

## Unpacking dasymetric modelling to correct spatial bias in environmental model outputs

Marko Kallio<sup>a,c,\*</sup>, Joseph H.A. Guillaume<sup>b</sup>, Peter Burek<sup>c</sup>, Sylvia Tramberend<sup>c</sup>, Mikhail Smilovic<sup>c</sup>, Alexander J. Horton<sup>a</sup>, Kirsi Virrantaus<sup>a</sup>

<sup>a</sup> Aalto University, Otakaari 24, 02150, Espoo, Finland

<sup>b</sup> The Australian National University, Canberra ACT, 2600, Australia

<sup>c</sup> International Institute of Applied Systems Analysis, Schlossplatz 1, A-2361, Laxenburg, Austria

### ARTICLE INFO

#### Keywords:

Bias correction  
Dasymetric modelling  
Downscaling  
Areal interpolation  
Environmental modelling  
meta-Modelling

### ABSTRACT

Complex environmental model outputs used to inform decisions often have systematic errors and are of inappropriate resolution, requiring downscaling and bias correction for local applications. Here we provide a new interpretation of dasymetric modelling (DM) as a spatial bias correction framework useful in environmental modelling. DM is based on areal interpolation where estimates of some variable at target zones are obtained from overlapping source zones using ancillary information. We explore DM by downscaling runoff output from a distributed hydrological model using two meta-models and describe the properties of the methodology in detail. Consistent with properties of linear scaling bias correction, results show that the methodology 1) reduces errors compared to the source data and meta-models, 2) improve the spatial structure of the estimates, and 3) improve the performance of the downscaled estimates, particularly where meta-models perform poorly. The framework is simple and useful in ensuring spatial coherence of downscaled products.

### Software and data availability

The necessary R code and data to produce the synthetic example, [Table 1](#), and [Figs. 4–7](#) are stored in a Zenodo repository at <https://zenodo.org/record/5857189> (last access: 16 Jan 2022).

### 1. Introduction

Modern environmental decision making depends on the results of various outputs from environmental modelling ([Jakeman et al., 2008](#); [Schmolke et al., 2010](#)). The need for impact modelling to support decision making is particularly important in the face of unprecedented changes the world is currently experiencing in climate, land use and other environmental conditions ([IPCC, 2021](#)). The application of complex environmental models is, however, not trivial, and the easily available outputs may be biased or of inappropriate scale or configuration, leading to a need for further processing particularly for local applications.

Mismatch between scales of analysis and the source data has been commonly addressed using downscaling, upscaling, or both.

Downscaling methods attempt to estimate the state of a variable at a finer resolution by relating it to the state of the variable at a coarser resolution ([von Storch and Zorita, 2019](#)). They are useful when we wish to estimate the internal variability within some temporal or spatial unit. In dynamic downscaling, a fine-resolution model is applied within areas defined by product under downscaling, and where the coarse data provides boundary conditions for the fine-scale model. This is common when, e.g., global climate models are downscaled using regional climate models. In empirical downscaling, the link between the finer and coarser scales is determined by statistical methods. There is a large body of literature on different methods of empirical downscaling, for instance, using interpolation methods (e.g. [Mennis, 2009](#); [Kallio et al., 2019](#); [Wang et al., 2015](#); [Lima et al., 2021](#)) or different machine learning methods (e.g. [Latombe et al., 2018](#); [Stevens et al., 2015](#); [Bardossy et al., 2005](#); [Ahmed et al., 2013](#)). The outputs from the downscaling methods may exhibit systematic errors (i.e. they may be *biased*; [Tabari et al., 2021](#)) when compared to observations. These systematic errors may arise from imperfect model conceptualization and data aggregation ([Teutschbein and Seibert, 2012](#)) or from errors in input data (e.g. [Sperna Weiland et al., 2015](#)), and often need to be corrected ([Ahmed et al.,](#)

\* Corresponding author. Aalto University, Otakaari 24, 02150, Espoo, Finland.  
E-mail address: [marko.k.kallio@aalto.fi](mailto:marko.k.kallio@aalto.fi) (M. Kallio).

<https://doi.org/10.1016/j.envsoft.2022.105511>

Received 16 January 2022; Received in revised form 29 August 2022; Accepted 31 August 2022

Available online 11 September 2022

1364-8152/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

2013; J. C. Chen et al., 2021; Ibarra et al., 2021; Teutschbein and Seibert, 2010). A large number of different approaches to bias correction are available (see e.g. Teutschbein and Seibert, 2012) and are of particular importance for climate change impact modelling (e.g. Hempel et al., 2013; Lange, 2019; Warszawski et al., 2014) which requires unbiased meteorological forcing data (Tramberend et al., 2021). These approaches commonly deal with the temporal dimension with less attention given to correction of spatially autocorrelated errors in distributed environmental models (Nahar et al., 2018).

Nonetheless, some methods to correct spatial bias exist. Hnilica et al. (2017) corrected precipitation with a combination of principal component analysis and quantile mapping. Nahar et al. (2018) used a similar methodology, but apply independent component analysis instead of principal component analysis. Kim et al. (2021) developed a Bayesian Kriging-Based Spatial Disaggregation-Quantile Delta Mapping method, where precipitation distribution parameters are interpolated on to a fine grid with Kriging and subsequently used to bias correct downscaled timeseries. Lange (2019) conducted multivariate bias correction using modified version of Cannon (2018) which ensures that downscaled estimates are consistent with the original climate simulations. Furthermore, data assimilation methods are used to optimally combine observations with model outputs in order to reduce their systematic errors (Reichle, 2008).

This study presents a novel interpretation of an advanced areal interpolation method *Dasymetric Modelling* (DM) as a method for spatial bias correction of downscaled data. Areal interpolation consists of all those methods which use values from a set *source zones* to estimate the same variables at a set of intersecting *target zones* (Comber and Zeng, 2019; Goodchild and Lam, 1980). The target zones are commonly of a higher resolution than source zones. In areal interpolation, the value of the variable is distributed among target zones within a source zone, thus preserving the aggregate value in the area represented in the source zone. DM is an extension to areal interpolation where the distribution of the source zone value is guided by some model. DM and bias correction share a similar objective of adjusting values to satisfy some constraint – e.g., the distributed values at target zones must match that of their source zone, or scaling a daily timeseries to match a monthly reference value.

We further note that when applying DM to downscale available coarse-resolution model outputs, the application can be described as meta-modelling. Meta-modelling is the process of building a model to emulate another model (also known as surrogate modelling; e.g. Razavi et al., 2012). In meta-modelling, the output of an original model is emulated by some statistical or process-based model from a set of input data with the common aim to reduce computational costs. Downscaling can be considered meta-modelling when the relationship between explanatory variables, for instance relating to topography, and e.g., a climate model output is established at a coarser scale and applied at a finer granularity.

We conduct an experiment in the Upper Bhima Basin (UBB), central

India, downscaling *local runoff* generation (i.e., without accumulation to *streamflow*). Several other methods have been developed to handle bias in runoff and streamflow. Skøien et al. (2006) extended kriging interpolation to account for topological relationships within a stream network necessary for the interpolation of streamflow records, and further developed a spatio-temporal topological kriging method in Skøien and Blöschl (2007). Paiva et al. (2015) provided another spatio-temporal kriging method to estimate river discharges along a river network. Loonat et al. (2020) used data assimilation of streamflow records to a distributed rainfall-runoff model by kriging and global bias correction. Naz et al. (2019) downscaled runoff data using a data assimilation scheme using soil moisture as an ancillary variable. Bennett et al. (2021) developed a mechanism for streamflow which specifically bias correct runoff and apply river routing to make it possible to use streamflow as the reference to correct against.

In our case study, we examine the bias correcting property of dasymetric modelling in detail. We train two meta-models to emulate the local runoff output of the Community Water Model (CwatM; Burek et al., 2020) – a distributed physically based hydrological model – based on a global model run at  $0.5^\circ$  (approx.  $50 \times 50$  km) resolution. The meta-models are then applied to a 5 arc-minute (approx.  $10 \times 10$  km) grid within the same area. The  $0.5^\circ$  grid serves as the source zones giving a spatial reference of the magnitude of the values within the basin, and the 5 arc-minute grid serves as the target zones. The meta-model outputs at the target zones are then bias corrected using the source zone values as a reference. The comparison to CwatM at higher resolution is appropriate because the meta-models are trained on CwatM at lower resolution, emulating the behavior of the model.

The experiment we conduct here extends previous work in Kallio et al. (2021, 2019), where global runoff products were downscaled with *Dasymetric Mapping*, i.e. DM with a single ancillary variable instead of a *model*. In the earlier studies, the bias correcting property is confounded due to A) use of irregular polygons which may be influenced by several intersecting source zones, B) the uncertainties related to river routing (modelling of accumulation of runoff in the stream network), C) the difficulty of assigning a low resolution streamflow value to the more complex, higher-resolution river network (see e.g. Bennett et al., 2021, for a discussion of this problem), and D) the use of a monthly timestep, which is large considering the size of the basins. The experiment here focuses on the bias correcting property by using a nested grid (controls the confounding by areal interpolation), comparing local runoff production (controlling for uncertainties in routing models and aggregation effects), and is applied at a daily rather than a monthly timestep. In this way, we can isolate the influence of bias correction from the other factors influencing the performance.

The remainder of the paper is structured as follows. Section 2 provides the background describing similarities and equivalence of DM to meta-modelling and bias correction, respectively. Section 3 describes our case study giving information on CwatM, the Upper Bhima Basin, meta-models and how we assess performance. Results and detailed

**Table 1**

Global performance metrics computed from the entire dataset covering all source or target zones, and all timesteps.  $\alpha$  is the variability error term,  $\beta$  is the bias term, and  $r$  is dynamics term in the non-parametric  $KGE_{np}$ .  $KGE_{ss}$  is the skill score wrt. the benchmark AW  $\hat{y}_t$ . NRMSE refers to the Normalized Root Mean Square of Error where the standardization is given by the mean of the timeseries.

Phase	Model	Set	Performance					
			NRMSE %	$\alpha$	$\beta$	$r$	$KGE_{np}$	$KGE_{ss}$
Meta-model training	RF	Source zone, train set	12.5	0.99	1.00	0.97	0.96	–
	RF	Source zone, test set	12.4	0.99	1.00	0.97	0.96	–
	LM	Source zone, train set	26.5	0.84	1.00	0.74	0.70	–
	LM	Source zone, test set	26.0	0.85	0.99	0.75	0.71	–
Benchmark	AW $\hat{y}_t$	Source zone values used at target zones	56.8	0.86	1.08	0.69	0.65	–
Meta-model application	RF $x_t$	Target zones	45.6	0.86	1.06	0.59	0.56	–0.24
	LM $x_t$	Target zones	49.7	0.62	0.99	0.53	0.40	–0.72
Bias correction	RF $\hat{y}_t$	Target zones	42.4	0.89	1.08	0.67	0.64	–0.01
	LM $\hat{y}_t$	Target zones	41.0	0.89	1.08	0.70	0.67	0.07

discussion of the case study outputs are given in Section 4. The bias correction framework and the case study are discussed in Section 5. Finally, we conclude the paper in Section 6.

## 2. Dasymetric modelling, meta-modelling, and bias correction

This section first introduces DM in Section 2.1. and proceeds to describe the similarities of the common DM workflow and meta-modelling in Section 2.2. In the next Section 2.3 we establish the equivalence of DM and linear scaling bias correction. In Section 2.4 we bring them all together in a single framework.

### 2.1. Dasymetric modelling

Areal interpolation – a form of spatial interpolation dealing with interpolating values from a source zone to another set of overlapping arbitrary target zones – is a standard practice in many fields (Comber and Zeng, 2019). In its simplest form, Area Weighted Interpolation (AW), a value from a source zone is assigned to target zones based on their overlapping areas as shown in Eq. 1

$$\hat{y}_t = \sum_{s \in S} y_s \frac{A_{t \cap s}}{A_s} \quad (1)$$

where  $\hat{y}_t$  is the estimated value at a target zone  $t$ ,  $y_s$  denotes the value associated with a source zone  $s$ , and  $A$  is the area of the zone. The target zone  $t$  is therefore assigned values from all source zones it intersects (denoted as the set  $S$ ) according to the proportion of the area of the target zone (within each source zone,  $A_{t \cap s}$ ) to the area of each overlapping source zone. The formulation allows estimation of values in arbitrary target zones which do not neatly conform to the (also arbitrary) shape of the source zones. Areal interpolation has a *pycnophylactic* – mass preserving – property because it is, in essence, distributing the source zone values onto the overlapping target zones. This procedure leads to the preservation of the source zone values in the target zones (Mennis, 2009; Tobler, 1979).

Dasymetric mapping is extension of AW where the interpolation is refined by the use of ancillary information (Comber and Zeng, 2019; Mennis, 2009). DM has originally been developed for improved cartographic presentation of population density (Mennis, 2009; Wright, 1936). A large body of literature has been written on different applications of DM and its extensions, particularly in the context of state-of-the-art population mapping (e.g. Stevens et al., 2015; Tatem, 2017), demographic attributes (e.g. Tatem et al., 2014), and to lesser extent in other fields such as environmental modelling (Kallio et al., 2019, 2021; Chen et al., 2019). In dasymetric mapping, the task of the ancillary data is to provide information on the distribution of the interpolated value within a source zone. With the ancillary information  $x_t$ , Eq. (1) becomes

$$\hat{y}_t = \sum_{s \in S} y_s \frac{A_{t \cap s} x_t}{\sum_{t \in T} A_{t \cap s} x_t} \quad (2)$$

which uses information about target zones  $t$  belonging to a set  $T$ , consisting of all target zones intersecting with source zone  $s$ , to scale the ancillary data. In Dasymetric Modelling (we will use DM to refer to Dasymetric Modelling for the remainder of the paper), the ancillary information  $x_t$  is given by some model describing the target zone level spatial distribution of the variable given in source zones (rather than a single variable as in dasymetric mapping; Nagle et al., 2014). For a review of areal interpolation methods we refer to the review by Comber and Zeng (2019).

### 2.2. Dasymetric modelling and meta-modelling

In statistical downscaling the relationship between ancillary variable

(s) and the downscaled quantity is commonly regressed at the source zone level (von Storch and Zorita, 2019). This is a common workflow in DM as well (Mennis, 2009; Nagle et al., 2014). The model trained at source zones is subsequently applied at the target zone level to obtain the downscaled estimates. In DM, the estimates are used as the ancillary information  $x_t$ . The outputs of the trained model are used to derive weights according to which the source zone value is distributed to contributing target zones. If the values associated with the source zones are an output of a model (e.g., a climate-, an impact- or some empirical statistical model), the workflow can be described as a meta-modelling approach. A meta-model attempts to emulate the output of the original model with the aid of explanatory variables – in this case the meta-model attempts to estimate what the full model would have output at the target zone level if it was run at that resolution. The motivation of using different types of meta-models is often to reduce the complexity of a full model, to reduce the computational cost, and sometimes to reduce input data requirements (Asher et al., 2015; Razavi et al., 2012). The common workflow in DM can therefore be seen as meta-model assisted distribution of the model outputs evaluated at source zones on to the overlapping target zones. Areal interpolation (DM) allows this to be done with the objective of upscaling, downscaling, or representing the values in similar resolution, but different zonal arrangement (Kallio et al., 2021).

### 2.3. Dasymetric modelling and bias correction

In a special case of DM where all target zones are located within a single source zone (e.g. a nested grid, a common setup in downscaling earth science data) and each target zone is of the same area (e.g. a raster grid with an equal area projection), the area terms in Eq. (2) can be eliminated resulting in Eq. (3a).

$$\hat{y}_t = y_s \frac{x_t}{\sum_{t \in T} x_t} \quad (3a)$$

This equation can be read as solution where the value associated with a source zone,  $y_s$ , is distributed to the target zones according to the proportion of some meta-model output  $x_t$  from the sum of meta-model results within the source zone. If we swap the position of  $y_s$  and  $x_t$ , the resulting Eq. (3b) is equivalent, but with the change in grouping of terms, we can now read the equation as the scaling of the meta-model outputs  $x_t$ .

$$\hat{y}_t = x_t \frac{y_s}{\sum_{t \in T} x_t} \quad (3b)$$

If the denominator is divided by the number  $n$  of target zones within a source zone, we obtain the mean output  $\mu_{x_t}$  of the meta-model within the source zone. Using the source zone as our reference, the equivalent mean across  $n$  target zones within the source zone is  $\mu_{y_t} = \frac{1}{n} y_s$ , and thus we can simplify Eq. (3b) to Eq. (5) with the intermediate step given in Eq. (4),

$$\hat{y}_t = x_t \frac{\frac{1}{n} y_s}{\frac{1}{n} \sum_{t \in T} x_t} \quad (4)$$

$$\hat{y}_t = x_t \frac{\mu_{y_t}}{\mu_{x_t}} \quad (5)$$

where  $y_s$  is the value associated with the source zone in which the target zone is located, and  $\mu_{x_t}$  is the mean value of all meta-model outputs within the source zone. This formulation is identical to the linear scaling bias correction method shown in Eq. (6) (Lenderink et al., 2007; Teutschbein and Seibert, 2012)

$$y_{model}^* = y_{model} \frac{\mu_{obs}}{\mu_{model}} \quad (6)$$

where  $y_{model}^*$  is the bias corrected model output.

Although bias correction is commonly applied to timeseries data, it may also be applied to a spatial domain where target zone values are corrected against the source zone data. Thus, when stripped of the area terms under the conditions described above, DM is mathematically equivalent to the common linear scaling bias correction method.

An example of linear scaling bias correction with one dimensional synthetic data is provided in Fig. 1, showing a cosine function and a hypothetical meta-model output  $x_t$ , and its bias corrected counterpart  $\hat{y}_t$ , along with goodness-of-fit statistics. For illustration purposes, the hypothetical meta-model output for each point  $t$  was obtained here by adding random variation into the cosine series, dividing the values into groups, and subtracting a random amount from each group. Linear scaling is applied group-wise for each source  $s$  using Equation (5) rather than to the dataset as a whole, akin to the DM approach having multiple source zones for which to perform bias correction separately. From this example we note a few important properties of linear scaling: 1) by design, bias is eliminated group-wise and subsequently for the entire data series. 2) Pearson correlation coefficient is preserved within group, but changes for the entire series. 3) the magnitude of standard deviation is changed.

The reason for preservation of Pearson correlation coefficient  $\rho$  (Eq. (7)) within a group is because  $\rho$  depends on how the two data series vary around their respective means.

$$\rho_{x,y} = \frac{E[(X - \mu_x)(Y - \mu_y)]}{\sigma_x \sigma_y} \quad (7)$$

Since each data point in a group is subjected to the same linear transformation, the relationship between the standard deviation  $\sigma$  and the mean  $\mu$  remains unchanged. In other words, coefficient of variation  $CV (\sigma/\mu)$  remains unaltered, even when the absolute magnitude of the standard deviation is changed. As a consequence, bias correction preserves the dynamics provided by the model *within* each group, but because each group is subjected to different scaling operation, the overall dynamics (correlation) is altered; in this synthetic example Pearson correlation coefficient improves from 0.74 to 0.96. Fig. 1B

shows autocorrelation functions (correlation of a variable with itself) for the true original, the meta-model output  $x_t$ , and bias corrected meta-model output  $\hat{y}_t$ . Autocorrelation is significantly improved with the method. Given the equivalence between linear scaling and dasymetric mapping on nested grids, a similar behavior is expected in a spatial bias correction context.

We note that for the linear scaling bias correction to work as described here, all meta-model output values  $x_t$  and the reference values associated with the source zones  $y_s$  should be strictly finite-positive values to prevent erratic behavior.

#### 2.4. Dasymetric modelling as a combined downscaling-bias correction framework

The previous sections describe how meta-modelling and bias correction relate to DM. Fig. 2 shows workflows for each of these three methods separately and combined together into a single framework. Each of the three methods have a specific purpose in the workflow. The meta-model outputs  $x_t$  evaluated at the target zones is responsible for representing the *spatial dynamics* of how the value is distributed within a source zone, i.e., the meta-model output should have a high correlation with the ground truth values within the source zones. Bias correction ensures that the spatial information associated with source zones are preserved in the target zones, leading to *spatially debiased* estimates. DM provides flexibility in the spatial configuration of source- and target zones. It describes how non-conforming source and target zones are handled – how should the meta-model output in each target zone be divided during the intersection with source zones, and how to combine the values back to the original target zones after bias correction. This operation may differ for instance for *intensive* (standardized variables, e. g., population density, or water use per capita) and *extensive* variables (counts, e.g., the amount of population, or volume of water).

The workflow is agnostic as to how  $x_t$  is estimated – whether given by another similar model, a meta-model, or measured data. Similarly, the method is agnostic as to where the source zone reference values  $y_s$  are obtained. In principle, with the redistribution of conventional DM

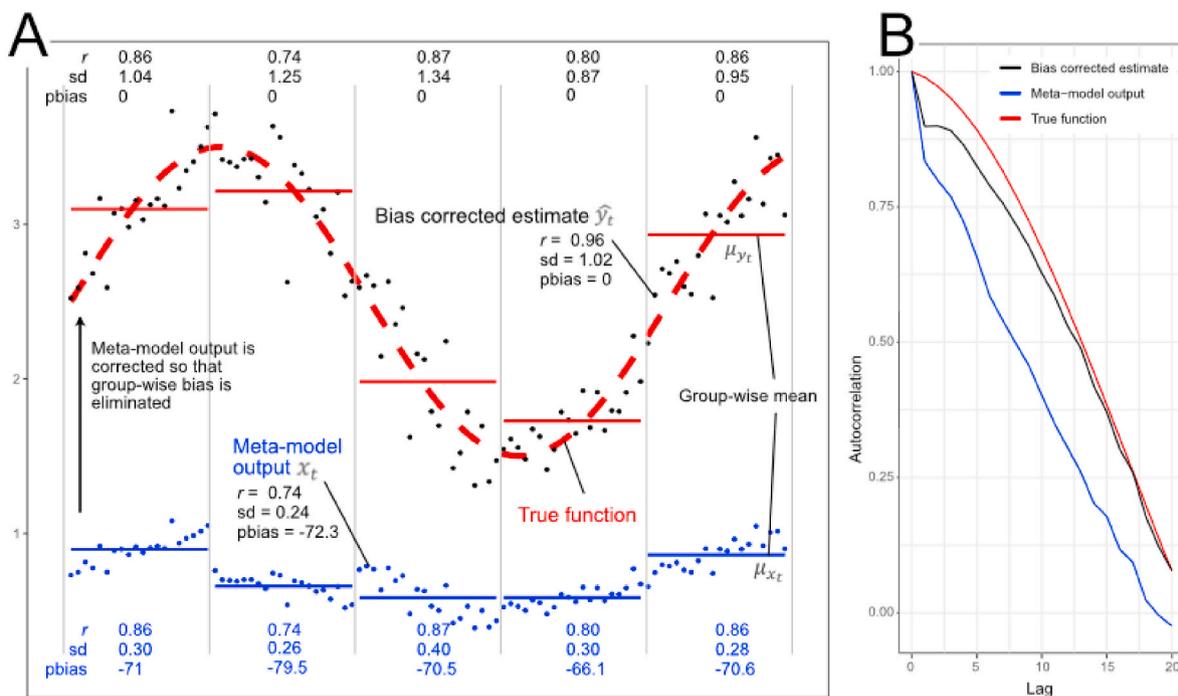


Fig. 1. Panel A) shows a 1-dimensional schematic of linear scaling bias correction and demonstrates some properties of the method. Statistics  $r$ ,  $sd$ , and  $pbias$  are the components Kling-Gupta Efficiency (Pearson correlation coefficient, ratio of standard deviations, and percent bias, respectively) and are evaluated against the true function shown in dashed red line. Panel B) shows an autocorrelation function computed for each of the three data series.

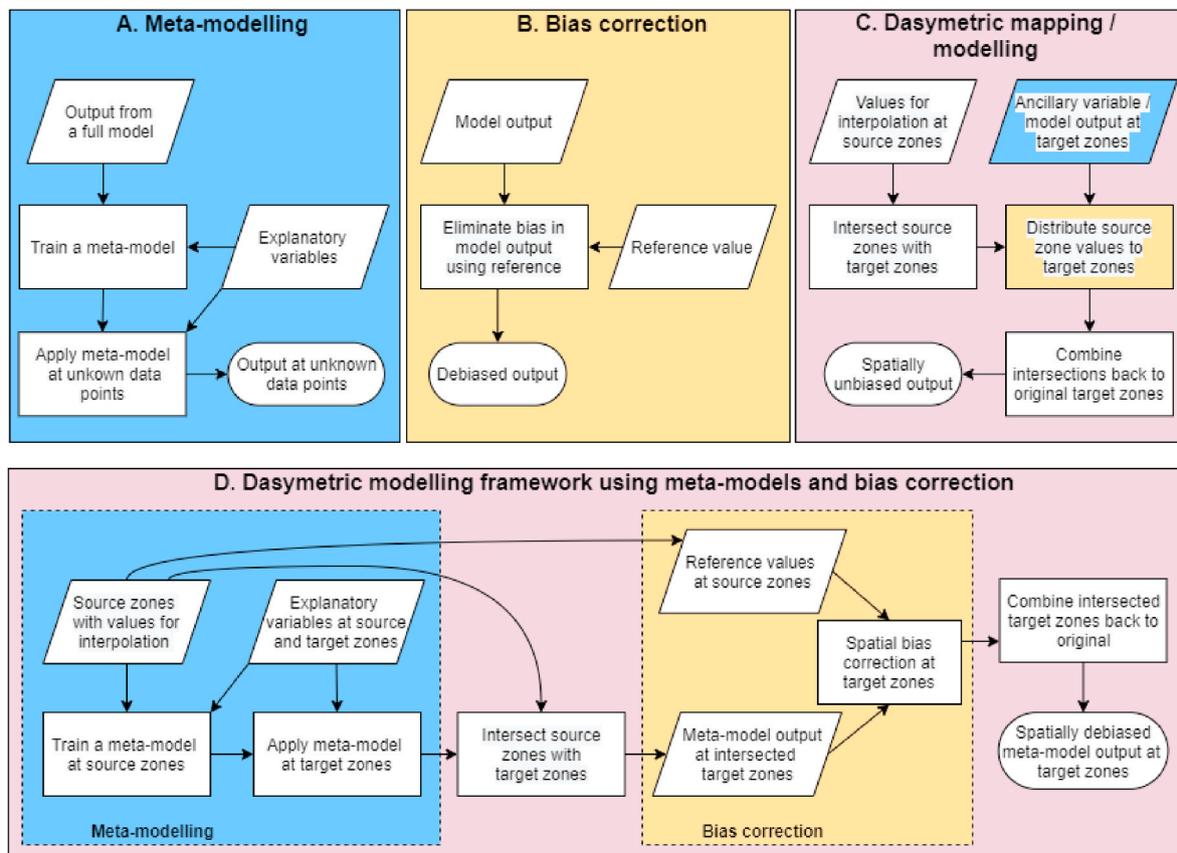


Fig. 2. The workflows in A) meta-modelling, B) bias correction, C) dasymetric modelling, and D) the spatial bias correction framework consisting of A-C.

interpreted here as bias correction, linear scaling may be replaced by more advanced bias correction methods. This leads to the framework being modular, consisting of three distinct parts.

### 3. Case study

We explore the dasymetric modelling framework as interpreted in Fig. 2, and its properties, by a case study where we downscale a coarse resolution ( $0.5^\circ$ , source zones) CWatM simulation with a global extent onto a higher resolution, 5 arc-minute grid (target zones). We make use of meta-models trained at the source zone level and applied at the target zones. The results are compared to a run of CWatM at the target zones in order to evaluate the usefulness of the methodology and the ability to emulate a more spatially detailed model output. The experiment is setup so that the source and target zones are nested, which reflects the simplification described in Section 2.3 – this allows us to examine the efficiency of the spatial bias correction without the added complication of non-conforming source and target zones.

In Section 3.1 we give a brief introduction to CWatM, and Section 3.2 contains a description of the study area, the Upper Bhima Basin. Next, Section 3.3 introduces the two meta-models used, and Section 3.4 describes how we assess the downscaling performance.

#### 3.1. Community Water Model

The Community Water Model is an open-source integrated hydrological and channel model that calculates water availability (surface and groundwater), environmental flow requirements, and socio-economic water demands and impacts from water infrastructures such as reservoirs, groundwater pumping, and irrigation (Burek et al., 2020). The global applications of the model are based on a  $0.5^\circ$  or 5 arc-minute resolution grid and are run at a daily temporal resolution. The model

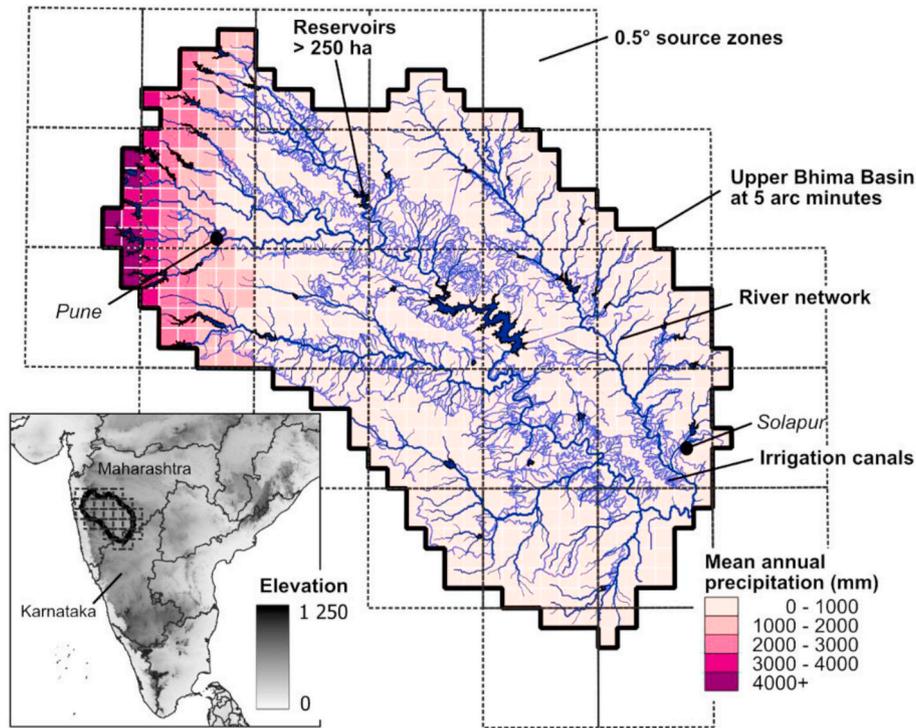
can be used for regional applications using a daily temporal and 30 arc-sec spatial resolution. For topography and land cover, it uses a sub-grid approach and for soil and routing processes, it uses sub-daily timesteps.

The conceptual framework and structure of CWatM are similar to that of other hydrological models such as H08 (Hanasaki et al., 2006, 2008, 2010), WaterGAP (Alcamo et al., 2003; Flörke et al., 2013), PCR-GLOBWB (van Beek et al., 2011; Wada et al., 2014) and others. The model accounts for future water demands based on socio-economic change and the impacts of water availability in response to climate change. Irrigation water demand is estimated for paddy and non-paddy crops separately by dynamically linking irrigation water demand with the surface and soil water balance. For the Upper Bhima basin CWatM is forced by the meteorological variables including temperature, wind speed, relative humidity, incoming longwave and shortwave radiation, and surface air pressure from W5E5 v2.0 (Cucchi et al., 2020). For precipitation, we use data from Pai et al. (2014). Please see Burek et al. (2020) for a more detailed description of CWatM hydrological processes.

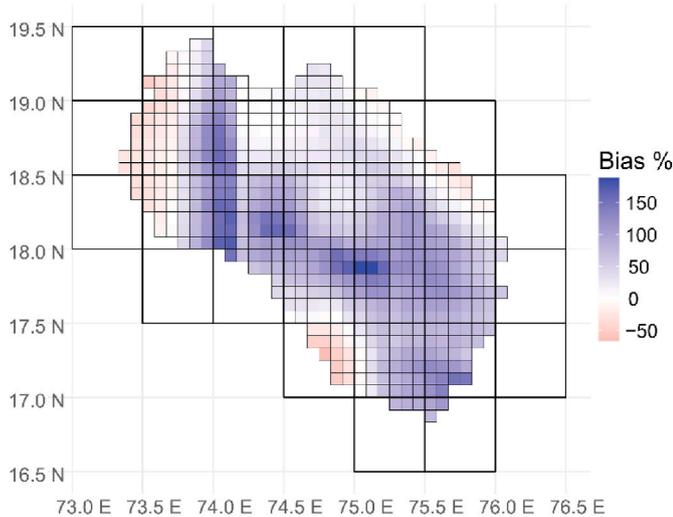
#### 3.2. Upper Bhima Basin

The Bhima Basin in India, shown in Fig. 3, is a highly managed and densely populated watershed with important rain-fed and irrigated crop production. Rainfall occurs towards the west and concentrates through 5 months of monsoon, with a total annual rainfall of around  $35 \text{ km}^3$  (approx. 760 mm per year).

The Bhima basin is an upstream basin of the Krishna Basin, originating in the mountains of the western Ghats, with a river length of  $\sim 325 \text{ km}$ . The Bhima basin supports a population of 18.7 million (2015), distributed over  $45,800 \text{ km}^2$  of largely agricultural land, flowing mostly within the state of Maharashtra. The largest urban agglomeration within the Bhima Basin is Pune with nearly 7 million residents, followed



**Fig. 3.** The Upper Bhima Basin on the western peninsular of India. The map shows the mean annual precipitation over 2001–2010 at the target zone level, major hydrological features of the basin as well as the source and target zones used in the case study.



**Fig. 4.** Local bias of CWatM simulated source zones (0.5°) runoff  $y_s$ , against target zones (5 arc-min) runoff  $y_t$ , computed over  $3 \times 3$  moving window.

by Solapur with approximately 1.2 million residents in 2015. CWatM has been successfully applied at the Bhima Basin (Guillaumot et al., 2022).

### 3.3. Meta-models

Two meta-models were developed to estimate CWatM modelled runoff output over Upper Bhima Basin during 2001–2010. The first meta-model is a standard Ordinary Least Squares Regression, referred to as the linear model, LM. The second meta-model is based on Random Forest Regression (RF; Breiman, 2001). Both models were trained at the source zone level (0.5° resolution) with the same model equation based on derived variables from a Digital Elevation Model (DEM), daily

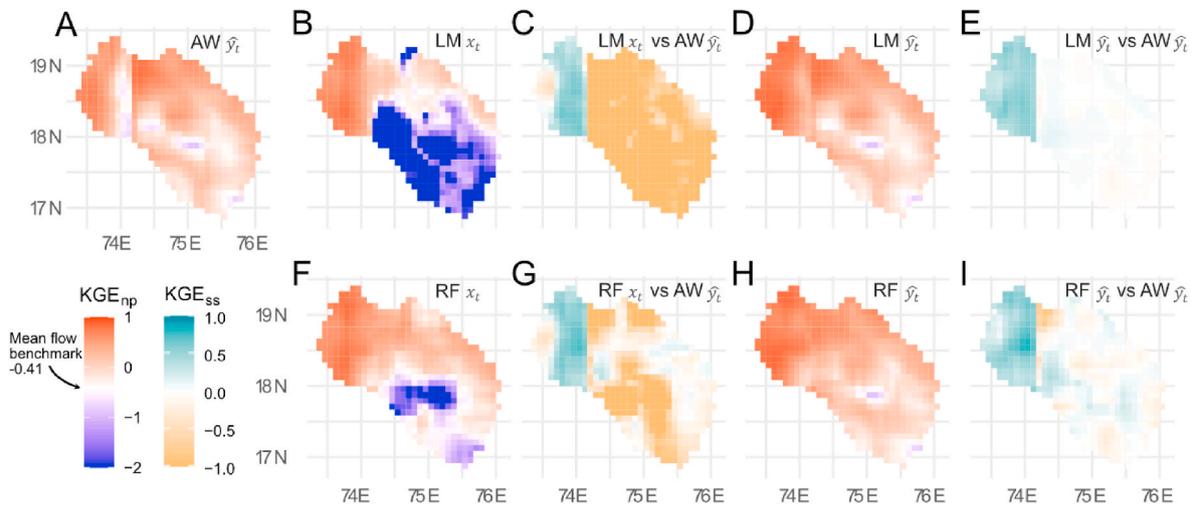
precipitation ( $P$ ) and daily temperature ( $T$ ) measurements

$$x_{ts} = P_{ts} + P_{ts-1} + P_{ts-2} + \text{sum}(P_{ts-1, \dots, -7}) + \text{sum}(P_{ts-1, \dots, -30}) + \text{sum}(P_{ts-1, \dots, -180}) + DSP + T_{ts} + T_{ts-1} + T_{ts-2} + \text{mean}(T_{ts-1, \dots, -7}) + \text{elevation} + DUNE \quad (8)$$

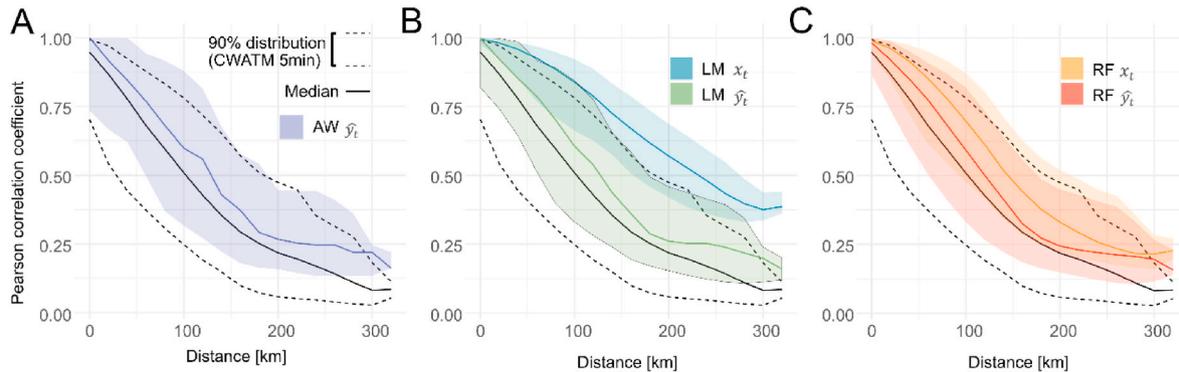
where  $x_{ts}$  is the output of meta-model at timestep  $ts$ ,  $DSP$  is the number of Days Since previous Precipitation event, and  $DUNE$  is a topographic index (Loritz et al., 2019) computed from the HydroSHEDS 15 arc second DEM (Lehner et al., 2008).  $DUNE$  describes runoff production through a combination of gravitational potential energy of water (height above nearest drainage), and dissipation of that potential energy as water moves along the landscape (distance to nearest drainage). Loritz et al. (2019) find that it can distinguish between regions with different runoff producing regimes, and Kallio et al. (2019) found it useful in a downscaling context with variable terrains. Both elevation and  $DUNE$  were aggregated from the 15 arc-second elevation model to 0.5° degree and 5 arc-minute resolution by taking their mean value. RF was applied using the ranger package for R (Wright and Ziegler, 2017) with 500 trees, unlimited tree depth, and variance-based split rule.

We chose the LM and RF meta-models to have a comparison of a simple linear meta-model, and a more complex machine learning method which can deal with non-linear relationships between the predictand and predictors (Cutler et al., 2007), and which has been successfully used in downscaling studies (e.g. C. Chen et al., 2021; Hutengs and Vohland, 2016), including DM (Stevens et al., 2015). Both meta-models were trained with the output of CWatM global simulation run at 0.5° resolution using those computational cells which intersect the study area of Upper Bhima Basin (Fig. 3). The source zone dataset was divided so that a random 75% of data points (all 30 source zones and all 3652 daily timesteps over 2001–2010 together yields 109 054 data points) were used for training, and the remaining 25% datapoints were used for testing.

When applying bias correction, we first shift the meta-model outputs so that the minimum value within each source zone is  $> 0$  (Reibel and



**Fig. 5.** Local  $KGE_{np}$  performance computed from a  $3 \times 3$  moving window and all timesteps for A) source zone values  $AW \hat{y}_t$ , B) LM meta-model output  $x_t$ , C)  $KGE_{ss}$  between LM  $x_t$  and  $AW \hat{y}_t$ , D) bias corrected LM meta-model output  $\hat{y}_t$ , and E)  $KGE_{ss}$  between LM  $\hat{y}_t$  and  $AW \hat{y}_t$ . The lower row shows F) RF meta-model  $x_t$ , G)  $KGE_{ss}$  between RF  $x_t$  and  $AW \hat{y}_t$ , H) bias corrected RF  $\hat{y}_t$ , and I)  $KGE_{ss}$  between RF  $\hat{y}_t$  and  $AW \hat{y}_t$ . The mean flow benchmark performance with  $KGE = -0.41$  is obtained by using temporally averaged mean runoff for the entire timeseries as a prediction (Knoben et al., 2019).



**Fig. 6.** Spatial correlation computed in 20 km lags, showing the 90% distribution and the median of the reference model run (CWatM at 5 arc minute target zones) and the distributions and the median of the meta-model outputs  $x_t$  and bias corrected outputs  $\hat{y}_t$ . Panel A) shows AW, a benchmark given by the source zone values, B) shows the distributions for LM meta-model, and C) for RF meta-model.

Agrawal, 2007), if there are any negative values. Shifting the values is done here because 1) for the linear scaling to work as described, all values should be positive, and 2) negative runoff values make no physical sense. While there may be better alternatives to removing negative values from meta-model outputs, we shift because shifting preserves the variability provided by the model (i.e., the standard deviation is preserved), which is task given to the meta-models in Section 2.4. We note that as a result, CV of any shifted timeseries is not preserved, because the mean changes, but standard deviation does not. Bias correction is applied to each daily timestep separately.

### 3.4 Assessment

We assess the performance of meta-models and bias corrected meta-models in downscaling coarse resolution CWatM output (at the source zone level) against a high resolution model run of CWatM (at the target zone level) using Normalized Root Mean Square of Error (NRMSE, normalized using the mean) and a form of the Kling-Gupta Efficiency (KGE; Gupta et al., 2009). KGE is a multi-objective function consisting of measures of bias, variability, and dynamics shown in Eq. 9

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \quad (9)$$

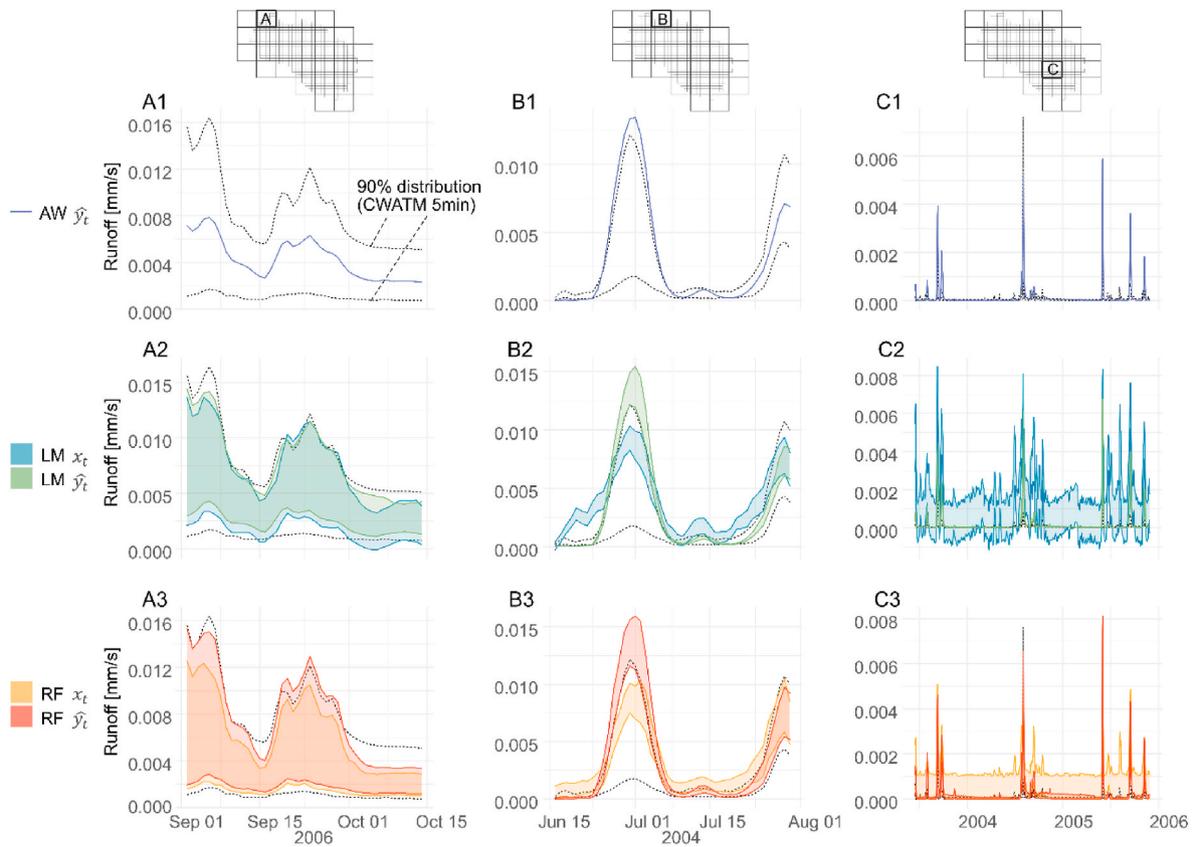
where  $r$  is the Pearson correlation coefficient between simulations  $sim$  and observations  $obs$ ,  $\alpha$  is a measure of variability error  $\sigma_{sim}/\sigma_{obs}$ , and  $\beta$  is the bias  $\mu_{sim}/\mu_{obs}$ .  $\sigma$  stands for the standard deviation, and  $\mu$  for the mean.

We use a non-parametric variant of KGE where the variability  $\alpha$  is described using the average error in normalized Flow Duration Curves (FDC) and Pearson correlation coefficient is replaced with Spearman rank correlation (Pool et al., 2018). We use the non-parametric KGE because standard deviation is sensitive to the mean of the timeseries, and thus the term is not independent of the bias, whereas the normalized FDC is independent of a possible bias in the timeseries. Further, Spearman rank correlation is desired here because it is less sensitive to extreme values (Pool et al., 2018).

In order to assess an improvement in the overall skill of the downscaling, we also compute a KGE skill score  $KGE_{ss}$  using Eq. (10) (Knoben et al., 2019),

$$KGE_{ss} = \frac{KGE_{model} - KGE_{benchmark}}{1 - KGE_{benchmark}} \quad (10)$$

where  $KGE_{model}$  refers to the meta-model, and  $KGE_{benchmark}$  refers to the performance of source zone output of CWatM as an estimate of the target zone output of CWatM.  $KGE_{ss}$  is interpreted so that any positive value means improvement over the benchmark,  $KGE_{ss} = 0$  signifies equal



**Fig. 7.** Selected time periods from three source zones showing the 90% distribution of target zone values within the source zones for AW, LM and RF and their bias corrected counterparts. Panel A and B are located where meta-model performance is relatively high (see Fig. 5), and panel C where meta-models perform poorly. The timeseries distribution shown here are not influenced by the neighboring source zones.

model performance, and negative values indicate poorer model performance.  $KGE_{ss}$  can take any value between  $+\infty$  (when  $KGE_{benchmark}$  approaches  $-\infty$ ) and  $-\infty$  (when  $KGE_{model}$  approaches  $-\infty$ ). For more details, we refer the reader to [Knoben et al. \(2019\)](#).

#### 4. Results and interpretation

We discuss three different aspects of the performance of DM with linear scaling bias correction. First overall performance of the meta-models, benchmark, and the results of the bias correction (Section 4.1), second, how DM is affecting the spatial correlation (4.2), and third, we consider temporal efficiency (4.3).

##### 4.1. Improved global and local performance after bias correction

Both meta-models trained at the source zone level show good overall performance and a comparison between a training and testing datasets show no sign of overfitting (Table 1). Of the two meta-models, RF shows, however, considerably better performance ( $KGE_{np}$  0.96) than LM ( $KGE_{np}$  0.71) in the testing dataset. When the meta-models trained at the source zones are applied to the target zones, their performance considerably deteriorates, as expected for out of sample prediction. Error increases and the performance in all components of  $KGE_{np}$  decrease. A global comparison of the meta-models benchmarked against use of the source zone values for the target zones ( $AW \hat{y}_t$ ) shows considerably worse performance for both meta-models. The deteriorating performance is likely due to 1) the simplicity of the meta-models; they do not consider land use or soil properties (which increase in importance in higher resolutions) and thus are not able to represent the heterogeneous conditions existing within the basin, and 2) the meta-model training using a lower resolution (spatially averaged) model run does not represent all

the conditions found when modelling is performed in higher resolution. Despite the meta-models showing inferior performance in the  $KGE_{np}$  components than  $AW \hat{y}_t$ , overall errors are smaller, as shown in Table 1.

The global goodness-of-fit statistics reveal that after applying the DM spatial bias correction, performance is similar to the benchmark with further reduced errors. We note that because the source model runs have 8% higher runoff than target zone model runs (reflected in the bias component of  $AW \hat{y}_t$  in Table 1, and shown in Fig. 4 as comparison between source and target zone runoff), the maximum attainable KGE, even when the dynamic and variability terms were perfectly matched, is 0.92. That is, the assumption that source zone information needs to be preserved means that we have potentially accepted a discrepancy compared to the target zone model runs. This level of KGE shows a very high match between two timeseries. In DM, bias is corrected for each source zone separately and consequently for the entire area, however, the spatial differences in the bias lead to different maximum attainable KGE in each area. We note here that the two CWatM model runs use different climate forcing data, which is a major source of errors in hydrological model runs ([Sperna Weiland et al., 2015](#)), and that it is known that model runs at different levels of spatial aggregation often result in differences in the simulated water balance ([Wen et al., 2021](#)). Due to this model-to-model comparison we do not assess which one more correctly represents the true runoff production within UBB. The merits of the presented downscaling method (without meta-models) against observed streamflow records are assessed in [Kallio et al. \(2019, 2021\)](#).

Using a global evaluation, however, only provides a partial understanding of the model performance. Fig. 5 shows how local performance is distributed across the basin in small neighborhoods of  $3 \times 3$  target zones. The maps reveal that the meta-models perform more poorly than the benchmark in the central and eastern part of the basin (Fig. 5C, G), areas characterized by low rainfall resulting in low runoff. However,

both meta-models improve on the performance in the western basin where runoff production has a high spatial gradient such that downscaling is more important. Bias correcting the meta-model output with DM eliminates nearly all differences between LM  $\hat{y}_t$  and AW  $\hat{y}_t$  in the area where the meta-model LM  $x_t$  was found to be the poorest, while still maintaining the improvement of the meta-model within the high-gradient zone. The poor performance artefact found in AW  $\hat{y}_t$  along longitude 74° E becomes almost eliminated in LM  $\hat{y}_t$ . The artefact is entirely eliminated in RF  $\hat{y}_t$ , but has much higher variability in performance elsewhere compared to AW  $\hat{y}_t$  than LM  $\hat{y}_t$ . While the RF meta-model shows extremely high performance at the source zone level, this finding may indicate that while it performs well on source zone testing data, it has still overfitted on the training data and is not as well transferable to the higher resolution target zones as the simpler LM meta-model. We note here, however, that the LM meta-model predicts nearly half of data points as negative runoff values, and the good performance can only be achieved after restricting the outputs to strictly positive values (see Section 2.3).

#### 4.2. Improved spatial correlation

One of the main benefits of spatial bias correction is that the spatial structure of outputs can substantially improve, similarly to the autocorrelation function shown in Fig. 1B. Fig. 6 shows the median and the 90% distribution of spatial correlation of the benchmark, meta-models, bias corrected meta-models, and the reference model run of CWatM<sub>5min</sub>, computed between all pairs of target zones. The source zone values at target zones AW  $\hat{y}_t$  show a reasonably good agreement with the reference model run. Both runs are made with the same model (albeit with different climate forcing) and thus it should be expected that when modelling the same output variable, their large-scale spatial correlation would be similar. However, the effect of the coarse correlation can really be seen at the short distances (the first two bins in Fig. 6A), since the resolution of the 0.5° source zones is approximately 50 km at UBB, a major proportion of all correlation pairs show a perfect match (correlation = 1). This is not the case for the meta-models in Fig. 6B and C. The figures also show that the spatial correlation structure of the meta-models in the study area substantially improves in both cases, though for RF the improvement is more modest than using LM due to the better structure in RF. Apart from small differences, spatial correlation in AW  $\hat{y}_t$ , LM  $\hat{y}_t$ , and RF  $\hat{y}_t$  are all highly similar.

#### 4.3. Assessment of timeseries

Evaluations based on global and local performance and spatial correlations suggest that using solely source zones seems to be a viable option. However, it does not accomplish the goal of downscaling as achieved using DM. This is highlighted in Fig. 7, which shows examples of the distribution of runoff values against time within three different source zones. The first row in the figure visualizes the fact that AW  $\hat{y}_t$  does not show any distribution of values within the source zone, while LM (row two) RF (row three) show a large distribution of values. The timeseries also illustrates many of the properties of linear scaling bias correction: 1) Panels A1 to A3 show that when the meta-model estimates are reasonably good, and with their means close to the reference value of the source zone (AW  $\hat{y}_t$ ), there are only minor adjustments to the distribution. 2) In panels B1 to B3 in mid-June and mid-July 2004, bias correcting meta-model outputs improves the distribution estimate. However, around July 1st, source zone value AW  $\hat{y}_t$  is outside the envelope of the CWatM reference model run, and thus the distribution of bias corrected meta-model outputs have also moved outside the envelope. 3) Panels C1 to C3 show multiple years of runoff production in the area where the performance of all  $\hat{y}_t$  estimates are poor. The LM meta-model produces a significant number of negative runoff values and a too wide a range of values. Bias correcting brings a major improvement

in the values although the performance is still poor overall as evidenced by the maps in Fig. 5. Similar improvement is seen also for RF, however without the need to shift values. 4) The absolute magnitude of the distribution changes according to the direction of the adjustment; if the meta-model mean is higher than the reference value, the distribution shrinks because the adjustment  $\mu_{obs}/\mu_{model} < 1$  (see Eqs. (5) and (6)). The effect is reversed if the mean of meta-model outputs is smaller than the reference.

As the examples in Fig. 7 portray, the final performance of the timeseries depends on the accuracy of the source zone reference values for providing an unbiased estimate, and on the ability of the meta-model to accurately portray the dynamics within the source zone.

## 5. Discussion

### 5.1. Differences to alternative methods and high-resolution modelling

Our study describes spatial bias correction with a method that complements methods found in the literature, mainly based on point-observations, in that it specifically addresses downscaling of data with (arbitrary) areal spatial support. Closest areal interpolation alternative, area-to-area kriging, results in areal output, but require areal features to be first cast to point representation for estimation of the semi-variogram (Hu and Huang, 2020). This intermediate step is unambiguous. Additionally, kriging is unbiased at the observed point locations, but the consequence of using point representation for areal features is that estimates may no longer be unbiased across areal features, which is a feature of DM. Kriging-based runoff interpolation furthermore require observations with a high spatial density (Parajka et al., 2015), not generally available in data-scarce contexts. On the other hand, data assimilation methods estimate an optimal combination of model simulations and reference data, but do not result in entirely unbiased estimates. They also require rigorous *a priori* uncertainty estimation for both model outputs and observations. The standard method of DM is mathematically and conceptually simple and relatively easy to implement: statistical (meta-)modelling of arbitrary complexity, linear scaling bias correction based on means, and spatial intersections between source and target zones.

Furthermore, the DM methodology presented herein provides a number of advantages over a full CWatM simulation at the higher resolution: 1) substantial computational improvement compared to running the full model. In our case study consisting of 30 source zones, 567 target zones and 3652 timesteps, the unoptimized DM implementation in R language (R Core Team, 2020) runs within a few minutes (including training of the meta-models), compared to approximately an hour (Kallio, 2020) for a full CWatM run at 5 arc-minute resolution in the area (without calibration procedure), 2) statistical models such as the LM and RF models used here can use a wide range of explanatory variables while hydrological models have strict requirements for inputs, 3) faster calibration to the study area compared to calibrating a distributed hydrological model against observed data or regionalizing parameters from another area, and 4) a wider pool of experts is familiar with statistical modelling than with physically based, distributed hydrological modelling. Each of the advantages listed here stack up when uncertainty management requires the use of multiple estimates, as is often the case (Saxe et al., 2021).

### 5.2. Sensitivities and limitations

In light of the presented results, source zone values AW  $\hat{y}_t$ , despite not being downscaled, seem to reflect the runoff outputs of the target zone reasonably well. The question therefore arises whether downscaling with the methodology of meta-models and bias correction bring any added value? Our analysis clearly suggests that there is little benefit in downscaling in general when the area within a source zone is ho-

mogenous with no large differences in the output. In these regions, using the coarse resolution data may be justified against the added complexity of training, applying and bias correcting a meta-model. Similar concepts have been long explored in hydrological literature; e.g. the concept of Hydrological Response Units in which areas with similar hydrological functioning are lumped together, or model applications with variable modelling resolution (Gharari et al., 2020). There is, however, strong improvement in the performance in areas with a high gradient in the modelled outputs, as seen in the western part of the UBB (Fig. 5).

Furthermore, the standard DM methodology used here assumes that the reference values provided by source zones are unbiased and accurate for the areas they describe. There are very few variables which can be measured to certainty over large spatial domains or variable spatial units (The Modifiable Area Unit Problem, MAUP; Manley, 2014), and this uncertainty can affect the regressed model parameters (Fotheringham and Wong, 1991). Our case study provides a simplified (but useful) use case with well-defined source and target zones as nested grids defined by a CWatM simulations, and where the configuration or sizes of target zones play no role. This allowed us to have a detailed look at the properties of DM as a downscaling and spatial bias correction method without an added complexity of non-conforming target and source zones. As an areal interpolation method, DM can, however, deal with arbitrarily shaped source and target zones, and therefore can be used to address MAUP (Dark and Bram, 2007; Goodchild and Lam, 1980; Kallio et al., 2021). Performance of the downscaling to non-conforming target zones in hydrology has been assessed in Kallio et al. (2019, 2021) and the uncertainty in different realizations of target zones in Virkki (2019). We note that modified DM methodologies able to deal with uncertain source and ancillary data are also available (Leyk et al., 2013; Nagle et al., 2014).

However, while DM methodology can handle arbitrary overlapping configurations of source and target zones, it does not handle spatial dependencies across source zones. In environmental modelling, this particularly means processes such as streamflow accumulation, or air pollution: the entire upstream catchment influences streamflow at any particular location, and atmospheric processes can transport pollutants from far-away regions. The meta-model used may include these spatial dependencies, but independent bias correction at each source zone means that discontinuities and mass-balance violations may be introduced at source zone boundaries. Methods for bias correction of streamflow have been developed which handle the tree-like structure of stream networks (see e.g. Bennett et al., 2021; Gottschalk, 1993; Paiva et al., 2015; Skøien et al., 2006). The standard DM methodology is suitable for any process which may be considered *local*, such as runoff generation (without accumulation), soil moisture, land use and land cover, or soil properties, or when the source data covers the entire area of generating process (see e.g. the application of the framework to migration data in Niva et al., 2022).

### 5.3. Potential advancements for DM as a bias correction framework

In our case study, bias correction is carried out individually for each timestep, and therefore the distributions of LM  $\hat{y}_t$  and RF  $\hat{y}_t$  closely follow the source zone values AW  $\hat{y}_t$ . However, extension to spatio-temporal variant of DM (Mennis, 2016) means that the reference source zone values could be expressed in a *spatio-temporal* domain, e.g. providing a monthly reference value for the source zone. In such a case, the daily bias corrected meta-model outputs would better retain temporal dynamics provided by the meta-model – an aspect which is lost with daily adjustments.

While the linear scaling method used in this study does not always result in the correct distribution of runoff values within source zones, other more advanced bias correction methods can be implemented. Knowledge about the distribution within the source zone, say the standard deviation of values found within, variance scaling (Chen et al.,

2011) could be used to correct for both mean, and standard deviation. If the actual distribution is known or can be estimated *a priori* for each source zone, more advanced methods like quantile-quantile mapping (Teutschbein and Seibert, 2012) can be used instead (and is recommended, if possible; Teutschbein and Seibert, 2013). We can expect these methods to improve the spatial distribution of runoff values, and thus further improve downscaling performance. In principle, the used bias correction method is limited by what information is available at the source zone level. We note here, however, that any method which modifies the distribution of the meta-model outputs, also modifies their spatial correlation. In DM, the spatial dynamics (correlation) of the meta-model is preserved within source zones, and only modified across source zones. This behavior is different from methods which apply bias corrections continuously.

We close the discussion of the presented methodology with a remark that while areal interpolation, and as an extension, DM, are based on area standardization, this is not a strict requirement. The methodology can extend to other type of variables as well, such as water use per capita. The modification from standard areal interpolation that needs to be done to support other standardization schemes is to determine how the values are divided when intersecting source and target zones, and when combining the target zone intersections back to original (i.e., the areal interpolation component of DM workflow in Fig. 2D). This is straightforward when standardization is by area but can be more complicated for alternative standardizations.

## 6. Conclusions

In this paper we presented an interpretation of dasymetric modelling as a flexible spatial bias correction framework for downscaling environmental model outputs. We showed that the common workflow in dasymetric modelling can be described using concepts of meta-modelling and that the distribution of source zone values in dasymetric modelling is in fact equivalent to the common linear scaling bias correction method. In a case study we then downscale coarse resolution runoff estimates from CWatM (source zones) and compare the outputs against a higher-resolution model run (target zones).

By exploring the properties of linear scaling in a spatial bias correction setting within the DM framework, we find that the DM methodology works well in areas where runoff production exhibits a strong gradient, but that there is limited benefit in more homogenous areas. The meta-models based on precipitation, temperature and topographic data have poor performance in most of the study area. Correcting for the spatially autocorrelated errors (spatial bias), however, corrects most of the problems exhibited by the meta-model outputs and significantly improve spatial correlation structure. The benefit is strongly dependent on unbiased reference values associated with the source zones, and correctly represented dynamics by the meta-models.

Dasymetric modelling is a mature technique developed in population studies with multiple variants in the literature. It offers great potential for environmental impact studies where large amount of data reflects the future uncertainty space and downscaling to local scales is imperative for policy making. As the presented downscaling and spatial bias correction method, it offers a flexible, simple, and effective method which can accommodate arbitrarily shaped areal units for both source and target zones.

### 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.

### Data availability

The data and code to reproduce the results are shared in an open

Zenodo repository. The link is provided in the data and software availability statement within the manuscript file.

## Acknowledgements

Part of the research was developed in the Young Scientists Summer Program 2020 at the International Institute for Applied Systems Analysis, Laxenburg (Austria) with financial support from the Academy of Finland (National Member Organization). The work was additionally supported by Maa-ja Vesiteknikan Tuki ry, the Doctoral Programme of the School of Engineering in Aalto University, by the Academy of Finland project WATVUL (Grant No. 317320). The work on the Upper Bima basin was conducted as part of the Belmont Forum Sustainable Urbanisation Global Initiative (SUGI)/Food-Water-Energy Nexus theme for which coordination was supported by the US National Science Foundation under grant ICER/EAR-1829999 to Stanford University.

## References

- Ahmed, K.F., Wang, G., Silander, J., Wilson, A.M., Allen, J.M., Horton, R., Anyah, R., 2013. Statistical downscaling and bias correction of climate model outputs for climate change impact assessment in the U.S. northeast. *Global Planet. Change* 100, 320–332. <https://doi.org/10.1016/j.gloplacha.2012.11.003>.
- Alcamo, J., Döll, P., Henrichs, T., Kaspar, F., Lehner, B., Rösch, T., Siebert, S., 2003. Development and testing of the WaterGAP 2 global model of water use and availability. *Hydrol. Sci. J.* 48, 317–338. <https://doi.org/10.1623/hysj.48.3.317.45290>.
- Asher, M.J., Croke, B.F.W., Jakeman, A.J., Peeters, L.J.M., 2015. A review of surrogate models and their application to groundwater modeling. *Water Resour. Res.* 51, 5957–5973. <https://doi.org/10.1002/2015WR016967>.
- Bardossy, A., Bogardi, I., Matyasovszky, I., 2005. Fuzzy rule-based downscaling of precipitation. *Theor. Appl. Climatol.* 82, 119–129. <https://doi.org/10.1007/s00704-004-0121-0>.
- Bennett, A., Stein, A., Cheng, Y., Nijssen, B., Mcguire, M., 2021. A process-conditioned and spatially consistent method for reducing systematic biases in modeled streamflow (preprint). *Hydrology*. <https://doi.org/10.1002/essoar.10507877.3>.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Burek, P., Satoh, Y., Kahil, T., Tang, T., Greve, P., Smilovic, M., Guillaumot, L., Zhao, F., Wada, Y., 2020. Development of the Community Water Model (CWatM v1.04) – a high-resolution hydrological model for global and regional assessment of integrated water resources management. *Geosci. Model Dev. (GMD)* 13, 3267–3298. <https://doi.org/10.5194/gmd-13-3267-2020>.
- Cannon, A.J., 2018. Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables. *Clim. Dynam.* 50, 31–49. <https://doi.org/10.1007/s00382-017-3580-6>.
- Chen, J., Broussard, W.P., Borrok, D.M., Speyrer, F.B., 2019. A GIS-Based Framework to Identify Opportunities to Use Surface Water to Offset Groundwater Withdrawals. *Water Resour. Manag.* <https://doi.org/10.1007/s11269-019-02298-5>.
- Chen, C., Hu, B., Li, Y., 2021. Easy-to-use spatial Random Forest-based downscaling-calibration method for producing high resolution and accurate precipitation data. *Hydrol. Earth Syst. Sci. Discuss.* 1–50. <https://doi.org/10.5194/hess-2021-332>.
- Chen, J., Arsenault, R., Brissette, F.P., Zhang, S., 2021. Climate change impact studies: should we bias correct climate model outputs or post-process impact model outputs? *Water Resour. Res.* 57, e2020WR028638 <https://doi.org/10.1029/2020WR028638>.
- Chen, J., Brissette, F.P., Leconte, R., 2011. Uncertainty of downscaling method in quantifying the impact of climate change on hydrology. *J. Hydrol.* 401, 190–202. <https://doi.org/10.1016/j.jhydrol.2011.02.020>.
- Comber, A., Zeng, W., 2019. Spatial interpolation using areal features: A review of methods and opportunities using new forms of data with coded illustrations. *Geogr. Compass* 13, e12465. <https://doi.org/10.1111/gec3.12465>.
- Cucchi, M., Weedon, G.P., Amici, A., Bellouin, N., Lange, S., Müller Schmied, H., Hersbach, H., Buontempo, C., 2020. WFD5: bias-adjusted ERA5 reanalysis data for impact studies. *Earth Syst. Sci. Data* 12, 2097–2120. <https://doi.org/10.5194/essd-12-2097-2020>.
- Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. Random forests for classification in ecology. *Ecology* 88, 2783–2792. <https://doi.org/10.1890/07-0539.1>.
- Dark, S.J., Bram, D., 2007. The modifiable areal unit problem (MAUP) in physical geography. *Prog. Phys. Geogr. Earth Environ.* 31, 471–479. <https://doi.org/10.1177/0309133307083294>.
- Flörke, M., Kynast, E., Bärlund, I., Eisner, S., Wimmer, F., Alcamo, J., 2013. Domestic and industrial water uses of the past 60 years as a mirror of socio-economic development: a global simulation study. *Global Environ. Change* 23, 144–156. <https://doi.org/10.1016/j.gloenvcha.2012.10.018>.
- Fotheringham, A.S., Wong, D.W.S., 1991. The modifiable areal unit problem in multivariate statistical analysis. *Environ. Plann.* 23, 1025–1044. <https://doi.org/10.1068/a231025>.
- Gharari, S., Clark, M.P., Mizukami, N., Knoben, W.J.M., Wong, J.S., Pietroniro, A., 2020. Flexible vector-based spatial configurations in land models. *Hydrol. Earth Syst. Sci.* 24, 5953–5971. <https://doi.org/10.5194/hess-24-5953-2020>.
- Goodchild, M.F., Lam, N.S.N., 1980. Areal interpolation: a variant of the traditional spatial problem. *Geo Process.* 1, 297–312.
- Gottschalk, L., 1993. Interpolation of runoff applying objective methods. *Stoch. Hydrol. Hydraul.* 7, 269–281. <https://doi.org/10.1007/BF01581615>.
- Guillaumot, L., Smilovic, M., Burek, P., de Bruijn, J., Greve, P., Kahil, T., Wada, Y., 2022. Coupling a large-scale hydrological model (CWatM) with a high-resolution groundwater flow model to assess the impact of irrigation at regional scale. *Geosci. Model Dev. Discuss. (GMD)* 1–37. <https://doi.org/10.5194/gmd-2022-161>.
- Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. *J. Hydrol.* 377, 80–91. <https://doi.org/10.1016/j.jhydrol.2009.08.003>.
- Hanasaki, N., Inuzuka, T., Kanae, S., Oki, T., 2010. An estimation of global virtual water flow and sources of water withdrawal for major crops and livestock products using a global hydrological model. *J. Hydrol.* 384, 232–244. <https://doi.org/10.1016/j.jhydrol.2009.09.028>.
- Hanasaki, N., Kanae, S., Oki, T., 2006. A reservoir operation scheme for global river routing models. *J. Hydrol.* 327, 22–41. <https://doi.org/10.1016/j.jhydrol.2005.11.011>.
- Hanasaki, N., Kanae, S., Oki, T., Masuda, K., Motoya, K., Shirakawa, N., Shen, Y., Tanaka, K., 2008. An integrated model for the assessment of global water resources – Part 1: model description and input meteorological forcing. *Hydrol. Earth Syst. Sci.* 12, 1007–1025. <https://doi.org/10.5194/hess-12-1007-2008>.
- Hempel, S., Frieler, K., Warszawski, L., Schewe, J., Piontek, F., 2013. A trend-preserving bias correction – the ISI-MIP approach. *Earth Syst. Dyn.* 4, 219–236. <https://doi.org/10.5194/esd-4-219-2013>.
- Hnilica, J., Hanel, M., Puš, V., 2017. Multisite bias correction of precipitation data from regional climate models. *Int. J. Climatol.* 37, 2934–2946. <https://doi.org/10.1002/joc.4890>.
- Hu, M., Huang, Y., 2020. atakrig: an R package for multivariate area-to-area and area-to-point kriging predictions. *Comput. Geosci.* 139, 104471 <https://doi.org/10.1016/j.cageo.2020.104471>.
- Hutengs, C., Vohland, M., 2016. Downscaling land surface temperatures at regional scales with random forest regression. *Remote Sens. Environ.* 178, 127–141. <https://doi.org/10.1016/j.rse.2016.03.006>.
- Ibarra, D.E., David, C.P.C., Tolentino, P.L.M., 2021. Technical note: evaluation and bias correction of an observation-based global runoff dataset using streamflow observations from small tropical catchments in the Philippines. *Hydrol. Earth Syst. Sci.* 25, 2805–2820. <https://doi.org/10.5194/hess-25-2805-2021>.
- IPCC, 2021. Climate Change 2021: the Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.
- Jakeman, A.J., Chen, S.H., Rizzoli, A.E., Voinov, A.A., 2008. Chapter one modelling and software as instruments for advancing sustainability. In: Jakeman, A.J., Voinov, A. A., Rizzoli, A.E., Chen, S.H. (Eds.), *Developments in Integrated Environmental Assessment, Environmental Modelling, Software and Decision Support*. Elsevier, pp. 1–13. [https://doi.org/10.1016/S1574-101X\(08\)00601-7](https://doi.org/10.1016/S1574-101X(08)00601-7).
- Kallio, M., 2020. *Process-Aware Interpolation Technique for Downscaling Hydrological Variables*, IIASA YSSP Report. IIASA, Laxenburg, Austria.
- Kallio, M., Guillaume, J.H.A., Virkki, V., Kumm, M., Virranta, K., 2021. Hydrostreamer v1.0 – improved streamflow predictions for local applications from an ensemble of downscaled global runoff products. *Geosci. Model Dev. (GMD)* 14, 5155–5181. <https://doi.org/10.5194/gmd-14-5155-2021>.
- Kallio, M., Virkki, V., Guillaume, J.H.A., van Dijk, A.J.J.M., 2019. Downscaling runoff products using areal interpolation: a combined pycnophylactic-dasymeric method. In: El Sawah, S. (Ed.), *MODSIM2019, 23rd International Congress on Modelling and Simulation*. Presented at the 23rd International Congress on Modelling and Simulation (MODSIM2019), Modelling and Simulation Society of Australia and New Zealand. <https://doi.org/10.36334/modsim.2019.K8.kallio>.
- Kim, Y.-T., Kwon, H.-H., Lima, C., Sharma, A., 2021. A novel spatial downscaling approach for climate change assessment in regions with sparse ground data networks. *Geophys. Res. Lett.* 48, e2021GL095729 <https://doi.org/10.1029/2021GL095729>.
- Knoben, W.J.M., Freer, J.E., Woods, R.A., 2019. Technical note: inherent benchmark or not? Comparing nash–sutcliffe and kling–gupta efficiency scores. *Hydrol. Earth Syst. Sci.* 23, 4323–4331. <https://doi.org/10.5194/hess-23-4323-2019>.
- Lange, S., 2019. Trend-preserving bias adjustment and statistical downscaling with ISIMP3BASD (v1.0). *Geosci. Model Dev. (GMD)* 12, 3055–3070. <https://doi.org/10.5194/gmd-12-3055-2019>.
- Latombe, G., Burke, A., Vrac, M., Levvasseur, G., Dumas, C., Kageyama, M., Ramstein, G., 2018. Comparison of spatial downscaling methods of general circulation model results to study climate variability during the Last Glacial Maximum. *Geosci. Model Dev. (GMD)* 11, 2563–2579. <https://doi.org/10.5194/gmd-11-2563-2018>.
- Lehner, B., Verdin, K., Jarvis, A., 2008. New global hydrography derived from spaceborne elevation data. *Eos Trans. Am. Geophys. Union* 89, 93–94. <https://doi.org/10.1029/2008EO100001>.
- Lenderink, G., Buishand, A., van Deursen, W., 2007. Estimates of future discharges of the river Rhine using two scenario methodologies: direct versus delta approach. *Hydrol. Earth Syst. Sci.* 11, 1145–1159. <https://doi.org/10.5194/hess-11-1145-2007>.
- Leyk, S., Nagle, N.N., Buttenfield, B.P., 2013. Maximum entropy dasymeric modelling for demographic small area estimation. *Geogr. Anal.* 45, 285–306. <https://doi.org/10.1111/gean.12011>.

- Lima, C.H.R., Kwon, H.-H., Kim, Y.-T., 2021. A Bayesian Kriging model applied for spatial downscaling of daily rainfall from GCMs. *J. Hydrol.* 597, 126095 <https://doi.org/10.1016/j.jhydrol.2021.126095>.
- Loonat, N.I., Van Dijk, A.I.J.M., Hutchinson, M.F., Weerts, A.H., 2020. Anomaly kriging helps to remove bias in spatial model runoff estimates. *Water Resour. Res.* 56, e2019WR026240 <https://doi.org/10.1029/2019WR026240>.
- Loritz, R., Kleidon, A., Jackisch, C., Westhoff, M., Ehret, U., Gupta, H., Zehe, E., 2019. A topographic index explaining hydrological similarity by accounting for the joint controls of runoff formation. *Hydrol. Earth Syst. Sci.* 23, 3807–3821. <https://doi.org/10.5194/hess-23-3807-2019>.
- Manley, D., 2014. Scale, aggregation, and the modifiable areal unit problem. In: Fischer, M.M., Nijkamp, P. (Eds.), *Handbook of Regional Science*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1157–1171. [https://doi.org/10.1007/978-3-642-23430-9\\_69](https://doi.org/10.1007/978-3-642-23430-9_69).
- Mennis, J., 2016. Dasytmetric spatiotemporal interpolation. *Prof. Geogr.* 68, 92–102. <https://doi.org/10.1080/00330124.2015.1033669>.
- Mennis, J., 2009. Dasytmetric mapping for estimating population in small areas. *Geogr. Compass* 3, 727–745. <https://doi.org/10.1111/j.1749-8198.2009.00220.x>.
- Nagle, N.N., Buttenfield, B.P., Leyk, S., Speilman, S., 2014. Dasytmetric modeling and uncertainty. *Ann. Assoc. Am. Geogr. Assoc. Am. Geogr.* 104, 80–95. <https://doi.org/10.1080/00045608.2013.843439>.
- Nahar, J., Johnson, F., Sharma, A., 2018. Addressing spatial dependence bias in climate model simulations—an independent component analysis approach. *Water Resour. Res.* 54, 827–841. <https://doi.org/10.1002/2017WR021293>.
- Naz, B.S., Kurtz, W., Montzka, C., Sharples, W., Goergen, K., Keune, J., Gao, H., Springer, A., Hendricks Franssen, H.-J., Kollet, S., 2019. Improving soil moisture and runoff simulations at 3 km over Europe using land surface data assimilation. *Hydrol. Earth Syst. Sci.* 23, 277–301. <https://doi.org/10.5194/hess-23-277-2019>.
- Niva, V., Horton, A., Heino, M., Kosonen, M., Kallio, M., Kinnunen, P., Muttarak, R., Taka, M., Varis, O., Virkki, V., Kummu, M., 2022. World's human migration patterns in 2000–2019 unveiled by high-resolution data (preprint). In: Review. <https://doi.org/10.21203/rs.3.rs-1827424/v1>.
- Pai, D.S., Sridhar, L., Rajeevan, M., Sreejith, O.P., Satbhai, N.S., Mukhopadhyay, B., 2014. Development of a new high spatial resolution (0.25° × 0.25°) long period (1901–2010) daily gridded rainfall data set over India and its comparison with existing data sets over the region. *Mausam* 65, 1–18.
- Paiva, R.C.D., Durand, M.T., Hossain, F., 2015. Spatiotemporal interpolation of discharge across a river network by using synthetic SWOT satellite data. *Water Resour. Res.* 51, 430–449. <https://doi.org/10.1002/2014WR015618>.
- Parajka, J., Merz, R., Sköien, J.O., Viglione, A., 2015. The role of station density for predicting daily runoff by top-kriging interpolation in Austria. *J. Hydrol. Hydromechanics* 63, 228–234. <https://doi.org/10.1515/johh-2015-0024>.
- Pool, S., Vis, M., Seibert, J., 2018. Evaluating model performance: towards a non-parametric variant of the Kling-Gupta efficiency. *Hydrol. Sci. J.* 63, 1941–1953. <https://doi.org/10.1080/02626667.2018.1552002>.
- R Core Team, 2020. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Razavi, S., Tolson, B.A., Burn, D.H., 2012. Review of surrogate modeling in water resources. *Water Resour. Res.* 48 <https://doi.org/10.1029/2011WR011527>.
- Reibel, M., Agrawal, A., 2007. Areal interpolation of population counts using pre-classified land cover data. *Popul. Res. Pol. Rev.* 26, 619–633. <https://doi.org/10.1007/s11113-007-9050-9>.
- Reichle, R.H., 2008. Data assimilation methods in the Earth sciences. *Adv. Water Resour., Hydrologic Remote Sensing* 31, 1411–1418. <https://doi.org/10.1016/j.advwatres.2008.01.001>.
- Saxe, S., Farmer, W., Driscoll, J., Hogue, T.S., 2021. Implications of model selection: a comparison of publicly available, conterminous US-extent hydrologic component estimates. *Hydrol. Earth Syst. Sci.* 25, 1529–1568. <https://doi.org/10.5194/hess-25-1529-2021>.
- Schmolke, A., Thorbek, P., DeAngelis, D.L., Grimm, V., 2010. Ecological models supporting environmental decision making: a strategy for the future. *Trends Ecol. Evol.* 25, 479–486. <https://doi.org/10.1016/j.tree.2010.05.001>.
- Sköien, J.O., Blöschl, G., 2007. Spatiotemporal topological kriging of runoff time series. *Water Resour. Res.* 43 <https://doi.org/10.1029/2006WR005760>.
- Sköien, J.O., Merz, R., Blöschl, G., 2006. Top-kriging – geostatistics on stream networks. *Hydrol. Earth Syst. Sci.* 10, 277–287.
- Sperna Weiland, F.C., Vrugt, J.A., van Beek, R., L, P.H., Weerts, A.H., Bierkens, M.F.P., 2015. Significant uncertainty in global scale hydrological modeling from precipitation data errors. *J. Hydrol.* 529, 1095–1115. <https://doi.org/10.1016/j.jhydrol.2015.08.061>.
- Stevens, F.R., Gaughan, A.E., Linard, C., Tatem, A.J., 2015. Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PLoS One* 10, e0107042. <https://doi.org/10.1371/journal.pone.0107042>.
- Tabari, H., Paz, S.M., Buekenhout, D., Willems, P., 2021. Comparison of statistical downscaling methods for climate change impact analysis on precipitation-driven drought. *Hydrol. Earth Syst. Sci.* 25, 3493–3517. <https://doi.org/10.5194/hess-25-3493-2021>.
- Tatem, A.J., 2017. WorldPop, open data for spatial demography. *Sci. Data* 4, 1–4. <https://doi.org/10.1038/sdata.2017.4>.
- Tatem, A.J., Campbell, J., Guerra-Arias, M., de Bernis, L., Moran, A., Matthews, Z., 2014. Mapping of maternal and newborn health: the distributions of women of childbearing age, pregnancies and births. *Int. J. Health Geogr.* 13, 2. <https://doi.org/10.1186/1476-072X-13-2>.
- Teutschbein, C., Seibert, J., 2013. Is bias correction of regional climate model (RCM) simulations possible for non-stationary conditions? *Hydrol. Earth Syst. Sci.* 17, 5061–5077. <https://doi.org/10.5194/hess-17-5061-2013>.
- Teutschbein, C., Seibert, J., 2012. Bias correction of regional climate model simulations for hydrological climate-change impact studies: review and evaluation of different methods. *J. Hydrol.* 456–457, 12–29. <https://doi.org/10.1016/j.jhydrol.2012.05.052>.
- Teutschbein, C., Seibert, J., 2010. Regional climate models for hydrological impact studies at the catchment scale: a review of recent modeling strategies. *Geogr. Compass* 4, 834–860. <https://doi.org/10.1111/j.1749-8198.2010.00357.x>.
- Tobler, W.R., 1979. Smooth pycnophylactic interpolation for geographical regions. *J. Am. Stat. Assoc.* 74, 519–530. <https://doi.org/10.2307/2286968>.
- Tramberend, S., Burtcher, R., Burek, P., Kahil, T., Fischer, G., Mochizuki, J., Greve, P., Kimwaga, R., Nyenje, P., Ondiek, R., Nakawuka, P., Hyandye, C., Sibomana, C., Luoga, H.P., Matano, A.S., Langan, S., Wada, Y., 2021. Co-development of East African regional water scenarios for 2050. *One Earth* 4, 434–447. <https://doi.org/10.1016/j.oneear.2021.02.012>.
- van Beek, L.P.H., Wada, Y., Bierkens, M.F.P., 2011. Global monthly water stress: 1. Water balance and water availability. *Water Resour. Res.* 47 <https://doi.org/10.1029/2010WR009791>.
- Virkki, V., 2019. *The Value of Open-Source River Streamflow Estimation in Southeast Asia* (Master's Thesis). Aalto University.
- von Storch, H., Zorita, E., 2019. The history of ideas of downscaling—from synoptic dynamics and spatial interpolation. *Front. Environ. Sci.* 7, 21. <https://doi.org/10.3389/fenvs.2019.00021>.
- Wada, Y., Wisser, D., Bierkens, M.F.P., 2014. Global modeling of withdrawal, allocation and consumptive use of surface water and groundwater resources. *Earth Syst. Dyn.* 5, 15–40. <https://doi.org/10.5194/esd-5-15-2014>.
- Wang, Q., Shi, W., Atkinson, P.M., Zhao, Y., 2015. Downscaling MODIS images with area-to-point regression kriging. *Remote Sens. Environ.* 166, 191–204. <https://doi.org/10.1016/j.rse.2015.06.003>.
- Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O., Schewe, J., 2014. The inter-sectoral impact model intercomparison project (ISI-MIP): project framework. *Proc. Natl. Acad. Sci. USA* 111, 3228–3232. <https://doi.org/10.1073/pnas.1312330110>.
- Wen, H., Brantley, S.L., Davis, K.J., Duncan, J.M., Li, L., 2021. The Limits of homogenization: what hydrological dynamics can a simple model represent at the catchment scale? *Water Resour. Res.* 57, e2020WR029528 <https://doi.org/10.1029/2020WR029528>.
- Wright, J.K., 1936. A method of mapping densities of population: with cape cod as an example. *Geogr. Rev.* 26, 103–110. <https://doi.org/10.2307/209467>.
- Wright, M.N., Ziegler, A., 2017. Ranger: a fast implementation of random forests for high dimensional data in C++ and R. *J. Stat. Software* 77, 1–17. <https://doi.org/10.18637/jss.v077.i01>.