodate non-linear relationships and high order interactions between predictors.

RF models are by nature large and unwieldy data structures, consisting of a large number of individual trees (500 for this study), each of which typically contains several hundred branches. Rather than inspect the entire forest, which would be both laborious and uninformative, it is much more useful to assess model performance using graphical and tabular methods to rank the predictors in order of importance, and create partial dependence plots. Predictor importance is measured by a numerical score (henceforth denoted I_{score}) which typically ranges from 10 or less for the least important predictors to 20 or more for the most important. Partial dependence plots show the marginal effect of a variable on the response after accounting for the average effects of the other variables in the model. These plots do not perfectly represent the effects of each variable, particularly when predictors are highly correlated or strongly interacting, but provide useful information for interpretation.

[5] Classification and regression trees (CART) use a decision tree as a predictive model which maps observations about an item to conclusions about the item's target value. In these structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels. See http://en.wikipedia.org/wiki/Decision_tree_learning for an overview.

An additional feature of RF models is that, because observations and predictors are randomly selected throughout the fitting process, successive models fitted to the same data set will exhibit subtle differences in structure and diagnostics. Consequently, statistics such as total explained deviance, mean square error, partial dependence plots, and the rank order of predictors of similar importance, will vary each time a model is estimated.

2.4.3 Predictor sets

RF models were developed using two complementary predictor subsets, yielding four sets of models in total. The first subset consisted of the first eight variables listed in Appendix A, i.e., those relating solely to catchment topography and source of flow, together with the 20 remaining variables from the REC, FWENZ, and LCDB2. The second subset consisted of the same first eight variables, together with the eight water quality variables from the 2010 NIWA/MfE study. My rationale for treating the water quality predictors separately was that these were modelled using essentially the same subset of REC/FWENZ/LCDB2 predictors to those listed in Appendix A, so that – in principle – they did not introduce any new information. By comparing results for the two predictor sets when applied to each data set, I hoped to establish whether or not models incorporating predictors specifically related to the aquatic environment (i.e., water quality) would outperform models based solely on catchment-scale predictors. For ease of reference, I refer to these two predictor sets as the REC predictors, and the WQ predictors, respectively.

2.4.4 Computational methods

All calculations were performed using Version 2.12.1 of the software environment R (R Development Core Team 2010) via the *randomForest* function library. Results for each model are summarised in Appendix B as a full-page figure showing a scatterplot of observed vs. predicted values; a normal quantile plot to assess the distribution of residuals for each model and characterise any departures from normality; and partial residual plots for the six leading predictors (see Appendix B for details).

For technical reasons, the algorithms for fitting RF models do not yield a precise measure of the percentage of variance explained by each model. As a surrogate, I have interpreted the squared correlation (r^2) between the observed and predicted values for each model as an approximate measure of percentage explained variance, with the caveat that this is in fact conservative, and hence that model performance is in general slightly better than indicated by the reported r^2.

3 Results

3.1 REC classification of river fisheries

The REC climate and landcover classes provide a broad-scale overview of the rivers fished by FGNZ anglers (Figures 1, 2). With respect to climate (Figure 1), South Island rivers are almost completely dominated by the CX (cold extremely wet), CW (cold wet), and CD (cold dry) classes, with the rivers in each class forming three bands lying more or less parallel to the Southern Alps. The main exceptions to this trend occur in South Otago, reflecting the wetter climate of the Catlins area, and parts of Canterbury, where rivers such as the Waiau, Hurunui, and Ashburton are classified as CW. North Island angling rivers are generally classified as CW south of the Waikato and W over the upper third of the island, although a few rivers in coastal Hawkes Bay also fall into class W. Class CX is primarily limited to the Tararua Ranges, Mt Taranaki, Mt Ruapehu, and inland Gisborne, although isolated pockets are apparent around local mountainous areas such as Pirongia. Class CD is limited to the Wairarapa, southern Hawkes Bay, and the Rangitikei catchment.

The distribution of rivers by source of flow class (Figure 2) shows considerable overlap with climate class, particularly in the upper North Island and along the Southern Alps, but reveals finer detail in areas such as Westland, Nelson, and the lower North Island. For example, Westland rivers are distributed across all four source of flow classes, particularly in the Grey Valley and Buller regions where the L (lowland) and H (hill) classes are at least as well represented as class M (mountain). In the Wellington region, the CX climate class separates naturally into two source of flow classes, with most such rivers classified as hill rather than mountain. Source of flow also clearly delineates rivers which are primarily lake fed, such as the Waikato, Waitaki, Clutha, upper Hurunui, and Arnold.

Overlaying the climate and source of flow classes creates a total of 16 possible two-factor groups, of which only W/M (warm, mountain) is not represented by any angling river. These groups (e.g., Figure 3) generally identify clusters of rivers at spatial scales of ~50-100 km within each climate class. South of the Clutha River, for example, source of flow helps to define up to three clusters of class L rivers in South Otago, central Southland, and western Southland, together with a more homogenous group of class H associated with the upper Mataura, Oreti, Aparima, and Waiau Rivers. Similar groupings (not shown) are apparent for the other three climate classes.

A comparison between the informal river classification system developed for the 1994/95 FGNZ survey (Unwin & Brown 1998) confirms that the FGNZ classes (lowland, back country, headwater, mainstem) are generally consistent with the REC model, but are much more ambiguous, and fall well short of capturing as much detail (Table 3). For example, source of flow class L includes 327 rivers, of which 250 (76%) were classified as lowland in the 1994/95 survey. However, class L includes 49 rivers which were classified as back country, and five which were classified as headwater. Conversely, source of flow class H includes 360 rivers, of which 91 (25%) were classified in 1994/95 as lowland, 141 (39%) as back country, and 95 (26%) as headwater.

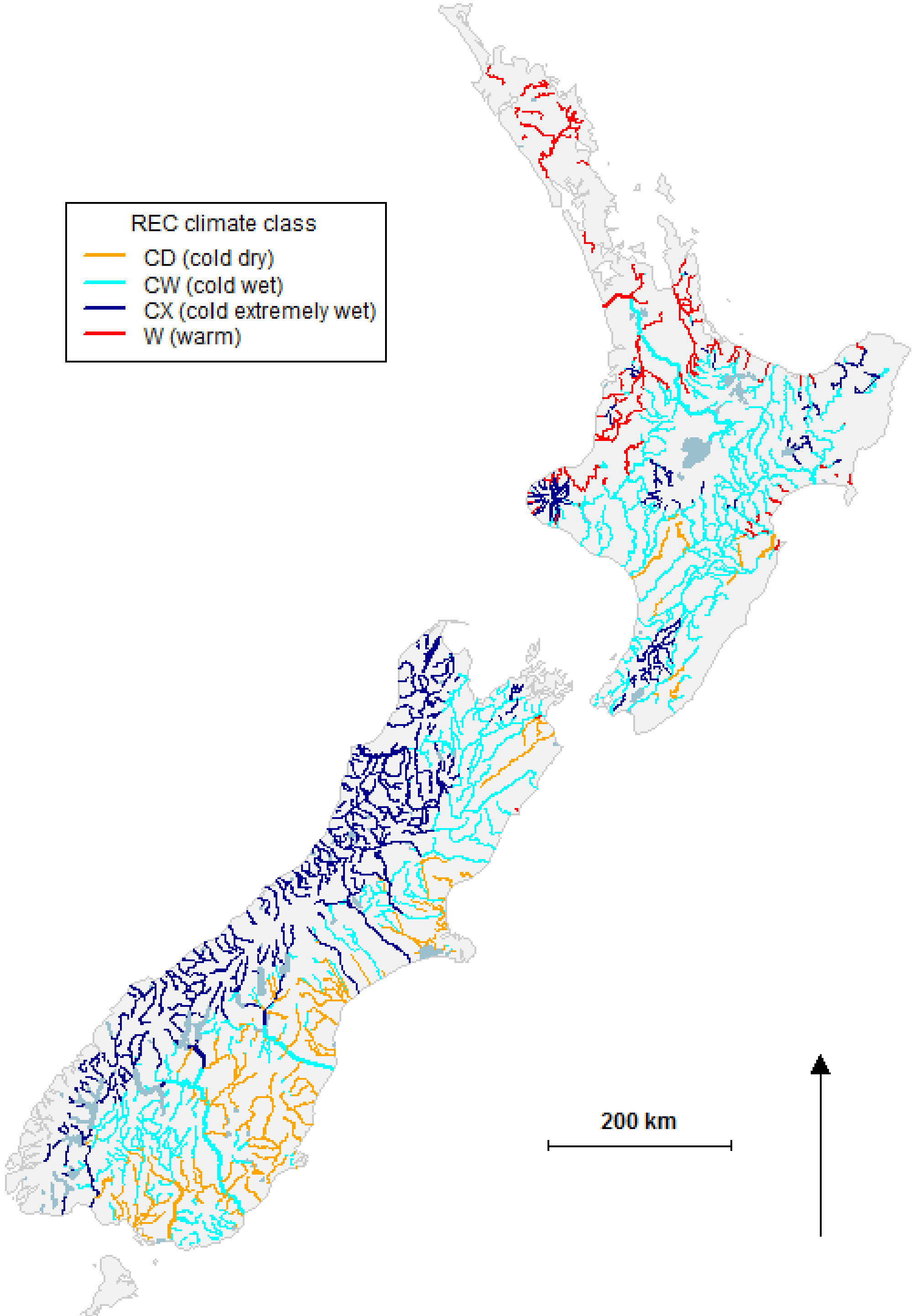


Figure 1: **FGNZ angling rivers by REC climate class.**

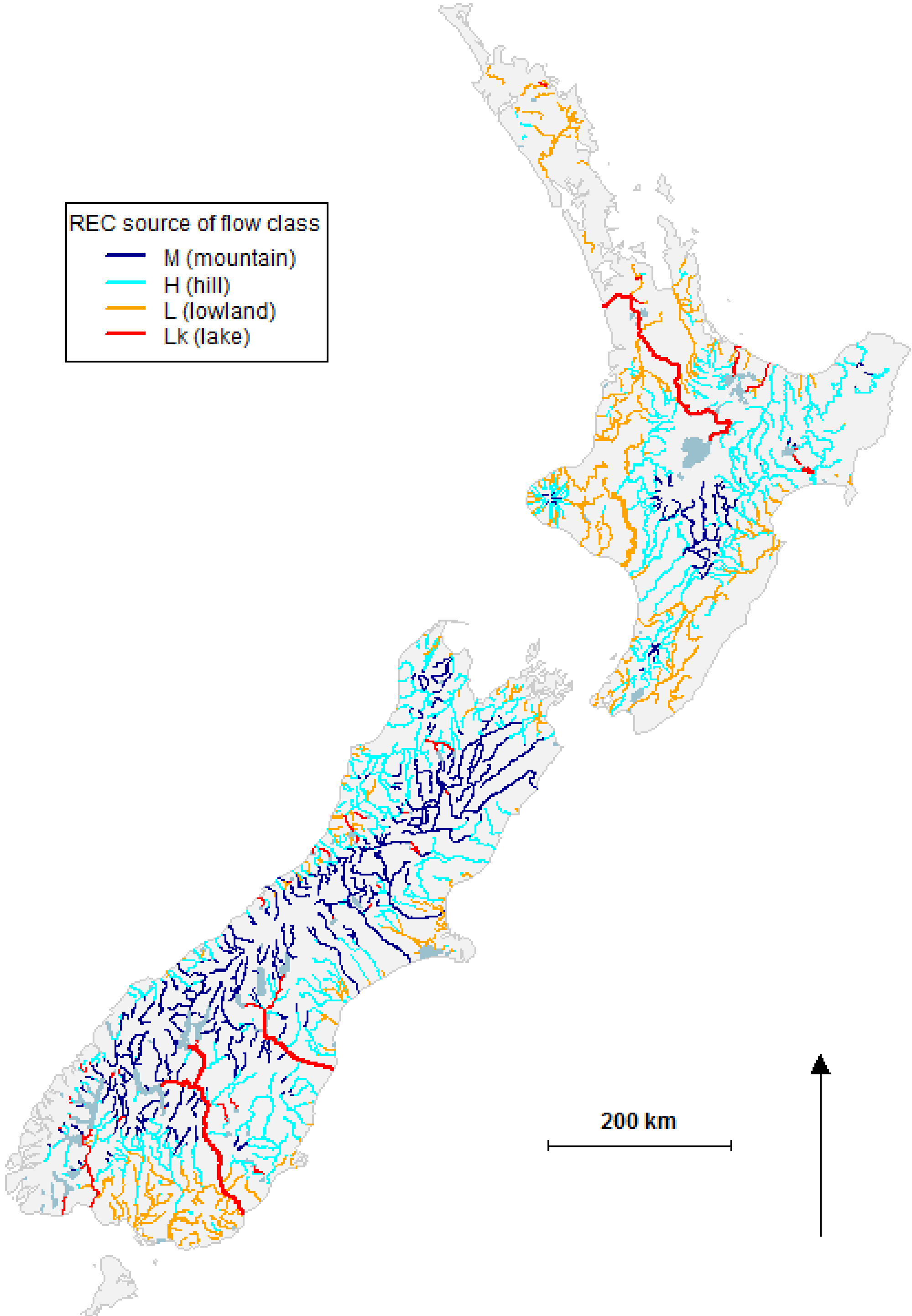


Figure 2: FGNZ angling rivers by REC source of flow class.

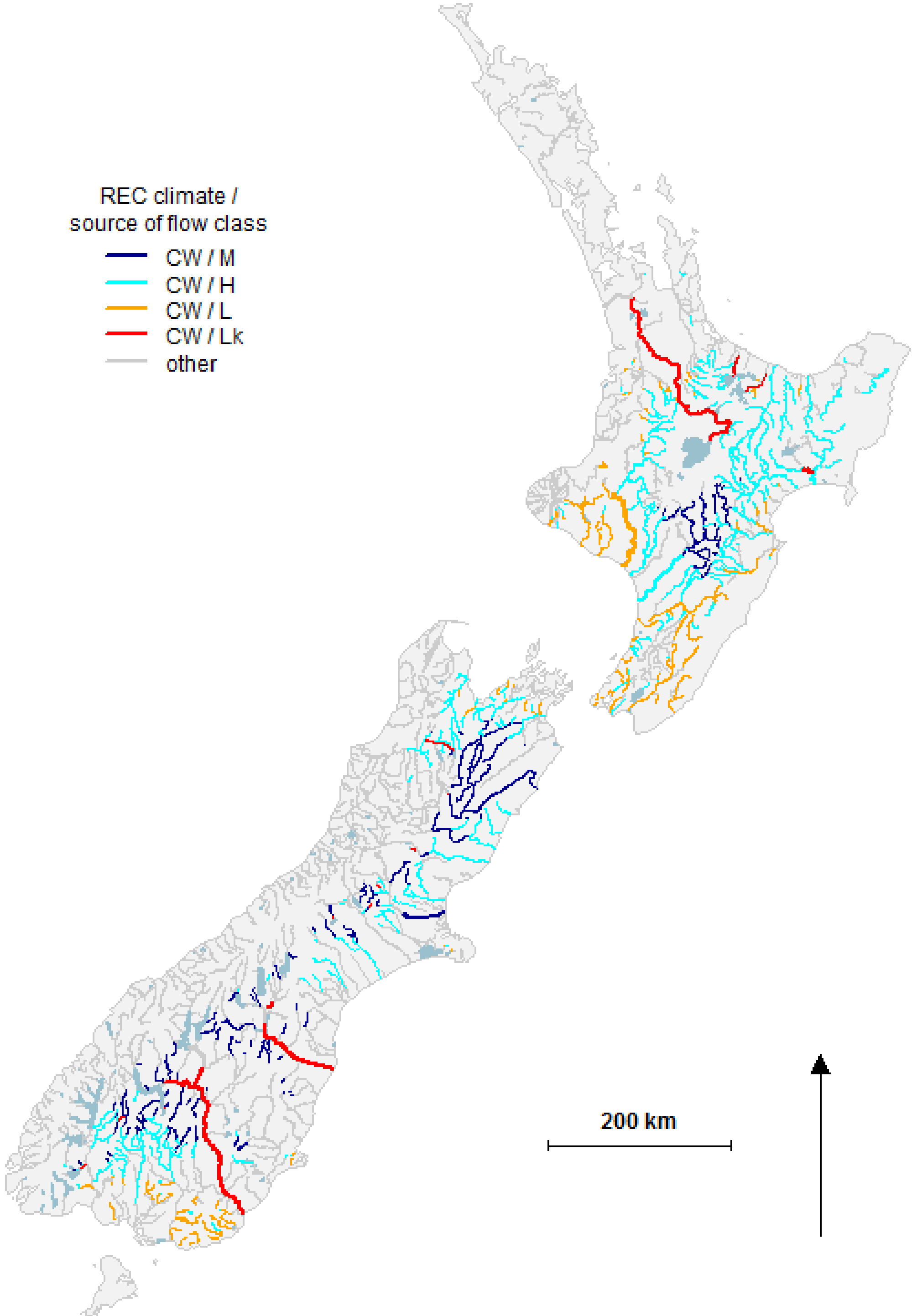


Figure 3: FGNZ angling rivers in the REC CW (cold wet) climate class, by REC source of flow class.

Table 3: Distribution of New Zealand angling rivers by REC climate and source of flow classes, relative to the earlier "water type" classification. Values in each cell show the number and total fishable length (km) of rivers by climate/source of flow/water type category, as absolute values (top row, plain font) and as percentages of the total for each column (bottom row, italic font).

Source of flow	Water type	Cold dry (CD)	Cold wet (CW)	Cold extremely wet (CX)	Warm (W)	Total
Lowland (L)	Lowland	69 / 1,446 *46.6% / 29.3%*	59 / 1,311 *18.9% / 12.8%*	20 / 171 *8.0% / 3.7%*	102 / 2,372 *83.6% / 72.8%*	250 / 5,300 *30.1% / 23.0%*
	Back country	8 / 295 *5.4% / 6.0%*	16 / 248 *5.1% / 2.4%*	20 / 149 *8.0% / 3.2%*	5 / 74 *4.1% / 2.3%*	49 / 765 *5.9% / 3.3%*
	Headwater	1 / 30 *0.7% /0.6%*	1 / 40 *0.3% /0.4%*	3 / 34 *1.2% /0.7%*	-	5 / 103 *0.6% /0.4%*
	Mainstem	11 / 1,231 *7.4% / 24.9%*	11 / 788 *3.5% / 7.7%*	1 / 54 *0.4% / 1.2%*	-	23 / 2,073 *2.8% / 9.0%*
Total, lowland		**89 / 3,001** ***60.1% / 60.7%***	**87 / 2,386** ***27.9% / 23.2%***	**44 / 408** ***17.7% / 8.9%***	**107 / 2,446** ***87.7% / 75.1%***	**327 / 8,241** ***39.4% / 35.7%***
Hill (H)	Lowland	14 / 541 *9.5% / 10.9%*	51 / 1,264 *16.3% / 12.3%*	24 / 320 *9.6% / 7.0%*	2 / 21 *1.6% /0.7%*	91 / 2,146 *11.0% / 9.3%*
	Back country	32 / 644 *21.6% / 13.0%*	64 / 1,697 *20.5% / 16.5%*	43 / 1,185 *17.3% / 25.8%*	2 / 88 *1.6% / 2.7%*	141 / 3,615 *17.0% / 15.7%*
	Headwater	2 / 16 *1.4% /0.3%*	43 / 578 *13.8% / 5.6%*	48 / 600 *19.3% / 13.0%*	2 / 24 *1.6% /0.8%*	95 / 1,219 *11.4% / 5.3%*
	Mainstem	5 / 596 *3.4% / 12.1%*	22 / 2,535 *7.1% / 24.7%*	6 / 436 *2.4% / 9.5%*	-	33 / 3,567 *4.0% / 15.5%*
Total, hill		**53 / 1,797** ***35.8% / 36.4%***	**180 / 6,074** ***57.7% / 59.2%***	**121 / 2,542** ***48.6% / 55.3%***	**6 / 134** ***4.9% / 4.1%***	**360 / 10,546** ***43.3% / 45.7%***
Mountain (M)	Lowland	1 / 2 *0.7% /0.04%*	-	-	-	1 / 2 *0.1% /0.01%*
	Back country	4 / 93 *2.7% / 1.9%*	22 / 944 *7.1% / 9.2%*	15 / 250 *6.0% / 5.4%*	-	41 / 1,287 *4.9% / 5.6%*
	Headwater	-	12 / 192 *3.8% / 1.9%*	47 / 622 *18.9% / 13.5%*	-	59 / 814 *7.1% / 3.5%*
	Mainstem		1 / 152 *0.3% / 1.5%*	3 / 308 *1.2% / 6.7%*	-	4 / 460 *0.5% / 2.0%*
Total, mountain		**5 / 95** ***3.4% / 1.9%***	**35 / 1,289** ***11.2% / 12.6%***	**65 / 1,180** ***26.1% / 25.7%***	**-**	**105 / 2,564** ***12.6% / 11.1%***
Lake (Lk)	Lowland	-	1 / 1 *0.3% /0.01%*	2 / 21 *0.8% /0.4%*	3 / 74 *2.5% / 2.3%*	6 / 96 *0.7% /0.4%*
	Back country	1 / 49 *0.7% / 1.0%*	3 / 34 *1.0% /0.3%*	12 / 178 *4.8% / 3.9%*	3 / 115 *2.5% / 3.5%*	19 / 376 *2.3% / 1.6%*
	Headwater	-	1 / 100 *0.3% / 1.0%*	2 / 2 *0.8% /0.04%*	-	3 / 102 *0.4% /0.4%*
	Mainstem	-	5 / 378 *1.6% / 3.7%*	3 / 269 *1.2% / 5.8%*	3 / 489 *2.5% / 15.0%*	11 / 1,136 *1.3% / 4.9%*
Total, lake		**1 / 49** ***0.7% / 1.0%***	**10 / 513** ***3.2% / 5.0%***	**19 / 470** ***7.6% / 10.2%***	**9 / 678** ***7.4% / 20.8%***	**39 / 1,710** ***4.7% / 7.4%***
Total		**148 / 4,942**	**312 / 10,261**	**249 / 4,599**	**122 / 3,258**	**831 / 23,061**

Climate class CW accounts for 312 (37.5%) of the 831 rivers listed in Table 3, representing 44.5% of total fishable river length. Classes CX (249 rivers) and CD (148 rivers) are the next most frequent, each representing 20-21% of total fishable length. Class W contains 122 rivers (14% of total fishable length). Source of flow classes H (360 rivers, 10,546 km) and L (327 rivers, 8,241 km) jointly account for 82.7% of rivers and 81.5% of total fishable length, followed by classes M (105 rivers, 2,564 km) and Lk (39 rivers, 1,710 km). Considered as class pairs, the five most common climate/source of flow classes are CW/H (180 rivers, 6,074 km); CX/H (121 rivers, 2,542 km); CD/L (89 rivers, 3,001 km); CW/L (87 rivers, 2,386 km); and W/L (107 rivers, 2,446 km). Collectively, these five classes represent 584 (70.3%) rivers, and 71.3% of total fishable length.

3.2 Trends in annual usage, 1994/95 – 2007/08

Total annual effort expended on rivers by New Zealand-resident anglers ranged from 710,200 angler days in 1994/95 to 627,900 angler days in 2001/02 (Table 4), with the relatively low numbers in 2001/02 largely driven by an unusually poor salmon fishing season in 2002 (Unwin & Image 2003). Removing the effect of the salmon fishery reveals a more consistent trend over the period of record, with effort devoted to trout fishing declining by 37,400 angler-days from 1994/95 to 2001/02, and by 36,600 angler-days from 2001/02 to 2007/08. These figures are tentative, given that the assumptions which underlie my estimates of salmon fishing effort are generic rather than season-specific, but – if taken at face value – suggest that total effort expended on FGNZ rivers by anglers fishing for trout has declined by 74,000 angler-days (13%) over the period of record. Assuming a uniform rate of decline, this represents an annual decrease of 1.05% over 13 years.

Table 4: Estimated total annual effort (angler-days ± 1 SE) expended on rivers by New Zealand-resident FGNZ licence holders, 1994/05 - 2007/08.

Angling season	Total days (all New Zealand anglers)	Total days (New Zealand trout anglers)
1994/95	710,200 ± 14,000	578,400 ± 10,400
2001/02	627,900 ± 11,900	541,000 ± 10,400
2007/08	657,400 ± 14,000	504,400 ± 11,000

The strength of this trend is strongly influenced by REC climate class (Figure 4), being most apparent for rivers in class CD (28.5%, from 206,200 to 147,400 angler-days), and class W (34.2%, from 41,000 to 27,000 angler-days). Usage also decreased by 8.3% for rivers in class CW (from 257,700 to 236,400 angler-days), but increased by 27.3% for rivers in class CX (from 73,500 to 93,500 angler-days). The general pattern thus indicates a strong decrease in rivers draining dryer and warmer areas of New Zealand, a weaker decrease in rivers draining areas of intermediate rainfall, and an increase in rivers draining areas of high rainfall. Total annual usage is also strongly related to stream order (Figure 5), increasing more or less uniformly from the smallest (2-3rd order) to the largest (7-8th order) streams. Stream order also appears to influence the strength of the underlying temporal trend, which is strongly apparent in high order streams (6th and above), weak or equivocal in 4th and 5th order streams, and possibly reversed in 2nd and 3rd order streams.

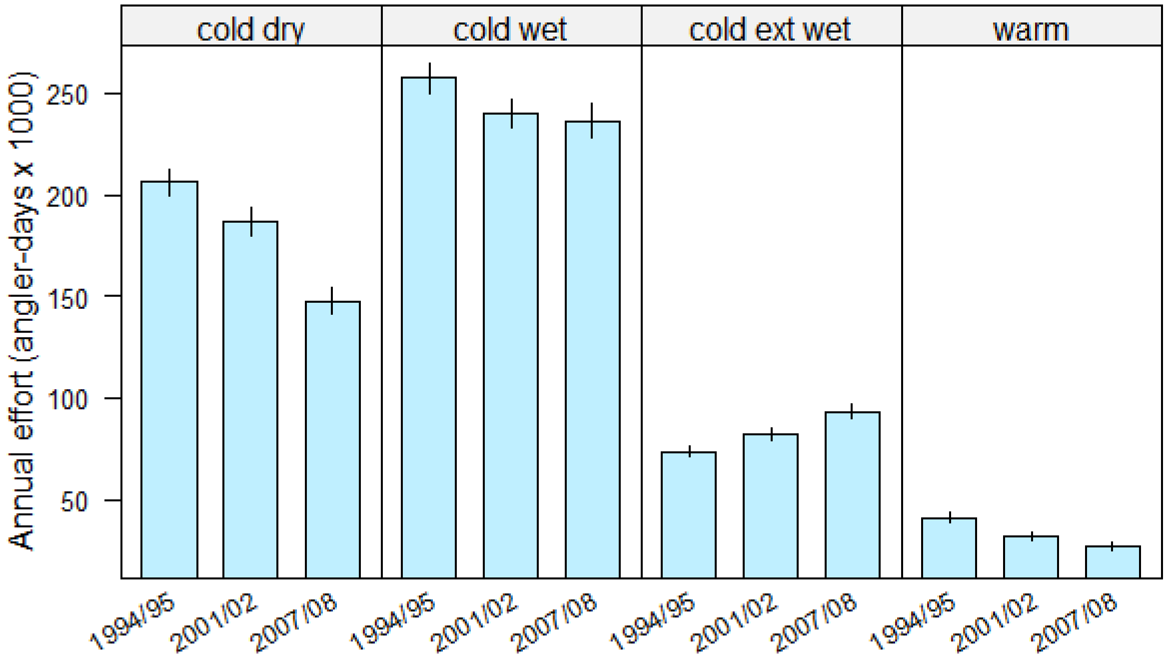


Figure 4: Total effort (angler-days ± 1 SE) expended on rivers, by New Zealand-resident FGNZ licence holders, by year and REC climate class.

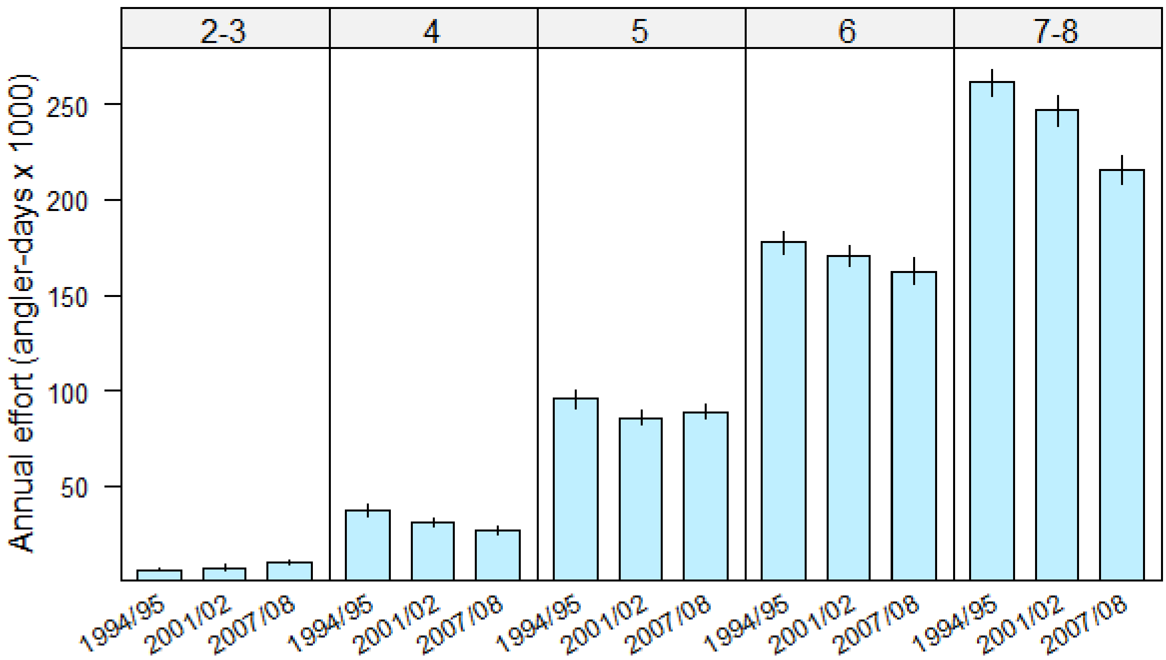


Figure 5: Total effort (angler-days ± 1 SE) expended on rivers, by New Zealand-resident FGNZ licence holders, by year and stream order.

Adding REC source of flow class to the above analysis confirms the strength of the decline in climate class CD, but also suggests a strong interaction with source of flow (Figure 6). Within climate class CD, the decline is almost entirely confined to source of flow class L, with usage declining by 37.7% (154,900 to 96,600 angler-days) in class CD/L, compared to 3.3% in class CD/H. In absolute terms, the latter figure represents a decrease from 49,500 ± 2,600 to 47,900 ± 4,100 angler-days, and is not statistically significant. A similar pattern is evident within climate class CW, where the results show a strong (36.7%) decline in class CW/L, no significant change in class CW/H, a significant increase in class CW/M (27.1%, p > 0.95), and a moderate (14.6%) but not statistically significant increase in class CW/Lk. Collectively, these results suggest that the observed decline in effort since 1994/95 has been largely confined to lowland rivers in areas of low and intermediate rainfall, which has been partly offset by in increased effort in rivers draining catchments in areas of higher rainfall.

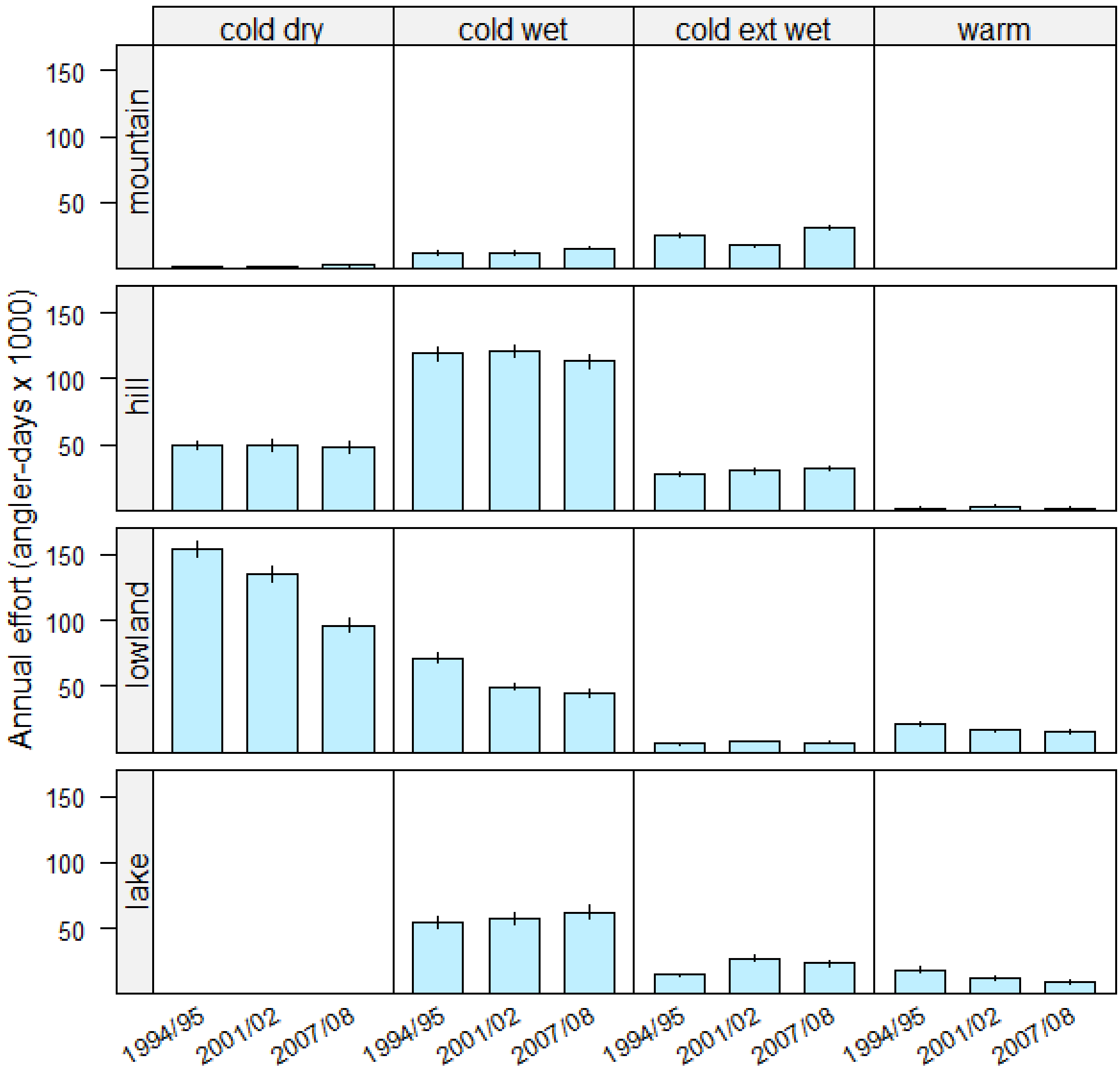


Figure 6: Total effort (angler-days ± 1 SE) expended on rivers, by New Zealand-resident FGNZ licence holders, by year and REC climate / source of flow class.

Including REC landcover class in these analyses further highlights the contrast between differing climate and source of flow classes (Figure 7). In relation to climate, angling effort on rivers draining catchments in which natural land cover predominates is largely confined to classes CW and CX. For these rivers, annual usage either shows no consistent trend (for class CX), or has increased over the period of record (class CX). Among rivers draining catchments in which land cover has been significantly modified, usage has declined by 28.5% for class CD (195,500 to 140,500 angler-days), and by 34.5% for class W (38,500 to 25,200 angler-days), but has shown no consistent trend for class CW. Class CD/N is represented by 24 rivers, which collectively attract between 5,900 and 9,700 angler-days per year but show no consistent long-term trend. Classes W/N and CX/M are the least used climate/landcover classes, collectively accounting for between 4,400 and 5,700 angler-days per year.

Usage trends in relation to source of flow and landcover show much less evidence for a strong interaction between these two classes (Figure 7, lower). A consistent trend is apparent only for source of flow class L, within which usage has declined by 60.6% for landcover class N (31,800 to 12,500 angler days), and by 31.4% for landcover class M (221,700 to 152,000 angler days). Annual usage levels for other classes, notably M/N, H/N, Lk/M, and H/M, appear to be relatively stable, although there is some evidence for an increase in usage within class M/N (20.0% from 1994/95 to 2007/08) and class Lk/N (24% over the same period). However, the dominant pattern suggested by this analysis is that decreasing usage of lowland fisheries is related more to source of flow than to catchment land cover.

3.3 Overseas visitor usage, 2007/08

Overseas visitors showed a moderate but clearly defined preference, relative to New Zealand anglers, for rivers in REC source of flow classes M and H (Figure 8). Visitors expended 69% of their river fishing effort (33,800 of 48,900 angler-days) on mountain- and hill-fed rivers, compared to 53.6% (277,400 of 516,900 angler-days) for New Zealand residents.

Further analysis of these data, taking into account the influence of REC climate class, suggests a tendency for visitors to preferentially favour the CW and CX classes, particularly for mountain and lake-fed rivers (Figure 9). Usage data for overseas visitors are confounded by a high proportion of zeros, but the trend towards increasing usage from CD to CW to CX for these two source of flow classes is unambiguous.

3.4 Model predictions

3.4.1 General performance

Model performance varied widely, with the percentage of variance explained ranging from 21.4% for the poorest fitting model (*overall importance*, WQ model) to 51.8% (*scenic beauty*, REC model; Tables 5, 6, Appendix 2). Overseas angler usage proved to be more or less intractable to the RF model approach, evidently because of the high proportion of zeroes, so these results (*days.os* and *pos*) were discarded. Explained variance exceeded 40% for 10 of the 20 remaining models, and 50% for the top four models.

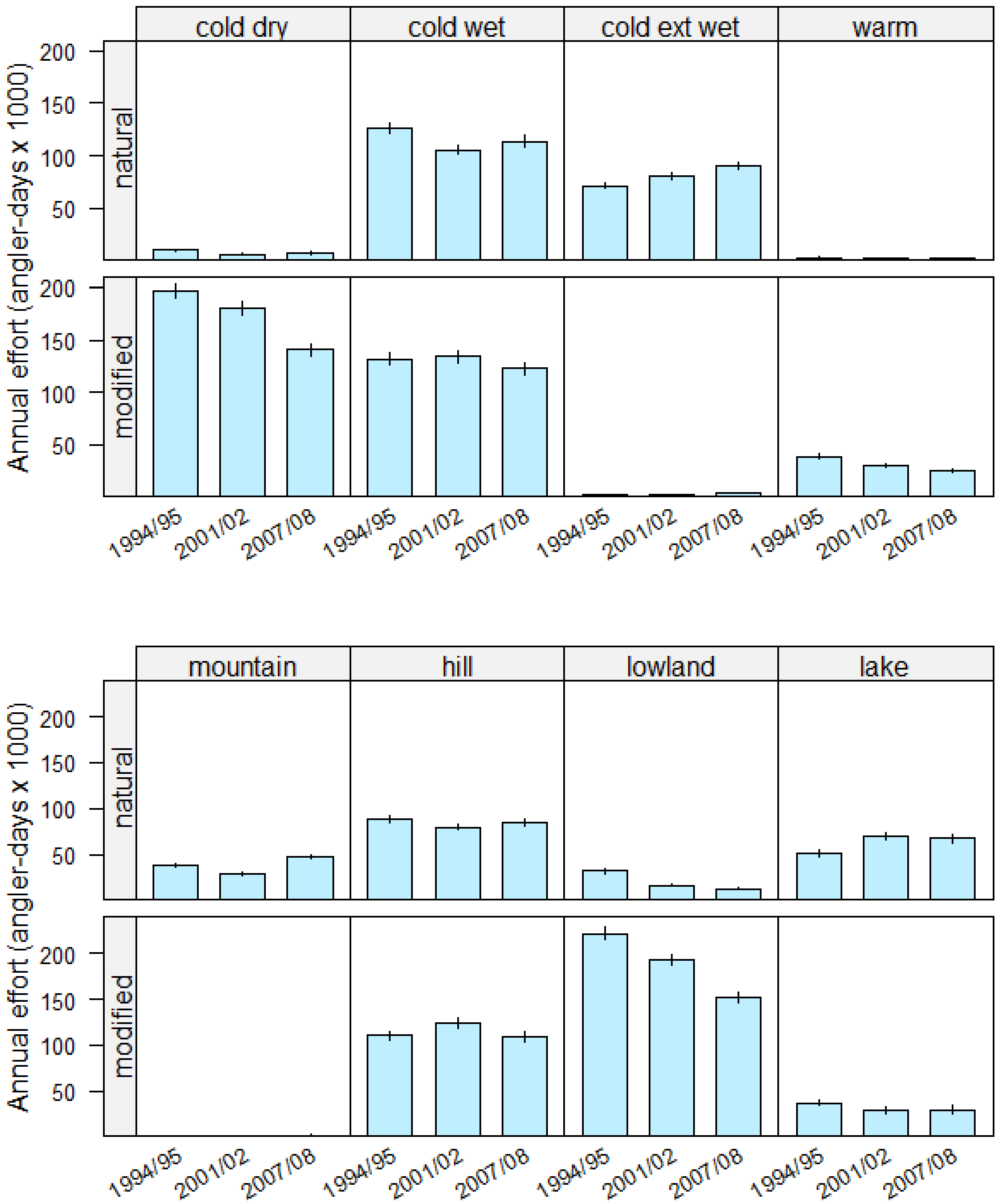


Figure 7: Total effort (angler-days ± 1 SE) expended on rivers, by New Zealand-resident FGNZ licence holders, by year and REC climate / source of flow / landcover class. The two panels show, respectively, annual usage grouped by REC climate / landcover class (upper), and REC source of flow / landcover class (lower). Landcover classes are pooled into two groups, representing natural and modified land cover, as described in Section 2.2.3. The source data are the same as for Figures 4-6.

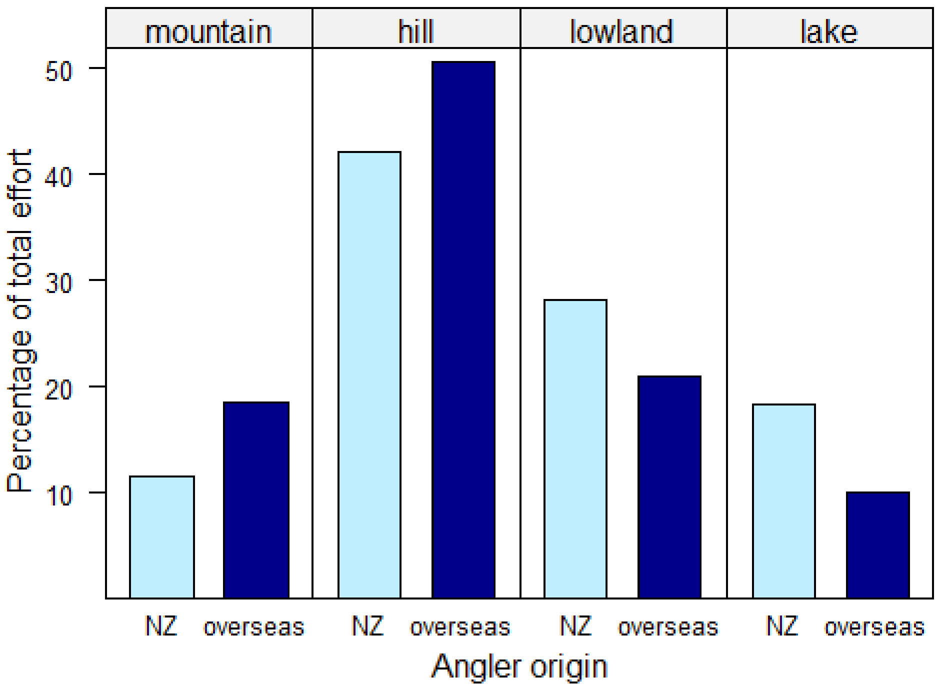


Figure 8: Distribution of angling effort by REC source of flow class and angler origin, 2007/08.

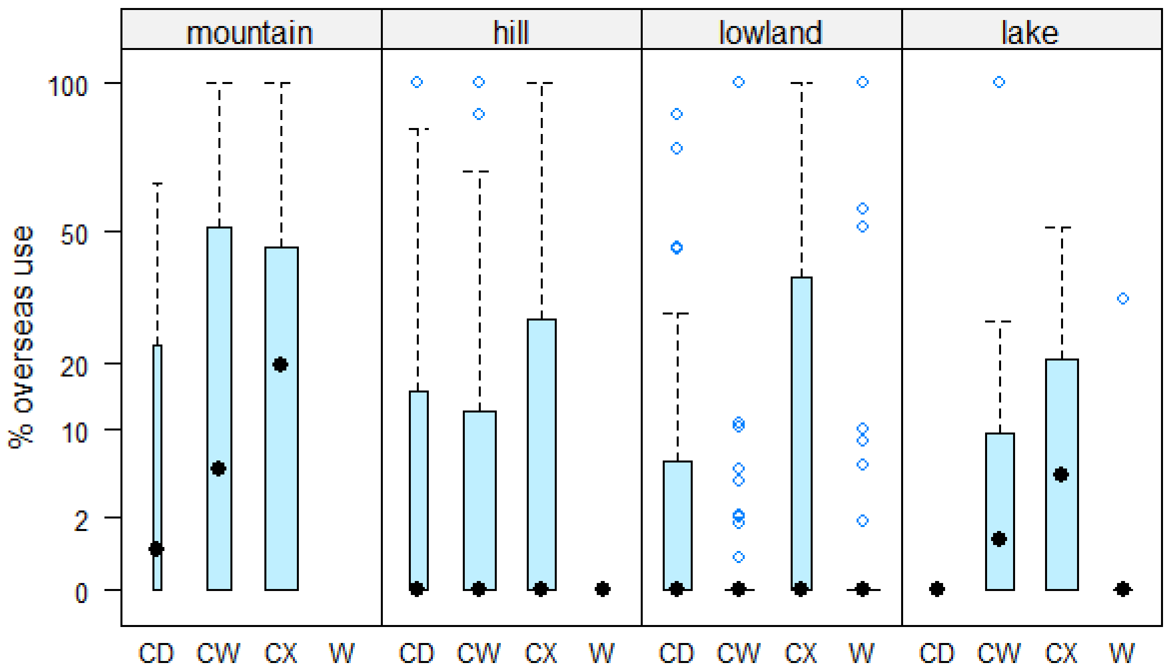


Figure 9: Overseas visitor usage of New Zealand rivers by REC climate and source of flow class in 2007/08, expressed as a percentage of total annual usage for each class . Filled and open symbols represent class medians, and outliers, respectively. Box widths are proportional to the number of rivers in each class; note that the vertical axis is square-root transformed.

Table 5: Importance scores[6] for predictors of water quality analytes, derived from the REC predictor set. Columns are ordered so as to facilitate comparisons between angling metrics representing estimated usage, importance scores, geographical attributes, scenic and wilderness values, and fish-related attributes. The approximate % of variance explained is shown for each model.

Predictor type	Predictor	days total (50.2%)	days trout (49.2%)	overall importance (27.2%)	ease of access (32.6%)	area fishable (34.7%)	close to home (44.4%))	scenic beauty (51.8%)	peace &solitude (43.1%)	catch rate (24.9%)	size of fish (42.4%)
base	Fishable length	**23.7**	**24.4**	9.0	-1.0	15.9	6.3	3.9	4.0	12.5	3.0
	Catchment area	**31.9**	**33.8**	14.3	6.4	14.7	9.0	4.6	7.9	10.5	5.1
	Distance to sea	5.1	5.1	9.8	6.7	1.7	7.4	8.9	12.8	13.6	7.0
	Downstream elevation	8.2	7.2	11.0	8.9	4.3	14.0	12.1	16.2	15.6	15.7
	Mean flow	17.5	18.2	15.2	5.1	20.0	5.7	5.5	7.2	13.2	8.8
	Mean slope	8.1	6.7	6.0	7.4	4.7	9.3	13.1	10.2	9.2	9.7
	Catchment elevation	12.6	12.2	8.0	15.0	4.4	18.0	14.8	13.2	8.4	**20.4**
	Lake index	6.0	7.6	2.5	2.9	2.2	0.9	3.7	4.1	2.1	2.3
climate	Rain variability	3.5	2.6	6.9	7.2	4.0	3.9	6.7	4.2	7.4	8.1
	Min temperature	13.2	12.3	7.7	8.6	5.9	7.8	7.0	8.7	9.9	8.1
	Max temperature	12.5	13.7	6.3	8.6	8.4	6.8	8.4	5.7	5.8	8.4
	Rain days > 10 mm	10.3	10.5	9.7	5.2	7.0	11.0	14.4	10.1	6.6	11.7
	Rain days > 50 mm	9.9	7.7	8.2	5.2	7.0	8.1	8.8	5.1	5.7	6.5
	Rain days > 200 mm	8.5	7.0	8.8	7.8	9.2	8.3	7.8	9.3	6.3	10.5
	Evapotranspiration	9.9	9.9	8.4	6.2	8.3	9.9	7.7	9.1	6.5	10.1
geology	%alluvium	6.5	5.6	7.3	7.4	6.4	6.7	12.3	7.1	5.9	3.7
	%glacial	4.5	3.1	0.5	2.3	0.2	2.2	2.0	2.2	-1.3	2.7
	%peat	5.0	4.9	0.2	2.2	-0.5	2.3	2.2	0.4	1.2	1.8
	Mean calcium	7.5	9.2	9.4	5.1	3.1	5.4	5.7	5.4	2.9	7.2
	Mean hardness	8.0	9.5	4.5	8.8	3.8	7.7	6.7	8.7	4.8	3.9
	Mean particle size	11.5	9.3	9.3	5.0	2.3	6.9	7.9	8.8	6.5	4.7
	Mean phosphorous	9.4	9.5	6.8	9.1	14.2	3.6	5.6	4.1	5.3	5.4
LCDB2	%bare	11.2	10.7	4.8	7.5	7.3	7.1	3.3	4.4	8.2	6.0
	%exotic forest	8.9	8.3	4.6	9.7	2.7	13.8	15.5	13.2	2.7	10.0
	%indigenous forest	7.1	8.5	6.6	8.7	5.3	12.3	**20.5**	12.8	6.5	5.9
	%pastoral heavy	8.1	10.8	4.5	19.8	8.4	17.4	19.4	16.4	6.4	9.9
	%pastoral light	9.1	9.9	7.0	7.0	8.4	4.9	6.6	3.8	5.3	8.0
	%scrub	6.9	9.7	6.1	10.0	3.9	8.0	12.7	11.0	4.1	3.3

[6] Importance core (I_{score}) is highlighted so as to identify $I_{score} \geq 20.0$ (bold red); $15 \leq I_{score} < 20$ (red); and $10 \leq I_{score} < 15$ (blue). See Appendix 1 for a more detailed description and interpretation of each predictor. Scores are indicative only, particularly for lower ranked predictors in weak models. Predictor order and I_{score} can vary slightly each time the model is fitted, due to the random component built in to the RF model fitting process.

Table 6: Importance scores for predictors of water quality analytes, derived from the WQ predictor set. See the caption to Table 5 for further details of the conventions used to construct this table.

Predictor type	Predictor	days total (46.8%)	days trout (45.7%)	overall importance (21.4%)	ease of access (22.8%)	area fishable (28.4%)	close to home (37.1%))	scenic beauty (44.8%)	peace &solitude (40.3%)	catch rate (21.6%)	size of fish (39.4%)
base	Fishable length	**25.0**	**24.4**	10.3	0.3	14.5	5.8	2.1	2.3	11.3	4.1
	Catchment area	**34.1**	**34.3**	14.6	8.2	14.8	8.4	6.3	10.5	11.3	7.8
	Distance to sea	3.4	6.0	9.2	9.8	2.8	4.6	6.4	13.1	15.0	9.9
	Downstream elevation	8.7	8.4	13.4	10.1	4.7	7.9	3.4	11.2	17.6	13.4
	Mean flow	19.3	18.3	19.6	8.4	**26.1**	8.0	4.5	6.4	15.8	13.0
	Mean slope	12.6	12.4	7.9	8.2	8.5	9.1	11.7	9.5	9.2	8.9
	Catchment elevation	17.2	18.4	10.4	15.0	8.2	13.3	8.8	10.8	10.0	18.8
	Lake index	5.3	4.9	1.6	3.3	1.0	-0.3	7.7	5.0	1.1	2.0
water quality	CLAR (m)	11.2	12.7	8.3	5.0	7.8	8.1	7.5	10.9	10.1	14.9
	COND (µS/cm)	13.2	13.6	12.2	11.4	12.9	13.9	11.0	13.3	5.4	**20.5**
	DRP (mg/l)	11.3	11.9	10.0	11.7	6.5	9.5	13.1	12.0	8.7	11.0
	ECOLI (/ 100 ml)	13.6	13.4	8.9	11.0	8.9	12.6	13.4	15.1	11.4	11.1
	NH4N (mg/l)	9.7	10.8	12.3	8.9	6.9	15.2	17.8	16.1	11.7	15.7
	NO3N (mg/l)	17.3	16.0	6.9	13.5	9.3	12.2	12.8	12.4	7.3	7.5
	TN (mg/l)	14.4	14.7	8.8	13.5	8.1	15.5	18.1	17.3	10.5	8.4
	TP (mg/l)	11.9	12.0	8.4	9.0	7.9	10.3	14.4	10.4	6.4	15.2

The strongest models were for *total usage*, irrespective of whether or not this was adjusted for salmon angling (explained variance 45.7-50.2%), and for *scenic beauty* (scenic; 44.8-51.8%). Attribute scores for *distance from home* (close), *peace & solitude* (wild), *area fishable* (area), and *size of fish* (size) were moderately well predicted (explained variance 34.7-44.4% for seven of the eight models for these four variables). The weakest models were for *overall importance* (imp), *ease of access* (access), and *catch rate* (crate), with explained variance 21.4-32.6%. Models based on the REC predictor set generally outperformed those based on the WQ predictor set.

3.4.2 Predictor variables

General trends

For models based on the REC predictor set, the leading predictors were consistently related either to catchment geometry and topography, or catchment land cover (Table 5). Predictors related to climate and geology generally appeared either weakly (e.g., number of rain days > 10 mm per month), or not at all (% of catchment runoff from glacial and peat areas). Models for the same variables based on the WQ data set generally shared the same topography/ geometry predictors as for the corresponding REC predictor model, often in the same order of importance (Table 6). However, water quality predictors with $I_{score} \geq 10$ were important in all models, and featured strongly ($I_{score} \geq 15$) in the models for *total usage*, *scenic beauty*, *peace & solitude*, and *size of fish*. Two models (*days.trout*, *peace & solitude*) were noteworthy in that I_{score} exceeded 10 for all eight WQ predictor variables. One water quality variable (conductivity) appeared as a leading predictor

(I_{score} ≥ 10) in nine of the ten models, and was the leading predictor (I_{score} = 20.5) for *size of fish*. Measures of dissolved nitrogen (NH4N, NO3N, TN) also featured relatively strongly (I_{score} ≥ 15) in at least two models, and were the leading predictors for *scenic beauty* and *peace & solitude*.

In the following sections, I briefly review the results for each model. In all cases, readers are encouraged to cross-reference to the corresponding graphic in Appendix 2, particularly the response curves in the lower section of each panel.

Angling usage

Model fits for *days total* and *day .trout* were almost identical. The following discussion focuses on the model for *days trout*, which had higher I_{score} values than the corresponding model for *days total* despite being slightly weaker.

The leading predictor, by a wide margin, was catchment area (I_{score} = 33.8), followed by fishable length (I_{score} = 24.5) and flow (I_{score} = 18.2). All three of these variables are directly related to river size. However, the individual response curves (Figure A2.2) suggest that – in each case – usage increases very rapidly as a function of increasing river size only for relatively small streams, levelling off abruptly once a certain threshold is attained. Detailed inspection of the response curves (not shown) indicates that this corresponds approximately to a catchment area of 2,500 km^2, a fishable length of 50 km, and a mean flow of about 25 m^3s^{-1}. These criteria correspond almost exactly to the distinction between 5th and 6th order rivers, collectively representing 84 rivers of which 75 (89%) are of 6th or higher order.

The next three predictors (maximum and minimum catchment temperature, and catchment elevation), are much more weakly related to usage, but consistently suggest a tendency for usage to increase in response to decreasing temperature, and increasing elevation. The RF fitting process ensures that these responses take account of the responses to other predictors in the model, so – for example – the response curve for elevation indicates that given two rivers with similar values for catchment area, fishable length, and mean flow, usage will tend to increase slightly with increasing elevation.

Results for the WQ predictor set were similar, but included two measures of stream nitrogen (NO_3N and TN) as the 5th (I_{score} = 16.0) and 6th (I_{score} = 14.7) most important predictors. In both cases the response curve (Figure A2.12) suggested a slight tendency for usage to increase as nitrogen increased. This result appears paradoxical, but may simply reflect a tendency for streams with elevated nitrogen levels to be closer to population centres, and hence more readily available to anglers, than streams with lower nitrogen.

Overall importance

Models for *overall importance* for both predictor sets were weak and unconvincing. The model for the REC predictor set was notable in that I_{score} exceeded 10 for only three of the 28 available predictors, with a maximum score of 15.2 for mean flow. The corresponding model for the WQ predictor set included a higher number of predictors with I_{score} ≥ 10, but was the weakest of all the models in terms of the percentage of explained variance (21.4%).

Ease of access

Prediction accuracy for *ease of access* was fair (explained variance 32.6%) when modelled using the REC predictor set, but weak (22.8%) when modelled using the WQ predictor set. The REC-based model included only three predictors with I_{score} ≥ 10, being positively related to the percentage of the catchment under heavy pastoral cover (I_{score} = 19.8), and negatively related to catchment elevation (I_{score} = 15.0). As with the results for angling usage in relation to river size, the shape of the response curve for *ease of access* in relation to pastoral cover suggests a very rapid increase as this percentage rises from 0% to about 10%, and very little effect thereafter (Figure A2.4). The corresponding curve for catchment elevation suggests a more continuous response, levelling off only when mean elevation exceeds 1000 m.

Area of fishable water

Prediction accuracy for *area of fishable water* was fair (explained variance 34.7%) using the REC predictor set, but somewhat weaker (28.4%) using the WQ predictor set. The REC-based model was essentially a weakened version of the *days.trout* model for the same data set, being driven almost completely by the same three leading predictors albeit in a different order (mean flow: I_{score} = 20.0; fishable length: I_{score} = 15.9; catchment area: I_{score} = 14.7). The response curves (Figure A2.5) were also similar to those for the corresponding *days.trout* model, suggesting that (as noted in Section 2.4.1) it may be more informative to treat *area of fishable water* as a predictor rather than a variable to be modelled.

Close to home

Prediction accuracy for *close to home* was relatively high (explained variance 37.1-44.4%) for both predictor sets. The leading two predictors for the REC-based model were catchment elevation (I_{score} = 18.0), and the percentage of the catchment under heavy pastoral cover (I_{score} = 17.4). The response curves for both variables (Figure A2.6) have a natural interpretation in terms of population demographics, and probably mean nothing more than that most anglers live in low-altitude urban centres, surrounded by pastoral farmland (c.f. Unwin & Deans 2003). However, this result is still encouraging, in that the very weak presence of predictors more directly related to in-river characteristics (e.g., fishable length, mean flow, lake index) confirms that the RF fitting process is effective in identifying and discarding those which *a priori* considerations suggest are unlikely to be meaningful.

Leading predictors for the WQ data set were dominated by WQ variables, led by TN (I_{score} = 15.5), and NH_4N (I_{score} = 15.2). Catchment elevation was less important than for the REC-based model (I_{score} = 13.3), although the shape of the response curve (Figure A2.16) was almost identical. The extent of heavy pastoral cover has previously been identified as the leading predictor of TN (Unwin *et al.* 2010), so this result also provides evidence that the RF fitting process yields credible results.

Scenic beauty

Scenic beauty was well modelled (explained variance 51.8%) by the REC predictor set, and was notable for being the only attribute with three leading predictors related solely to catchment land cover. These showed a consistent tendency for perceived scenic beauty to increase as the percentage of indigenous forest cover increased (I_{score} = 20.5), and to decrease as the percentage of heavy pastoral (I_{score} = 19.4) and exotic forest (I_{score} = 15.5) increased. The corresponding response curves (Figure A2.7) suggest contrasting patterns as land cover became increasingly modified, with a more or less linear increase in perceived scenic beauty with increasing indigenous forest cover, but a much more rapid decrease for even a relatively modest increase (e.g., from 0-5%) in heavy pasture or exotic forest cover. Other leading predictors included catchment elevation, mean number of rain days per month, and slope. A striking feature of the model was the almost complete absence of any relationship between *scenic beauty* and the extent of light pastoral cover, suggesting that the underlying drivers are related to the intensity of pastoral grazing rather than merely to its presence or absence.

Scenic beauty was also reasonably well modelled (explained variance 44.8%) by the WQ data set, with the striking difference that all six leading predictors were related to water quality rather than topography or location. Responses for the leading two predictors (TN and NH_4N; Figure A2.17) suggested a strong tendency for perceived scenic beauty to decline rapidly over a relatively restricted portion of the observed range for these two variables. However, response curves for the remaining four WQ predictors (TP, ECOLI, DRP, and NO_3N) are much less consistent, and suggest no clear underlying trend.

Given the strong association between TN and increasing heavy pastoral cover (Unwin *et al.* 2010), these results may amount to no more than a restatement of the inverse relationship between scenic beauty and pastoral cover apparent in the corresponding RF model for the REC predictor set. In particular, it is highly likely that the relationship between perceived scenic beauty and catchment-scale variables is more directly expressed in terms of land cover than in terms of water quality.

Peace & solitude

Peace & solitude was less well modelled than *scenic beauty*, with explained variance (40.3%-43.1%) similar to the result for *close to home*. For the REC predictor set, three of the six leading predictors were related to catchment topography, in ways that suggested a strong tendency for peace & solitude to increase with catchment altitude and increasing distance inland. The other three predictors, including the first, were the percentage of heavy pastoral landcover (I_{score} = 16.5), exotic forest landcover (I_{score} = 13.2), and indigenous forest landcover (I_{score} = 12.76). Even more strongly than for *scenic beauty*, response curves for the landcover predictors (Figure A2.8) suggested a tendency for *peace & solitude* to decrease very rapidly in catchments with even a small percentage (< 5%) of pastoral or exotic forest landcover. Effects associated with the extent of light pastoral landcover were conspicuous by their absence.

Results for the WQ predictor set were also similar to the corresponding results for scenic beauty, with the leading two predictors being TN (I_{score} = 17.3) and NH4N (I_{score} = 16.1). Again, the response curve for TN suggested a very strong negative response as soon as TN increased above baseline levels (Figure A2.18), followed by a much more muted decline thereafter. The most novel feature of this result was the response curve for conductivity, which suggested a step response, with peace & solitude scores declining abruptly, from about 100-120 µS cm^{-1}.

Catch rate

Catch rate was the most poorly predicted of any of the 1979/81 survey attributes, with explained variance 24.9% for the REC predictor set, and 21.6% for the WQ predictor set. Both models were dominated by variables from the base set, with downstream elevation, distance to sea, flow, consistently making up the leading three predictors.

Size of fish

Prediction accuracy for *size of fish* was above average (explained variance 39.4-42.4%) for both predictor sets, although well below that for the most successful models. Both predictor sets suggested a strong tendency for *size of fish* to increase with altitude, with catchment elevation either the most important (I_{score} = 20.4) or second most important (I_{score} = 18.8) predictor, and downstream elevation also consistently among the top six predictors. Response curves for these two variables (Figures A2.10, A2.20) were essentially identical across both models. For the REC predictor set, the remaining predictors were predominantly related to climate, although their importance scores are relatively low (I_{score} ≤ 11.7) and the corresponding response curves tend to be flat. The WQ-based model is notably mainly for the appearance of conductivity as the leading predictor (I_{score} = 20.5), with evidence of a well-defined threshold decrease in fish size as conductivity increases from 50 to 80 µS cm^{-1}. This result is consistent with a similar, although weaker, result for NH4N, but response curves for the remaining two WQ predictors (TP and black disc clarity) are much less clear cut. In particular, the response for clarity shows some evidence of an initial decline in fish size with increasing clarity (i.e., increasing black disc visibility), a condition which would normally be associated with more rapid growth (Hayes *et al.* 2000), and hence larger fish. However, fish size (as indexed by the 1979/81 survey) is not well-defined, as noted in Table 1, and this result could potentially be linked to confounding factors such as the increased difficulty of catching trout in clear water.

4 Discussion

The REC appears to have considerable potential to add value to existing FGNZ angler survey databases. Both lines of inquiry explored in this study, analysing long term usage trends in relation to REC classes, and developing predictive models of angler usage metrics, yield results which appear to be consistent and meaningful, and are relevant to FGNZ management objectives. Several obvious difficulties remain, but the results to date are encouraging. Many of the RF modelling results are suggestive rather than conclusive, but should – at the very least – serve as useful pointers when considering future priorities.

4.1 River fishery classification

The primary virtue of using REC-based measures to classify individual river fisheries is that the resulting groupings transcend regional boundaries and hence provide a more coherent national perspective. Two-factor classes (e.g., climate + source of flow, climate + landcover) appear to be most consistent with the underlying spatial scales which characterise recreational angling, although further subdivision is possible. For example, in the present study I simply grouped the REC landcover designations into two groups, representing natural and modified vegetation, ignoring finer distinctions such as between tussock and indigenous forest. The classes explored in this study appear to be broadly compatible with the previous *ad hoc* groupings (i.e., headwater, back country etc.) developed from the 1994/95 survey, but are more objectively based and are likely to be easier to defend when used in an advocacy context.

A limitation of the approach developed here is that assigning each river fishery to a single REC class becomes an increasingly blunt instrument as stream order increases, and sub-catchments become increasingly diverse. The most meaningful classifications are likely to be those for catchments of intermediate (e.g., 4^{th} – 6^{th}) order, which are large enough to sustain a viable fishery but small enough to be considered essentially homogenous. Rivers draining higher order catchments, by contrast, are more likely to vary in character over spatial scales large enough to be relevant to anglers considering where to direct their effort, and may be more appropriately subdivided into discrete reaches. The most recent angler survey data allow for this possibility only on some large mainstem rivers. Current survey techniques allow for rivers to be subdivided as necessary, but only at the expense of adding further complexity to the telephone interview processes used to collect the raw sample data.

4.2 Trends in usage

Trends in annual usage of river fisheries from 1994/95 to 2007/08 are strongly aligned to climate and source of flow class, with clear evidence of a consistent decline in usage of lowland rivers. This finding is not new, but the evidence presented here is perhaps more objective than equivalent results derived from earlier studies (e.g., Jellyman *et al.* 2003), and will also allow the true strength of any underlying long-term trends to be monitored into the future as data from future surveys come to hand.

The decline in angler usage of lowland river fisheries is almost entirely confined to rivers in the CD/L (cold dry / lowland) and CW/L (cold wet / lowland) REC classes. Geographically, most of these rivers fall into one of four main clusters (Figure 10), representing rivers on the southern Taranaki coast; southern Hawkes Bay, Wairarapa, and Manawatu; coastal Canterbury; and south Otago / Southland. Most of these areas are characterised either by

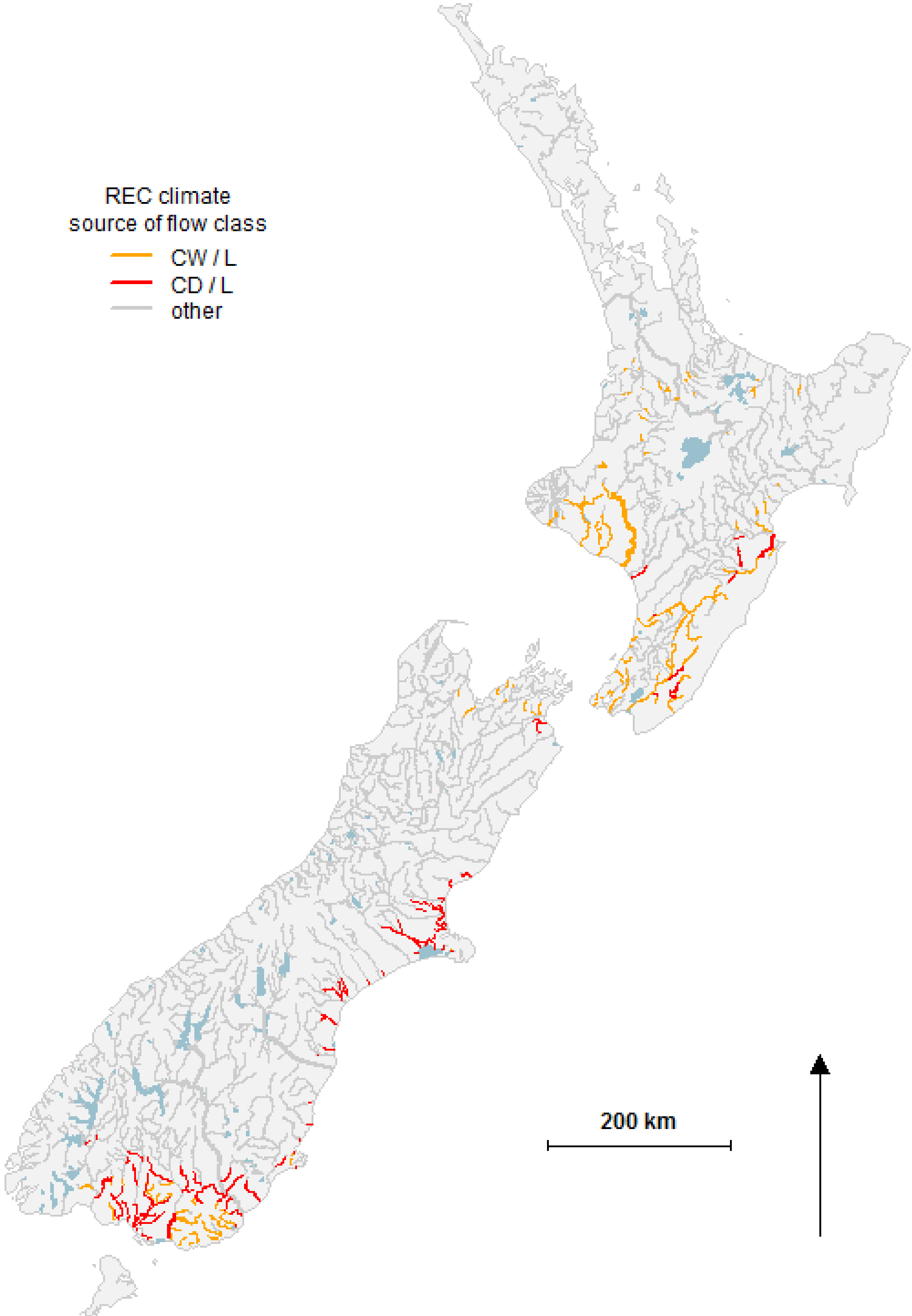


Figure 10: FGNZ angling rivers in the CW/L (cold wet / lowland) and CD/L (cold dry / lowland) REC classes, the two classes where angler usage has declined most strongly since 1994/95.

high demand for surface waters or intensive pastoral agriculture, and in some regions (e.g., coastal Canterbury, Hawkes Bay) both of these circumstances apply. The long-term trend may be partially confounded by the appearance of the invasive diatom *Didymosphenia geminata* in some rivers since 2004 (Kilroy & Unwin 2011), although its effect has generally been more strongly felt in rivers such as the Waiau/Mararoa, and the lower Waitaki River, both of which are lake-fed. In addition, the decline is also clearly evident between 1994/95 and 2001/02, three years before *D. geminata* was first recorded in New Zealand.

4.3 Random forest models

Results for predictive models of total annual usage, and attribute scores from the 1979/81 National Angler Survey, were weaker than for the MfE water quality models which were the prototype for the present study (Unwin *et al.* 2010). The strongest models were for total annual usage, irrespective of whether or not this included salmon angling, scenic beauty, and feelings of peace & solitude. Of the two attributes (catch rate, size of fish) directly related to trout populations, only size of fish yielded a credible model. Overall importance scores for each river, essentially a single index measuring their angling appeal, were very poorly predicted.

Total usage was essentially a function of river size: larger rivers, in larger catchments, are fished more heavily than smaller rivers in small catchments. This result may appear unremarkable, but is noteworthy for the relative weakness of predictors related to catchment land cover, which might reasonably be expected to influence angler's decisions on where to fish. It is possible that a more detailed analysis of the RF models, taking into account interactions between predictors such as land cover and topography, may yield further insight into the underlying mechanisms, but such analyses are beyond the scope of this study. In particular, the tendency for usage to increase slightly in streams with elevated nitrogen levels may be driven as much be population demographics as by water quality. A possible addition to the predictor set would be catchment-scale measures of population density, so as to include angler demographics in a natural way. The only such measure currently available is the travel distance index for each river developed from the 2001/02 survey responses (Unwin & Deans 2003), but this is unsuitable as a predictor of usage because it is derived from the same underlying data set.

An important caveat regarding the parallel analyses for the REC and WQ predictor sets is that the WQ predictor set is essentially a subset of the REC predictor set, and provides no new information. In essence, the WQ predictor set is merely a compressed copy of the REC predictor set, somewhat analogous to the results of a principal components analysis, whereby a large and unwieldy data set is replaced by a smaller data set which is optimised so as to most efficiently capture the relationship between the REC predictors and water quality. This interpretation is consistent with the general tendency for the REC-based models to outperform their WQ-based counterparts, simply by virtue of including multiple low-order predictors to incrementally improve the model fit. The WQ model for *size of fish* offers some encouragement that it may ultimately be possible to demonstrate credible relationships between water quality and fish size as indexed by angler survey data, but falls well short of establishing that any such relationship actually exists.

4.4 Future directions

On the strength of the results reported here, I believe the REC offers FGNZ a powerful set of tools for informing management of the freshwater angling resource. These tools are freely available, and are already widely used by agencies such as regional councils and DoC. FGNZ's angler survey databases are gaining increasing recognition as a source of objective, long-term data on usage of New Zealand rivers (Booth *et al.* 2009, Unwin 2011), and this reputation can only be strengthened by explicitly linking these databases to the REC. Some suggestions for furthering this work, and identifying future priorities, are as follows.

4.4.1 Rivers database

The pivotal data set in FGNZ's angler survey databases lists all waterbodies, including lakes as well as rivers, fished by FGNZ licence holders. This data set currently includes 1,226 named waterbodies, comprising 978 rivers and 248 lakes. Attributes associated with each river include up to four pairs of coordinates (downstream reach, upstream reach, upstream angling limit, angling centroid), and the corresponding NZReach for each point. These reach identifiers are the basis for all subsequent REC linkages, and for derived attributes such as total river length and fishable length.

The current database has grown incrementally since it was first developed for the 1979/81 survey, and – while it has been checked to the extent possible – is neither complete nor 100% error-free. It is therefore timely to review the entire database, on a region by region basis, both to correct any errors or omissions, and to consider additional variables which could usefully be associated with each river.

At the time of writing, the REC is being reformulated so as to correct some long-standing anomalies (such as poor representation of lakes) in the original version. In addition, the LCDB2 database has recently been updated, based on satellite images taken in 2008, and these data are now available as LCDB3. Once the REC update is complete, it will be necessary to remap the rivers database onto the new REC network, and the corresponding LCDB3 records. Much of this can be done automatically, but it is likely that some manual checking will be required.

4.4.2 Usage databases

The 1979/81 angler survey database is now over 30 years old, and therefore likely to be dated with respect to at least some of the attributes associated with each river fishery. Recognising this, FGNZ has already developed plans to update the survey (Unwin 2009b), and this is likely to proceed within the next 1-2 years. If so, an important consideration when developing a sampling frame for each region will be managing the trade-off between increasing sample size (and hence data capture rate for individual rivers) while minimising resource costs.

Similar considerations apply to what will become the fourth survey conducted since 1994/95, in or about 2014. Many of the recommendations which followed the 2007/08 survey, particularly those relating to the need for thorough manual checking of any potentially ambiguous river names by regional FGNZ staff (Unwin 2009a), apply with even greater force if the data are to be reliably matched against the REC. At present, this match depends entirely on using the river name, as recorded by the survey interview, to locate each river in REC space. Several regions have identified what are almost certainly errors in usage summary statistics caused by misidentified rivers. These are rarely if ever significant in terms of national or regional totals, but are nevertheless highly undesirable as they undermine confidence in the survey integrity.

5 Acknowledgements

Thanks to Maurice Rodway, Phil Jellyman, and Ton Snelder for constructive reviews of an earlier draft of this report. This study was funded by Fish & Game New Zealand.

6 References

Booth, K.; Deans, N.; Unwin, M.J.; Baker, M.-A. (2009). Salmonid angling: application of the significance assessment method. 8 p.

Breiman, L. (2001). Random Forests. *Machine Learning 45*: 5-32.

Cutler, D.R.; Edwards, T.C., Jr.; Beard, K.H.; Cutler, A.; Hess, K.T.; Gibson, J.; Lawler, J.J. (2007). Random forests for classification in ecology. *Ecology 88*: 2783–2792.

Hayes, J.W.; Stark, J.D.; Shearer, K.A. (2000). Development and test of a whole-lifetime foraging and bioenergetics growth model for drift-feeding brown trout. *Transactions of the American Fisheries Society 129*: 315-332.

Jellyman, D.J.; Unwin, M.J.; James, G.D. (2003). Anglers' perceptions of the status of New Zealand lowland rivers and their trout fisheries. *NIWA Technical Report 122.* 61 p.

Kilroy, C.; Unwin, M.J. (2011). The arrival and spread of the bloom-forming, freshwater diatom, *Didymosphenia geminata*, in New Zealand. *Aquatic Invasions 6*: 349–362.

McDowall, R.M. (1994). Gamekeepers for the nation. Canterbury University Press, Christchurch.

R Development Core Team. (2010). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Snelder, T.; Hurren, H.; Dey, K. (2005). Conversion of land-cover categories of the River Environment Classification to Land-Cover Database Version 2. *NIWA Client Report CHC2005-065.* 7 p.

Snelder, T.H.; Biggs, B.J.F. (2002). Multi-scale river environment classification for water resources management. *Journal of the American Water Resources Association 38*: 1225-1240.

Stankey, G.; Wood, J. (1982). The Recreation Opportunity Spectrum: an introduction. *Australian Parks and Recreation 22*: 5-13.

Teirney, L.D.; Richardson, J. (1992). Attributes that characterize angling rivers of importance in New Zealand, based on angler use and perceptions. *North American Journal of Fisheries Management 12*: 693-702.

Teirney, L.D.; Unwin, M.J.; Rowe, D.K.; McDowall, R.M.; Graynoth, E. (1982). Submission on the draft inventory of wild and scenic rivers of national importance. *Fisheries Environmental Report 28.* 122 p.

Unwin, M.J. (2009a). Angler usage of lake and river fisheries managed by Fish & Game New Zealand: results from the 2007/08 National Angling Survey. *NIWA Client Report CHC2009-046.* 48 p.

Unwin, M.J. (2009b). Attributes characterising river fisheries managed by Fish & Game New Zealand: a pilot survey of the Otago and Nelson/Marlborough regions. *NIWA Client Report CHC2009-090.* 37 p.

Unwin, M.J. (2009c). Who fishes where and why. *Fish & Game New Zealand 65*: 26-29.

Unwin, M.J. (2011). Sustainability Performance Criteria for River Mouths: Freshwater Angling. Presented to Waterways Centre for Freshwater Management Seminar, Canterbury University, Christchurch, 29 November 2011.

Unwin, M.J.; Brown, S. (1998). The geography of freshwater angling in New Zealand: A summary of results from the 1994/96 National Angling Survey. *NIWA Client Report CHC98/33.* 78 p.

Unwin, M.J.; Deans, N. (2003). Travel distance as an index of angling value: a preliminary study based on the 2001/02 National Angling Survey. *NIWA Client Report CHC2003-113.* 24 p.

Unwin, M.J.; Image, K. (2003). Angler usage of lake and river fisheries managed by Fish & Game New Zealand: results from the 2001/02 National Angling Survey. *NIWA Client Report CHC2003-114.* 48 p.

Unwin, M.J.; Snelder, T.; Booker, D.; Ballantine, D.; Lessard, J. (2010). Predicting water quality in New Zealand rivers from catchment-scale physical, hydrological and land cover descriptors using random forest models. *NIWA Client Report CHC2010-037.* 21 p.

Wild, M.; Snelder, T.; Leathwick, J.; Shankar, U.; Hurren, H. (2005). Environmental variables for the freshwater environments of New Zealand River Classification. *NIWA Client Report CHC2004-086.* 31 p.

Woods, R.; Hendrikx, J.; Henderson, R.; Tait, A. (2006). Estimating mean flow of New Zealand rivers. *Journal of Hydrology (New Zealand) 45*: 95-110.

Appendix A Predictor variables available for modelling angling usage metrics as a function of local- and catchment-scale descriptors.

Type	Description	Runoff weighted	Units	Name	Minimum / median / maximum §
Geography/ topography/ flow (Source: REC)	Total catchment area	no	m^2	Catchment area	2 / 180 / 20606
	Downstream distance to sea	no	km	Distance to sea	0.1 / 53.3 / 328
	Downstream elevation	no	m	Downstream elevation	0 / 80 / 880
	Mean discharge at site	no	$m^3 s^{-1}$	Mean flow	0 / 6 / 646
	Fishable length	no	km	Fishable length	0.5 / 16.3 / 288
	Mean catchment elevation	no	m	Catchment elevation	7 / 506 / 1569
	Mean catchment slope	yes	degrees	Mean slope	0.1 / 17.2 / 38.6
	Lake index	no	-	Lake index	0 / 0 / 0.26
Climate (Source: REC / FWENZ)	Mean minimum July air temperature	yes	°C × 10	Min temperature	-47.1 / 1.5 / 73.7
	Mean maximum January air temperature	yes	°C × 10	Max temperature	74.8 / 138.6 / 191.6
	CV of annual catchment rainfall	yes	mm	Rain variability	133 / 165 / 257
	Catchment rain days > 10 mm / month	Yes	days yr^{-1}	Rain days > 10	1 / 3.8 / 9.1
	Catchment rain days > 50 mm / month	Yes	days yr^{-1}	Rain days > 50	0 / 0.2 / 1.3
	Catchment rain days > 200 mm / month	Yes	days yr^{-1}	Rain days > 200	0 / 0 / 0.1
	Annual potential evapotranspiration	Yes	mm	Evapotranspiration	271 / 868 / 1321
Geology (Source: REC / FWENZ)	% of runoff from LRI category alluvium	Yes	% / 100	%alluvium	0 / 0.09 / 1
	% of runoff from LRI category glacial	Yes	% / 100	%glacial	0 / 0 / 0.35
	% of runoff from LRI category peat	Yes	% / 100	%peat	0 / 0 / 0.4
	Catchment average of calcium	Yes	Ordinal	Calcium	0.61 / 1.44 / 3.15
	Catchment average of hardness	Yes	Ordinal	Hardness	0.99 / 3.23 / 4.94

Type	Description	Runoff weighted	Units	Name	Minimum / median / maximum §
	Catchment average of particle size	Yes	Ordinal	Particle size	1 / 3.31 / 4.97
	Catchment average of phosphorous	Yes	Ordinal	Phosphorous	0.93 / 2.71 / 5
Landcover (Source: LCDB2)	% of catchment, LCDB2 bare ground	No	%	%bare	0 / 0.5 / 66.5
	% of catchment, LCDB2 exotic forest	No	%	%exotic forest	0 / 1.9 / 94.2
	% of catchment, LCDB2 indigenous forest	no	%	%indigenous forest	0 / 31.1 / 100
	% of catchment, LCDB2 heavy pastoral	no	%	%pastoral heavy	0 / 18.8 / 98.7
	% of catchment, LCDB2 light pastoral	no	%	%pastoral light	0 / 4.2 / 90.1
	% of catchment, LCDB2 scrub	no	%	%scrub	0 / 2.5 / 37.4
Water quality (Source: NIWA/ MfE)	Black disc visibility (clarity)	no	m	CLAR	0.3 / 1.9 / 7.4
	Electrical conductivity	no	µS/cm	COND	11.4 / 81.8 / 442.9
	Escherichia coli	no	/100 mL	ECOLI	2 / 53 / 742
	Dissolved reactive phosphorus	no	mg/l	DRP	0.001 / 0.007 / 0.084
	Ammoniacal nitrogen	no	mg/l	NH4N	0.002 / 0.007 / 0.145
	Oxidised nitrogen	no	mg/l	NO3N	0.01 / 0.102 / 3.095
	Total nitrogen	no	mg/l	TN	0.05 / 0.247 / 3.488
	Total phosphorus	no	mg/l	TP	0.004 / 0.015 / 0.16

§ Summary statistics for each variable are based on the set of 632 rivers for which usage estimates for the 2007/2008 angling season were available.

Appendix B Graphical summaries of Random Forest (RF) models for the ten angling usage metrics considered in this report

Results for each usage metric are represented by a full page panel showing eight diagnostic statistics for the corresponding RF model. These plots are as follows, from top left:

1. Observed vs. predicted values for all sites, using a “jack knife” procedure whereby the prediction error for each observation is derived by successively fitting RF models for the full data set minus the observation of interest, and then comparing the model prediction for that point with the observed value. Both axes are plotted to the same scale, with the diagonal dashed line representing agreement between observation and prediction. The number of observations and the nominal r^2 are also shown. These values are close to, but not necessarily identical to, the percentage of explained variance for each model as listed in Tables 5-6.

2. Normal Q-Q (quantile) plot, contrasting the observed distribution of residuals for the fitted data (Sample Quantiles) to the theoretical distribution if the residuals were distributed normally (Theoretical Quantiles, diagonal line). The best fitting models perform well over the entire observed data range, but weaker models are characterised by large residuals for the most extreme values, indicating a general tendency to overestimate low values and underestimate high values.

3. Smoothed partial plots (using the default “3RS3R” algorithm as implemented in the smooth() function of R Version 2.12.1) for the six most important predictors in each model indicating the modelled response of the dependent variable to each predictor, plotted to a common vertical scale. The “rug” at the bottom of each plot represents the distribution of each predictor variable. Additional insight into the influence of each predictor can be gained by comparing the vertical response range for each plot with the vertical scale on the plot of observed vs. predicted values at top left.

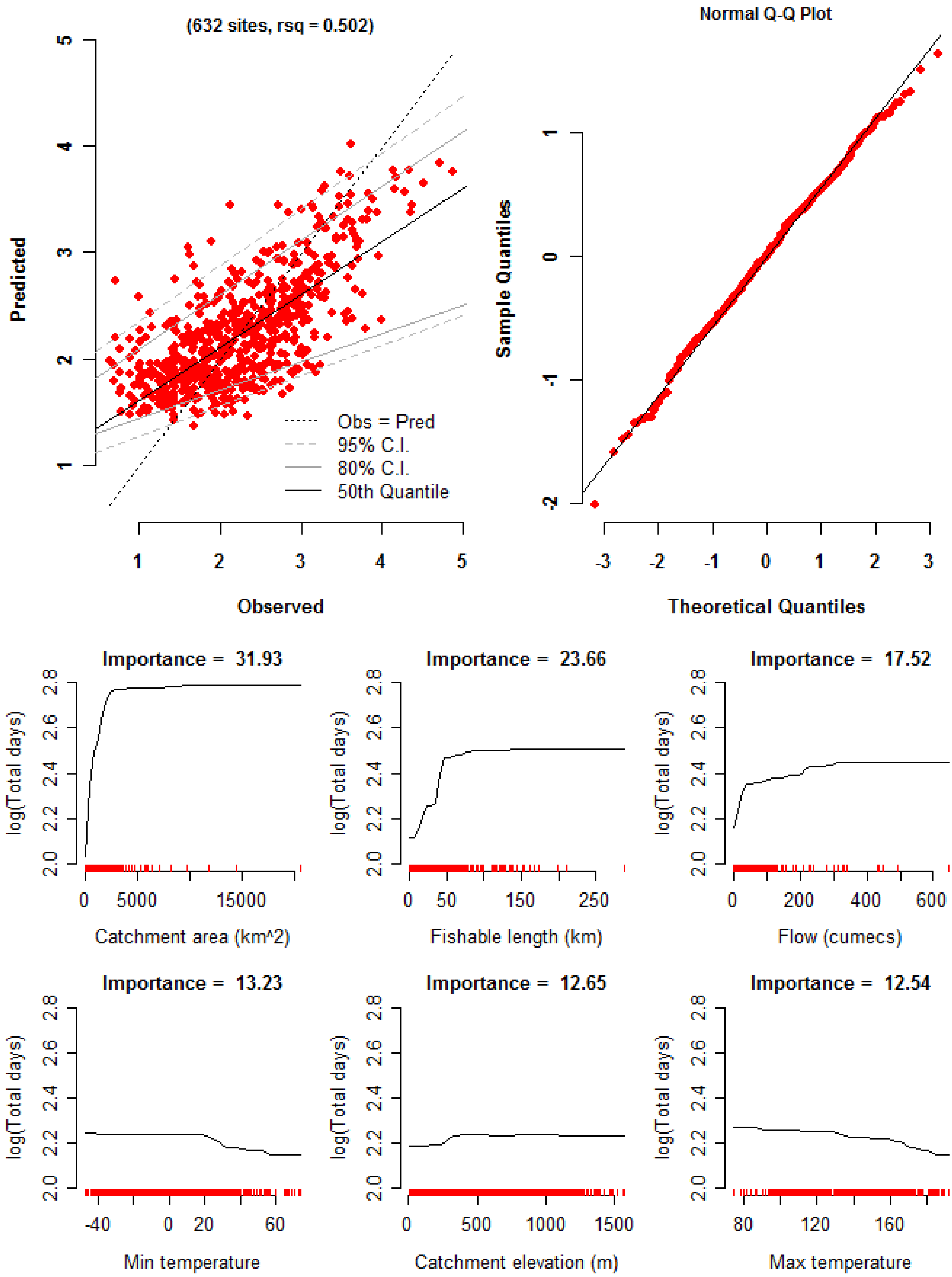


Figure B-1. Diagnostic plots for a RF model of annual angling usage (angler-days, log transformed), based on the REC predictor set. Note that horizontal axes for flow-weighted variables (e.g., minimum temperature) in all appendix plots are scaled in flow-weighted rather than raw units.

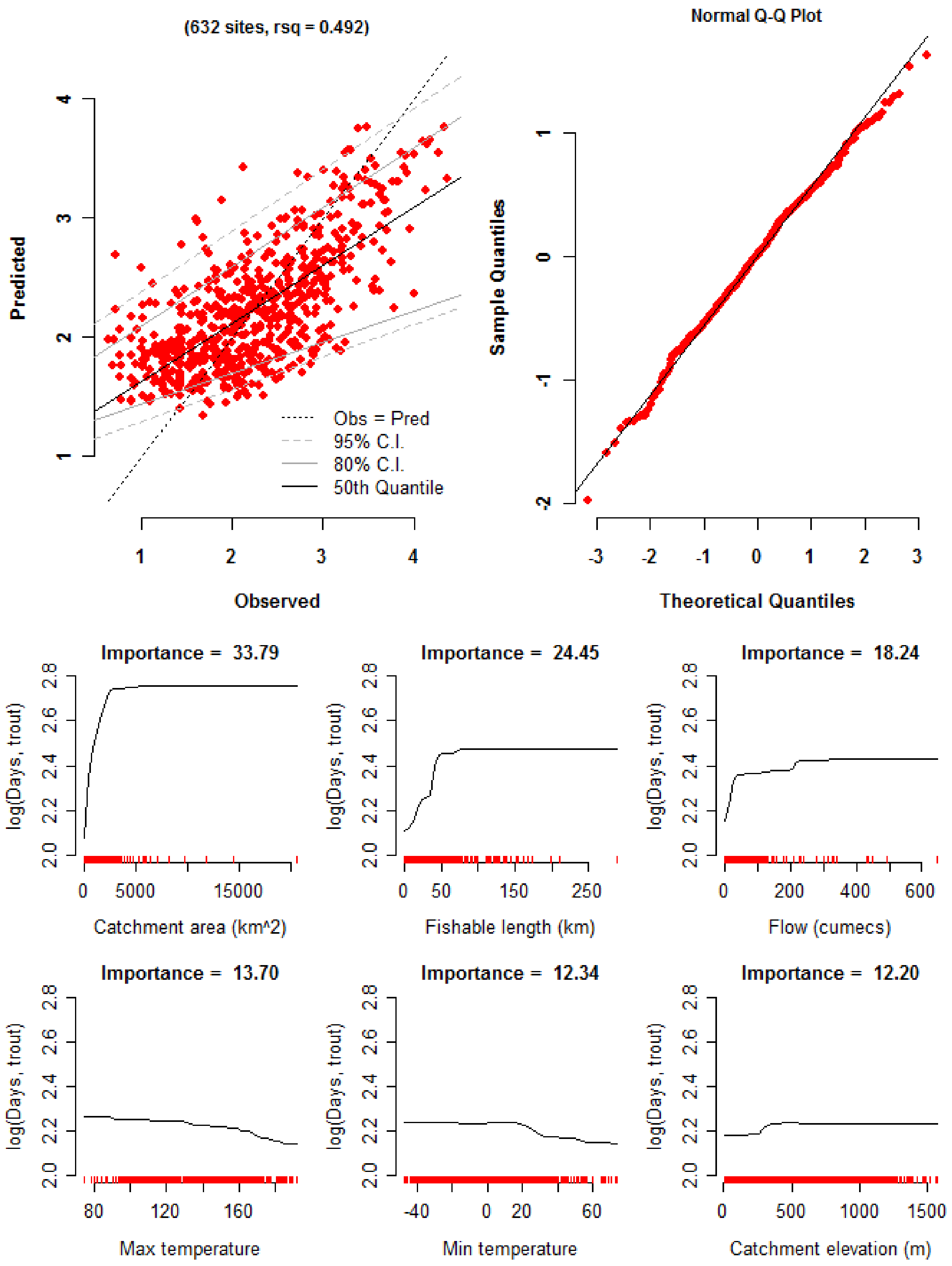


Figure B-2. Diagnostic plots for a RF model of annual angling usage by trout anglers (angler-days, log transformed), based on the REC predictor set.

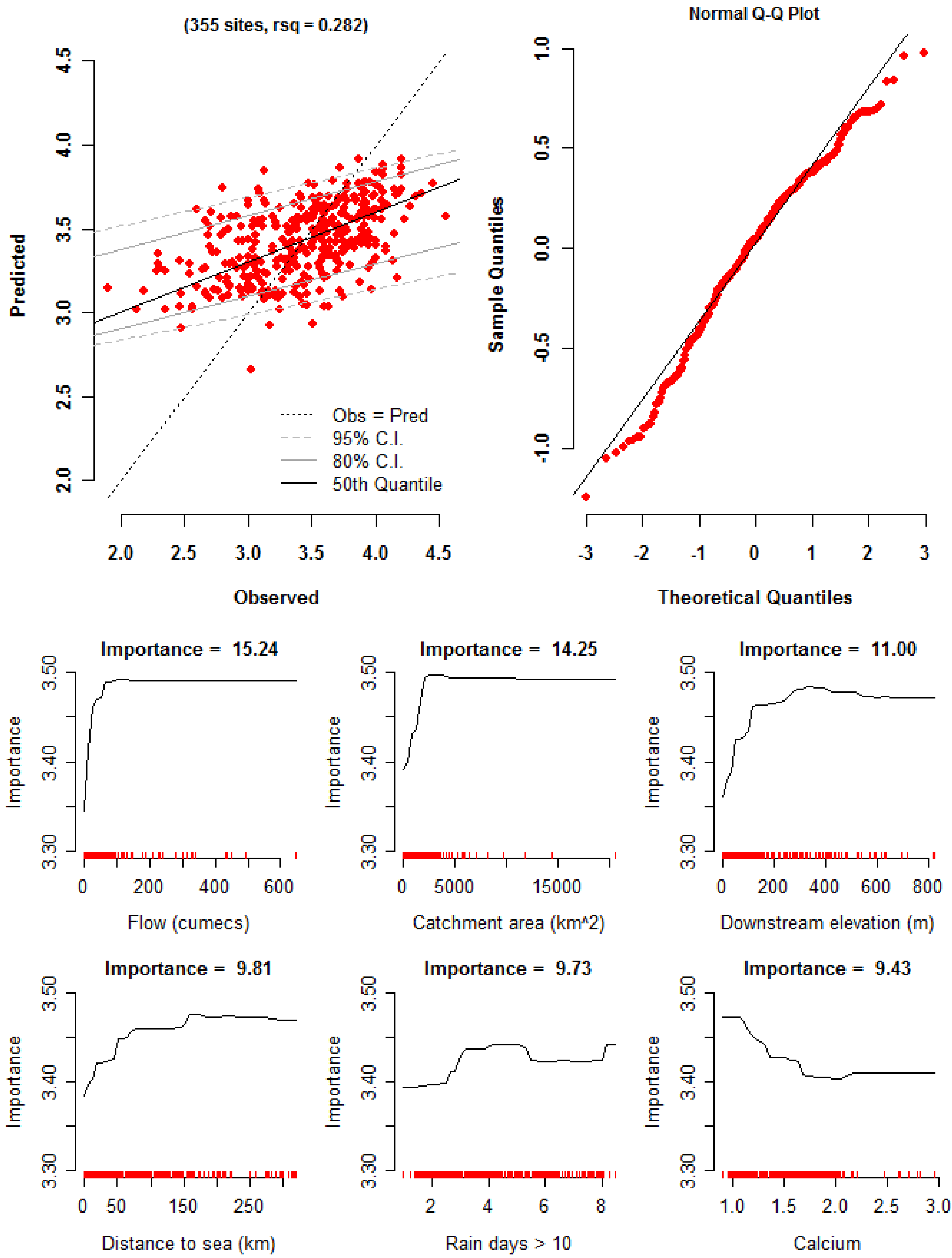


Figure B-3. Diagnostic plots for a RF model of overall importance scores (derived from the 1979/81 National Angling Survey), based on the REC predictor set.

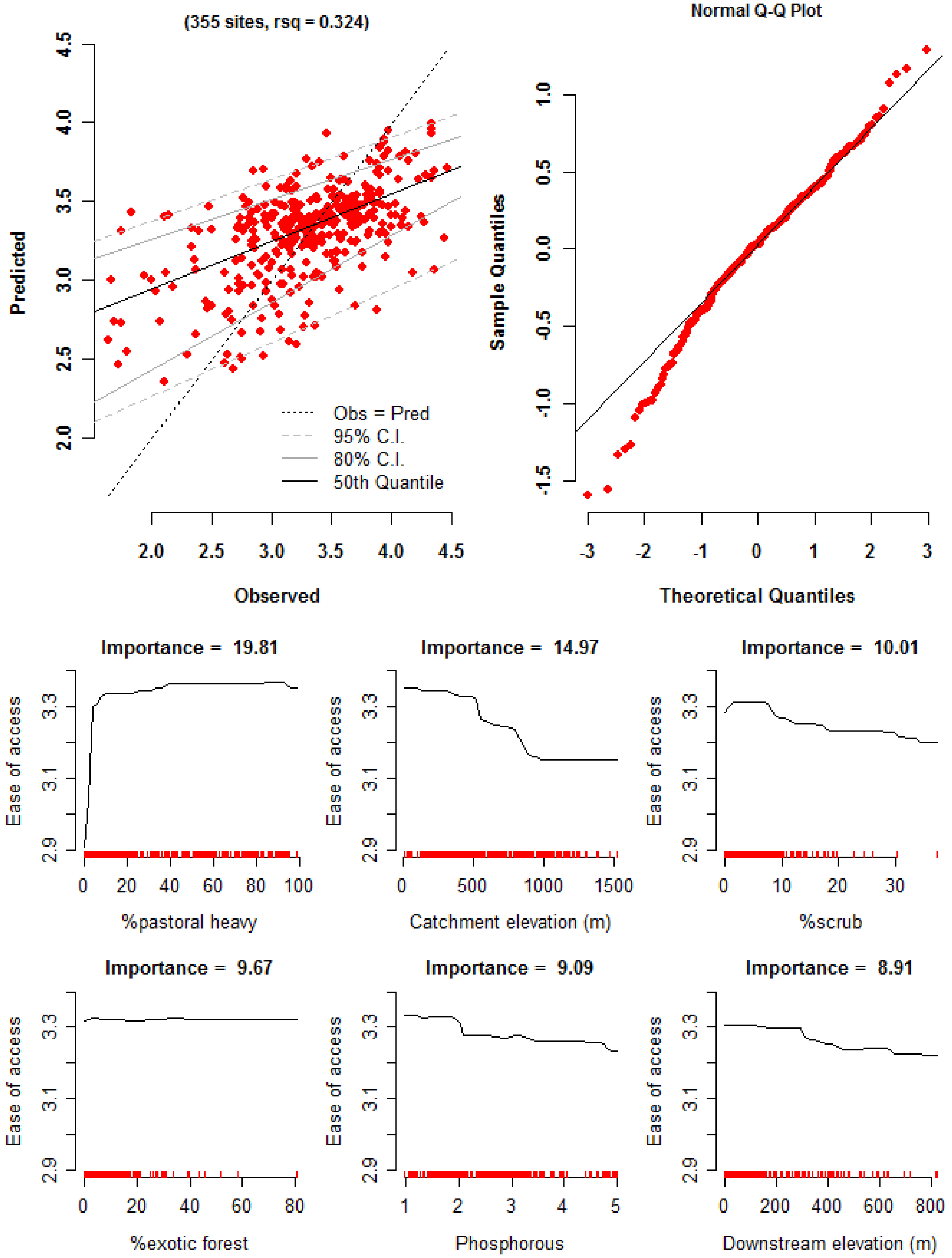


Figure B-4. Diagnostic plots for a RF model of ease of access scores (derived from the 1979/81 National Angling Survey), based on the REC predictor set.

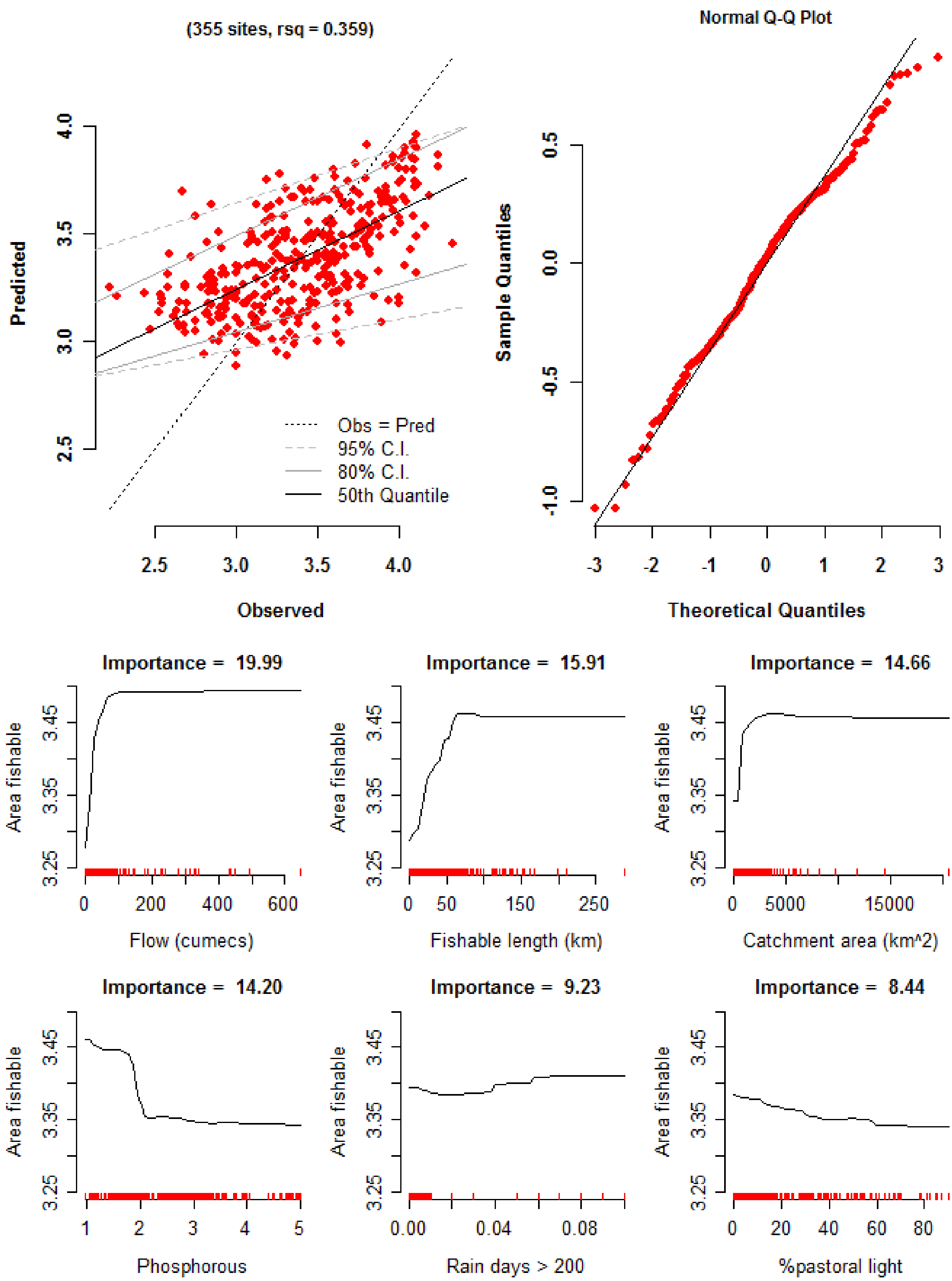


Figure B-5. Diagnostic plots for a RF model of area fishable scores (derived from the 1979/81 National Angling Survey), based on the REC predictor set.

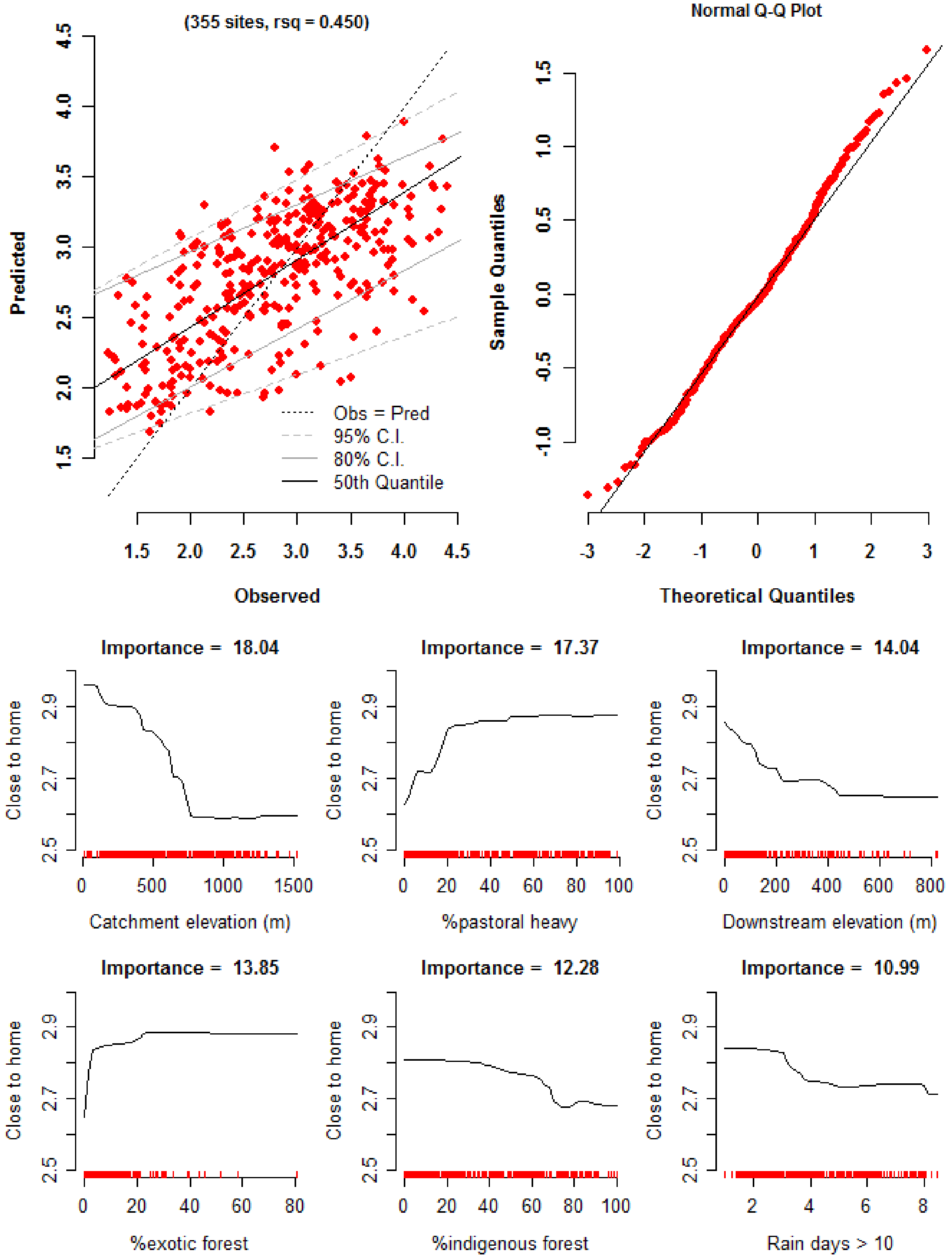


Figure B-.6. Diagnostic plots for a RF model of close to home scores (derived from the 1979/81 National Angling Survey), based on the REC predictor set.

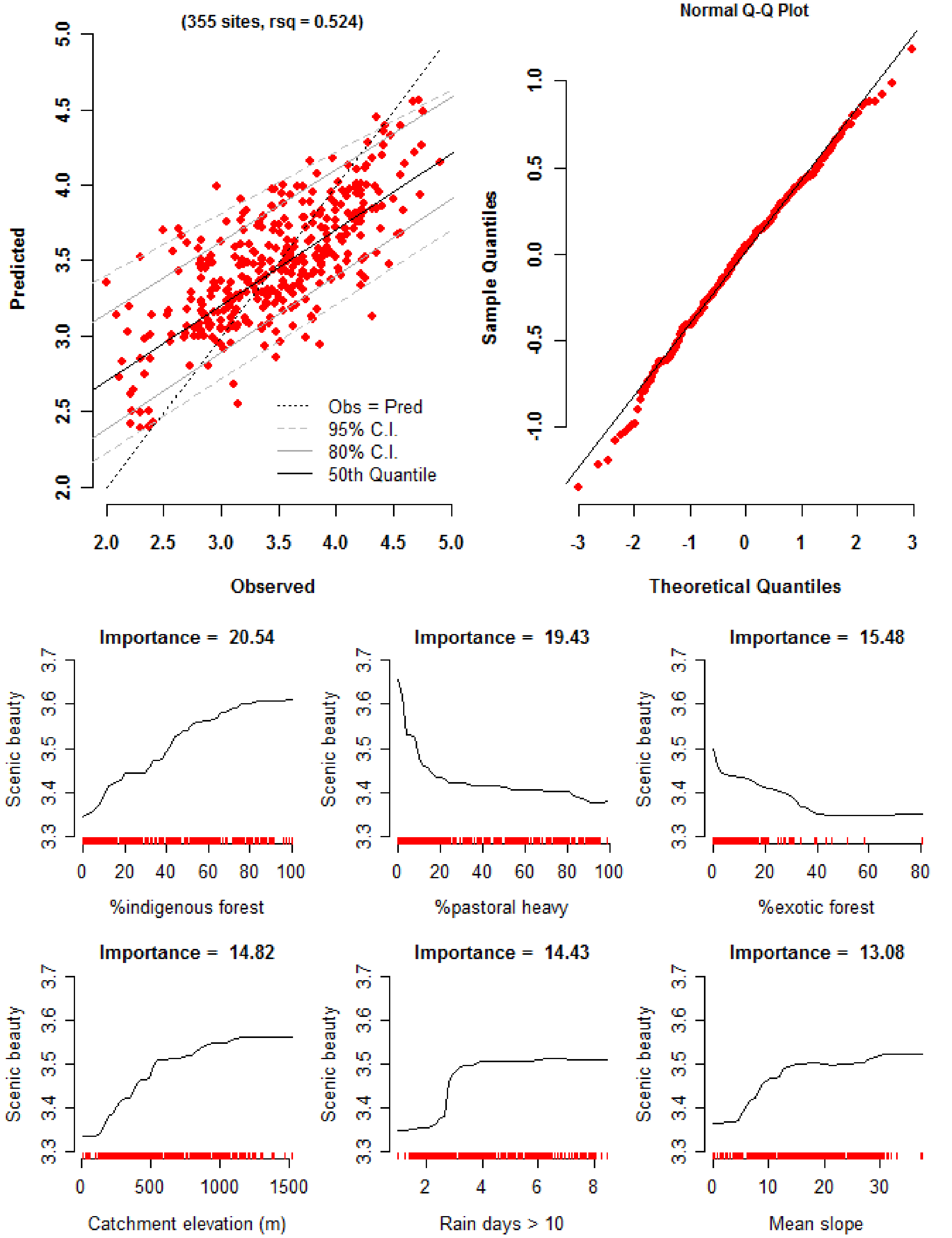


Figure B-7. Diagnostic plots for a RF model of scenic beauty scores (derived from the 1979/81 National Angling Survey), based on the REC predictor set.

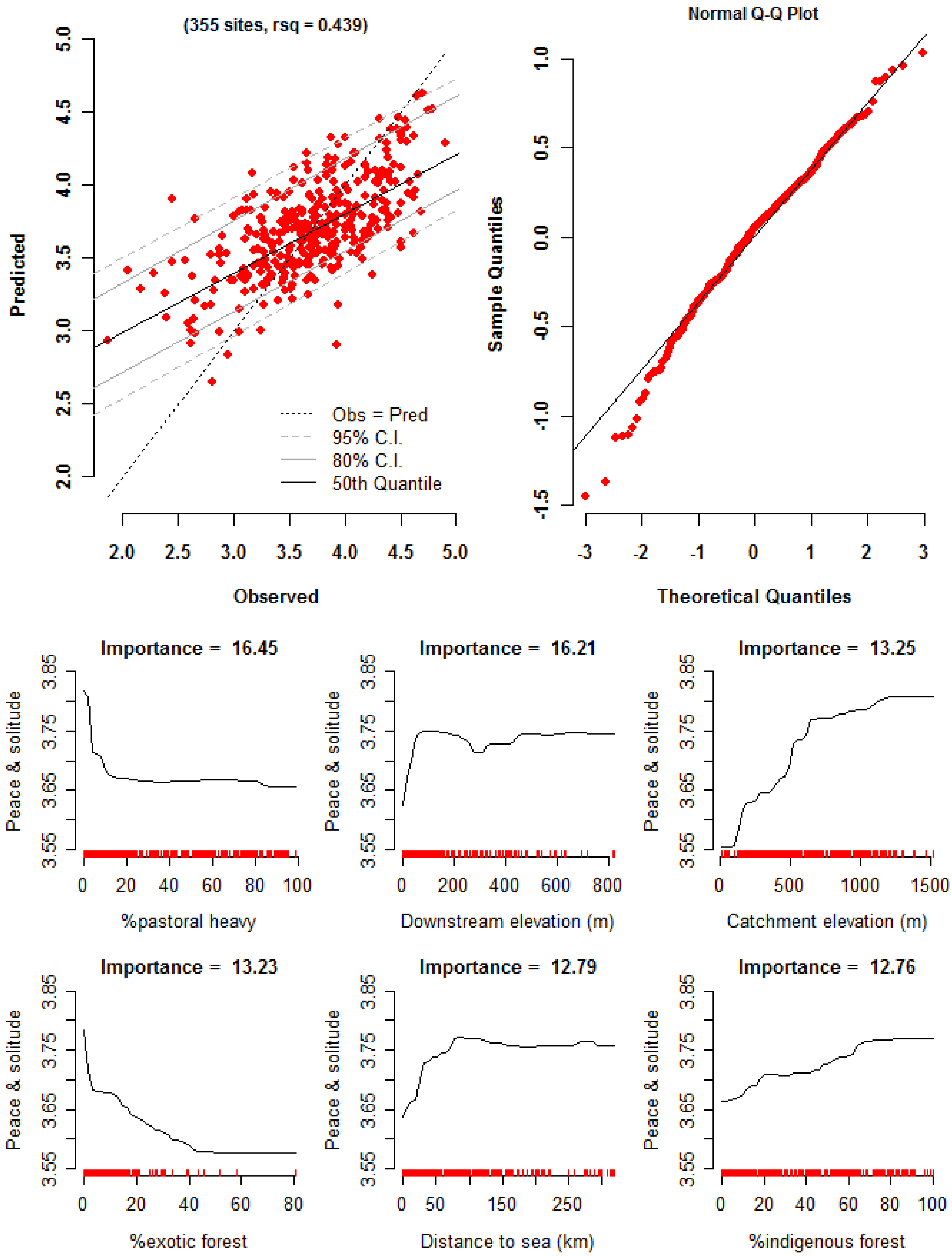


Figure B-8. Diagnostic plots for a RF model of peace & solitude scores (derived from the 1979/81 National Angling Survey), based on the REC predictor set.

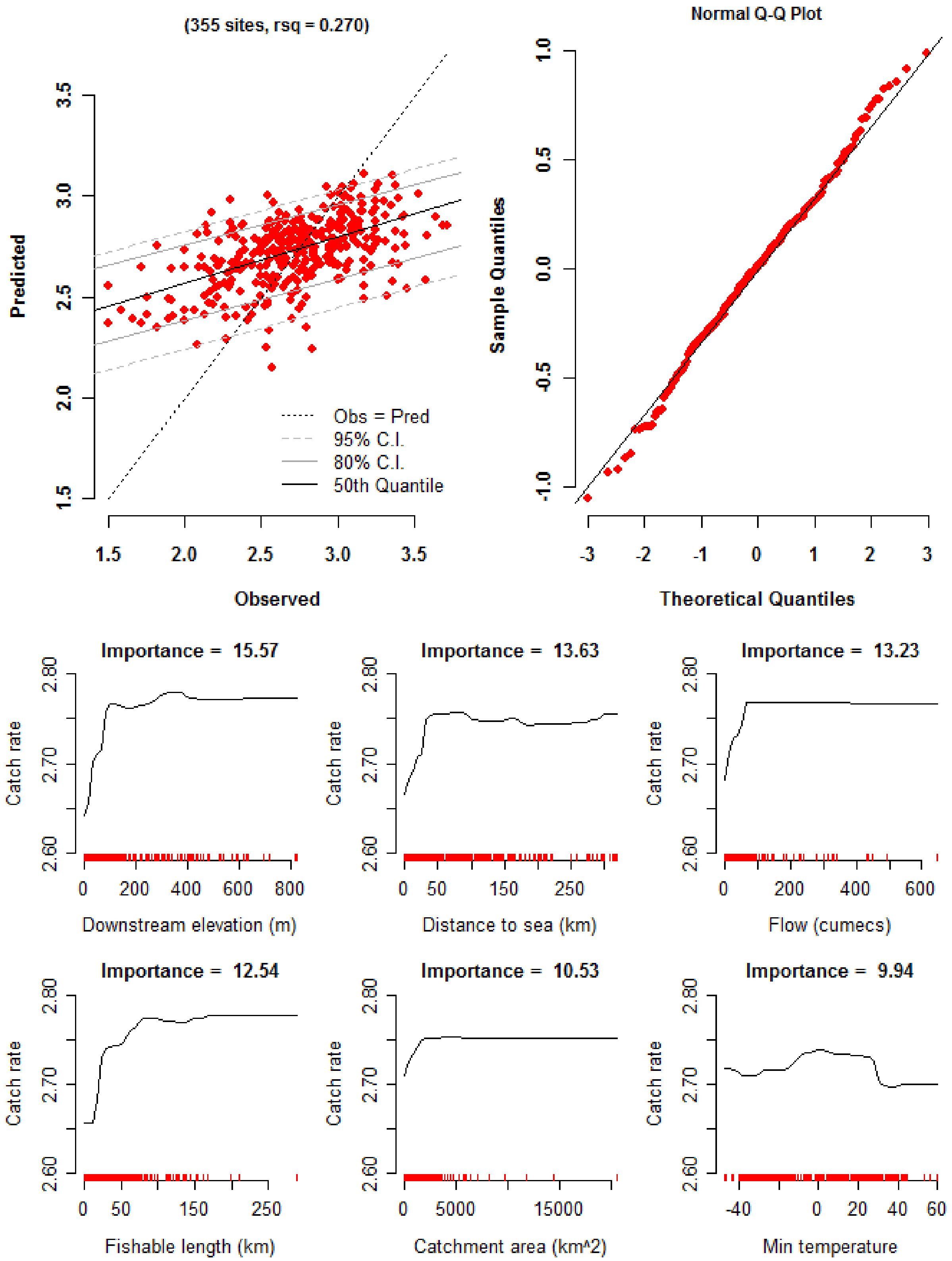


Figure B-9. Diagnostic plots for a RF model of catch rate scores (derived from the 1979/81 National Angling Survey), based on the REC predictor set.

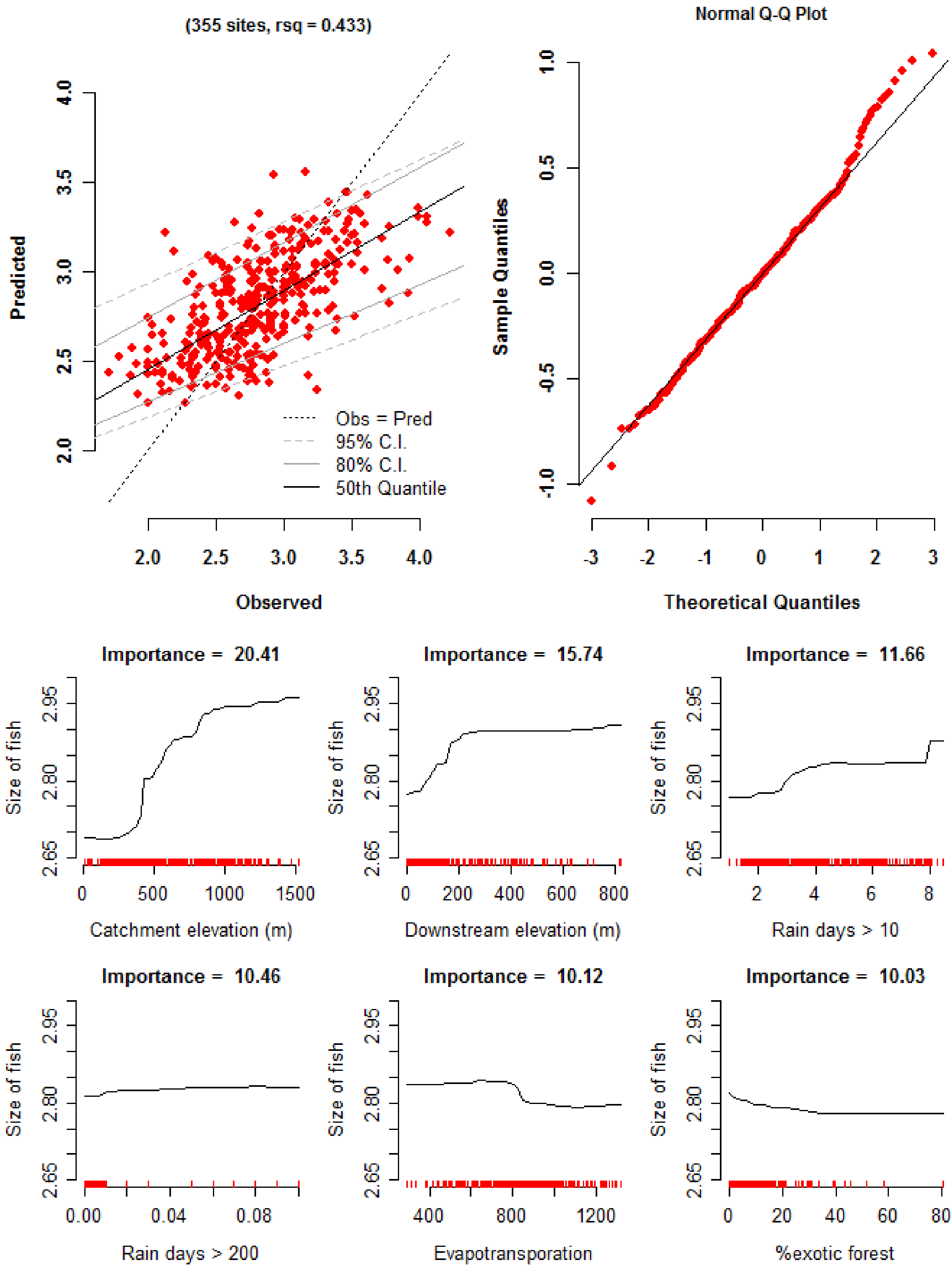


Figure B-10. Diagnostic plots for a RF model of size of fish scores (derived from the 1979/81 National Angling Survey), based on the REC predictor set.

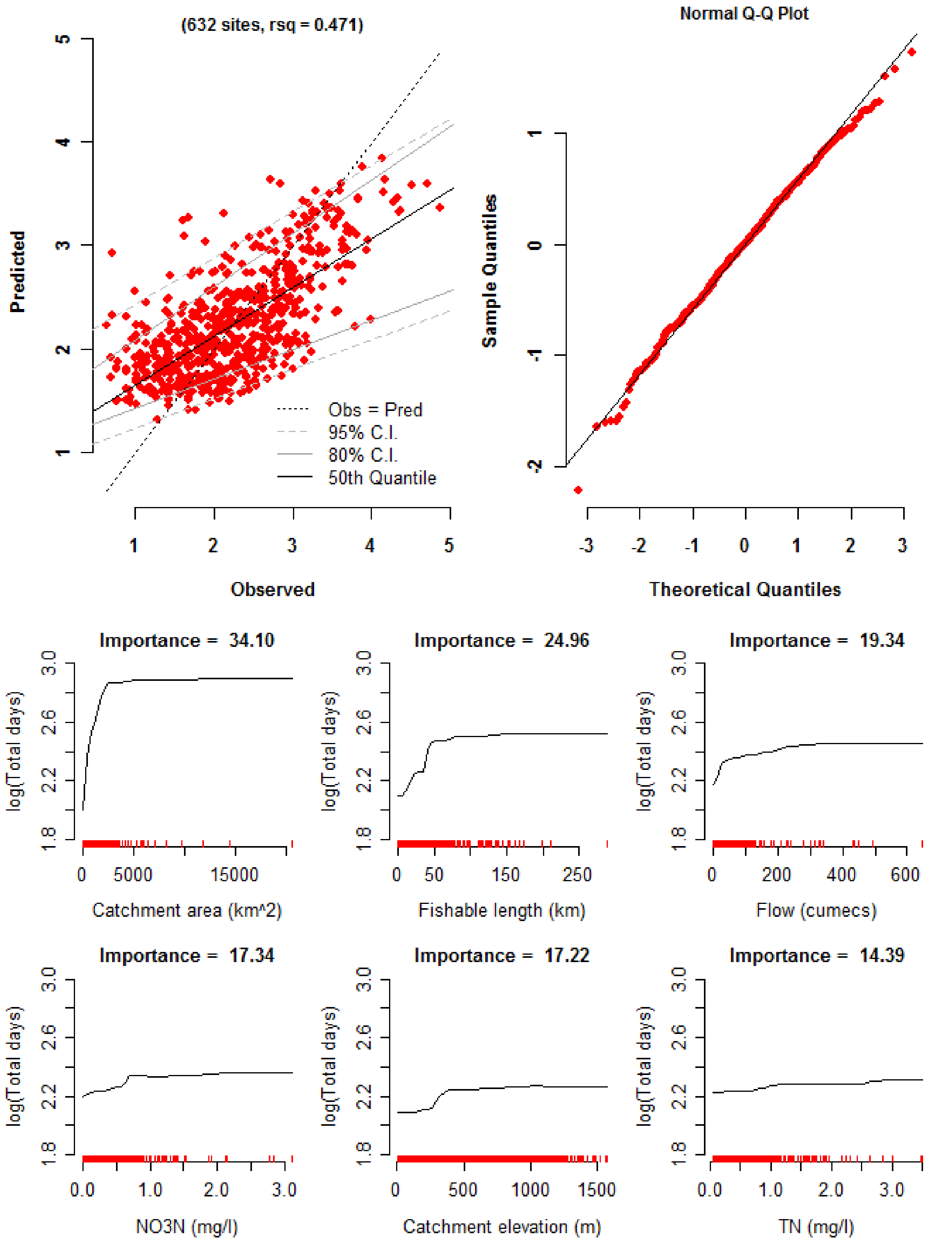


Figure B-11. Diagnostic plots for a RF model of annual angling usage (angler-days, log transformed), based on the WQ predictor set.

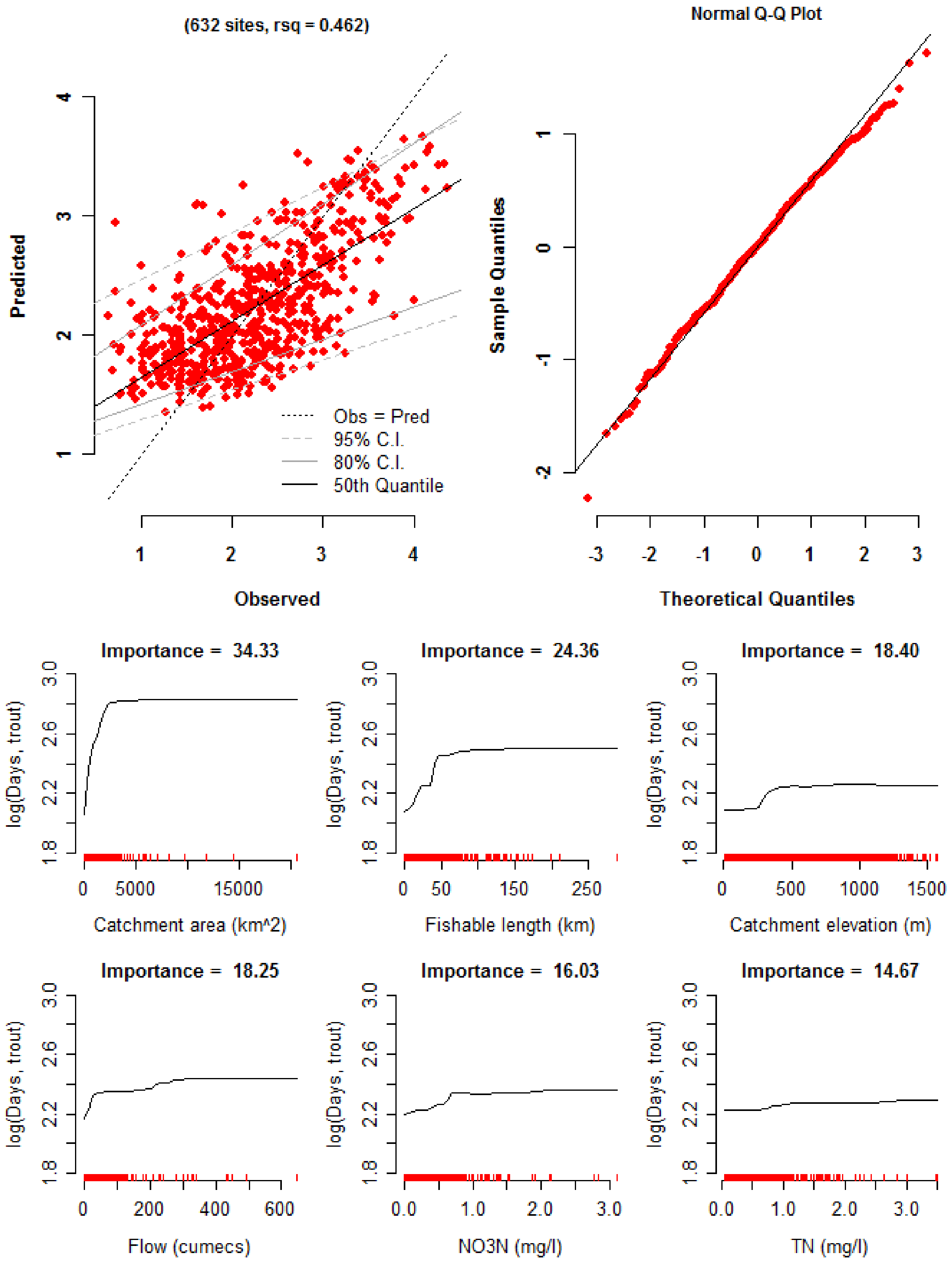


Figure B-12. Diagnostic plots for a RF model of annual angling usage by trout anglers (angler-days, log transformed), based on the WQ predictor set.

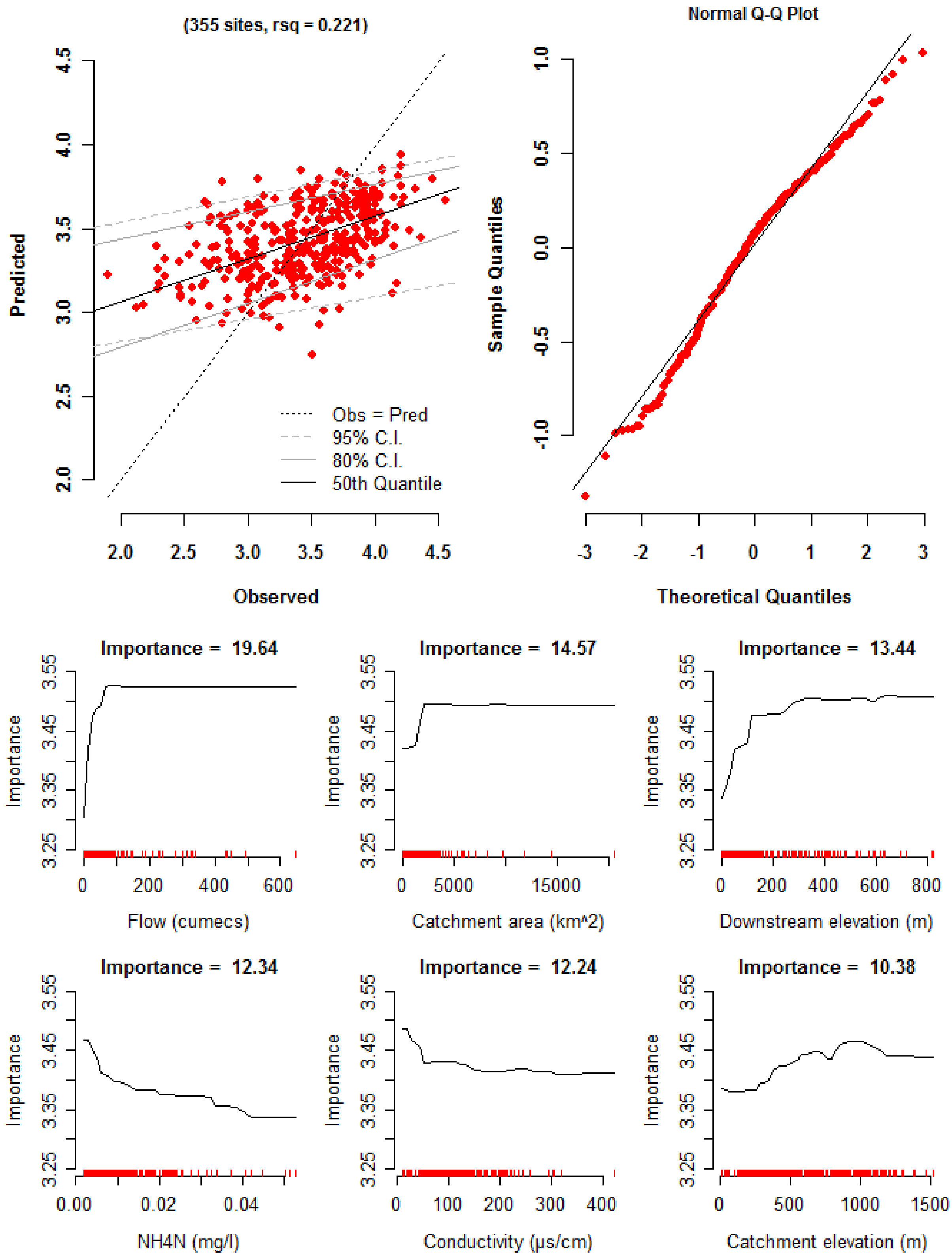


Figure B-13. Diagnostic plots for a RF model of overall importance scores (derived from the 1979/81 National Angling Survey), based on the WQ predictor set.

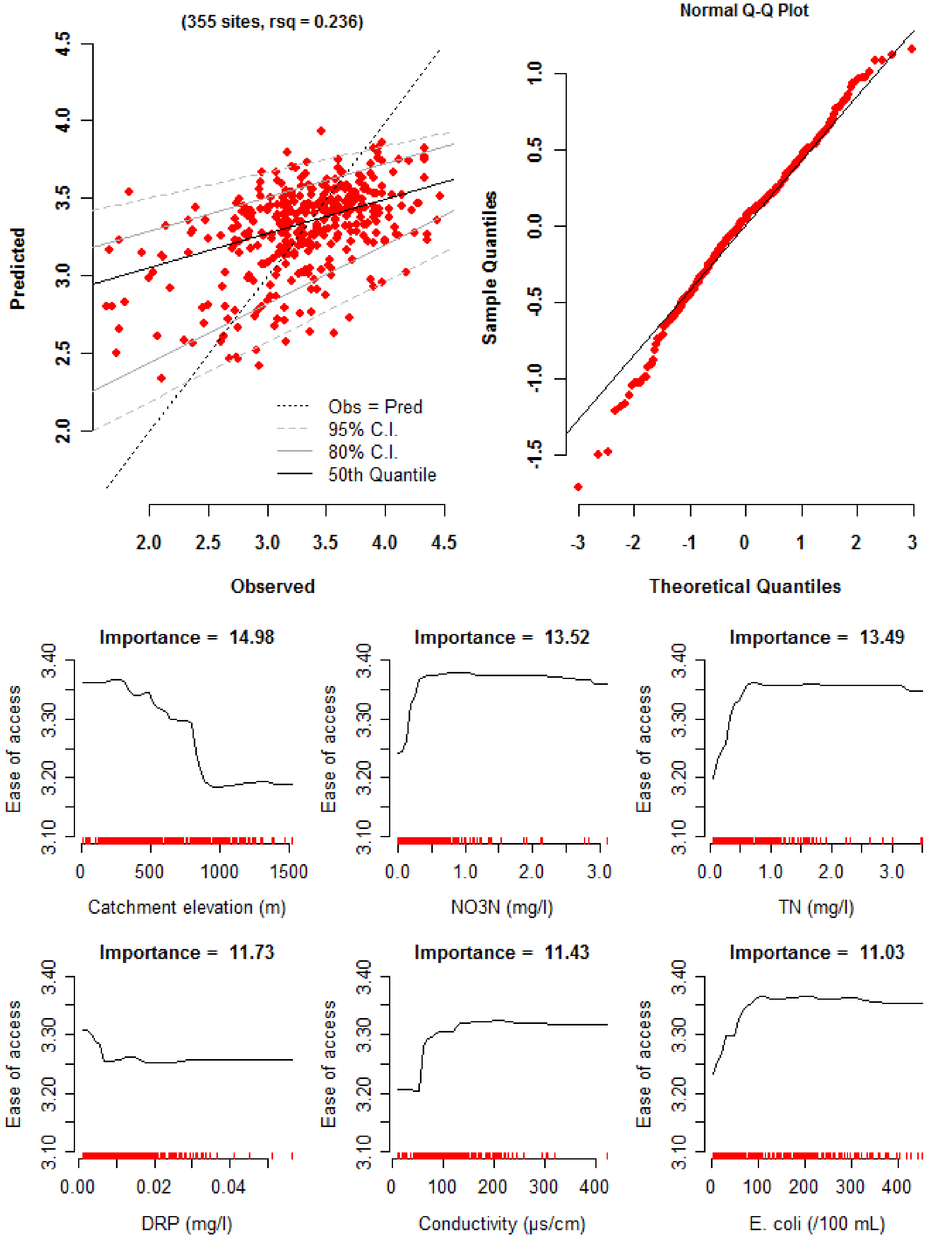


Figure B-14. Diagnostic plots for a RF model of ease of access scores (derived from the 1979/81 National Angling Survey), based on the WQ predictor set.

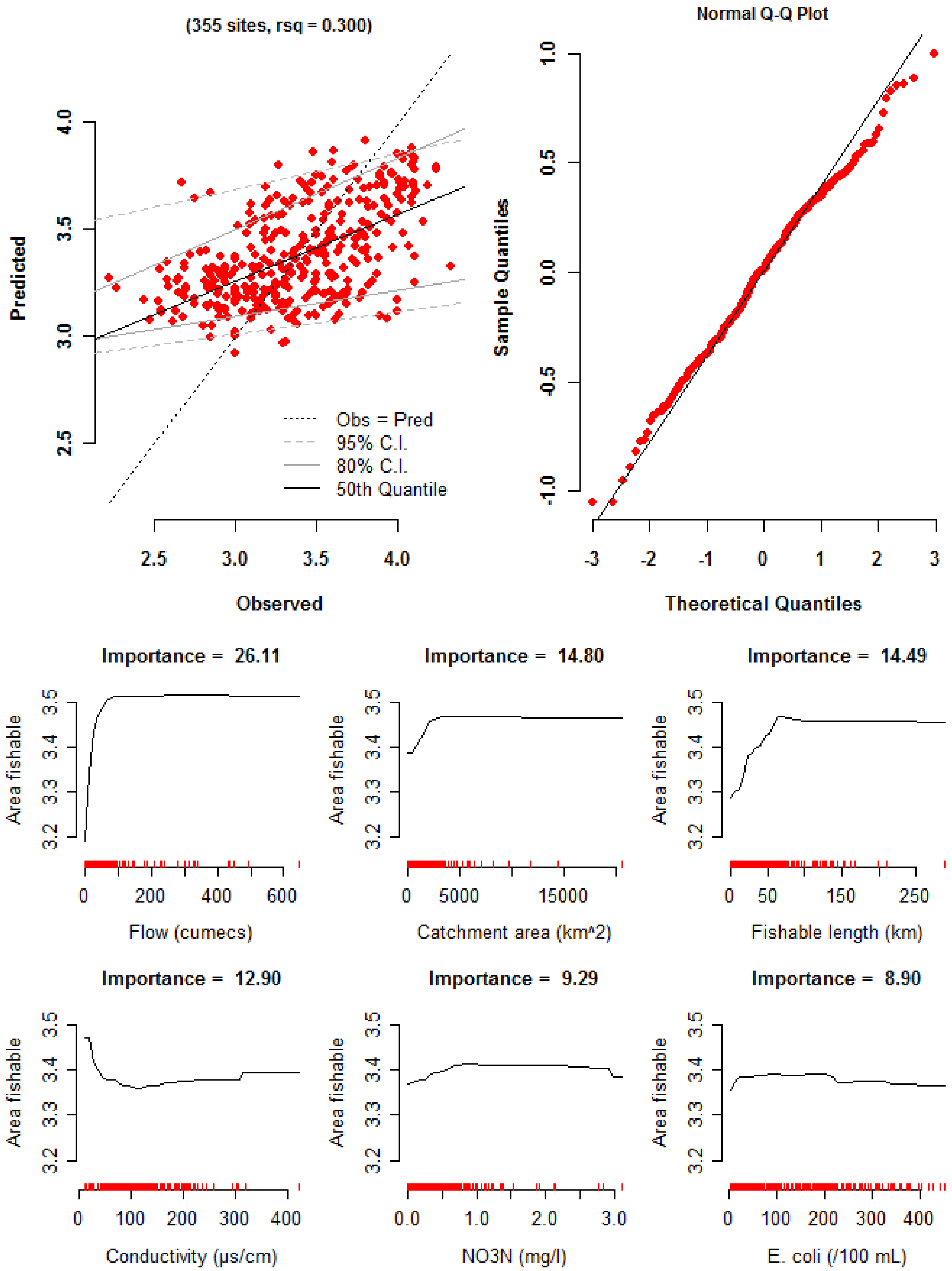


Figure B-15. Diagnostic plots for a RF model of area fishable scores (derived from the 1979/81 National Angling Survey), based on the WQ predictor set.

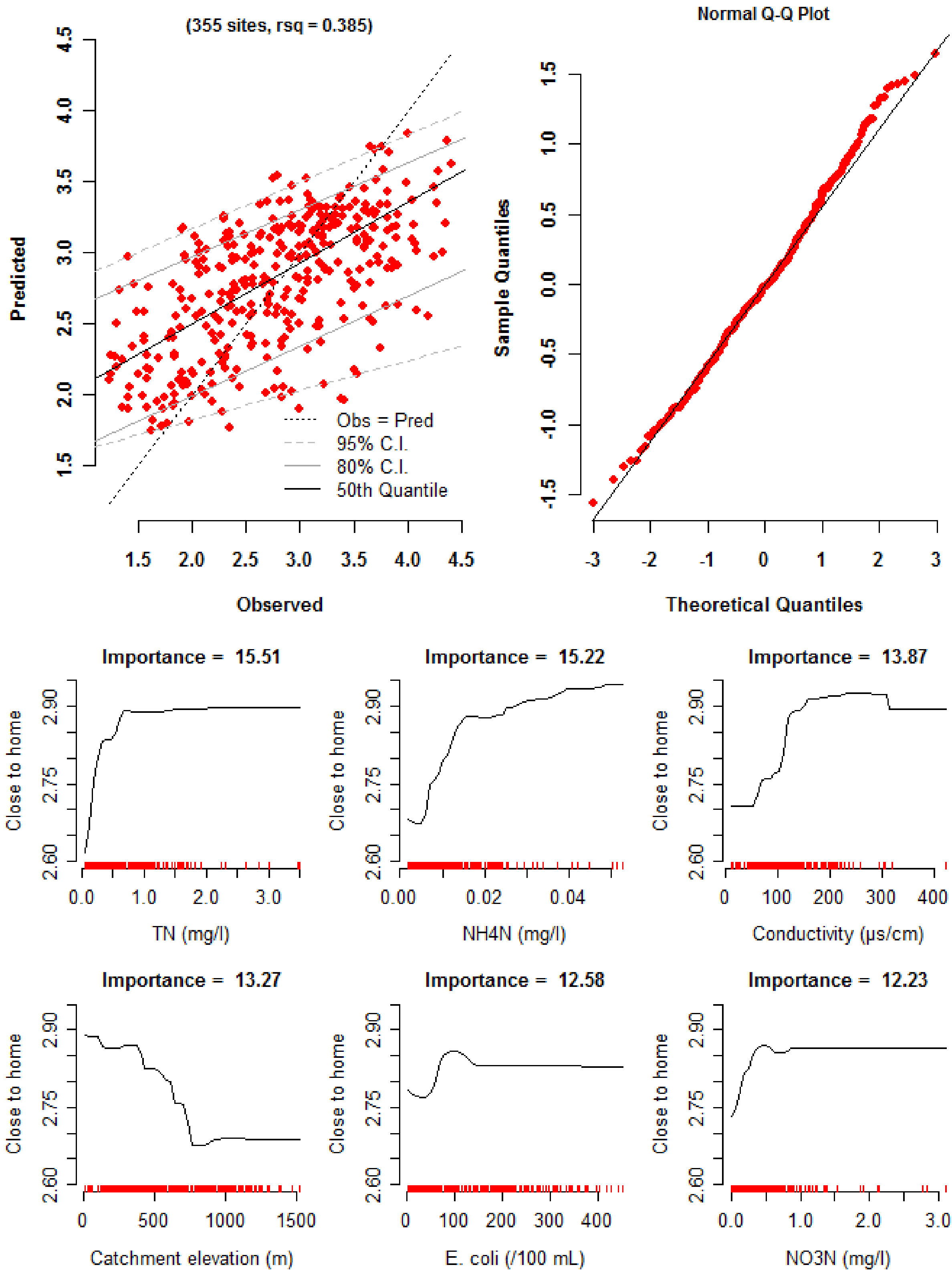


Figure B-16. Diagnostic plots for a RF model of close to home scores (derived from the 1979/81 National Angling Survey), based on the WQ predictor set.

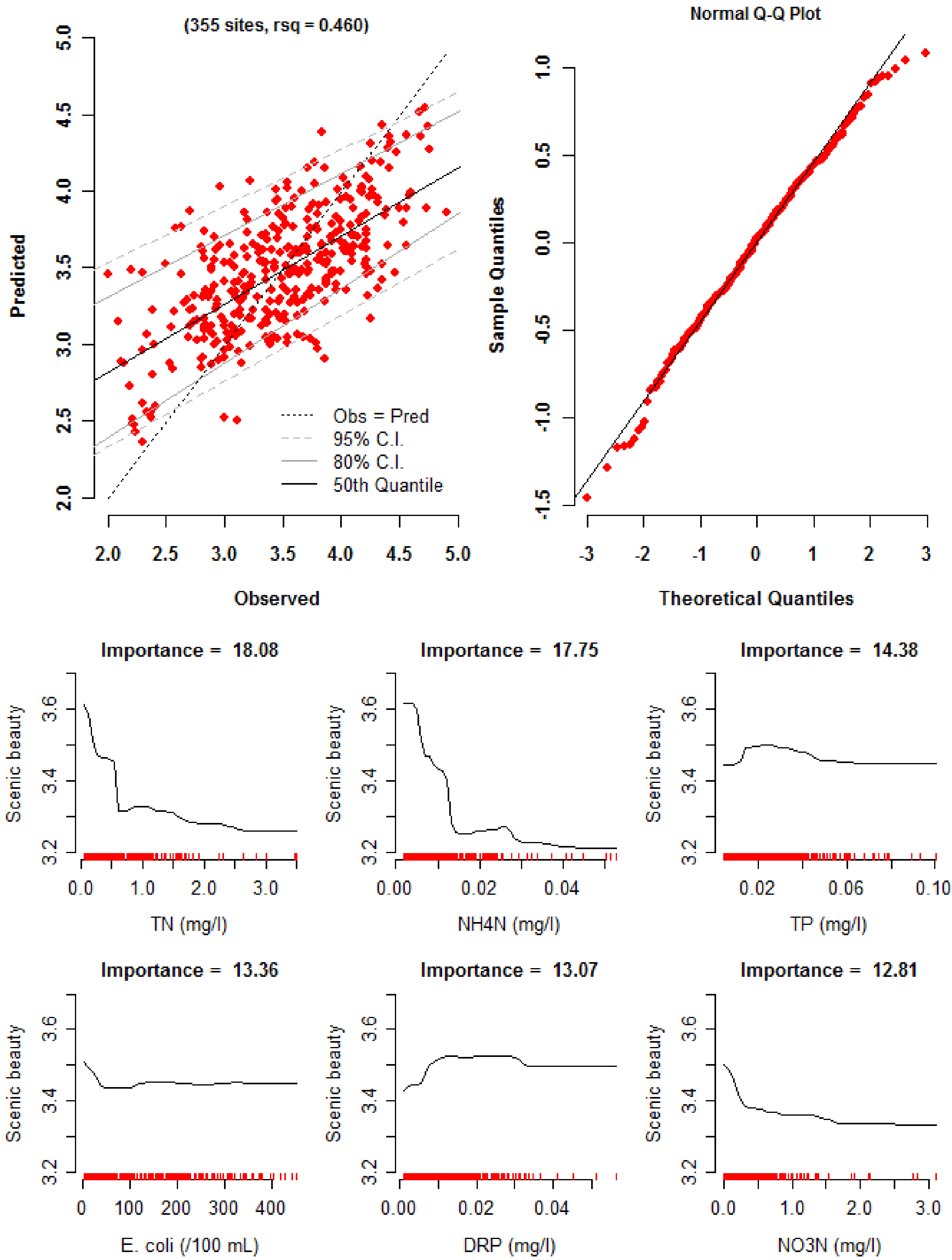


Figure B-17. Diagnostic plots for a RF model of scenic beauty scores (derived from the 1979/81 National Angling Survey), based on the WQ predictor set.

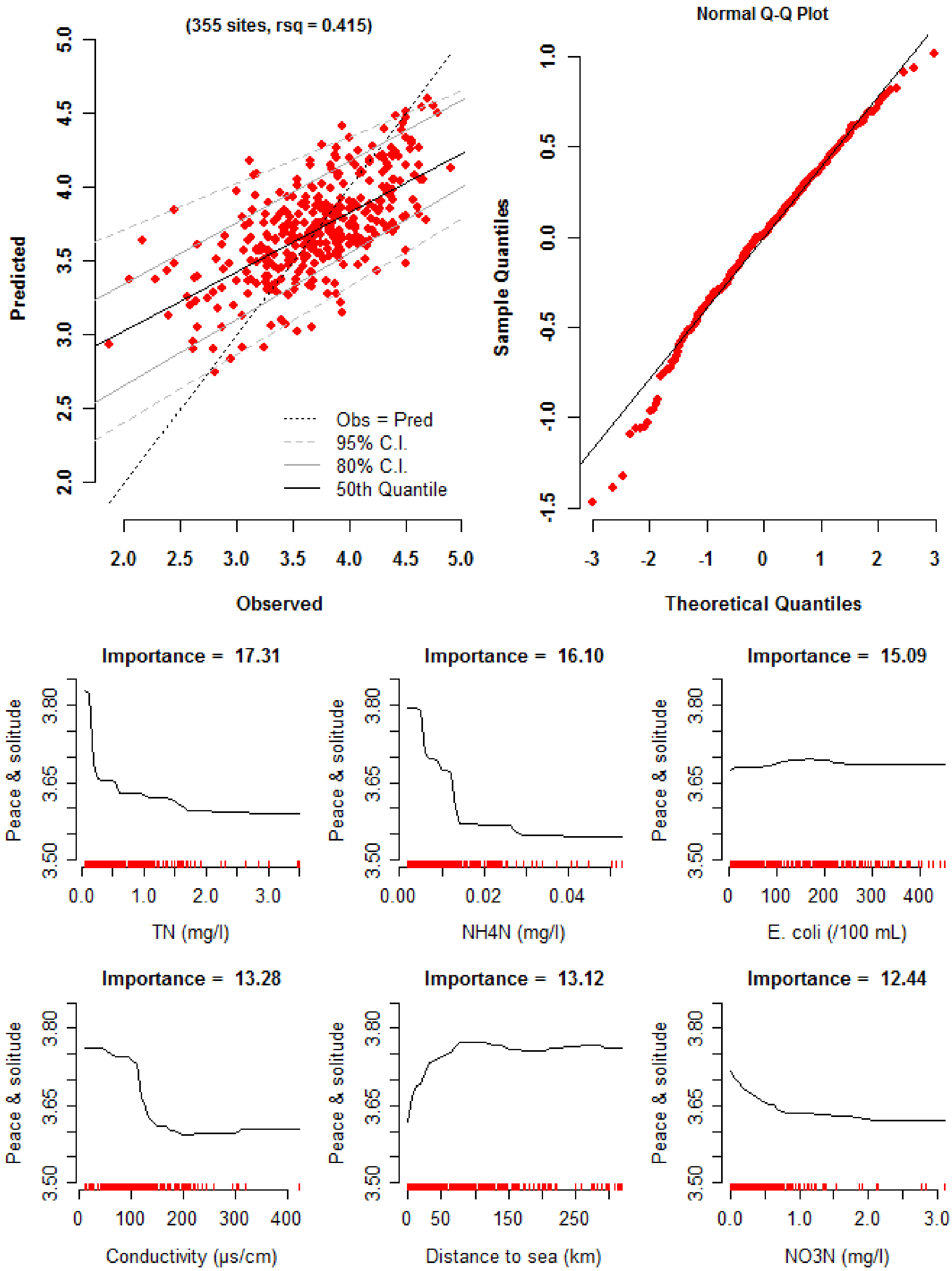


Figure B-18. Diagnostic plots for a RF model of peace & solitude scores (derived from the 1979/81 National Angling Survey), based on the WQ predictor set.

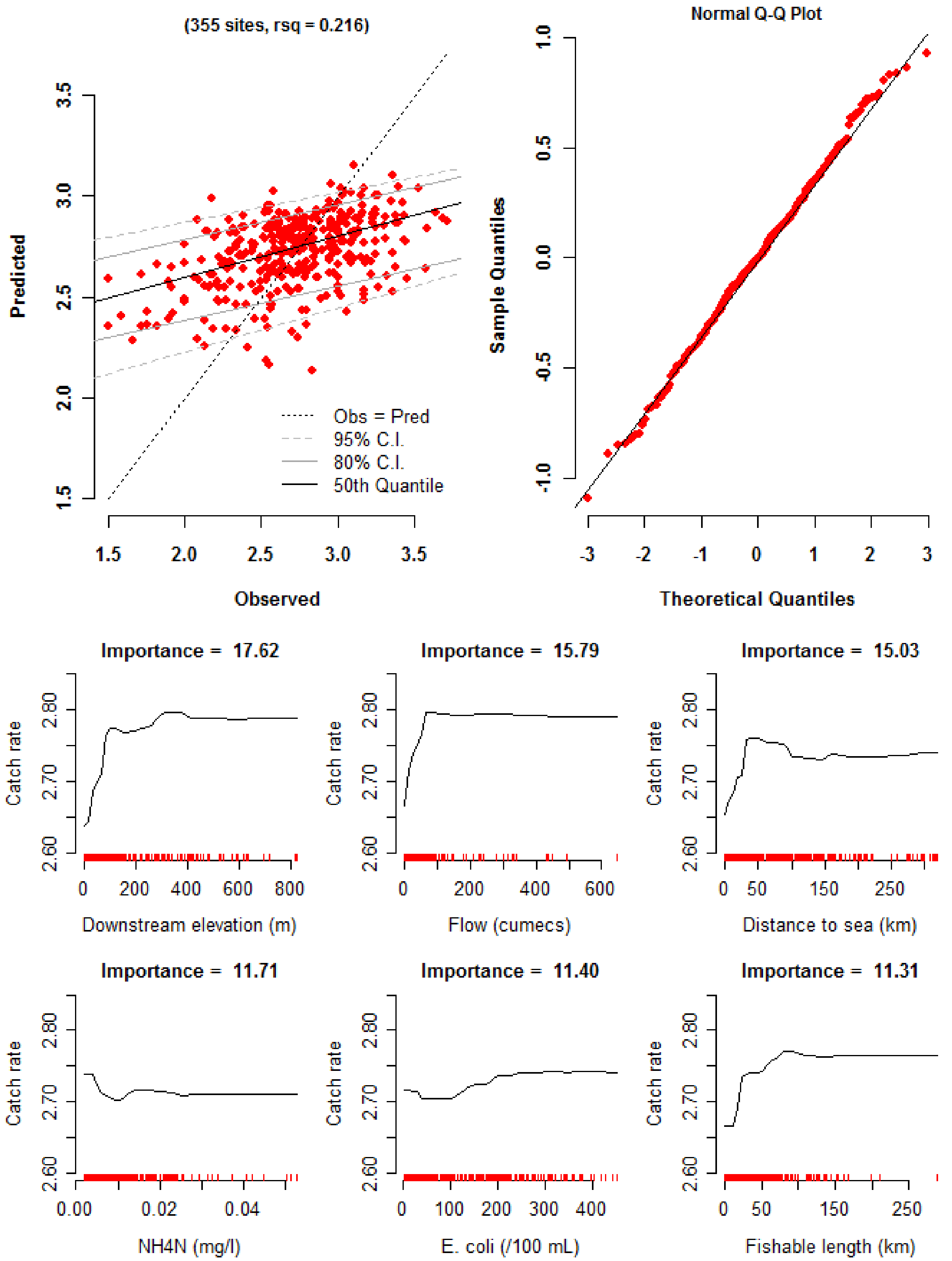


Figure B-19. Diagnostic plots for a RF model of catch rate scores (derived from the 1979/81 National Angling Survey), based on the WQ predictor set.

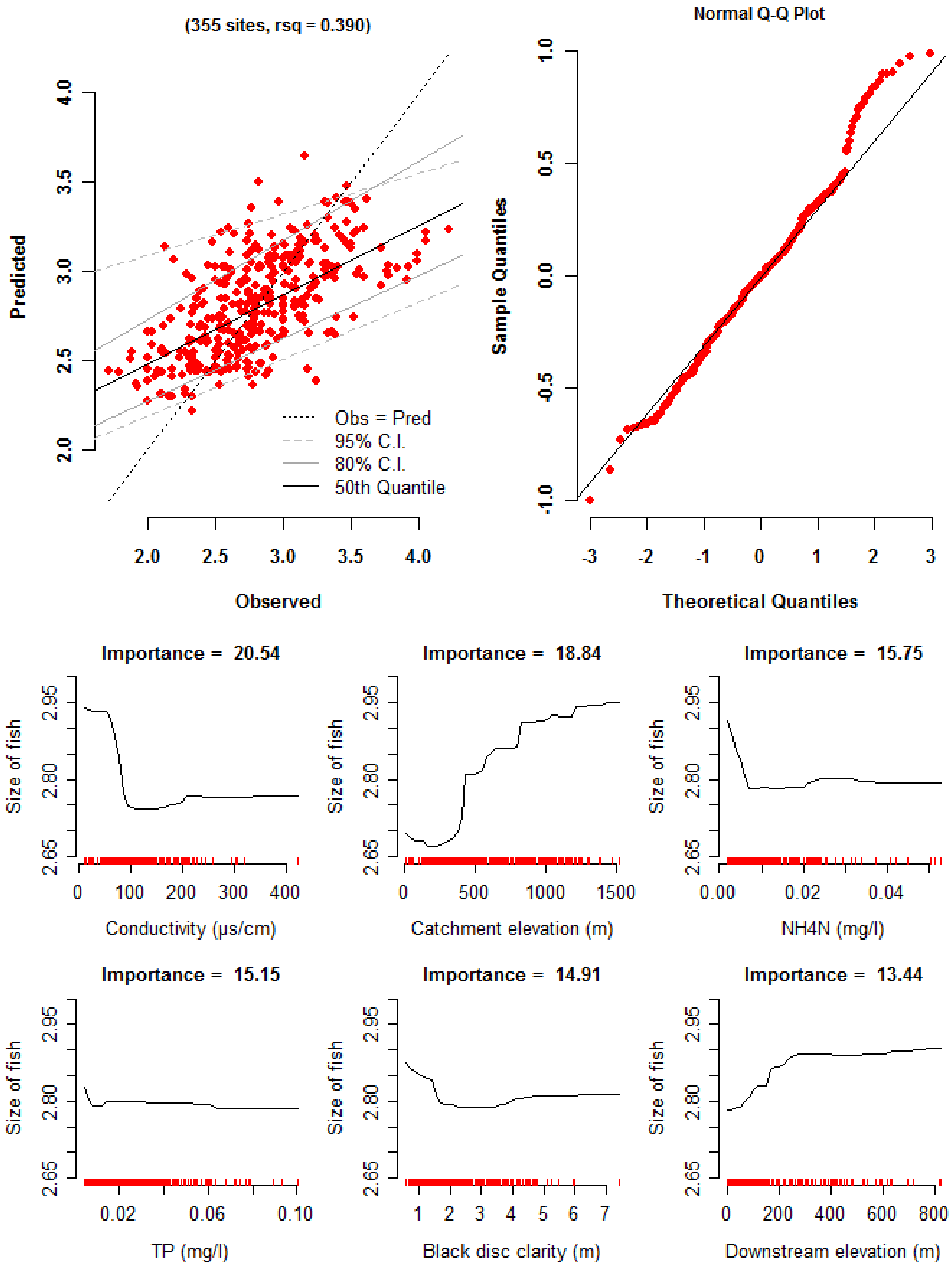


Figure B-20. Diagnostic plots for a RF model of size of fish scores (derived from the 1979/81 National Angling Survey), based on the WQ predictor set.