OTAGO FISH AND GAME COUNCIL

Development of a Salmonid Spawning Layer

July 2024

J Couper – Otago Fish & Game Officer

Photo 1: Brown trout digging a redd in a previously un-surveyed Catlins tributary, identified by an early iteration of the spawning model.

Summary

A model of salmonid spawning habitat in the Otago Region has been created. The model was trained using the extensive Otago Fish & Game spawning database and records of juvenile salmonids from the New Zealand Freshwater fish database. The model has been ground tested over the summer of 2023-24, scrutinised by expert staff and found to perform well.

As the model is based on environmental datasets and predicts a biological activity, it is not perfect, however it appears to do an excellent job overall at predicting spawning habitat across a very large region. It is expected that the model will serve as a "living document", being improved over time as new data emerges.

The model can be explored here.

Defining Spawning water

Salmonid spawning occurs in flowing waterways throughout the Otago region but is concentrated in places with suitable gravel composition, flow profiles, and temperatures. Spawning requirements vary with the size of the fish, so there must be enough variability to accommodate different sizes.

The Otago region has four species of salmonids: brown trout, rainbow trout, brook char, and Chinook salmon. A fifth species, Atlantic salmon, is likely locally extinct. The spawning requirements of the four main species are likely to be similar although Chinook salmon are known to have stronger preference for spring or lake fed waterways and favour larger substrate sizes and faster flows. It's also likely that rainbows will be more affected by temperature and flow issues due to their later spawning period.

The need to identify spawning water

The need to produce a spatial database of spawning habitat has come from multiple work streams. Primarily, it is helpful to communicate the range of spawning habitat to stakeholders. An important aspect of this will be to support policy and planning work, such as the upcoming Otago Land and Water Plan process, the next sports fish and game management plan, and defining salmonid spawning areas under the Conservation Act.

Ideally, mapping spawning habitat would be conducted by ground survey alongside spawning surveys. However, spawning surveys are labour-intensive. Fish & Game staff can complete around 10 km of surveying in a day. Due to the size of the Otago Region and the small number of staff, ground surveying the whole region is unfeasible. Instead, models have been set up to extrapolate from publicly available records of juvenile salmonids and the extensive spawning database that Otago Fish & Game has built over many decades.

Alongside the Otago spawning modelling, there has been a National Fish & Game project to model spawning habitat nationally. As the timeframes of this project were unclear and relied on many external parties, Otago Fish & Game staff decided to push ahead with Otago-based modelling. The nationwide project is led by Dr Adam Canning of James Cook University. It is expected that Dr Canning's work will help to improve or supersede the model created by Otago Fish & Game staff.

Model design

Several methods to model spawning in the Otago region were trialled. The use of maximum entropy or Maxent (Phillips et al., 2006) based models fit our data the best. Maxent, based on machine learning principles, emphasises habitat variables that are good predictors and minimises those that aren't. This is important as habitat variables can contribute to species distributions in subtle and often non-intuitive ways. Maxent employs a technique that predicts species occurrences by finding the distribution that is most spread out or closest to uniform, while considering the limits of environmental variables at known locations (Baldwin, 2009). Another key requirement in choosing a model was our need for one which worked using "presence only" data. True absence data for spawning habitat is difficult to obtain, as even a spawning survey that didn't find any evidence of spawning doesn't mean an area is not potentially suitable for spawning, or that it won't happen there in future.

Model Training

The model was trained using a list of waterway reaches defined as suitable for spawning (presence data). We confirmed whether a river reach was suitable for spawning in multiple ways:

- Salmonid redd (nest) presence recorded by Fish & Game staff (n=4,600).
- Presence of adult salmonids during spawning season (n=3,700). These salmonids were almost always present alongside redds. These records only added sections where redds could not be identified due to extremely low algal levels.
- Presence of juvenile salmonids the following spawning season (n=14,200 records). We defined this as salmonids 100mm or less, determined by communications with staff from around the larger organisation. There was consensus that this size class would be found close to their natal habitat. This data was sourced from the NIWA freshwater fish database.
- Fish and Game staff surveying a reach and providing a professional opinion on whether the site was suitable (n=76).

In practice, many of these observations occurred in the same reaches. Overall, 1,550 waterway segments were used as training presence data.

The model, trained with data from all salmonid species, can predict the spawning of these species due to their similar habitat preferences.

A limitation of the presence data is that the survey effort to obtain it was not randomised. The data is biased by location, as sites that are practical to access and closer to the organisations conducting electric fishing or spawning surveys are more likely to be sampled. There is also a further bias in that survey effort is more likely to be concentrated on areas assumed to be important, such as streams running into popular and prolific lake fisheries. There is no practical way to fix this bias, so it is recognised that models created from this dataset are likely to bias towards easy-to-access waterways and recognised fisheries.

Habitat data inputs

The key habitat dataset used to model salmonid spawning was the FENZ (Freshwater Ecosystems of New Zealand) river GIS database (J. R. Leathwick et al., 2010). This database consists of a line geometry with associated habitat variables, described in Appendix 1. The line geometry of FENZ matches that of RECv1 (River Environment Classification).

Several habitat variables were also taken directly from the RECv1 dataset (https://niwa.co.nz/freshwater/management-tools/environmental-flow-tools/river-environment-classification). The variables are defined in Appendix 2. These data were linked to the FENZ habitat models using the NZ river reach number.

There were limitations with these datasets, including:

- The habitat variables in these datasets are created by various ground-tested modelling techniques and are not entirely accurate.
- The data are stored at a relatively coarse level. Stream segments in Otago are 0.66 km on average, but can be as long as 7.6 km.
- Lines do not always follow waterways exactly, particularly in places where streams have been diverted.
- Some waterways are not present in the database, most apparent for spring-fed streams. A notable example is Bullock Creek in Wanaka, a recognised spawning habitat.

- Some waterways are not well depicted by the FENZ/REC GIS layer. Notably, the lower Kyeburn, Pig Burn, and the lower reaches of the upper Clutha which is sometimes inundated by Lake Dunstan.
- Braided river habitat is represented by a single line, meaning it is not well described by the models.

Model Testing

Initial modelling

The creation of the current iteration of the spawning layer consisted of multiple parts. The initial stages involved setting up a predictive model using only presence records from the national freshwater fish database (NZFFD (Stoffels, 2022)). This model was then informally tested by overlaying the Otago database of salmonid redds. We found that almost all recorded spawning occurred in areas the initial model predicted as suitable salmonid spawning habitat.

This result validated the method and showed that records of salmonids under 100 mm were a valid proxy for spawning habitat.

A key issue noted at this stage was that the modelling method, which uses presence-only data, can focus on variables that indicate survey effort rather than species presence. The main problem variable was the downstream distance to the coast (“DSDist2Coast”, Appendix 1). This was an issue because previous electric fishing work tended to be clustered closer to the organisations that carried out the work (Fish & Game, Otago University, and DOC) and often targeted particular, mostly coastal species. This meant the initial models unreasonably penalised sites further from the coast despite other variables being suitable. The “DSDist2Coast” variable was removed in future models, which significantly improved them.

Second Stage

A second model was set up using data from both the NZFFD juvenile records and redd locations from the Otago spawning database. A small number of redds had their positions estimated by evenly placing the number of redds seen along the riverbed between the recorded start and end points. This was necessary for some older spawning surveys that only had start, end, and total number of redds, rather than GPS-marked individual redds. This method was only used when more recent spawning data was unavailable. Although not ideal, it appeared to be suitable for the scale at which the model operates.

Ideally, the model would have been finalised earlier, allowing the 2023/2024 summer programme to focus entirely on testing. Unfortunately, due to delays caused by the need to improve the model and enter historic spawning data, the summer work programme designed mainly to test the model also had to include work to fill out areas assumed to be spawning habitat but not identified by the model.

Second Stage – Testing

Figure 1 shows the Otago region and the Otago Fish & Game region, which are largely aligned except for north of the Waikouaiti catchment, which lies in the Central South Island Fish & Game Region. Overlaid on the maps are the 130 ground testing sites surveyed over the 2023-24 summer work programme. Data were entered directly into a phone application designed specifically for this project to allow for easier analysis. The data will be uploaded into the freshwater fish database in the future.

Sites were selected mainly by looking at areas without previous records of spawning or juvenile presence, with a slight priority given to sites likely to be spawning habitats. This selection was informed by staff knowledge and previous models.

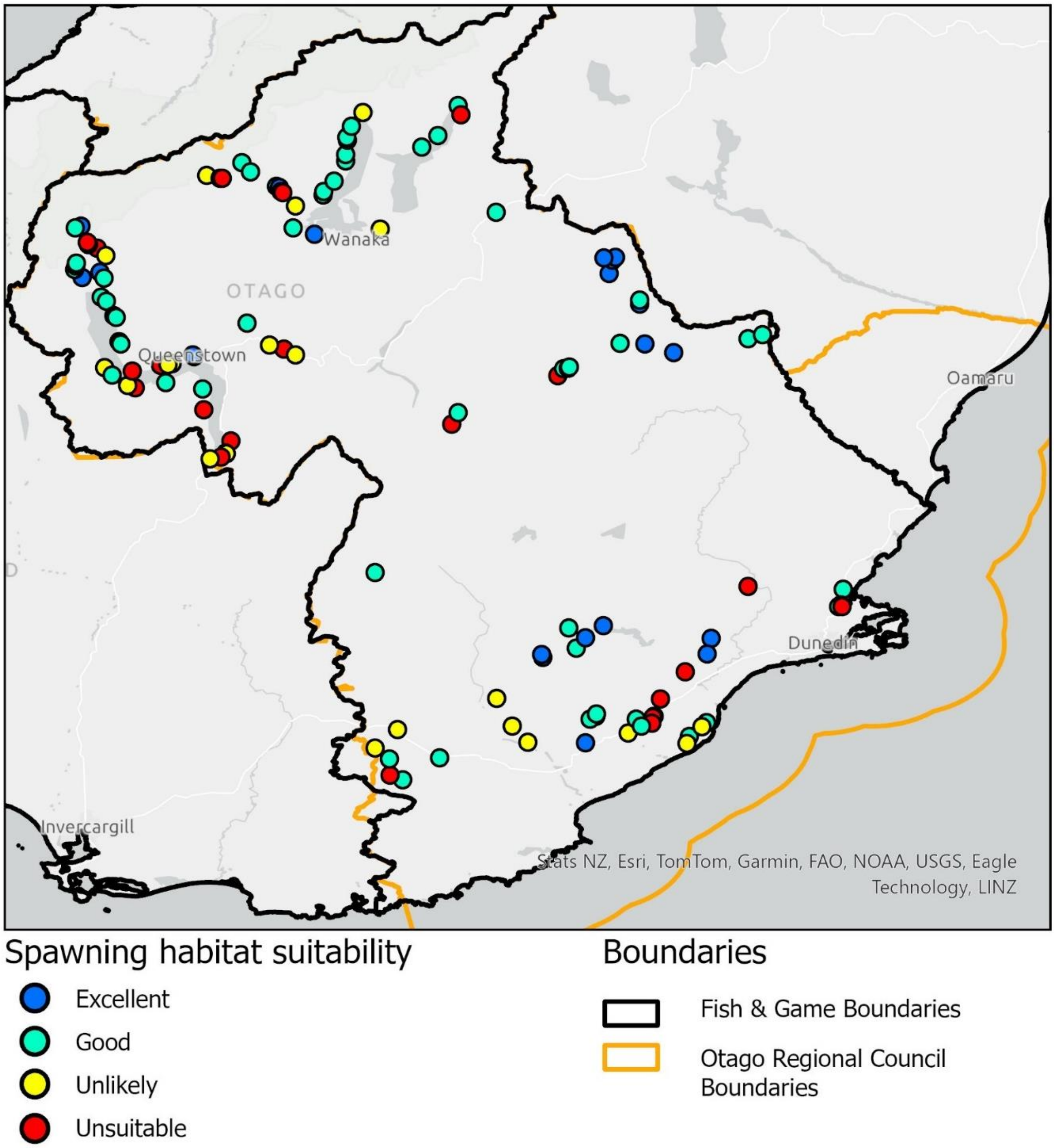

Figure 1: Map showing the sites that were assessed for spawning suitability (n=130) in the 2023-24 ground testing program

The coloured dots in Figure 1 show the staff assessment of how suitable each site was for spawning. The total number in each category is noted in the axis titles of Figure 2 below. Figure 2 also shows the average value and variation in model output for each spawning suitability band that staff categorised.

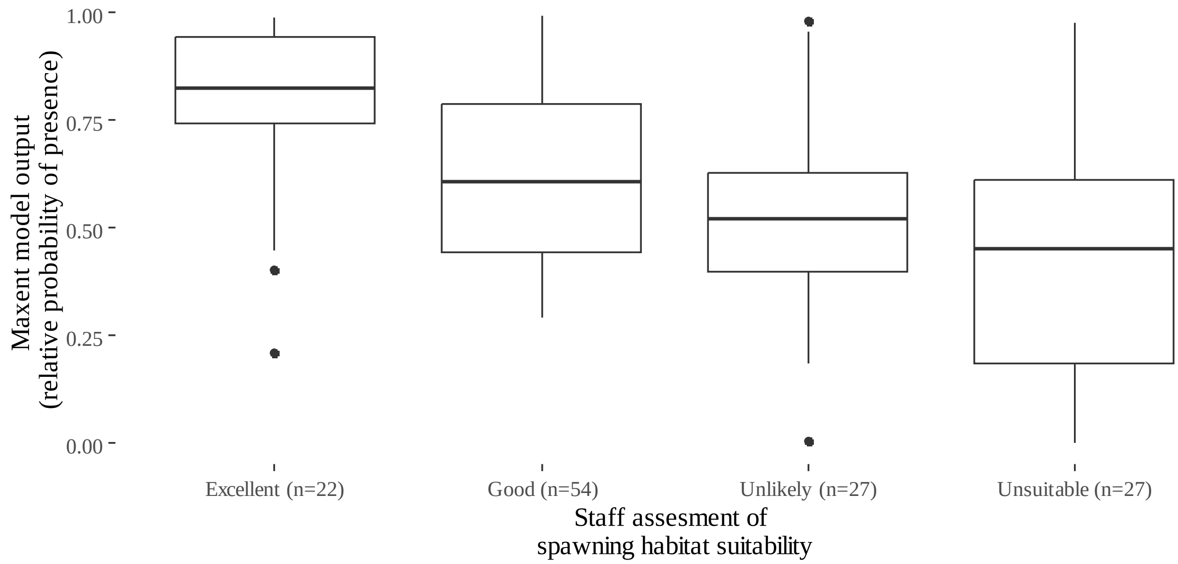

Figure 2: Staff assessment of a water ways suitability for spawning against the output of the second round of modelling.

The ground testing aligned well with the model, as the highest-rated spawning habitat categories had the highest median model outputs, and this pattern continued sensibly into the lower categories. This provided further evidence of the method's suitability for modelling spawning habitat in Otago.

Despite the overall fit of the staff assessments against the second-stage model, there was variability in outputs across each category. Lower model outputs in the higher categories could generally be attributed to issues with the modelling process, which were addressed in subsequent iterations. Conversely, higher model outputs in the lower categories were often due to staff identifying a reach as poor spawning habitat, typically because of a known downstream barrier—a factor not well-modelled by this process. Another contributing factor was low flow on the day of the survey; however, many of these sites are expected to have better flow during winter or spring, when most salmonid spawning occurs.

Of the 130 sites assessed for spawning habitat suitability, 95 were also electric fished. Reasons for not electric fishing a site varied, but generally included unsuitability due to low flow or inability to arrange access to the land in time. Table 1 shows the total number of sites where each juvenile species was detected.

Table 1: The number of sites that were electric fished in summer 2023-24 with the number that had salmonids present.

Variable	Number of Sites
No juvenile salmonids present	22
Juvenile brown trout present	56
Juvenile rainbow trout present	28
Juvenile brook char present	3
Juvenile Chinook salmon present	0
Any juvenile salmonid present	73
Total number of sites electric fished	**95**

Just over three-quarters of the sites that were fished contained juvenile salmonids, with brown trout being found at the majority. Unfortunately, no juvenile Chinook salmon were found across the summer.

The presence of salmonids was tested against the output from the second round of modelling, as shown in Figure 3.

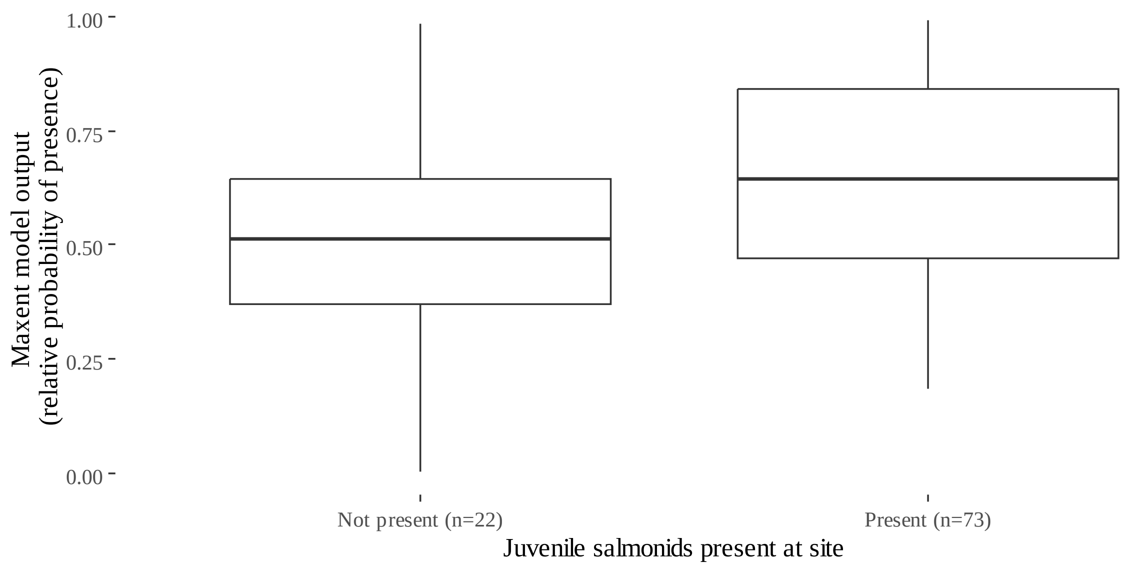

Figure 3: The presence of salmonids at ground testing sites against the output of the second round of modelling.

Comparing the model output to the presence of juvenile salmonids presented more challenges than the staff suitability assessments, as there were several reasons why juveniles were not found despite favourable habitat characteristics. One key factor was the scale of the ground testing project, which meant that some sites couldn't be electric fished until late in the season. During this time, juveniles may have moved downstream for growth or been displaced by floods.

Despite juvenile presence not being as suitable for model testing as staff assessments of habitat suitability, it provided additional evidence of the method's effectiveness, contributed to staff assessments, and enhanced understanding of salmonid presence across the region.

Final Model Creation

Key issues with previous models

Previous stages of model creation identified issues with juvenile and spawning records being incorrectly allocated to river sections. This was addressed in the final model by manually verifying records that corresponded to very low flows in the FENZ dataset and low outputs from previous models. The main issue often stemmed from GPS points being closer to nearby minor tributaries (see Figure 4), particularly notable during helicopter surveys.

Figure 4: Screenshot showing two redds marked in red, overlaid with waterway shapefiles labelled with their modelled flow. Note that the more southern redd would have been incorrectly allocated to the very small tributary below it in earlier models.

Models were tested by removing some variables, focusing on those likely to be highly correlated. However, removing variables did not improve the models, so the advice of Elith et al. (2011) was followed. This paper outlines that the machine learning process reduces the effect of correlated variables by amplifying important ones, suggesting no need to remove variables unless they are biologically irrelevant.

Model parameters

Feature classes (the types of relationships allowed: lineal, quadratic, product, threshold and hinge) were assessed in multiple iterations of the models. Restricting the number of feature classes didn't appear to significantly improve the models. As our number of presence sample reaches (n=1,549) is far higher than the minimum of 80 suggested in Merow et al. (2013),all feature classes were retained for the final model.

The modelling requires a prior assumption about the percentage of stream lengths that contain spawning. This relied on staff expertise based on multiple spawning surveys around the region. Adjusting this factor made very little difference to the model output. Various percentages were tested, with five percent selected for the final model as it provided the output that fit testing data the best.

The final crucial modelling variable is the regularisation parameter. Regularisation smooths the models, reducing the risk of overfitting to the training data and mitigating biases in the collection of presence data. Essentially, this means the model fits the data trends rather than the noise in the data. A regularisation parameter of four was chosen for the final model. This parameter balances the overall precision of the model (expressed as AUC, described below) with ensuring that individual habitat variables contribute sensibly to the model without fluctuating wildly.

Final Model Assessment

Figure 5 presents an ROC curve (receiver operating characteristic curve) illustrating the model's performance across different thresholds. Traditionally, an ROC curve assesses how well a model distinguishes between presence and absence points. However, since our model uses presence-only data, it distinguishes between presences and the background data, which could also be suitable for spawning. Therefore, the model is not expected to completely differentiate between these types even under optimal conditions.

The y-axis (sensitivity) indicates the model's ability to correctly identify a spawning reach at a specific threshold. The x-axis (1-specificity) shows how much of the background is identified as spawning at that same threshold. For instance, a green circle on the chart suggests that to correctly identify 90% of spawning grounds, approximately 30% of the total reaches in the region need to be identified.

The model was validated using a random 5% of sample and background points, with their classification shown in red, while the remaining 95% of training data is depicted in red. For comparison, the curve for a model that randomly guesses whether a reach contains spawning is shown in black.

The area under each curve (AUC) is also provided. As stated in Merow et al. (2013), "AUC is interpreted as the probability that a randomly chosen presence location is ranked higher than a randomly chosen background point."

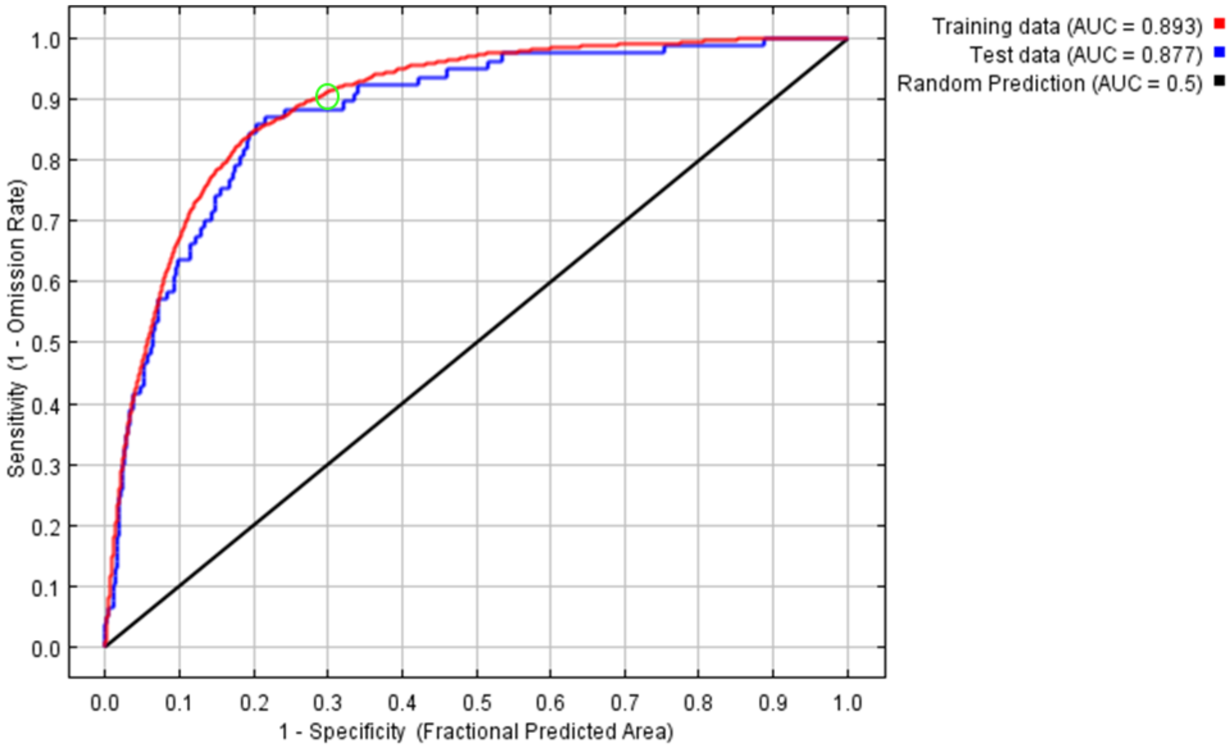

Figure 5: ROC curve for the finalised model.

The curve shows that overall the model performs exceptionally for a presence only based model. AUC reached as high as .904 in a similar model without a regularisation multiple (smoother), however this model was chosen as a good balance.

Figure 6 illustrates the relative importance of each environmental variable in the final model. The values, shown as "Regularized training gain," indicate how well each model can differentiate between the presence and background records. For example, seg flow has a regularized training gain of 0.7. As this is a logarithmic value, it converts to 2.01 (e^0.7=2.01), meaning the model based solely on flow is twice as likely to correctly allocate spawning reaches compared to random chance. The regularization factor, discussed earlier, is applied here to prevent the model from overfitting.

The most influential factor was the flow of the reach, followed by the stream order (as explained in Appendix 2). Since many flow variables scored highly, it suggests that the size, velocity, and stability of a waterway are crucial factors in determining its suitability for spawning, albeit complicated by correlations among flow variables.

Among the environmental variables, the maximum slope downstream of the section and the prevalence of pasture upstream of the section had the most significant impact on model gain when omitted. This reduction in gain indicates that these variables provide unique information not captured by other variables.

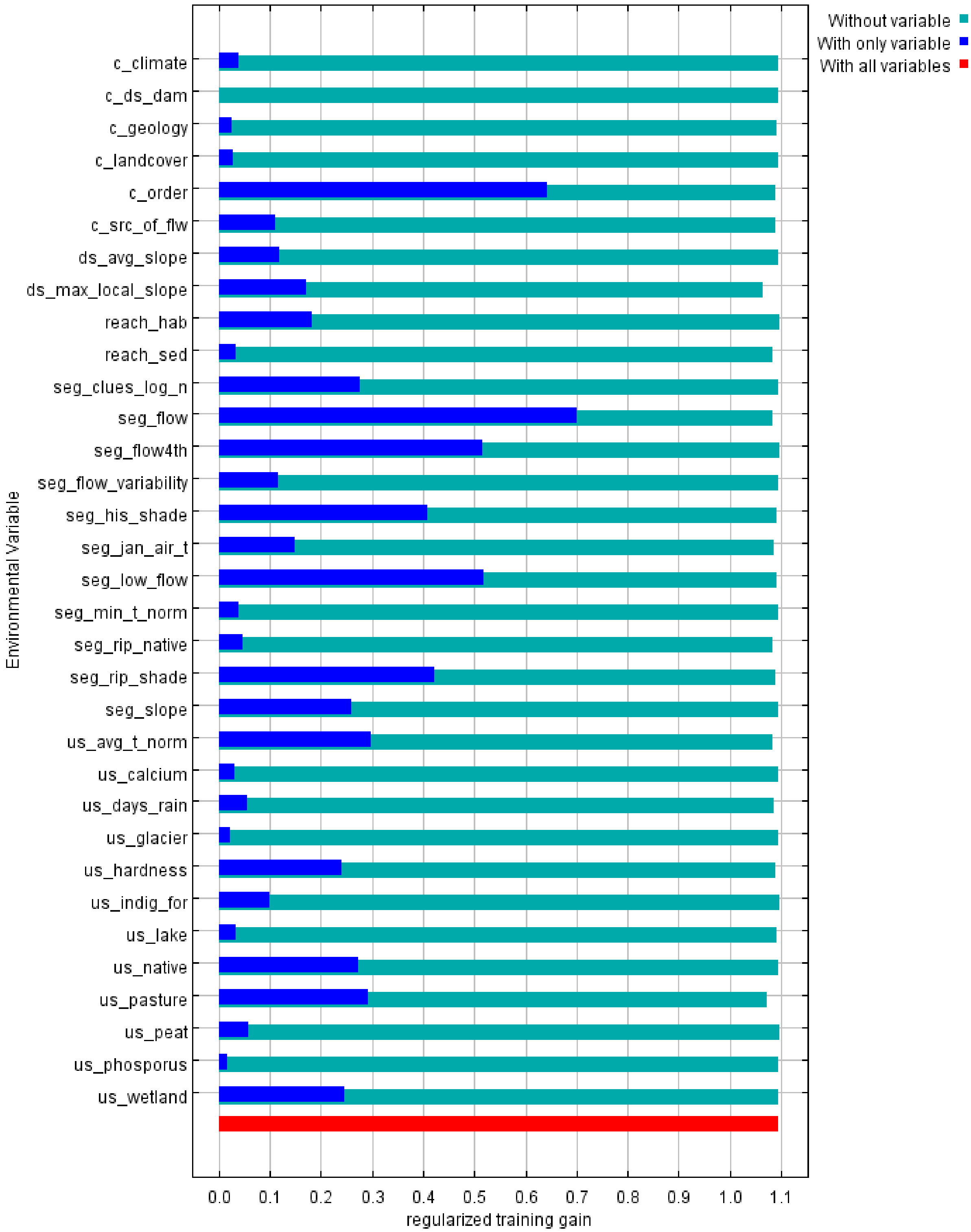

Figure 6: Jackknife chart illustrating the importance of each environmental variable. Teal bars indicate the model's performance when each factor is removed, while blue bars represent the effectiveness of a model created by adjusting only that variable. Note that some variables are prefixed with "c_" to denote they are categorical variables.

Figure 7 presents a series of charts, each showing a model set up with only one environmental variable. This setup provides insight into how each factor contributes to spawning habitat independently of other variables and their correlations. The high values near 1 on the Y-axis indicate the model predicts a higher probability of encountering conditions suitable for spawning, while low values near 0 suggest a lower probability of suitable conditions based on that specific environmental variable. The values are relative, so actual numbers are less important than the patterns shown in the charts (increasing, decreasing, peaking for example).

Overall, the results generally align with intuition. For instance, the model output from a low maximum downstream local slope starts at a moderate value, increases as the distance above swampy low-lying areas grows, and then decreases at higher values in very steep waterways that likely impede trout passage or lack suitable physical characteristics.

Some variables exhibit patterns that may seem counterintuitive, which can sometimes be explained by the characteristics of the main stem of the Clutha/Mata-Au. Despite not matching traditional spawning stream characteristics (shallow, stable, clear), the river between lakes Wanaka and Dunstan and below Roxburgh Dam are significant spawning locations.

Both Figure 6 and Figure 7 highlight an interesting finding: the presence of a dam downstream surprisingly had minimal influence on the model. Waterways without downstream dams received only slightly higher model outputs than those with dams present. This is likely due to a significant proportion of Otago's waterways lying upstream of dams, particularly upstream of Roxburgh Dam, and the presence of salmonid populations in dammed reservoirs that seek spawning locations.

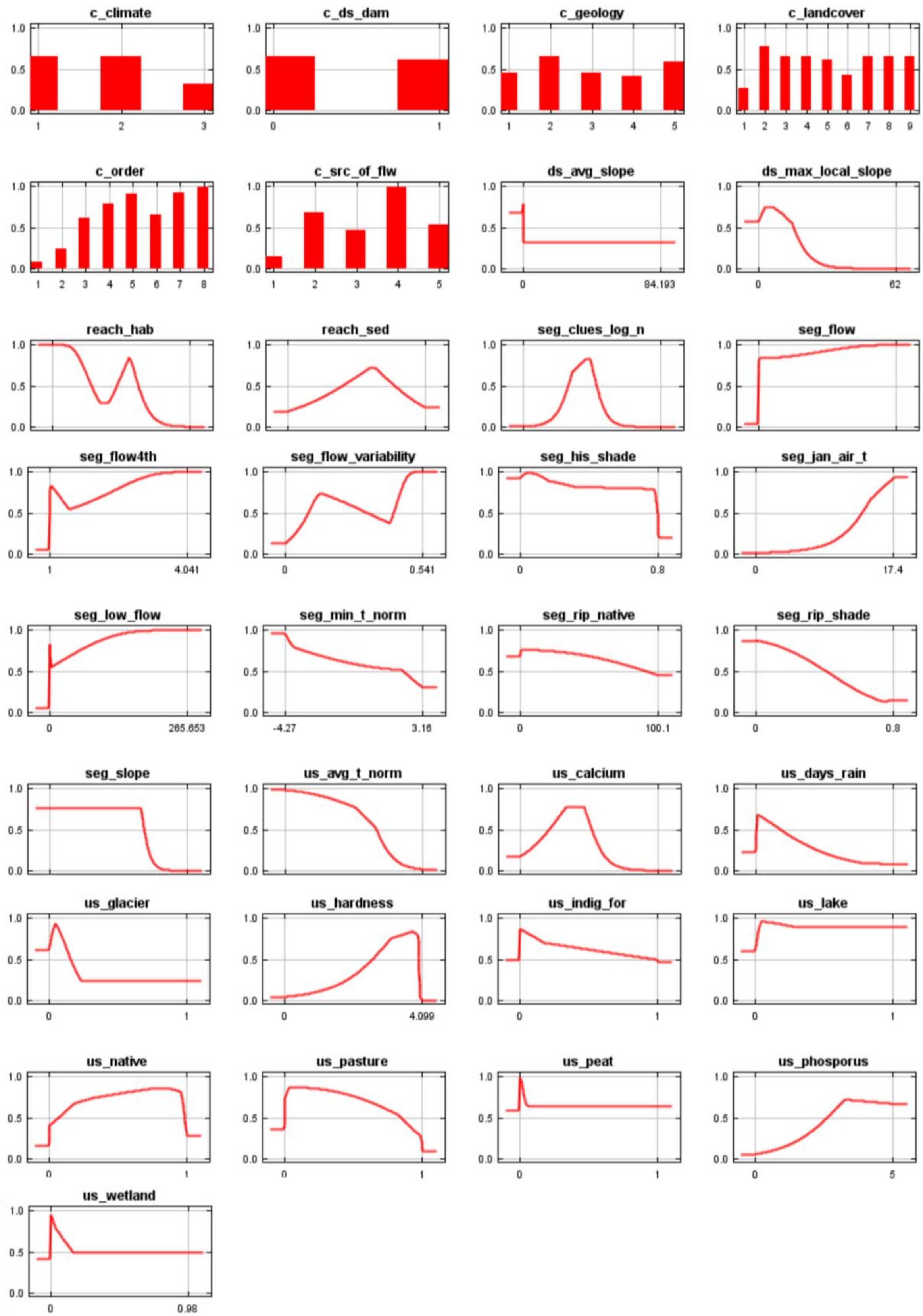

Figure 7: Charts showing how each variable contributes to model output.

Model Example

Suitability bands

The reaches have been color-coded: red indicates a high probability of being suitable spawning habitat, while yellow indicates a medium probability. It is estimated that the red sections will encompass approximately 60% of the suitable spawning reaches, with an additional 30% covered by the yellow sections. This calculation assumes unbiased habitat sampling for model creation and accurate underlying datasets.

The red reaches cover just under 10% of waterways in the region, and including the yellow sections extends this coverage to a further 15%.

Example of model

The model's detailed output is best viewed online due to the size of the Otago Region. You can access it via the following link (Link Here), however an example of the model output is shown in Figure 8.

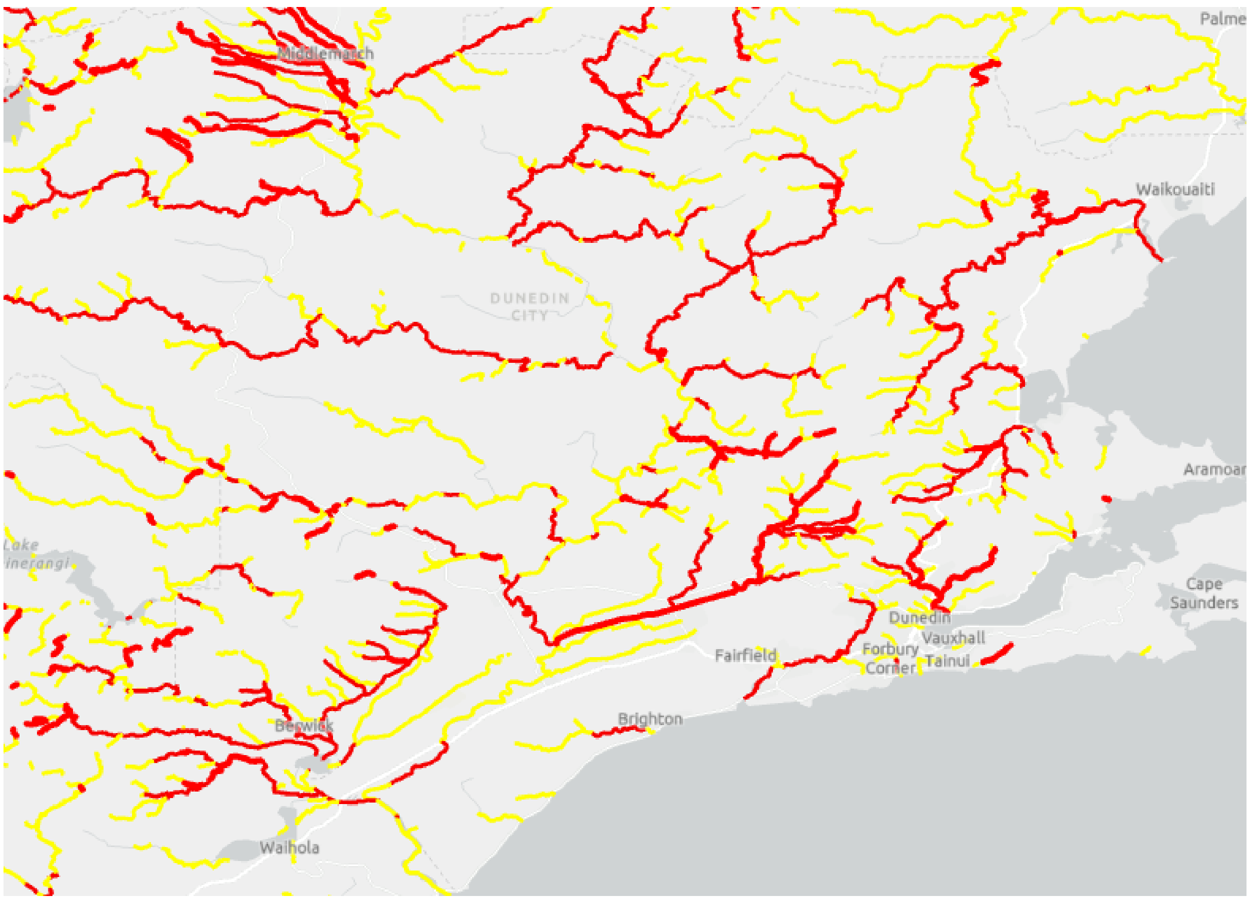

Figure 8: Example output of the spawning model. Dunedin to Middlemarch area.

Discussion/ Summary

Overall, the modelling work effectively portrays the extent of spawning habitat in the Otago Region. This conclusion is supported by several validations: excluding portions of the training data, ground testing, subsequent spawning surveys, and statistical analyses have collectively affirmed the model's accuracy. Continued improvement is anticipated as more data becomes available. It is important to acknowledge several limitations of the final model, which have been addressed in various ways:

Salmonid Barriers

There are numerous salmonid barriers throughout the Otago Region, often crucial for protecting vulnerable and important galaxiid populations from salmonids. The model partially addresses this by using the maximum downstream slope value, which penalises waterways with steep sections below them. Some barriers may have suitable spawning habitat characteristics but remain free of salmonids. Conversely, some barriers do have salmonids above them due to various reasons. Deciding whether a reach above a barrier is suitable for spawning and whether barriers should be maintained, improved or removed requires collaboration with key stakeholders.

The influence of rearing habitat

As the model is trained on juvenile presence data as well as adult spawning data, it consequently models juvenile rearing habitat to some extent. Rearing habitat is similar to spawning habitat and generally occurs in similar areas. Only salmonids under 100mm have been used for training, so this is not likely to drastically affect the model, especially given the spatial scale of the underlying datasets.

Braided rivers

The dataset for braided rivers is not well-suited for our purposes because they are represented by single lines. Spawning in braided rivers primarily occurs in springs and stable braids that shift with floods, appearing and disappearing across the braid plain. To account for this dynamic, polygons representing the probable spawning range have been positioned around braid plains.

Spring Fed Creeks

On average, spring-fed creeks provide better spawning habitat than non-spring-fed creeks, primarily due to their stable flow conditions. Unfortunately, spring-fed creeks are *not well* represented in the underlying datasets and consequently do not score well in the model, largely due to their low stream order and lack of training data. To address this gap, manual additions of important spring-fed spawning streams have been included. These additions encompass streams like Welcome Stream (Waitaki tributary), Diamond Creek (Rees), and Bullock Creek (Lake Wanaka). However, this list does not encompass all significant spring-fed streams and will require ongoing updates.

As a result, users of the model should take a precautionary approach when considering spawning in spring-fed creeks, as the model is more likely to provide false negatives in that context.

Waterways poorly mapped by the datasets

Several waterways have discrepancies between their actual positions and the shapefiles used in the model. To rectify this, these discrepancies, such as those found in the lower Kyeburn, sections of the Pig Burn, and the Kakanui River near State Highway 1, have been manually adjusted to align with their true geographical locations.

Site with previously surveyed spawning or juvenile salmonids

Care has been taken to avoid setting the model threshold for defining spawning too low, as this would cause it to identify an excessive area as potentially suitable for spawning. While this is likely to be more accurate, it would also be uninformative. To prevent this, the thresholds for defining areas have been set conservatively. Consequently, some areas with confirmed juvenile or spawning presence have not met the model's thresholds. To remedy this, an overlay layer marks these locations as confirmed spawning sites despite not being recognised by the model. This layer can be toggled to see its extent.

Multiple species

As the model is trained on data from all species and their spawning habitat preferences are similar, its output can be used to predict the spawning area of all the species, this does however require a knowledge of whether the species are present.

The model was trained on Chinook salmon data but only to a limited extent, only 0.6% of juvenile records were for Chinook. Salmon spawning requirements are notably different to the other salmonids, having a stronger preference for spring or lake fed waterways and favouring larger substrate sizes and faster flows particularly for large sea run salmon. The salmon fishery is at historically low levels in the lower Waitaki, the lower Clutha River Catchment and the Southern Lakes, meaning that protection of spawning is crucial to the return of these species and potentially their long-term survival. Despite this, the understanding of the extent of salmonid spawning grounds is limited, although improving. Its assumed that the model does model salmon spawning habitat to an extent, but that a manual process would improve the dataset.

Acknowledgments

Thanks to the Otago staff particularly Steve Dixon, Ben Sowry and Mason Court, who put in a huge effort over the summer to ground test the modelling work. Thanks also to Dr Adam Canning, Matt Kavermann, Adam Daniel and Nigel Paragreen, who provided datasets, advice and expertise on spawning habitats, modelling and ground testing.

Jayde Couper
Fish & Game Officer
July 2024

References

Baldwin, R. A. (2009). Use of maximum entropy modeling in wildlife research. In *Entropy* (Vol. 11, Issue 4, pp. 854–866). MDPI AG. https://doi.org/10.3390/e11040854

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, *17*(1), 43–57. https://doi.org/10.1111/j.1472-4642.2010.00725.x

Leathwick, J., Julian, K., Elith, J., Chadderton, L., Ferrier, S., & Snelder, T. (2008). *A biologically-optimised environmental classification of New Zealand rivers and streams: reanalysis excluding human impacts variables*. www.niwa.co.nz

Leathwick, J. R., West, D., Chadderton, L., Gerbeaux, P., Kelly, D., Robertson, H., & Brown, D. (2010). *Freshwater Ecosystems of New Zealand (FENZ) Geodatabase*.

Merow, C., Smith, M. J., & Silander, J. A. (2013). A practical guide to MaxEnt for modeling species' distributions: What it does, and why inputs and settings matter. *Ecography*, *36*(10), 1058–1069. https://doi.org/10.1111/j.1600-0587.2013.07872.x

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*(3–4), 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Snelder, T., Biggs, B., & Weatherhead, M. (2004). *New Zealand River Environment Classification User Guide*.

Stoffels R (2022). New Zealand Freshwater Fish Database (extended). The National Institute of Water and Atmospheric Research (NIWA). Sampling event dataset, https://doi.org/10.15468/jbpw92

Appendices

Appendix 1: Description of FENZ database attributes, reproduced from (J. R. Leathwick et al., 2010)

Field	Description
NZREACH	The unique identifier associated with each REC segment
SegJanAirT	Summer (January) air temperature (degrees C) – used in the absence of robust estimates of water temperature
SegMinTNorm	Average minimum daily air temperature (degrees C) normalised with respect to SegJanAirT – negative values indicate strongly seasonal climates and positive values indicate weakly seasonal climates
SegFlow	Mean annual flow (m3/sec), derived from hydrological models, provided by Jochen Schmidt, NIWA, 2006
SegLowFlow	Mean annual 7-day low flow (m3/sec), derived from hydrological models, provided by Jochen Schmidt, NIWA, 2006 – see http://wrenz.niwa.co.nz/webmodel/ for details.
SegFlow4th	4th root transformed mean annual 7-day low flow, i.e., (low flow+1)0.25, accommodates the strong skew in distribution of values when fitting statistical models, values are approximately linearly related to flow velocity
SegFlowVariability	Ratio of annual low flow/annual mean flow – indicates long-term stability of flow through the year
SegSlope	Segment slope (degrees), derived from GIS calculation using length and difference between upstream and downstream elevation for each segment
SegSlopeSqrt	Square-root transformed segment slope (slope +1)0.5
SegRipShade	Riparian shading (proportion), the likely degree of riparian shading derived by using national, satellite image-based vegetation classification to identify riparian shading in each segment, with the degree of shading then estimated from river size and expected vegetation height
SegHisShade	Estimated shade assuming complete vegetation cover as could be expected during pre-human conditions
SegRipNative	Proportion of native riparian vegetation within a 100 m buffer of the river, calculated using landcover information contained in version one of the Landcover Database (LCDB)

Field	Description
SegCluesN	Nitrogen concentration (ppb) as estimated from CLUES, a leaching model combined with a regionally-based regression model, implemented within a catchment framework (Woods et al., 2006)
SegCluesLogN	Log_{10} transformed values of nitrogen concentration
DSDist2Coast	Distance to coast (km), from mid-point of each river segment, recomputed in Hamilton and differing from the original REC estimates of downstream distance that were computed from the upstream end; change made so values indicate average distance from a segment to the coast, rather than the maximum
DSAvgSlope	Average slope (degrees), from mid-point of each river segment to the coast, differing from the original REC estimates of downstream slope that were computed from the upstream end; change made so values indicate average slope from within a segment
DSAvgSlopeSqrt	Square root transformed values of DSAvgSlope , i.e. (slope+1)0.5.
DSMaxLocalSlope	Maximum downstream slope (degrees), local slopes at 100m intervals along each river segment where calculated and maximum value encountered recorded. Each segment was traversed downstream from its mid-point to the coast to identify the maximum downstream value encountered
DSDam	Presence of downstream obstruction (mostly dams) are indicated by a ‘1’, absence by a ‘0’.
USAvgTNorm	Average air temperature (degrees C) in the upstream catchment, normalised with respect to SegJanAirT, with negative values indicating colder (higher elevation) headwaters than average, given the segment temperature, and positive values indicating warmer temperatures
USDaysRain	Days/year with rainfall greater than 25 mm in the upstream catchment to indicate the likely frequency of elevated flows, rainday frequencies were provided by Brett Mullan (NIWA) and were derived by averaging across estimated daily rainfalls over the 10 year period from 1990 to 2000 – indicates short-term stability of flow through the year
USAvgSlope	Average slope in the upstream catchment (degrees), describes catchment-driven modification of flow variability
USCalcium	Calcium concentrations in surface rocks using values derived from the underlying LENZ layers – refer LENZ documentation for details

Field	Description
USHardness	Average hardness (induration) of surface rocks using values derived from the underlying LENZ layers – refer LENZ documentation for details
USPhosporus	Phosphorus concentrations in surface rocks using values derived from the underlying LENZ layers – refer LENZ documentation for details
USPeat	Flow-weighted area of peat in upstream catchment (proportion)
USLake	Lake buffering in the upstream catchment, computed as described in the original REC manual
USWetland	Flow-weighted area of wetland in upstream catchment (proportion)
USIndigFor	Flow-weighted area of indigenous forest in upstream catchment (proportion), computed using cover estimates from LCDB1
USNative	Flow-weighted area of indigenous vegetation in upstream catchment (proportion), computed using cover estimates from LCDB1
USPasture	Flow-weighted area of pasture in upstream catchment (proportion), computed using cover estimates from LCDB1
USGlacier	Flow-weighted area of glacial cover in upstream catchment (proportion), computed using cover estimates from LCDB1
ReachSed	Weighted average of proportional cover of bed sediment using categories of: 1– mud; 2–sand; 3–fine gravel; 4–coarse gravel; 5–cobble; 6–boulder; 7–bedrock, predicted from a boosted regression tree model – details of model fitting are provided in (J. Leathwick et al., 2008)
ReachHab	Weighted average of proportional cover of local habitat using categories of: 1– still; 2–backwater; 3–pool; 4–run; 5–riffle; 6–rapid; 7–cascade, predicted from a boosted regression tree model – details of model fitting are provided in Leathwick et al. (2008)
Shape_Length	Segment shape length in metres as computed by ArcGis

Appendix 2: Description of REC database attributes, reproduced from (Snelder et al., 2004)

Field	Description
Climate	The climate classification procedure is based on two pieces of information, Mean Annual Temperature and Effective Rainfall. 1: Cool- dry 2: Cool- wet 3: Cool- extremely wet
Geology	The classification procedure for the Geology factor was based on an assessment of catchment rock type using the New Zealand Land Resource Inventory (LRI) database. 1: Alluvium 2: Hard-Sedimentary 3: Miscellaneous 4: Soft-Sedimentary 5: Volcanic-Basic
Landcover	The Land-Cover classification was based on information derived from the New Zealand Land Cover Database (LCDB). 1: Bare-Ground 2: Exotic-Forest 3: Indigenous-Forest 4: Miscellaneous 5: Pastoral 6: Scrub 7: Tussock 8: Urban 9: Wetlands
Order	A measure of stream or river size defined by the degree of branching in a drainage system. For example, a first-order stream has no tributaries, while a second-order stream has at least two first-order tributaries. A third-order stream must have at least 2 second-order tributaries (LAWA).
Src_of_flow	The classification of the Source-of-Flow factor for each river section was based on two sets of information: an estimate of the distribution of rainfall within the upstream 100 New Zealand River Environment Classification User Guide catchment and an estimate of the influence of lakes. 1: Glacial/ mountain 2: Hill 3: Low elevation 4: Lake 5: Mountain