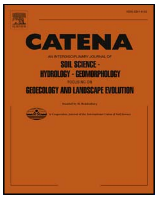
Catena 163 (2018) 130–146

Contents lists available at ScienceDirect

Catena

journal homepage: www.elsevier.com/locate/catena



Predicting river bed substrate cover proportions across New Zealand

Arman Haddadchi*, Doug J. Booker, Richard J. Measures

National Institute of Water and Atmospheric Research, Christchurch, New Zealand

ARTICLE INFO

Keywords:
River-bed
Substrate sizes
National predictions
Downstream fining

ABSTRACT

Predictions of river bed substrate cover are required for various purposes including delineating management zones, linking with ecological status and assessing river rehabilitation options. Three contrasting methods were tested for predicting the proportion of river bed covered by seven different substrate categories: generalised linear models (GLMs), machine learning regression models (random forest), and a summed normal distribution model (SND) which incorporates distribution of predictors and substrate covers throughout the modelling framework. Various predictors representing climate, geomorphology, land cover and geology were derived from existing environmental databases to generate predictive models. Model performance was assessed through a cross-validated comparison with substrate samples collected from 229 river sites distributed across New Zealand. Model performance for 10-fold cross-validated predictions showed that the SND model performed best in predicting the proportions of riverbed covered by bedrock, boulder, cobble and fine gravel categories. Random forest models performed best in predicting coarse gravel, sand and mud plus vegetation proportions. Therefore, combined random forest and SND methods were used for estimating substrate cover proportions at unsampled sites across New Zealand. Texture analysis of predicted substrate cover consistently showed downstream fining of sediment size. The national predictions of substrate cover proportions are key descriptors that can be linked with a wide range of national scale applications for ecological assessment of New Zealand Rivers. The techniques developed and tested are applicable to other locations but it is notable that relatively poor performance in regional cross-validation tests shows that transferability of substrate models to locations with no calibration data is challenging.

1. Introduction

There is a growing requirement for exploring the controls on, and prediction of, substrate cover in rivers. Aquatic biota show strong responses to substrate movement as a direct mechanistically linked indicator of bed disturbance (Jellyman et al., 2013). Sedimentary conditions influence macroinvertebrate community structure (Rempel et al., 2000). River bed grain size also influences suitability of spawning, rearing and feeding habitats for many fish species, particularly salmonids (Kondolf and Wolman, 1993; Armstrong et al., 2003; Hedger et al., 2006). Without suitable stream habitat a given species is unlikely to exist at that particular location (Reiser, 1998; Maddock, 1999). Obtaining a detailed knowledge about the characteristics and spatial distribution of river bed substrate cover over a variety of spatial scales is therefore essential for ecological assessment of rivers.

Understanding longitudinal variations in river bed grain size is important as it has a dominant control on geomorphological and sedimentological regimes. Rivers generally show a downstream fining of sediments (Church and Kellerhals, 1978; Rice, 1998; Morris and Williams, 1999; Ferguson, 2003; Costigan et al., 2014). River bed grain size affects abrasion rates (Frings, 2008), rate and mode of sediment transport (Wilcock and Crowe, 2003; Haddadchi et al., 2013), type and dimension of river bed forms (Buffington and Montgomery, 1997; de Almeida and Rodríguez, 2011), and the size of channel bank deposits (Ten Brinke et al., 2004). Downstream fining of bed material occurs in both gravel-bed and sand-bed rivers (Frings, 2008). However, this general trend can be interrupted by: (i) sedimentation processes in lakes, reservoirs and water conveyance structures; (ii) tributaries which introduce large sedimentary inputs to significantly punctuate this fining trend (Rice, 1998; Benda et al., 2004); (iii) dominated proximal sediment sources from surface soils with dissimilar characteristics established independently from upstream catchment surface soil sources (Haddadchi et al., 2015).

Fining of river bed sediments over the longitudinal profile is commonly modelled using a downstream exponential decrease in grain size:

$$D = D_0 e^{\alpha L} \qquad (1)$$

where D in Eq. (1) is particle size characteristics (i.e., median

---

* Corresponding author at: 10 Kyle Street, Riccarton, Christchurch 8011, PO Box 8602, Christchurch, New Zealand.
E-mail addresses: arman.haddadchi@niwa.co.nz (A. Haddadchi), doug.booker@niwa.co.nz (D.J. Booker), richard.measures@niwa.co.nz (R.J. Measures).

https://doi.org/10.1016/j.catena.2017.12.014
Received 6 January 2017; Received in revised form 27 November 2017; Accepted 13 December 2017

diameter), D_0 is initial particle size diameter (i.e., particle size of most upstream sediment), L is distance downstream (in km) and α is an empirical diminution coefficient (in km^{-1}).

In addition to the effect of chemical weathering and abrasion in situ (Miller et al., 2014; Menting et al., 2015), differential mobility of coarse and fine grains within the bed sediment mixtures leads to downstream fining (Parker and Toro-Escobar, 2002). Therefore, the diminution coefficient reflects cumulative effects of both abrasion and sediment sorting and, thus, it depends on lithology, channel morphology and flow and sediment transport conditions (Powell, 1998).

Measurements of river bed substrate proportions have been carried out for assessing stream habitat in the USA (Herbst and Suk, 2005), New Zealand (Harding et al., 2009) and elsewhere. However, because of the temporal and financial limitations of monitoring river bed grain size via direct observation (Wright et al., 1998), the application of indirect methods based on topographic mapping analyses and remote sensing are growing fast. Channel morphologic measurements derived from traditional digital elevation models together with empirical hydrologic methods have been used to predict bed grain size (Buffington et al., 2004; Gorman et al., 2011). Airborne LiDAR data has been used to identify potential habitat in catchments by estimating river bed grain size (Wilkins and Snyder, 2011; Carbonneau et al., 2012; Rinaldi et al., 2013; Snyder et al., 2013). The main limitation of these approaches is their low accuracy when applied to wetted areas of rivers, especially in rivers with high turbidity and large water depth (Groll et al., 2016).

New Zealand has strong gradients in climate, geology, topography and hydrological regime at the national scale. Various river and catchment characteristics have been mapped onto a national river network describing the spatial configuration of New Zealand's rivers (Snelder and Biggs, 2002). Each segment of the river network has characteristics assigned to it including: catchment area, stream length, elevation and slope derived from digital elevations models; catchment geology derived from geological maps; land cover from remote sensing data; and runoff, rainfall and potential evapotranspiration from climate station data (Leathwick et al., 2011). These characteristics have previously been used to predict the spatial distribution of invertebrate communities (Booker et al., 2015), various fish species (Crow et al., 2013), availability of physical habitat (Snelder et al., 2011a, 2011b; Booker, 2016), hydrological indices (Booker and Woods, 2014), and hydraulic geometry (Booker, 2010) in rivers across New Zealand.

The aim of this study was to predict spatial patterns in substrate characteristics of alluvial river channels across New Zealand from nationally available site and catchment characteristics. To do this, three models with different levels of complexity, data needs and user inputs were used to predict substrate proportions; a generalised linear model (GLM) using an ordinary linear regression, random forest (RF) using machine learning to fit a flexible regression, and summed normal distribution (SND) representing a complex model using distribution of predictors for selection procedure together with genetic algorithm procedure to optimise the results.

The study objectives were: (1) to apply various statistical techniques to elucidate the distribution of river bed substrate covers as a function of upstream catchment characteristics incorporating climate, geomorphology, land cover and hydrological factors; (2) to compare the predictive performance of these techniques when used to make predictions at unvisited sites; (3) to predict river bed substrate proportions at unsampled rivers across New Zealand based on the best performing models; and (4) to increase understanding of controls on sediment characteristics at the national scale.

Fig. 1 outlines the strategy used to predict the substrate cover proportions for river reaches across New Zealand. It involved calculating the areal proportions for each substrate category for each site, extracting predictors and selecting independent variables using combined expert opinion and chi-square tests, independency tests or automated procedures (depending on the type of model being fitted), fitting various types of model to predict each substrate category, and calculating predicted values across the entire river network. Substrate cover proportions calculated from each model type were compared. Several performance metrics were then used to quantify predictive performance.

2. Materials and methods

2.1. Site substrate observations

Field data were assembled from physical habitat studies applied by NIWA (National Institute of Water and Atmospheric Research, New Zealand) and various regional councils at 284 sites across New Zealand. At each site, areal proportions of bedrock (> 512 mm), boulder (256–512 mm), cobble (64–256 mm), gravel (8–64 mm), fine gravel (2–8 mm), sand (0.06–2 mm), mud (< 0.06 mm) and vegetation were observed visually at discrete observation locations across multiple cross-sections. Observation locations were centred at regular intervals across each cross-section except on sections with abrupt changes in bed height, where extra observations were added. Cross-sections were positioned to represent all meso-habitat types (e.g., pool, riffle, run) present within each site. Cross-section average sediment cover by each substrate category was calculated using a weighted mean, with weightings based on the separation of observation points. Reach averaged sediment cover was calculated as a weighted mean of cross-section cover with weightings based on the number of sampling cross-sections, and the proportion of the entire reach area, covered by each mesohabitat type. See Jowett et al. (2008) for further details of field procedures. In total, 73,550 observations were included in the data set (an average of 259 per site). The reach length surveyed at each site ranged from 30 to 3000 m, averaging 330 m per site. The average number of cross-sections at each site was 14, and the average spacing between observation points was 0.84 m. Sampling sites were located throughout the New Zealand river network (Fig. 2) and represented a wide range of river sizes, climatic, topographic and hydrological conditions. See Booker (2016) for further details. Particle size distribution of observed substrates varied between sites, with median diameter (D_{50}) of substrate materials ranging from < 0.06 mm to larger than 100 mm.

Grid co-ordinates and site descriptions from various data providers were used to identify which of the 570,000 reaches that comprise the New Zealand river network best represented the position of each site. There were 37 reaches that had more than one sampled site (two to five sites) assigned to them. The proportion for each substrate category averaged over all sites assigned to the same reach was used to represent substrate proportions at these reaches. This reduced the number of sampled sites from 284 to 229.

2.2. Nationally available predictors

Many environmental variables have previously been mapped onto the New Zealand river network (Leathwick et al., 2008; Booker et al., 2015) and were therefore available as potential explanatory variables for predictive models (Table 1). Climate is represented by various parameters representing different characteristics of precipitation (i.e., usRainDays10, usRainDays25, usRainDays50, usRainDays100, usRainDays200, usAnRainVar), hydrology (usFlow, SpecificMeanFlow, SpecificMALF, FRE3, SpecificAnnualFlood) and temperature (segEquiTSum, segEquiTwin, usPET). Geomorphology is represented by eight parameters such as upstream catchment area (usArea) which is strongly related to wetted width of the river segment (Booker, 2010), average slope of catchment (usAveSlope) calculated from 30-m digital elevation model (DEM), and distance from the coast (dsDistToSea) indicating the location of the site in the river network. Land cover is represented by the proportion of surface area occupied by five categories of land cover (usPastoral, usIndigForest, usExoticForest, usUrban, usScrub; see Table 1 for details). Geology of the upstream catchment, which has a strong influence on the bed material cover of downstream river reaches,