Environmental Modelling and Software 131 (2020) 104738

Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: http://www.elsevier.com/locate/envsoft

# *BaHSYM*: Parsimonious Bayesian hierarchical model to predict river sediment yield

Ottavia Zoboli [*], Gerold Hepp [1], Jörg Krampe, Matthias Zessner

*Institute for Water Quality and Resource Management, TU Wien, Karlsplatz 13/226, 1040 Vienna, Austria*

ARTICLE INFO

*Keywords:*
Bayesian hierarchical modelling
Erosion
Sediment transport
River sediment yield
Spatial prediction
Temporal prediction
Cluster analysis

ABSTRACT

The prediction and control of river sediment yield (SY) are critical but challenging tasks. Erosion and sediment transfer in river catchments are controlled by different processes, whose relative importance varies in space and time. We thus put forward that SY can be estimated more efficiently by using explicitly the information contained in the similarity within groups. To test this hypothesis, we developed a novel Bayesian hierarchical model, applied it to a sample of heterogeneous river catchments and compared its fixed-effects and mixed-effects performance incorporating different group levels, namely gauges, rivers, basins and clusters of catchments. With a parsimonious linear model consisting of four variables (specific and extreme discharge, elevation and retention coefficient), we achieved good performance criteria in the calibration (NSE: 0.79–0.85) and in the cross-validation for temporal and spatial prediction (NSE: 0.71 and 0.72, respectively). These results support the promising potential of this technique.

## 1. Introduction

The delivery of sediments to surface water bodies as a result of soil erosion can exert a critical effect on flood risk (Lane et al., 2007), on the lifetime of reservoirs (Kondolf et al., 2014) and on the health of benthic ecosystems (Greig et al., 2005). It can also be responsible for increased water treatment costs and for the decline of fisheries resources (Bilotta and Brazier, 2008). Further, the transport of sediments is mostly coupled with the transfer of organic carbon, phosphorus and a broad spectrum of particle-bound contaminants from soil into water, which further contribute to the degradation of aquatic environments (Long, 2006; Moran et al., 2017). Prediction and control of riverine sediment transport are thus fundamental goals for water managers worldwide. In this context, models are essential tools to estimate sediment yield (SY, e.g. t $y^{-1}$) at catchment outlets, to interpret spatial and temporal dynamics as well as to quantify and predict the consequences of climate and land use changes. However, the extreme complexity and variability of the processes linking soil erosion with river SY make these tasks very challenging.

The need for estimates of sediment yield and for the understanding of the major factors and processes controlling SY from watersheds across spatial scales is a field of research of long-standing nature (Lane et al., 1997). Yet, the reviews of Merritt et al. (2003) and de Vente et al. (2013), which compared and critically discussed existing models designed to predict soil erosion and sediment yield at catchment scale, revealed a scattered and still unsatisfactory situation. Based on the classification system proposed by Wheater et al. (1993), Merritt et al. (2003) distinguished between empirical, conceptual and physics-based models. Similarly, according to de Vente et al. (2013) models can be conceptually classified as follows: spatially lumped and spatially distributed, empirical, regression, physics-based, and factorial scoring models. LISEM (Roo et al., 1996), PESERA (Kirkby et al., 2008) and SWAT (Arnold et al., 1998) are examples of widely-used physics-based models, which are based on the numerical solution of fundamental physical equations, such as transfer of mass, momentum and energy. Empirical and conceptual models, e.g. WaTEM/SEDEM (Van Oost et al., 2000; Verstraeten and Poesen, 2001), do not make explicit inference on detailed physical processes and rely instead on observed or stochastic relationships between causal variables and sediment yield. This distinction should not be interpreted in absolute terms, because there are also models which consist of a mix of physics-based and empirical components (e.g. AGNPS, Young et al. (1989)). Factorial scoring models like FSM (Verstraeten et al., 2003) are semi-quantitative methods which characterise catchments through factors coupled with scoring and which

* Corresponding author.
*E-mail address:* ozoboli@iwag.tuwien.ac.at (O. Zoboli).
[1] co-first author.

https://doi.org/10.1016/j.envsoft.2020.104738
Received 23 October 2019; Received in revised form 10 May 2020; Accepted 13 May 2020
Available online 3 June 2020
1364-8152/

require in general an expert assessment in the field. Approaches for statistical modelling of SY, which are most typically spatially lumped, include multiple linear regression (e.g. de Vente et al. (2011)) and non-linear regression models (e.g. BQART, Syvitski and Milliman (2007)). The main outcome of Merritt et al. (2003) was that simpler empirical and conceptual approaches were more appropriate for the estimation of SY than physics-based or more complex conceptual models because these were limited by the lack of sufficient spatially distributed data, by the over-dependency of the results on the experience of the modeller and by high computational requirements. The more recent review of de Vente et al. (2013) similarly found that the elevated requirement of calibration parameters for most physics-based models often leads to equifinality and limits their applicability for spatial extrapolations and for scenario studies. Further outcomes of this review were, among others, that: i) the accuracy of existing models differs across spatial and temporal scales and that different models should be selected according to the size of the catchments of interest, ii) a drawback of many models lies in the fact that they depict only selected erosion and sediment transport processes, which limits their applicability to catchments where such processes are dominant and in turn requires an extensive prior knowledge of those (e.g. sheet, rill and ephemeral gully erosion, hillslope erosion and channel erosion), iii) there is definitely need for further model development and for balancing between model complexity and quality of input data.

In this respect, there is a powerful technique that has been so far overlooked in this field. We refer to hierarchical linear models, also known as multilevel models or random coefficient models. In such an approach, data is grouped in a hierarchy of successively higher-level units and, instead of considering observations as independent from each other, it is assumed that groups within each level (e.g. annual SY of gauges, gauges of rivers and rivers of basins) share certain attributes and show similarities. The key idea is to explicitly use this information by considering both the within and between group variances, with the goal to improve model efficiency and estimate accuracy. From the statistical point of view this means that model parameters vary at more than one level and that inferences made about one group affect inferences about another. In other words, the model operates a partial pooling within levels, providing a balanced approach between complete pooling (same intercept and slopes for all data points, i.e. underfitting) and no pooling (individual intercepts and slopes for each data point, i.e. overfitting). Major advantages of hierarchical models are improved estimates for repeated and imbalanced samples, the explicit modelling of the variation across individuals within groups of the data, the fact that there is no need to perform averaging and consequently no associated loss of information as well as an optimal trade-off between overfitting and underfitting (McElreath, 2016). Hierarchical models are a well-established method in social and medical sciences to divide subjects into groups, and they are increasingly used in environmental and ecological sciences, because they enable incorporating cross-scales interactions and thus enhance the model effectiveness both in understanding causal effects and in prediction (Quian et al., 2010). Further, thanks to the exceptional progress of computational power achieved over the last decades, it is now technically feasible to develop hierarchical models within a Bayesian framework. This offers important possibilities, among which the explicit incorporation of prior knowledge into the model and the obtainment of probability distributions of both model parameters and estimates. The latter in turn allows for a thorough analysis of the significance and the uncertainty of the results (Gelman and Hill, 2006).

Based on these characteristics, we hypothesise that this technique holds a considerable potential to efficiently and reliably predict SY at catchment level and that it might help to overcome, at least partially, the limitation of having to rely on different models depending on the scale and the properties of the catchments. The main consideration behind this hypothesis is that on the one hand there are common processes regulating soil erosion and transport of sediments in all catchments and that on the other hand their relative importance changes greatly in space and time. We can for instance expect that catchments with similar morphological traits, hydrological regimes and land use also show similarities in the dominant processes determining SY at their outlets. Further, individual catchments typically belong to larger groups, such as river systems or basins. This constitutes an ideal problem for Bayesian hierarchical models, which are designed to optimally use the information contained in the variability of the data across nested levels.

A widely-used technique in hydrology, Top-kriging (Skøien et al., 2006), relies on a similar idea. It also makes use of the fact that information provided by gauges of the same river system helps to predict a streamflow-related variable at an unobserved location better than information provided by gauges of other river systems. While Top-kriging incorporates similarity of topological relation and geographical location only, *BaHSYM* is also capable of incorporating similarity of other factors.

The goal of this work is to test our hypothesis by developing a Bayesian hierarchical model able to describe and predict SY in heterogeneous river catchments in Austria. For the development and validation of the model we consider both temporal and spatial cross-validation. According to the outcomes of their review, de Vente et al. (2013) discarded linear regression equations as suitable prediction models, since in a number of case studies they proved unstable and unsuitable to extrapolate sediment yield beyond the calibration datasets (Verstraeten and Poesen, 2001; Grauso et al., 2008; Haregeweyn et al., 2008; de Vente et al., 2011). We hypothesise that the Bayesian hierarchical approach holds the theoretical power to considerably improve the performance and reliability of otherwise unstable linear regressions. This is why in this work we develop and test parsimonious linear regression models consisting of few explanatory variables.

## 2. Methods

### 2.1. Study area

The study area includes 30 Austrian river catchments for which data of suspended solids transported at the outlet is available for multiple years with high temporal resolution. This sample consists of catchments that are highly heterogeneous in their total area (135–10,660 $km^2$), average elevation (256–2495 m a.s.l.), mean slope (9–61%), mean discharge (1256 $m^3 s^{-1}$) and land use. Most of the gauges are located in alpine or mountainous areas, whereas agricultural catchments in lowland are more sparse. Fig. 1 depicts the geographical location of the gauges, whereas the basic properties of the corresponding catchments are reported in Table 1. It is important to observe the enormous temporal variability of sediment transport in the dataset, with annual SY varying in some cases up to an order of magnitude at the very same gauge.

### 2.2. BaHSYM (Bayesian hierarchical sediment yield model)

*General model structure* . *BaHSYM* consists of a linear regression model embedded within a Bayesian hierarchical approach. The basic level that we consider in the hierarchical model structure are individual gauges, i.e. we assume that at each gauge SY observations in different years are not fully independent and that dominant processes are to a certain degree similar. Further, we hypothesise that a specific gauge shares more similarities with other gauges within the same river system or within the same basin than with gauges located in other river systems or basins. Therefore, we created two model variants in which we nested the first level into a second higher-level, namely the level of rivers or that of basins, respectively. To evaluate whether and how much this hierarchical structure improves the model performance, we also tested the model without any level, which would correspond to an ordinary linear regression model in a non-Bayesian context. We refer to the variant without levels as fixed-effects model and to the variants with hierarchical levels, i.e. random-effects, as mixed-effects models. In a

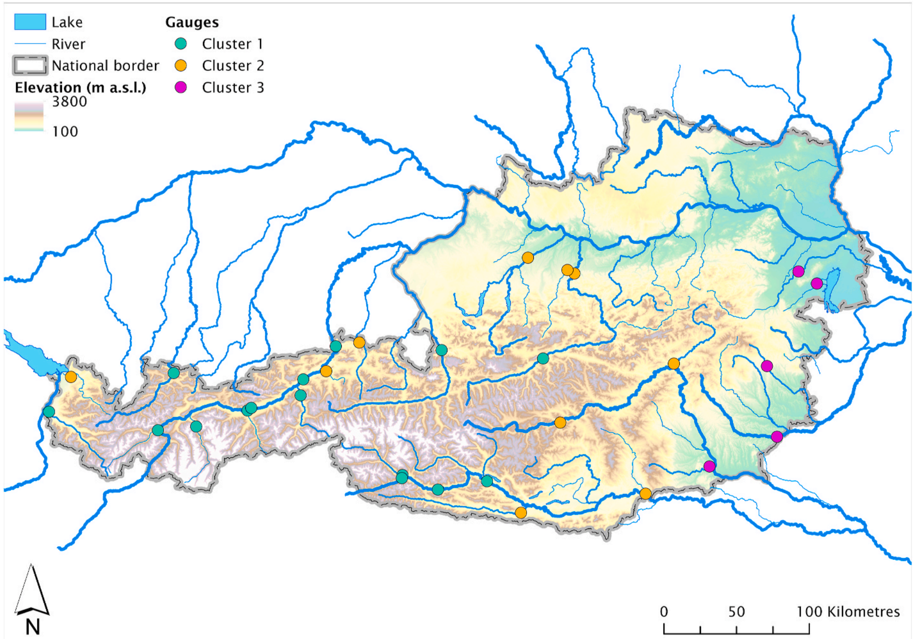


**Fig. 1.** Location of the catchment outlet gauges, from which the SY dataset stems. Gauges are reported with different colours depending on the cluster they belong to (please refer to Sections Section 2.5. and Section 3.1. for details regarding the clusters). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

**Table 1**
Basic properties of the catchments whose outlet gauge data is used to test the model: A (total area), E (average elevation), S (average slope), Q (long-term mean discharge), SY (range of annual sediment yield in the period 2009–2014).

| Basin | River | Gauge | A ($km^2$) | E (m a.s.l.) | S (%) | Q ($m^3s^{-1}$) | SY (1000 t $y^{-1}$) |
|---|---|---|---|---|---|---|---|
| Bregenzerach | Bregenzerach | Kennelbach | 838 | 1121 | 40 | 46 | 177–643 |
| Drau | Drau | Lienz-Falkensteinsteg | 658 | 1905 | 56 | 14 | 55–200 |
| Drau | Drau | Dellach | 2172 | 1981 | 56 | 63 | 286–622 |
| Drau | Drau | Amlach | 4779 | 1832 | 54 | 129 | 330–811 |
| Drau | Drau | Lavamünd | 10660 | 1379 | 43 | 256 | 70–498 |
| Drau | Isel | Lienz | 1202 | 2145 | 58 | 39 | 276–768 |
| Enns | Enns | Trautenfels | 1518 | 1468 | 48 | 49 | 49–335 |
| Enns | Enns | Jägerberg | 5942 | 1162 | 49 | 166 | 173–1048 |
| Drau | Gail | Federaun | 1297 | 1286 | 45 | 44 | 77–525 |
| Ill | Ill | Gisingen | 1230 | 1626 | 54 | 65 | 128–343 |
| Inn | Inn | Innsbruck | 5786 | 2148 | 57 | 164 | 863–2310 |
| Inn | Inn | Rattenberg | 8483 | 2027 | 56 | 257 | 1113–3167 |
| Inn | Inn | Oberaudorf | 9683 | 1917 | 54 | 305 | 1327–4103 |
| Inn | Brixentaler Ache | Bruckhäusl | 326 | 1318 | 44 | 11 | 48–478 |
| Inn | Großache | Kössen-Hütte | 706 | 1201 | 43 | 27 | 151–834 |
| Inn | Ötztaler Ache | Tumpen | 781 | 2495 | 57 | 26 | 338–916 |
| Inn | Salzach | Golling | 3620 | 1577 | 52 | 141 | 181–1332 |
| Inn | Sanna | Landeck-Bruggen | 728 | 2139 | 58 | 20 | 51–379 |
| Inn | Sill | Innsbruck-Reichenau | 856 | 1911 | 56 | 25 | 82–379 |
| Inn | Ziller | Hart im Zillertal | 1135 | 1897 | 57 | 45 | 96–434 |
| Lech | Lech | Lechaschau | 1012 | 1761 | 61 | 44 | 178–626 |
| Leitha | Leitha | Deutsch Brodersdorf | 1592 | 694 | 28 | 9 | 6–53 |
| Mur | Mur | Gestüthof | 1701 | 1637 | 44 | 36 | 23–134 |
| Mürz | Mürz | Kapfenberg-Diemlach | 1506 | 1071 | 44 | 22 | 16–83 |
| Pinka | Pinka | Pinkafeld | 135 | 743 | 25 | 1 | 4–22 |
| Raab | Raab | Neumarkt | 1009 | 492 | 20 | 7 | 20–33 |
| Steyr | Steyr | Pergern | 919 | 946 | 48 | 36 | 18–94 |
| Sulm | Sulm | Leibnitz | 1113 | 582 | 23 | 16 | 19–156 |
| Traun | Traun | Wels-Lichtenegg | 3379 | 847 | 31 | 129 | 56–294 |
| Wulka | Wulka | Schützen am Gebirge | 404 | 256 | 9 | 1 | 2–9 |

mixed-effects model the fixed-effects describe the effect sizes of the overall mean and the random-effects the individual effect sizes of the hierarchical levels, which have to be either added to or subtracted from the fixed-effects.

Annual SY was not modelled as such, but instead as the logarithmic transformation of the specific annual sediment yield (SSY, t $km^{-2}y^{-1}$). Choosing SSY instead of SY enables to overcome the masking effect of different catchment sizes. The logarithmic transformation was necessary to meet the assumption of normality.

The set of equations (1) presents the mathematical formulation of the mixed-effects model with two different group levels. It is formulated using non-centered parametrization, i.e. with subtraction of the mean (fixed-effects) and factored out standard deviations of the random-effects:

$$\log(SSY_i) \sim \text{Normal}(\mu_i, \sigma) \tag{1a}$$

$$\mu_i = X_{ij}\mathscr{B}_{j_i} \tag{1b}$$

$$\mathscr{B}_{j_i} = \beta_j + \beta_{j,Level_{1_i}} \odot \sigma_{j,Level_1} + \beta_{j,Level_{2_i}} \odot \sigma_{j,Level_2} \tag{1c}$$

$$\beta_j \sim \text{Normal}(0, 0.5) \tag{1d}$$

$$\beta_{j,Level_{1_i}} \sim \text{MVNormal}\left(0_j, \mathrm{P}_{jk,Level_1}\right) \tag{1e}$$

$$\beta_{j,Level_{2_i}} \sim \text{MVNormal}\left(0_j, \mathrm{P}_{jk,Level_2}\right) \tag{1f}$$

$$\mathrm{P}_{jk,Level_1} \sim \text{LKJcorr}(1) \tag{1g}$$

$$\mathrm{P}_{jk,Level_2} \sim \text{LKJcorr}(1) \tag{1h}$$

$$\sigma_{j,Level_1} \sim \text{Exponential}(1) \tag{1i}$$

$$\sigma_{j,Level_2} \sim \text{Exponential}(1) \tag{1j}$$

$$\sigma \sim \text{Exponential}(1) \tag{1k}$$

where log ($SSY_i$) is assumed to be normally distributed with mean $\mu_i$ and standard deviation σ. $X_{ij}$ represents the matrix of explanatory variables; $\mathscr{B}_{j_i}$ the combined fixed- and random-effects (i.e. mixed-effects); $\beta_j$ the slopes of the fixed-effect; $\beta_{j,Level_{1i}}$ and $\beta_{j,Level_{2i}}$ the slopes of the random-effects of the two different group levels; $\sigma_{j,Level_1}$ and $\sigma_{j,Level_2}$ the factored out standard deviations of the slopes of the random-effects. The symbol ⊙ represents the Hadamard product, which means variable-wise multiplication of the slopes of the random-effects with their corresponding standard deviation in the present case. Furthermore, $P_{jk,Level_1}$ and $P_{jk,Level_2}$ stand for the correlation matrices of the slopes of the random-effects. With respect to the matrix/vector indices, $i$ indexes observations (rows of the matrix of explanatory variables), whereas $j$ and $k$ index explanatory variables (columns of the matrix of explanatory variables, vectors of the effect sizes of the explanatory variables, rows and columns of the correlation matrices of the explanatory variables).

Due to the fact that the response variable has been centered and scaled and thus has a mean of zero, the model is formulated without any intercept. Adding an intercept in such a case would not significantly improve model predictions.

Equations (1d)–(1k) define standard priors for centered and scaled variables according to McElreath (2020) (see Section 2.3 for details on centering and scaling). While these standard priors do not improve model predictions, they prevent implausible parameter values. Furthermore, having only centered and scaled explanatory variables allows us to use the same standard priors for all of them. Apart from the possibility to incorporate subjective or prior information into the model, in our opinion an underestimated advantage of Bayesian statistics is that it can make certain modelling steps simpler. For example, back transformation of logarithmic response variables, i.e. exponentiation, is straight forward and does not require a correction such as the one explained by Laurent (1963).

### 2.3. Explanatory variables

We considered a pool of potential explanatory variables for *BaHSYM*, which are reported in Table 2. Since this work focuses on the methodological approach and not so much on identifying the best performing variables, we constrained the selection to a set of relatively few attributes, which have the theoretical potential to explain the spatial and/or temporal variability of erosion and sediment transfer at catchment scale. We thus have chosen average and extreme hydrological variables, morphometric traits of the catchments as well as main land uses. Additionally, we have considered the sediment retention coefficient ξ, a variable conceived by Gavrilovic (1976) as sediment delivery component to be combined with an erosion rate in a semi-quantitative model to predict SY at basin scale (de Vente and Poesen, 2005). It is thus based on characteristics which mainly influence sediment transport and retention processes, such as morphometric traits and waterway length. It is calculated as follows:

$$\xi = \frac{\sqrt{Pe \times E}\left(L_p + L_a\right)}{\left(L_p + 10\right)A} \tag{2}$$

where $Pe$ is the perimeter of the catchment (km), $E$ the average elevation in km a.s.l., $L_p$ the length of the principal waterway (km) and $L_a$ the cumulated length of secondary waterways (km).

The selection of these variables was driven by expert knowledge regarding the specific study area. In other regions, different variables might be more suitable and relevant. All variables, including the response variable, have been centered and scaled for modelling purposes. Centering refers to the practice of subtracting the sample mean from all values of a variable, whilst scaling describes the practice of dividing all values of a variable by its sample standard deviation. As a result, all variables have a mean of zero and a standard deviation of one and their effect sizes in the model can be compared independently of the scale of the original variables. The additional centering and scaling of the response variable mean that increasing or decreasing an explanatory variable by one standard deviation, while keeping all other explanatory variables constant, causes the response variable to change by one standard deviation times the effect size of the adjusted explanatory variable. Increasing an explanatory variable with, for example, an effect size of

**Table 2**
Description of selected potential explanatory variables for the model.

| Variable | Description (unit) |
|---|---|
| *Hydrological and morphometric attributes* | |
| A | Total area of the catchment ($km^2$) |
| E | Average elevation of the catchment (m a.s.l.) |
| S | Average slope of the catchment (%) |
| C | Compactness ratio: square root function of the ratio between A and the area of the circle having the same perimeter (-) |
| ξ | Sediment retention coefficient calculated according to equation (2) (-) |
| $l_p$ | Specific main channel length: total length of the principal waterway normalised by catchment area ($km^{-1}$) |
| $l_a$ | Specific tributary length: total length of secondary waterways normalised by catchment area ($km^{-1}$) |
| q | Specific discharge: average annual river discharge normalised by catchment area ($m^3s^{-1}km^{-2}$) |
| $Q_{p95}$ | Extreme discharge: maximum annual fraction of discharge above the 95th percentile of monthly discharge (%) |
| $P_{p95}$ | Extreme precipitation: maximum annual fraction of precipitation above the 95th percentile of monthly precipitation (%) (cf. Hanel et al., 2016) |
| *Land use* | |
| Glc | Percentage of total catchment area occupied by glaciers (%) |
| Agr | Percentage of total catchment area occupied by arable land (%) |
| $Agr_4$ | Percentage of total catchment area occupied by arable land with slope > 4 % (%) |
| $Agr_8$ | Percentage of total catchment area occupied by arable land with slope > 8 % (%) |
| Nat | Percentage of total catchment area with natural cover (mostly forests) (%) |
| Alp | Percentage of total catchment area occupied by bare alpine surfaces (%) |
| Grl | Percentage of total catchment area occupied by grassland (%) |
| Lakes | Percentage of total catchment area occupied by lakes (%) |

plus 0.5 by 0.5 would therefore cause the response variable to change by 0.25 standard deviations. Thus, centering and scaling eases the interpretation of the modelling results.

When testing different combinations of explanatory variables, we selected for each run a maximum of three to four variables to keep the model parsimonious. Apart from additive effects, we also tested multiplicative interactions between the variables.

### 2.4. Data

For 27 out of 30 gauges, data on daily loads of transported suspended solids were provided for the years 2009–2014 by the Directorate IV/4 – Water Balance (Hydrographical Central Office) of the Austrian Federal Ministry for Sustainability and Tourism (delivered in May 2017). For the rivers Pinka and Wulka, data on suspended solids concentrations corresponding to 48 h composite samples were provided for the same period of time by the Provincial Government of Burgenland. As for the river Raab, data for the years 2009–2014 stems from a station equipped with devices for continuous monitoring of water quality parameters, which is operated by the Institute for Water Quality and Resource Management of the TU Wien on behalf of the Austrian Federal Ministry for Sustainability and Tourism (Camhy et al., 2013; Fuiko et al., 2016). For all gauges, daily or more detailed available loads were summed up to obtain annual sediment yields for each year, i.e. the modelled response variable in the proposed *BaHSYM* approach.

Daily precipitation data with 1 × 1 km spatial resolution were extracted from the SPARTACUS dataset of the Central Institution for Meteorology and Geodynamics (Hiebl and Frei, 2017). Daily discharges were obtained from the Austrian Hydrographical Service (eHYD, 2017). Elevation and slope data stem from the official digital terrain model of Austria, which has a spatial resolution of 10 × 10 m (DEM, 2016). Detailed land use data at catchment scale for the period 2009–2014, including river network length and lakes surface, was made available by Clara et al. (2019).

### 2.5. Model combined with catchment clustering

Should the model be applied for spatial prediction, i.e. to estimate annual SY for unmonitored locations, it would not be meaningful to use specific gauges as group level. For this purpose we employed and tested a variation of the mixed-effects model, in which the group level consists of clusters of similar catchments only. In this way, it shall be possible to predict SY for new gauges by assigning them to one of the identified clusters.

We formed clusters of catchments by following a two-step procedure. In the first place, we carried out a Principal Component Analysis (PCA, Jolliffe, 1989) based on a subset of the variables reported in Table 2, including topographic attributes, traits of the surface water network, land use and river discharge. The subset of variables, together with their value for each catchment, are reported by Zoboli and Hepp (2020).

**Table 3**
Principal components obtained in the PCA: percentage and cumulative explained variance of SY.

| Component | Explained variance (%) | Cumulative explained variance (%) |
|---|---|---|
| 1 | 49.0 | 49.0 |
| 2 | 14.5 | 63.5 |
| 3 | 10.8 | 74.3 |
| 4 | 8.4 | 82.7 |
| 5 | 5.6 | 88.3 |
| 6–12 | 11.7 | 100.0 |

Thus, in a second step we used the first two principal components, which explain approximately 63% of the total variance (Table 3), to identify clusters of catchments. To perform the cluster analysis, we applied the Partitioning Around Medoids (PAM) algorithm (Park and Jun 2009; Kaufman et al., 1987). The identified clusters are described and discussed in Section 3.1.

### 2.6. Model evaluation

As described in the previous sections, we tested different combinations of group levels, which bring into being the *BaHSYM* variants depicted by the set of equations (3). These equations correspond to variants of equation (1c), whereas the rest of the general modelling approach is the same for all. For the reasons discussed previously, we tested equations (3a)–(3c) for temporal prediction (e.g. to fill yearly gaps), whereas equation (3d) was employed for spatial prediction (e.g. to extrapolate SY for catchments without monitoring of sediment transport).

$$\mathscr{B}_{ji} = \beta_j + \beta_{j,Gauge_i} \odot \sigma_{j,Gauge} \tag{3a}$$

$$\mathscr{B}_{ji} = \beta_j + \beta_{j,Gauge_i} \odot \sigma_{j,Gauge} + \beta_{j,River_i} \odot \sigma_{j,River} \tag{3b}$$

$$\mathscr{B}_{ji} = \beta_j + \beta_{j,Gauge_i} \odot \sigma_{j,Gauge} + \beta_{j,Basin_i} \odot \sigma_{j,Basin} \tag{3c}$$

$$\mathscr{B}_{ji} = \beta_j + \beta_{j,Cluster_i} \odot \sigma_{j,Cluster} \tag{3d}$$

The *BaHSYM* variants were tested through a k-fold cross-validation procedure. To test the use of the model for spatial prediction (to estimate SY for out-of-sample gauges), we applied a 10-fold cross-validation leaving out gauges stratified by cluster. The available data was split in ten training and test sets. Given the 30 available gauges, each training set consists of approximately 27 sites and each test set of approximately three different sites each time. It is important to note that all annual observations at one site stay together each time, i.e. it cannot happen that, for example, the years 2009, 2010, 2012 and 2014 of one site are in the training set and the years 2011 and 2013 of the same site are in the test set. This type of cross-validation is solely used to test the model's capability to predict SY at new sites. Furthermore, stratifying by clusters makes sure that each test set contains approximately one site of each cluster. The goodness of fit metrics are then calculated from the collected predictions of all test sets. Since the available dataset of SY comprises six years, to test the performance of the models for temporal prediction (to estimate SY for out-of-sample years), we applied a 6-fold leave-one-year-out cross-validation. The six years of available data was split into six training and test sets. Each training set consists of the data corresponding to all sites for five years, whereas each test set contains the data corresponding to all sites for the remaining year (in each fold a different year). This type of cross-validation is solely used to test the model's capability to predict SY for new years. The goodness of fit metrics are then likewise calculated from the collected predictions of all test sets.

Following criteria were selected to quantify the performance of the models in estimating annual SY: Nash-Sutcliffe Efficiency (NSE), Modified Nash-Sutcliffe Efficiency (mNSE) and R squared ($R^2$). NSE measures the goodness of fit of the model, with values ranging from $-\infty$ for poor predictive power to one for perfect match between modelled values and data (Nash and Sutcliffe, 1970). mNSE is a modification of NSE which is less sensitive to extreme values and is more influenced by low values (Krause et al., 2005). We applied it to better evaluate the model performance for catchments with relatively smaller SSY. In addition, we selected the Root Mean Square Error (RMSE) and percent Bias (PBIAS)

(Yapo et al., 1996). Whereas RMSE calculates the standard deviation of the model prediction error, PBIAS indicates the average tendency of modelled values to be larger or smaller than the observed ones. In addition, *BaHSYM* was applied to the whole dataset to identify the best-fit model. The general best-fit model is presented and discussed in Section 3.2, while the results for temporal and spatial prediction are described in Section 3.3.

### 2.7. Software

The model was developed and tested with *R* version 3.6.1 (R Core Team, 2019). In particular, we made use of the *brms* package version 2.10.0, which is specifically designed to implement Bayesian multilevel models in *R* using the probabilistic programming language *Stan* (Bürkner, 2017, 2018). Codes and dataset are publicly available on the Zenodo platform (Zoboli and Hepp, 2020).

## 3. Results and discussion

### 3.1. Clusters of catchments

The three groups of catchments obtained through the cluster analysis are depicted in Fig. 1 and their detailed composition and specific attributes are reported on the Zenodo platform (Zoboli and Hepp, 2020). Here, their main characteristics are described. Before analysing their distinctive traits, it is important to observe that they largely differ in size. Clusters No. 1, 2 and 3 are composed by 15, ten and five catchments respectively. This unbalance is mainly caused by the fact that available SY data stems in the majority from gauges located in alpine and mountainous regions, whereas lowland areas are rather under-represented. Fig. 2 shows for each cluster the range of variation of the variables selected for the PCA.

The first cluster (No. 1) is characterised by a median elevation of 1911 m and a median slope of 56%. Other distinctive attributes are the presence of glaciers and large shares of alpine bare areas and alpine grassland. The second cluster (No. 2) is mainly composed of mountainous catchments, although with lower median elevation (1181 m) and slope (44%). In this cluster, glaciers and bare areas are present as well, though they occupy a much smaller share of land, whereas forests and grassland are largely prevalent. The third and smallest cluster (No. 3) has a median elevation of 582 m and a median slope of 23%. There are obviously no glaciers, and bare areas and grassland are far less important. This cluster has a high average share of forest cover, but its most distinct trait is the significantly larger presence of steep arable land. With respect to specific discharge, the first two clusters are quite similar with a median value of 0.03 $m^3\ s^{-1}\ km^{-2}$, whereas the third cluster presents a significantly lower value of 0.01 $m^3\ s^{-1}\ km^{-2}$. Total catchment area is not a distinctive factor for these clusters. We can only observe that whereas cluster No. 3 presents a very narrow range of variation around small areas, the other two are characterised by broad ranges.

### 3.2. Best-fit model

The best model and cross-validation results were obtained with an additive model comprising four variables, namely average elevation (E), specific discharge (q), extreme discharge ($Q_{p95}$) and sediment retention coefficient (ξ). It performs best in its variant with two group levels: gauges and rivers. This model is described in the set of equations (4), where priors consist of standard priors for centered and scaled variables according to McElreath (2020) and are described by equations (1d)–(1k).

$$\log(SSY_i) \sim \text{Normal}(\mu_i, \sigma) \tag{4a}$$

**Table 4**
Estimate and significance of the slope parameters ($\beta_j$) for the input explanatory variables of *BaHSYM* (best variant with two group levels: gauges and rivers) applied to the whole dataset (SE: Standard Error; CI: Confidence Interval).

| Variable | Estimate | SE | 99% CI | 90% CI |
|---|---|---|---|---|
| E | 0.46 | 0.14 | 0.04–0.83 | 0.22–0.69 |
| q | 0.64 | 0.09 | 0.42–0.88 | 0.50–0.78 |
| $Q_{p95}$ | 0.27 | 0.07 | 0.10–0.46 | 0.16–0.38 |
| ξ | 0.25 | 0.13 | − 0.13–0.59 | 0.03–0.46 |

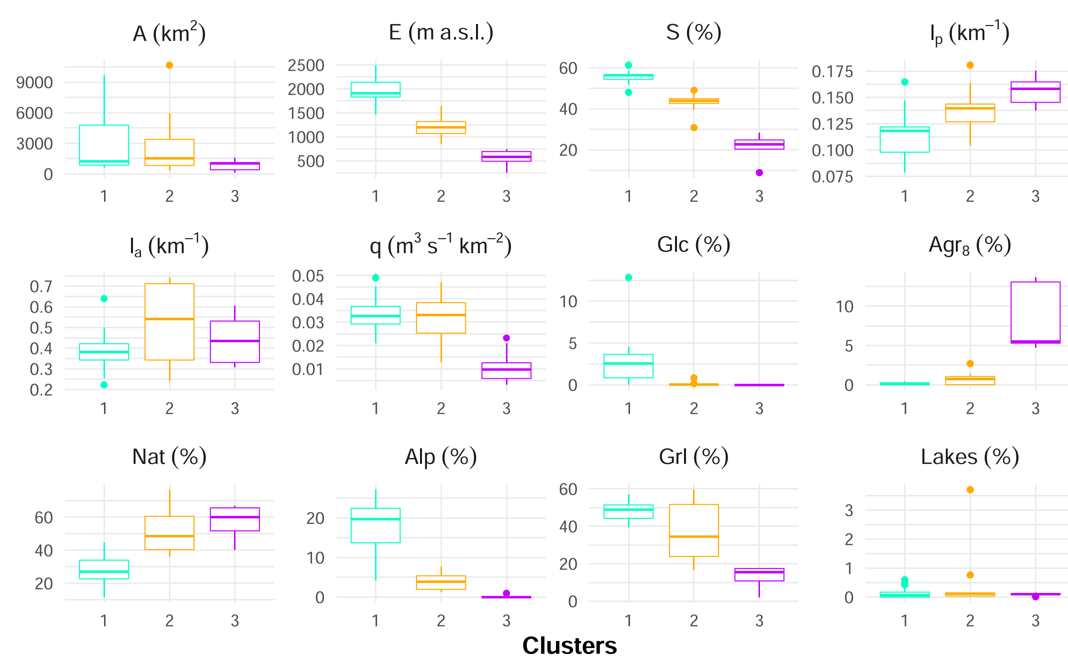


**Fig. 2.** Ranges of variation of the variables selected for the cluster analysis within the three resulting clusters.

$$\mu_i = X_{ij}\mathscr{B}_{j_i} \tag{4b}$$

$$X_{ij} = \left[E_i q_i Q_{p95_i} \xi_i\right] \tag{4c}$$

$$\mathscr{B}_{j_i} = \beta_j + \beta_{j,Gauge_i} \odot \sigma_{j,Gauge} + \beta_{j,River_i} \odot \sigma_{j,River} \tag{4d}$$

As reported in Table 4, the slope parameters ($\beta_j$) of the first three variables are statistically significant with a 99% confidence interval, whereas this is true for the sediment retention coefficient ξ with a 90% confidence interval. An investigation of the residuals furthermore confirmed that the assumption of normality was met and that the residuals do not show any signs of non-linearity or heteroscedasticity. The residuals plot is reported in Fig. A1 in the Appendix.

Although catchment area has been considered in the past to be a relevant predictor for SSY, the critical review of de Vente et al. (2007) concluded that, depending on scales and regional specificities, the relation between A and SSY can vary from positive to negative and is often non-linear. It is thus in general a poor predictor of SSY. This is in line with our outcomes. In fact, although A is indirectly considered in this work as component of the sediment retention coefficient, its inclusion as separate variable does not improve the model performance. Slope is almost interchangeable with elevation, although E performs slightly better. The fact that they hold a similar explanatory power can be explained through their very high correlation coefficient of 0.92. Likewise, high correlation coefficients between most land use variables and E explain to a large extent why these do not bring almost any improvement to the performance of the parsimonious model indicated above. The study of Gericke and Venohr (2012) on erosion in German mountainous catchments similarly found a strong correlation between SY and average elevation. For a complete overview of the correlation coefficients between the variables reported in Table 2, please refer to Fig. A2 in the Appendix.

Fig. 3 graphically shows the comparison between observed and modelled SSY obtained with the best *BaHSYM* variant with two group levels (gauges and rivers), whereas Table 5 reports the performance criteria for all *BaHSYM* variants applied to the whole dataset.

In support of our initial hypothesis, we achieve a notable improvement of the model performance through the technique of partial pooling. This is true for all variants of *BaHSYM* with distinct group levels,

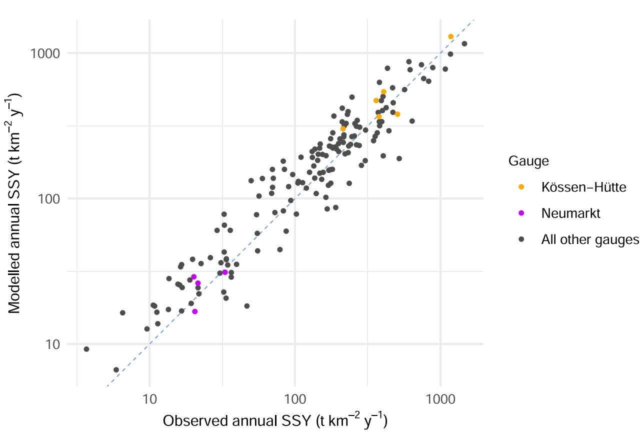


**Fig. 3.** Observed annual SSY vs. annual SSY modelled with *BaHSYM* (best variant with two group levels: gauges and rivers) applied to the whole dataset. The blue dashed line indicates the perfect match between modelled and observed values. The results for two gauges are highlighted in colour, as examples of different performance of the model for specific gauges. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

**Table 5**
Results of the different *BaHSYM* variants applied to the whole dataset for the estimation of annual SSY.

| Model | $R^2$ | NSE | mNSE | RMSE | PBIAS |
|---|---|---|---|---|---|
| | (-) | (-) | (-) | (t km$^{-2}$) | (%) |
| *Fixed-effects model* | 0.70 | 0.69 | 0.43 | 133 | 6.9 |
| *Mixed-effects model utilising equation (3a)* | | | | | |
| One group level: gauges | 0.84 | 0.83 | 0.62 | 96 | 5.5 |
| *Mixed-effects model utilising equation (3b)* | | | | | |
| Two group levels: gauges, rivers | 0.85 | 0.85 | 0.63 | 92 | 6.1 |
| *Mixed-effects model utilising equation (3c)* | | | | | |
| Two group levels: gauges, basins | 0.85 | 0.85 | 0.63 | 93 | 5.5 |
| *Mixed-effects model utilising equation (3d)* | | | | | |
| One group level: catchment clusters | 0.80 | 0.79 | 0.54 | 108 | 11.3 |

although they lead to partially different outcomes. The variant based on catchment clusters as group level brings a notable enhancement compared to the fixed-effects model, with $R^2$ raised from 0.70 to 0.80, NSE from 0.69 to 0.79 and mNSE from 0.43 to 0.54 respectively, whilst RMSE was lowered from 1.33 t ha$^{-1}$ to 1.08 t ha$^{-1}$. This improvement comes however at the expense of a bias increase from 6.9 to 11.3%. Nevertheless, it was by partially pooling over gauges and over the combination of gauges with rivers or basins that we achieved the real breakthrough in model performance. $R^2$ increased to 0.84–0.85, NSE to 0.83–0.85, mNSE to 0.62–0.63, whereas RMSE was decreased below 1 t ha$^{-1}$. Further, these three variants even reduced the bias.

Despite the improvements, we can observe that mNSE is consistently lower than NSE in all model variants. In other words, the model's estimates are more reliable for catchments with greater SSY, independently from their size. This is exemplified by the performance of the best *BaHSYM* variant with two group levels (gauges and rivers) for two gauges highlighted in colour in Fig. 3. The model performed very well for the gauge Kössen-Hütte, located in the mountainous river Großache and included in cluster No. 2 ($R^2$: 0.93, NSE: 0.88, mNSE: 0.56, RMSE: 106 t km$^{-2}$, PBIAS: 10.3%). The performance was however rather poor for the gauge Neumarkt, located in the rather lowland river Raab and included in cluster No. 3 ($R^2$: 0.32, NSE: −0.11, mNSE: −0.07, RMSE: 6 t km$^{-2}$, PBIAS: 9.7%). While their performance visually might appear comparable, the goodness of fit metrics clearly show the difference. Especially the NSE is a sensitive metric for the relative relationship of the magnitude of modelled residual variance and measured data variance. In this respect, viewed in isolation, the model captures the measured data variance of Kössen-Hütte way better than of Neumarkt, which is also clearly reflected by the $R^2$ metric.

### 3.3. Model for temporal and spatial prediction

Taking into account the huge temporal variability of SY observed in our dataset, the fixed-effects model does not perform badly for temporal predictions, with a NSE of 0.65 and a RMSE of 1.41 t ha$^{-1}$ (Table 6). This

**Table 6**
Results of the 6-fold leave-one-year-out cross-validation for different *BaHSYM* variants tested for temporal prediction of annual SSY.

| Model | $R^2$ | NSE | mNSE | RMSE | PBIAS |
|---|---|---|---|---|---|
| | (-) | (-) | (-) | (t km$^{-2}$) | (%) |
| *Fixed-effects model* | 0.66 | 0.65 | 0.39 | 141 | 6.9 |
| *Mixed-effects model utilising equation (3a)* | | | | | |
| One group level: gauges | 0.72 | 0.71 | 0.50 | 128 | 7.2 |
| *Mixed-effects model utilising equation (3b)* | | | | | |
| Two group levels: gauges, rivers | 0.73 | 0.71 | 0.49 | 127 | 8.6 |
| *Mixed-effects model utilising equation (3c)* | | | | | |
| Two group levels: gauges, basins | 0.73 | 0.71 | 0.49 | 127 | 8.4 |

means that the selected combination of variables has a high explanatory power with respect to the temporal variability of SSY. Two of the four variables, namely specific and extreme discharge, are time-dependent and especially the latter largely varies from one year to the next.

Table 6 shows the considerable improvement of the model performance for temporal prediction achieved through partial pooling. The three *BaHSYM* variants achieve the same NSE of 0.71. The combination of gauges with rivers or with basins in a mixed-effects model with two group levels slightly improves RMSE from 1.41 t ha$^{-1}$ to 1.27 t ha$^{-1}$, at the expense of a small bias increase. Taking individual gauges as single group level instead than the combination of two group levels leads to a better fit for individual catchments, also for the ones with smaller SSY, which is reflected in a slightly higher mNSE of 0.50. Nevertheless, this improvement comes at a small expense of the general performance, since $R^2$ and RMSE reflect to a greater extent the dominance of catchments with high SSY. The ability of the fixed-effects model for temporal prediction is thus enhanced by the structure of *BaHSYM.* Including random-effects allows the effect sizes of the explanatory variables to be correlated. Making use of at least one explanatory variable that varies in time can cause the effect sizes of the other explanatory variables to depend on the effect size of that time-dependent variable. For example, in a year with high extreme discharge, the effect size of the retention coefficient (which does not vary in time) can be different from the one it has in a year with low extreme discharge. While such correlations are often not statistically significant, they still affect the predictive power of models making use of random-effects. In other words, random-effects have the potential to add sometimes complicated "it-depends-structures" to a linear regression model.

de Vente et al. (2013) state that extrapolating SY for different years for catchments that were used for calibration generally leads to better validation results than extrapolating SY for out-of-sample catchments, since differences in e.g. land use and dominant erosion processes between calibration and validation datasets are relatively small. In our case, the model performs almost equally well for both purposes. The two hydrological variables in the model are fundamental to describe temporal variability, but their combination with elevation and the morphometric variables contained in the sediment retention coefficient also allow capturing to a great extent the spatial variability.

This model application further supports our hypothesis regarding the potential of Bayesian hierarchical models in this field. As reported in Table 7, adding partial pooling over clusters of catchments notably improves all performance criteria. Although lowland catchments with generally lower SSY are under-represented in the sample available for model training, partial pooling over the clusters significantly improves mNSE from 0.39 to 0.48. This implies that even though the cluster of this type of catchments is relatively small, it conveys a significant amount of information on the different erosion and sediment transport processes that distinguish these catchments from the mountainous and alpine ones. That is an exemplary benefit of this technique in case of unbalanced datasets. Nevertheless, it is clear from the difference between NSE (0.72) and mNSE (0.48) that the good performance of the model is dominated by mountainous and alpine catchments with greater erosion and larger transfer of sediments. Our outcomes show that the idea of combining *BaHSYM* with clusters of catchments as group level holds a great potential for spatial extrapolation, but they also reveal its limitations. In order to apply the model to make robust predictions, it is essential that new catchments share fundamental similarities with the available clusters. In our case, given the largely heterogeneous sample available, elevation, slope, land use and discharge were all important factors to determine the clustering. However, if a more homogeneous sample was available, more specific and targeted criteria could be used.

The outcomes of *BaHSYM* are very promising when compared to those of ordinary linear regressions. For example, de Vente et al. (2011) attempted to spatially extrapolate SY and SSY based on linear regression and on a wide number of variables for a sample of 61 catchments with areas comprised between 30 and 13,000 km$^2$ in Spain. Although they could achieve quite good results for calibration (NSE: 0.58), the model performed very poorly in the validation step, with a NSE of $-0.10$. For the same dataset, they did achieve better validation results (NSE of 0.35–0.67 with spatially distributed models and 0.72 with the Factorial Score Model), but such models are more complex and require considerably more data and expert assessments than the variants of *BaHSYM* presented in this paper (de Vente et al., 2008; de Vente and Poesen, 2005). Our benchmark is not the model by de Vente et al. (2011) per se, but rather the use of ordinary linear regressions of which their study is an example. Nevertheless, we tested the performance of their model for our study area. To do that, we reproduced the *BaHSYM* model with the variables selected by de Vente et al. (2011) to model SY and SSY: i) two topographic variables, namely average slope (%), also termed "mean slope gradient", based on a 25 × 25 m digital terrain model and "relief ratio" (m km$^{-2}$) calculated as $(E_{max} - E_{min})/A$; ii) the climatic variable "precipitation concentration index" (%) calculated as $\sum_{i=1}^{12} p_i^2/P^2 \times 100$, where $p_i$ is the average monthly precipitation (mm) and $P$ is the average annual precipitation (mm); iii) the land use variable "Matorral" and iv) three lithological and soil texture type variables, namely percentage of "acid metamorphic rock", "limestone" and "Fluvisols" (%, ESDB (2004)). With respect to the land use variable, there is no perfect match

**Table 7**
Results of the 10-fold cross-validation leaving out gauges stratified by cluster for different *BaHSYM* variants tested for spatial prediction of annual SSY.

| Model | $R^2$ | NSE | mNSE | RMSE | PBIAS |
|---|---|---|---|---|---|
| | (-) | (-) | (-) | (t km$^{-2}$) | (%) |
| *Fixed-effects model* | 0.66 | 0.64 | 0.39 | 142 | 6.1 |
| *Mixed-effects model utilising equation (3d)* | | | | | |
| One group level: catchment clusters | 0.74 | 0.72 | 0.48 | 128 | 10.5 |

**Table 8**
Results of *BaHSYM* variants tested for the estimation of annual SSY with the variables of the ordinary regression model developed by de Vente et al. (2011). Results are reported for the model with sparsely vegetated area as land use variable.

| Model | $R^2$ | NSE | mNSE | RMSE | PBIAS |
|---|---|---|---|---|---|
| | (-) | (-) | (-) | (t km$^{-2}$) | (%) |
| **Model fit to whole dataset** | | | | | |
| *Fixed-effects model* | 0.21 | 0.10 | 0.08 | 226 | 15.3 |
| *Mixed-effects model utilising equation (3a)* | | | | | |
| One group level: gauges | 0.61 | 0.60 | 0.47 | 150 | 6.1 |
| *Mixed-effects model utilising equation (3b)* | | | | | |
| Two group levels: gauges, rivers | 0.61 | 0.60 | 0.47 | 150 | 6.1 |
| *Mixed-effects model utilising equation (3c)* | | | | | |
| Two group levels: gauges, basins | 0.61 | 0.60 | 0.47 | 150 | 6.1 |
| *Mixed-effects model utilising equation (3d)* | | | | | |
| One group level: catchment clusters | 0.36 | 0.29 | 0.21 | 201 | 15.6 |
| **Cross-validation for temporal prediction** | | | | | |
| *Fixed-effects model* | 0.13 | $-0.11$ | $-0.05$ | 250 | 19.3 |
| *Mixed-effects model utilising equation (3a)* | | | | | |
| One group level: gauges | 0.43 | 0.38 | 0.29 | 187 | 9.7 |
| *Mixed-effects model utilising equation (3b)* | | | | | |
| Two group levels: gauges, rivers | 0.43 | 0.37 | 0.28 | 188 | 10.7 |
| *Mixed-effects model utilising equation (3c)* | | | | | |
| Two group levels: gauges, basins | 0.41 | 0.36 | 0.27 | 190 | 8.9 |
| **Cross-validation for spatial prediction** | | | | | |
| *Fixed-effects model* | 0.0009 | $-18.7$ | $-1.2$ | 1055 | 107 |
| *Mixed-effects model utilising equation (3d)* | | | | | |
| One group level: catchment clusters | 0.007 | $< -10^3$ | $< -10^3$ | $> 10^6$ | $> 10^6$ |

for "Matorral" in Austria, which corresponds to a Mediterranean and sub-Mediterranean evergreen bush and scrub land use type. Instead, we tested three variables closely resembling this land use type in the Austrian landscape, namely percentage of sparsely vegetated area, transitional woodland-shrub and natural grassland (%, Corine Land Cover codes 333, 324 and 321, respectively). The data and R code for this version of the model are fully available on the Zenodo platform (Zoboli and Hepp, 2020). The results of this comparative exercise are reported in Table 8. The original fixed-effects model overall did not perform well and an important reason might lie in the variables, which were selected for catchments very different from the ones included in our study area. Nevertheless, it is interesting to observe that adding group levels, i.e. turning it into a mixed-effects model, did improve the best-fit model (NSE of 0.29–0.60 instead of 0.10) as well as the model for temporal prediction (NSE of 0.36–0.38 instead $-$ 0.11) considerably. However, adding group levels failed for the use of the model for spatial prediction (NSE $< -10^3$ instead of $-$ 18.7). The comparison of these two models is thus useful to show that more complex structures, such as the one of *BaHSYM*, can indeed be very beneficial, but only in combination with suitable variables for each application. Otherwise, they might even worsen the model performance.

## 4. Conclusions and outlook

The outcomes of this study support the hypothesis that Bayesian hierarchical models hold a great potential to improve the prediction of sediment yield in rivers. We have shown that through the implementation of this technique even parsimonious linear regression models can provide relatively robust temporal and spatial extrapolations. This means that with a reduced amount of data availability for few variables, this technique enables filling annual gaps, performing predictions for future scenarios and extrapolating SY for catchments without monitoring of sediment transport. We have also shown that through *BaHSYM* the limitation of having unbalanced datasets for the model training is partially compensated for. Nevertheless, the power of this technique can overcome the lack of information only to a certain extent. The robustness and reliability of the predictions remain constrained by the availability of sediment transport data. For the Austrian case study, for example, it is evident that at present an enhanced monitoring network would be required in lowland catchments with dominant erosion on arable land.

What we put forward is the use of this technique for an enhanced extrapolation of sediment yield across scales, but the model per se will likely need to be adapted for each case study. *BaHSYM* shall be thus seen as a methodological approach. Specific purpose, data availability and required temporal and spatial scales shall determine in each application the most adequate variables and group levels to be used.

Future lines of research include upgrading *BaHSYM* via advanced correlation structures (e.g. Gaussian processes), formulating informative priors as well as extending the application of this technique to the investigation of the selected transfer of particulate-bound contaminants in river catchments.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

Financial support for this study was provided by the Austrian Federal Ministry of Sustainability and Tourism. We would like to thank the Associate Editor and three anonymous reviewers for providing exceptionally detailed and constructive feedback to a previous version of the manuscript.

## Appendix A. Residuals and correlation matrix

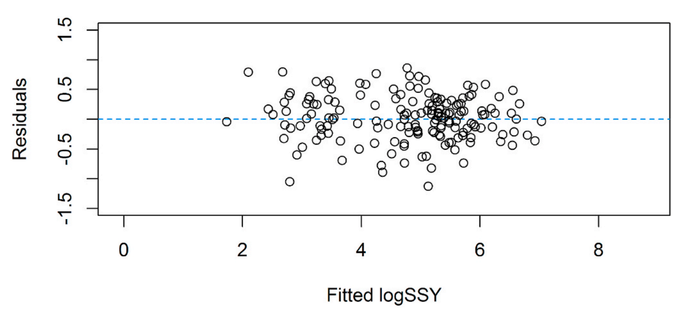


**Fig. A1.** Plot of residuals for the best-fit model.

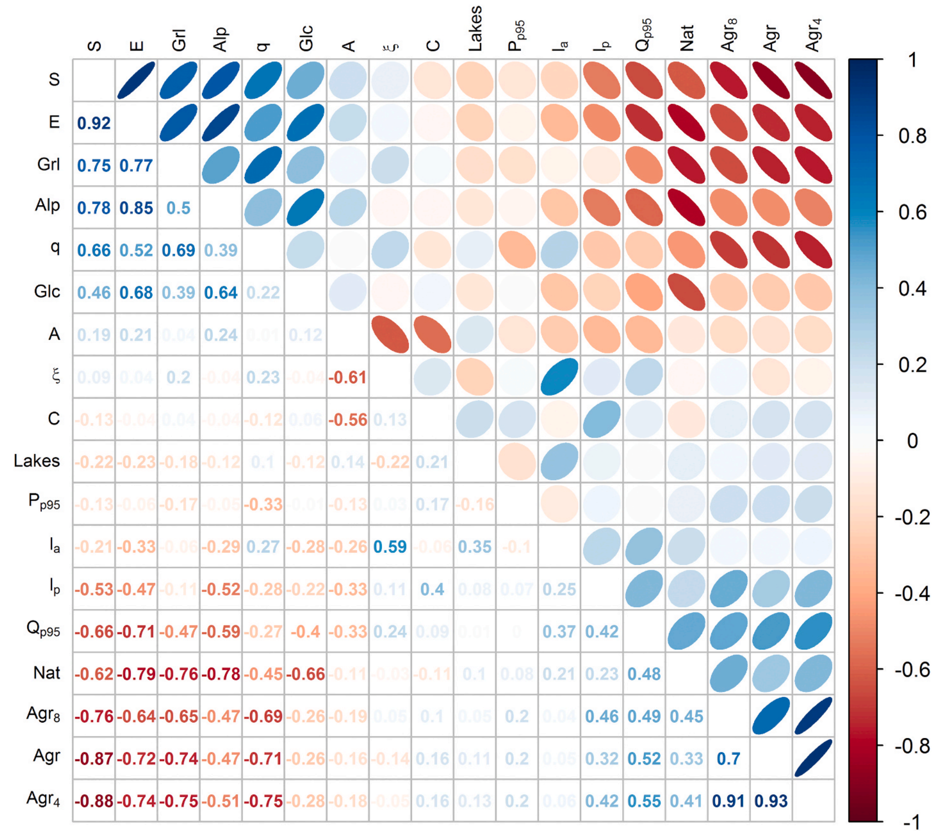


**Fig. A2.** Correlation matrix for the explanatory variables tested in the model.

## References


Arnold, J.G., Srinivasan, R., Muttiah, R.S., Williams, J.R., 1998. Large area hydrologic modeling and assessment Part I: model Development1. JAWRA J. Am. Water Resour. Assoc. 34, 73–89. https://doi.org/10.1111/j.1752-1688.1998.tb05961.x.

Bilotta, G.S., Brazier, R.E., 2008. Understanding the influence of suspended solids on water quality and aquatic biota. Water Res. 42, 2849–2861. https://doi.org/10.1016/j.watres.2008.03.018.

Bürkner, P.C., 2017. brms: an R package for Bayesian multilevel models using Stan. J. Stat. Software 80, 1–28. https://doi.org/10.18637/jss.v080.i01.

Bürkner, P.C., 2018. Advanced Bayesian multilevel modeling with the R package brms. The R J. 10, 395–411. https://doi.org/10.32614/RJ-2018-017.

Camhy, D., Ertl, T., Fuiko, R., Gamerith, V., Gruber, G., Hofer, T., Höller, M., Kinzer, C., Kleidorfer, M., Kornfeind, L., Leonhardt, G., Muschalla, D., Pressl, A., Rauch, W., Steffelbauer, D., Steger, B., Svardal, K., Urich, C., Winkelbauer, A., Winkler, S., 2013. Integrierte Betrachtung eines Gewässerabschnitts auf Basis kontinuierlicher und validierter Langzeitmessreihen. Technical Report. Bundesministerium für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft. Vienna.

Clara, M., Amann, A., Gabriel, O., Hochedlinger, G., Humer, M., Humer, F., Kittlaus, S., Kulcsar, S., Scheffknecht, C., Trautvetter, H., Zessner, M., Zoboli, O., 2019. STOBIMO Spurenstoffe. Stoffbilanzmodellierung für Spurenstoffe auf Einzugsgebietsebene. Technical Report. Bundesministerium für Nachhaltigkeit und Tourismus, Vienna.

DEM, 2016. Digital terrain model of Austria, based on airborne laserscanning. http://www.geoland.at.

eHYD, 2017. EHYD Hydrographic Data. Federal Ministry of Sustainability and Tourism. https://ehyd.gv.at.

ESDB, 2004. The European Soil Database Distribution Version 2.0. European Commission and the European Soil Bureau Network, CD-ROM. EUR 19945 EN. https://esdac.jrc.ec.europa.eu/content/european-soil-database-v20-vector-and-attribute-data.

Fuiko, R., Kornfeind, L., Schilling, K., Weilguni, S., Winkelbauer, A., 2016. Nachhaltige Wassergütewirtschaft Raab. Online – Monitoring. Technical Report. Bundesministerium für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft. Vienna.

Gavrilovic, S., 1976. Bujicni tokovi i erozija (Torrents and erosion). Gradevinski kal. 159–311.

Gelman, A., Hill, J., 2006. Data Analysis Using Regression and Multilvel/Hierarchical Models. Cambridge University Press, New York, New York (USA).

Gericke, A., Venohr, M., 2012. Improving the estimation of erosion-related suspended solid yields in mountainous, non-alpine river catchments. Environ. Model. Software 37, 30–40. https://doi.org/10.1016/j.envsoft.2012.04.008.

Grauso, S., Pagano, A., Fattoruso, G., De Bonis, P., Onori, F., Regina, P., Tebano, C., 2008. Relations between climatic–geomorphological parameters and sediment yield in a mediterranean semi-arid area (Sicily, southern Italy). Environ. Geol. 54, 219–234. https://doi.org/10.1007/s00254-007-0809-4.

Greig, S.M., Sear, D.A., Carling, P.A., 2005. The impact of fine sediment accumulation on the survival of incubating salmon progeny: implications for sediment management. Sci. Total Environ. 344, 241–258. https://doi.org/10.1016/j.scitotenv.2005.02.010.

Hanel, M., Máca, P., Bašta, P., Vlnas, R., Pech, P., 2016. The rainfall erosivity factor in the Czech Republic and its uncertainty. Hydrol. Earth Syst. Sci. 20, 4307–4322. https://doi.org/10.5194/hess-20-4307-2016.

Haregeweyn, N., Poesen, J., Nyssen, J., Govers, G., Verstraeten, G., de Vente, J., Deckers, J., Moeyersons, J., Haile, M., 2008. Sediment yield variability in Northern Ethiopia: a quantitative analysis of its controlling factors. Catena 75, 65–76. https://doi.org/10.1016/j.catena.2008.04.011.

Hiebl, J., Frei, C., 2017. Daily precipitation grids for Austria since 1961–development and evaluation of a spatial dataset for hydroclimatic monitoring and modelling. Appl. Climatol. 132, 327–345. https://doi.org/10.1007/s00704-017-2093-x.

Jolliffe, I., 1989. Principal Component Analysis. Springer, New York.

Kaufman, L., Rousseeuw, P., Dodge, Y., 1987. Clustering by means of medoids. In: Statistical Data Analysis Based on the L1 Norm and Related Methods. North-Holland, Amsterdam, pp. 405–416.

Kirkby, M.J., Irvine, B.J., Jones, R.J.A., Govers, G., 2008. The PESERA coarse scale erosion model for Europe. I. – model rationale and implementation. Eur. J. Soil Sci. 59, 1293–1306. https://doi.org/10.1111/j.1365-2389.2008.01072.x.

Kondolf, G.M., Gao, Y., Annandale, G.W., Morris, G.L., Jiang, E., Zhang, J., Cao, Y., Carling, P., Fu, K., Guo, Q., Hotchkiss, R., Peteuil, C., Sumi, T., Wang, H.W., Wang, Z., Wei, Z., Wu, B., Wu, C., Yang, C.T., 2014. Sustainable sediment management in reservoirs and regulated rivers: experiences from five continents. Earth's Future 2, 256–280. https://doi.org/10.1002/2013EF000184.

Krause, J., Boyle, D., Base, F., 2005. Comparison of different efficiency criteria for hydrological model assessment. Adv. Geosci. 5, 89–97.

Lane, L., Hernandez, M., Nichols, M., 1997. Processes controlling sediment yield from watersheds as functions of spatial scale. Environ. Model. Software 12, 355–369.

Lane, S.N., Tayefi, V., Reid, S.C., Yu, D., Hardy, R.J., 2007. Interactions between sediment delivery, channel change, climate change and flood risk in a temperate upland environment. Earth Surf. Process. Landforms 32, 429–446. https://doi.org/10.1002/esp.1404.

Laurent, A., 1963. The lognormal distribution and the translation method: description and estimation problems. J. Am. Stat. Assoc. 58, 231–235. https://doi.org/10.1080/01621459.1963.10500844.

Long, E.R., 2006. Calculation and uses of mean sediment quality guideline quotients: a critical review. Environ. Sci. Technol. 40, 1726–1736. https://doi.org/10.1021/es058012d.

McElreath, R., 2016. Statistical Rethinking: A Bayesian Course with Examples in R and Stan. CRC Press, Boca Raton, FL (USA).

McElreath, R., 2020. Statistical Rethinking: A Bayesian Course with Examples in R and Stan, second ed. CRC Press, Boca Raton, FL (USA).

Merritt, W., Letcher, R., Jakeman, A., 2003. A review of erosion and sediment transport models. Environ. Model. Software 18, 761–799. https://doi.org/10.1016/S1364-8152(03)00078-1.

Moran, P.W., Nowell, L.H., Kemble, N.E., Mahler, B.J., Waite, I.R., Van Metre, P.C., 2017. Influence of sediment chemistry and sediment toxicity on macroinvertebrate communities across 99 wadable streams of the Midwestern USA. Sci. Total Environ. 599–600, 1469–1478. https://doi.org/10.1016/j.scitotenv.2017.05.035.

Nash, J., Sutcliffe, J., 1970. River flow forecasting through conceptual models part I – A discussion of principles. J. Hydrol. 10, 282–290. https://doi.org/10.1016/0022-1694(70)90255-6.

Van Oost, K., Govers, G., Desmet, P., 2000. Evaluating the effects of changes in landscape structure on soil erosion by water and tillage. Landsc. Ecol. 15, 577–589. https://doi.org/10.1023/A:1008198215674.

Park, H., Jun, C., 2009. A simple and fast algorithm for K-medoids clustering. Expert Syst. Appl. 36, 3336–3341.

Quian, S., Cuffney, T., Alameddine, I., McMahon, G., Reckhow, K., 2010. On the application of multilevel modeling in environmental and ecological studies. Ecology 91, 355–361. https://doi.org/10.1890/09-1043.1.

R Core Team, 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/.

Roo, A.P.J.D., Wesseling, C.G., Ritsema, C.J., 1996. Lisem: a single-event physically based hydrological and soil erosion model for drainage basins. I: theory, input and output. Hydrol. Process. 10, 1107–1117. https://doi.org/10.1002/(SICI)1099-1085(199608)10:8<1107::AID-HYP415>3.0.CO;2-4.

Skøien, J.O., Merz, R., Blöschl, G., 2006. Top-kriging - geostatistics on stream networks. Hydrol. Earth Syst. Sci. 10, 277–287. https://doi.org/10.5194/hess-10-277-2006.

Syvitski, J., Milliman, J., 2007. Geology, geography, and humans battle for dominance over the delivery of fluvial sediment to the coastal ocean. J. Geol. 115, 1–19.

de Vente, J., Poesen, J., 2005. Predicting soil erosion and sediment yield at the basin scale: scale issues and semi-quantitative models. Earth Sci. Rev. 71, 95–125. https://doi.org/10.1016/j.earscirev.2005.02.002.

de Vente, J., Poesen, J., Arabkhedri, M., Verstraeten, G., 2007. The sediment delivery problem revisited. Prog. Phys. Geogr. 31, 155–178. https://doi.org/10.1177/0309133307076485.

de Vente, J., Poesen, J., Verstraeten, G., Van Rompaey, A., Govers, G., 2008. Spatially distributed modelling of soil erosion and sediment yield at regional scales in Spain. Global Planet. Change 60, 393–415. https://doi.org/10.1016/j.gloplacha.2007.05.002.

de Vente, J., Verduyn, R., Verstraeten, G., Vanmaercke, M., Poesen, J., 2011. Factors controlling sediment yield at the catchment scale in NW Mediterranean geoecosystems. J. Soils Sediments 11, 690–707. https://doi.org/10.1007/s11368-011-0346-3.

de Vente, J., Poesen, J., Verstraeten, G., Govers, G., Vanmaercke, M., Van Rompaey, A., Arabkhedri, M., Boix-Fayos, C., 2013. Predicting soil erosion and sediment yield at regional scales: where do we stand? Earth Sci. Rev. 127, 16–29. https://doi.org/10.1016/j.earscirev.2013.08.014.

Verstraeten, G., Poesen, J., 2001. Factors controlling sediment yield from small intensively cultivated catchments in a temperate humid climate. Geomorphology 40, 123–144. https://doi.org/10.1016/S0169-555X(01)00040-X.

Verstraeten, G., Poesen, J., de Vente, J., Koninckx, X., 2003. Sediment yield variability in Spain: a quantitative and semiqualitative analysis using reservoir sedimentation rates. Geomorphology 50, 327–348. https://doi.org/10.1016/S0169-555X(02)00220-9.

Wheater, H., Jakeman, A., Beven, K., 1993. Progress and directions in rainfall-runoff modelling. Model. Change Environ. Syst. 101–132.

Yapo, P.O., Gupta, H.V., Sorooshian, S., 1996. Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data. J. Hydrol. 181, 23–48. https://doi.org/10.1016/0022-1694(95)02918-4.

Young, R., Onstad, C., Bosch, D., Anderson, W., 1989. AGNPS: a nonpoint-source pollution model for evaluating agricultural watersheds. J. Soil Water Conserv. 44, 168–173.

Zoboli, O., Hepp, G., 2020. Dataset and codes for "BaHSYM: parsimonious bayesian hierarchical model to predict river sediment yield. URL: https://zenodo.org/deposit/3514721.