

This document contains our answers to the referees. Before each answer, the corresponding remark or issue raised by the referee is quoted in italics. All line numbers refer to the old manuscript (used for publication in HESSD), unless otherwise stated. A marked-up version of the manuscript, clearly showing the changes brought to the text in response to the referees' comments, can be found at the end of this document.

Answer to Anonymous Referee #1

We wish to thank the referee for his/her comments that will contribute to improve the quality of our manuscript. Below is our response to the remarks and issues raised by the referee.

General Comments

“Generally, the claim that the model is a net improvement over other regressions models is a bit of an overstatement. I see it more as a regression model for which the selection of predictors is based on known relationships between water temperature, climate, solar radiation and local physiography/hydrology.”

We certainly acknowledge this comment and fully agree with the referee. The analytical solution to the heat-balance equation leads to an expression, certain terms of which cannot be readily evaluated. These terms are approximated using linear (and non-linear) regression expressions of known physiographic properties of the catchments, hereby introducing statistical aspects into the model. Moreover, among the different regression expressions used to approximate each unknown term, the best one is selected using the Akaike information criterion, which is a well-established statistical technique. To address the referee's comment, we modified the general tone of the manuscript. We thus rewrote Sect. 1.5 and the first two paragraphs of the conclusion so as to describe our model as being statistical, but with a structure incorporating physical considerations. We also changed the denomination “physically-based model” into “physics-inspired statistical model”, and changed “regression model” (used for comparison with the physics-inspired model) into “standard regression model” to make the distinction clearer between the two. The assertion that our model avoids the typical limitations of statistical models (lines 228–229) was removed.

“The authors claim (line 210) that the selection of predictor variables relies on the experience of the authors in statistical models. This is not always necessarily the case. Some authors have used objective or semiobjective methods for this selection. In the submitted manuscript, the selection of proxy variables for the net radiative flux is also based on experience and known relationships.”

Although probably a bit overstated, our claim is, according to us, justified. We agree with the referee's comment regarding the fact that objective or semi-objective methods can be used for variable selection. This is exactly what we did when selecting the best model among the different variants considered. However, these methods can only be applied to a predefined set of variables to be tested. And the definition of this latter set actually relies on the experience and understanding of the author. We added a few sentences in the text in order to clarify this point:

“Although some objective methods can help to perform th[e predictor variable] selection (Burnham and Anderson, 2002), the original set of variables on which these methods act must initially be indicated by the user. In the end, the choice of predictor variables is necessarily affected to some extent by the training and experience of the authors, hereby introducing

some diversity in the sets of predictor variables.” (lines 217–221 of the marked-up manuscript below)

Moreover, the use of these automated techniques should, in our opinion, not be substituted to physical understanding whenever possible. Some variables may be selected which actually appear to incidentally have some correlation with the predictand, but do not have any physical relationship with it. One might rightfully argue that our discourse is ambiguous, criticizing the statistical techniques on the one hand, but relying on them on the other hand. In the present case, we are limited by the fact that we aim to develop a model for prediction in ungauged basins, and we are hereby limited to the use of the only variables which are available from classical spatial data sets. Not knowing the relationship between these regional scale variables and the unknown terms we wish to compute, we have to rely on statistical techniques.

“Parcimony was only used for model variants, but when the author’s model is compared to the regression approach, parcimony is evacuated from the comparison.”

From our understanding, the “model variants” mentioned by reviewer refer to the variants of the physics-inspired model (previously referred to as “physically-based model”). If this is indeed what was meant, we would like to kindly mention that this is actually not the case. As the referee pointed out in line 530 (see Specific Comments below), we limited the number of parameters also in the standard regression model. The maximum number of parameters was fixed to 6 in this model in order to match the maximum number of parameters in the physics-inspired model and hereby provide a more even comparison between the two models. This point was clarified in the text:

“This limitation [of the number of parameters] was introduced in order to avoid over-parametrization, but also to ensure that the number of parameters in the final standard regression model was about the same as in the physics-inspired model, hereby guarantying a more even comparison between the two.” (lines 757–759 of the marked-up manuscript below)

Contrary to what we wrote in our first response to the referees, we thought that further discussion on parcimony was not needed in Section 6, since comparison between the two models is fair.

Specific Comments

“Line 78: Last sentence appears to be a strong statement, perhaps not fully substantiated.”

The original statement,

“As such, no significant decrease in the prediction errors should be awaited from a change in the statistical modeling technique”

was modified into

“These studies tend to suggest that no significant decrease in the prediction errors should be awaited from a change in the statistical modeling technique.”

The new formulation insists more on the fact that the conclusion is drawn solely based on the reported studies and might not be generally valid.

“Line 185: ‘for each group, they computed the characteristics of the thermal regime by river class...’ ”

We used yet another formulation than the new one proposed by the referee:

“They computed the characteristics of the typical thermal regime of each group.” (lines 197–198 of the marked-up manuscript below).

“Line 210: As stated earlier, not all statistical models are built using a subjective selection of predictors. For instance, multivariate analysis has been used in some cases.”

See our response under General Comments above.

“Line 236: Linear functions of available data are being used. Why linear? Is there proof of linearity in all cases?”

As stated under General Comments, we completely rewrote Sect. 1.5. It is actually not true that we use only linear relationships of available quantities to approximate all unknown variables. As a matter of fact, a power function is tested to express the fraction of discharge originating from lateral inflow (Eq. (17)). We however decided not to enter into too much detail in the introduction, as this would distract the reader from our main point. We therefore decided to maintain the sentence stating that all unknown terms are approximated as multi-linear regression expressions of available data. We made clear that this choice was motivated by our lack of knowledge regarding the nature of the relationships between the predictand and the predictors (lines 269–270 of the marked-up manuscript below). We justified our choice with the fact that the main non-linearities may already have been captured by the general form of the model, namely by the analytical formula obtained by solving the heat balance equation. Such non-linearities appear mainly in the respective expressions of the weighting factors ω_1 , ω_2 , ω_3 (Eqs. (9)–(14)), and in the definition of the distance-weighted averaging operator $\langle \cdot \rangle_L$ (Eq. (A9)). Based on this, it might be assumed that the specific form of the expressions used to approximate the unknown terms have a minor effect compared to the general structure of the model.

Additionally to Sect. 1.5, we also added a paragraph in the discussion to further comment on this point (lines 1175–1186 of the marked-up manuscript below). We mention there that alternative formulations to a linear expression have been tested for the term $\gamma\phi_r$ accounting for the net heat flux at the air–water interface. A power-law function was considered, but could not be used because of convergence issues during model calibration. Similarly, a more physically-based expression was tested, which proved to decrease the model precision as compared to the linear expression.

“Line 315: ‘All sources of the network are supposed to have the same discharge’. What are the potential repercussions of this simplification, especially in small drainage basins?”

Contrary to our first response to the referee, we commented on this point in the discussion rather than in the section describing the simplifying assumptions:

“Assumption (5) stating that all sources in a given catchment have the same discharge rate is also disputable. This is particularly true for small catchments, where the short distance to the outlet and the low number of sources do not allow the averaging effect to be significant enough to compensate for the introduced error.”

“Line 404: Meteorological stations were selected based on the fact that they are located at 20 km or less from the outlet. What is the impact of this selection when L_c is larger than 20 km?”

The referee raises an interesting point. While computing air temperature as a function of altitude in each catchment, we limited ourselves to meteorological stations located at less than 20km from the

watershed outlet. On the other hand, $L_c = 32\text{km}$ in the model used to estimate the standard deviation of water temperature (see Table 6). This apparent discrepancy is actually not considered to be a limitation, due to the fact that only 2 of the 26 selected catchments have an area extending beyond the 20km-radius disk centered on the catchment outlet. As such, increasing the range of considered meteorological stations would not improve the quality of the interpolation but in 2 cases. This point was added in the manuscript (see lines 384–387 of the marked-up manuscript below). Moreover, the stream networks in those last 2 catchments have a marked sinuosity, implying that all their stream segments at a streamwise distance less than 32 km from the outlet are actually located within the 20-km-radius disk centered on the outlet point. The proposed methodology for air temperature interpolation is therefore not considered to be at odds with the rest of our work.

“Line 421: Speaking of subjective selection: setting at least two of the coefficients to zero. At least? Were there more? Why two?”

By “setting at least two [...] to zero”, we meant that all linear functions involving any possible combination of either one or two of the predictors variables were considered for approximating the term $\gamma\phi_r$. The arbitrary choice to consider expressions with at most two terms (plus the intercept) was made in order to restrict the number of model parameters. We acknowledge the paragraph was not very clear and completely rewrote it following the advices of Anonymous Referee #2 (see lines 602–666 of the marked-up manuscript below). We notably emphasized the fact that the restriction to a maximum of two terms is arbitrary.

“Line 530: Number of terms in the model was fixed to six to prevent over-parameterization. Given that the final model is a multiple regression? Why not use a stepwise approach?”

While determining the structure of the multilinear model, we restricted ourselves to a maximum of 6 predictor variables. This choice was made in order to avoid over-parametrization, but also to ensure that the number of parameters in both models (the physics-inspired one and the standard regression one) were about the same. This guaranteed a more even comparison between the two modeling techniques. As mentioned under General Comments, this point was clarified in the manuscript. We also discarded the step-wise approach, since it may fail to identify the best set of predictor variables (Miller, 2002) and is subject to some criticism (Harrell, 2001).

“Line 595 (and elsewhere): ‘...data are...’ ”

We thank the referee for the correction, which we applied in the whole manuscript.

“Line 620: Threshold of 30°C. Although this is a high value, might it not be feasible that some rivers occasionally reach this temperature in shallow, open areas?”

In response to this comment, we added the following lines in the manuscript:

“Although the limit of 30°C might be naturally reached in shallow areas, some temperature series showed clear evidence that such temperature was recorded as a result of the sensor being out of water. As a consequence, it was decided to remove all data points above 30°C, potentially discarding correct data.” (lines 364–367 of the marked-up manuscript below)

“Line 669: Subjective selection of 5 of the 26 catchments. Why not do a bootstrap on the stations excluded?”

The following lines were added in the manuscript to answer the referee's comment:

"A bootstrap on the validation stations was not possible because of too high computational requirements. Indeed, Burnham and Anderson (2002) recommend using at least 10'000 bootstrap samples, which led to a prohibitively high number of model evaluations in our case."
(lines 898–901 of the marked-up manuscript below)

"Line 713: Why not show AIC separately for calibration and validation?"

We actually followed Burnham and Anderson (2002) in computing the value of the corrected Akaike information criterion over the entire data set (calibration + validation sets). The reference was cited a few sentences above in the paragraph, but we acknowledge it was not clear that we followed their recommendation in doing so. We explicitly cited their work in the sentence to make it clear.

"Lines 1015–1020 read more like a discussion than a conclusion."

We actually suppressed lines 1015–1018, judging that they were not particularly relevant. Following the referee's suggestion, we also moved lines 1018–1024 to the discussion (see lines 1181–1186 of the marked-up manuscript below).

References

Burnham, K. P. and Anderson, D. R.: *Model selection and multimodel inference: a practical information theoretic approach*, Springer, 2002.

Harrell, F. E.: *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis*, Springer, New York, 2001.

Miller, A. J.: *Subset selection in regression*, Chapman & Hall, London, 2002.

Answer to Anonymous Referee #2

We wish to thank the referee for his/her comments that will help sharpening our manuscript, and for his/her interest in our work. Below is our response to the remarks and issues raised by the referee.

"1. Section 1.1 and 2.1. I recommend that the authors broaden the context by referring to studies that have applied deterministic models at the scale of a medium-scale catchment, including SNTemp (Bartholow, 1991; Mattax and Quigley, 1989), a model based on the HSPF (Hydrological Simulation Program – FORTRAN) model developed by the US Environmental Protection Agency and the US Geological Survey (Chen et al., 1998a, 1998b) and CEQUEAU (St.-Hilaire et al., 2000). These models all have the capacity to generate spatially distributed predictions of the annual cycle of stream temperature. Another study that should be referenced is Allen et al. (2004), who developed a deterministic model for spatially distributed stream temperature prediction. Although Allen et al. focused on summer maximum temperature, the approach could be adapted for year-round application."

We thank the referee for the list of references. We were actually already familiar with these studies, as we are currently working on the development and application of a spatially-distributed hydrological model with stream temperature prediction abilities. When presenting the results of our

current work in a publication, we will certainly not omit to cite the suggested models. In the context of the present paper, we however think the focus should remain on stream prediction in ungauged basins using statistical techniques. This also fits with the revisions we brought to the paper, namely describe our work as a statistical model whose structure derives from physical principles, rather than as a physically-based model (see points 4 and 9 below, and our response to Anonymous Referee #1).

“2. The review of stream temperature modelling in Switzerland seems a bit out of place in an international journal. I recommend that the authors add a sentence or two to set the Swiss experience into a broader international context. For example, what is different about Switzerland that sets it apart from other geographic settings in terms of what can be learned about stream temperature variability and modeling?”

We acknowledge that a section about Switzerland seemed a bit awkward in the general introduction. We therefore broadened the scope of Sect. 1.4 to the entire European Alps, although very little information could not be added as compared to Switzerland alone, due to the scarcity of the publications on this subject.

“3. Section 1.5. It should be mentioned that a major constraint on the inclusion of predictor variables is the availability and reliability of data sources. It is particularly challenging to quantify riparian shading based on the coarseness of vegetation data sets that are available in many jurisdictions, and the difficulty of representing the seasonally changing effects of deciduous vegetation. This point could also be reflected on in the discussion in the context of the model’s inability to distinguish among different buffer widths for characterizing riparian vegetation.”

We thank the referee for this suggestion. While completely rewriting Sect. 1.5 in response to the comments of Anonymous Referee #1, we included the following lines:

“It should be mentioned that th[e] diversity [of predictor variables] also largely results from the varying availability and reliability of data among different geographic areas. This is particularly true for riparian shading, which is never directly measured and can only be estimated based on the data at disposal. For example, Isaak et al. (2010) approximated riparian shading using a sophisticated combination of satellite orthoimages and ground hemispherical canopy pictures, whereas DeWeber and Wagner (2014) could only rely on country-wide landuse data.” (lines 233–239 of the marked-up manuscript below)

We also added a sentence in the discussion:

“This also points at the difficulty to adequately account for the effect of riparian vegetation using the available spatial data sets, which often lack important details such as the distinction between deciduous and coniferous forest.” (lines 1122–1125 of the marked-up manuscript below)

“4. p. 4091, line 6ff.¹ This reference to social science seems unnecessary. I would argue that the need to calibrate any model in any discipline is an admission of inadequate knowledge about a system or inability to characterize its boundary conditions. I recommend that the authors begin the paragraph with a statement that expresses the motivation for this work in the context of the relative strengths and weaknesses of statistical models (e.g., see my introductory comments, above).”

¹ The page and line numbers quoted by Anonymous Referee #2 refer to the online version of the discussion paper [note of the author]

We acknowledge that the reference to Social Sciences was unnecessary. While rewriting Sect. 1.5, we however found it more logical to start with a description of the standard statistical methods used to predict stream temperature in ungauged catchments, rather than a general discussion of the “relative strengths and weaknesses of statistical models.” We however appreciated the idea of the referee, and slightly modified our general introduction in order to motivate our work according to his/her suggestions (see lines 36, 51–52, 54–55).

“5. A number of assumptions are made that are likely not to be valid based on a priori reasoning. To what extent might the lack of validity of the assumptions have limited the model’s performance? Three examples follow.

a. p. 4094, point (v). In fluvial geomorphology, stream width is commonly modeled as a power-law function of discharge, with a typical exponent of 0.5, which would generate different scaling than the assumed proportional relation between w and Q .

b. The lateral inflow rate, q_l , would be expected to increase with elevation, especially during the spring, when upper elevations experience snowmelt while the lower elevations do not. This would introduce a covariance between q_l and T_l that would not be included in the spatial average along the stream network.

c. The radiation term is a linear function of several components, but process-based reasoning indicates that there should be interactions (i.e., product terms).”

We followed the referee’s advices by writing two new paragraphs in the discussion about the validity and limitations of each approximation (see lines 1159–1186 of the marked-up manuscript below).

“6. The authors state that monthly net radiation is dominated by solar radiation (p. 4100, line 1ff). Is this true? For example, in a heavily shaded stream, I would expect incident longwave radiation to be larger than incident solar radiation at the stream surface even”

We thank the referee for pointing out an erroneous statement. Citing the work of Johnson (2004), Webb et al. (2008) indeed report that in a watershed in Oregon, USA, “an open reach under full sun [was measured to] experience[...] a net energy gain of 580 Wm^{-2} but a reach under full shade [...] a net loss of 149 Wm^{-2} ” at midday in July. We suppressed the erroneous sentence.

“7. I like the authors’ approach to model testing, which, according to the hierarchical approach promoted by Vit Klemes (1986), could be termed ‘split sample,’ ‘proxy basin’ and ‘proxy basin split sample.’ The authors could consider including a differential split sample and a proxy basin differential split sample approach to provide a further assessment of model robustness. In the differential split sample approaches, the record is split based on climatic conditions. For example, the model could be calibrated for wet/cool conditions and then tested on dry/warm conditions. This test might be revealing about the relative merits of the analytical and empirical models.”

We thank the referee for this proposition. It would indeed have been valuable to calibrate the models on a particular period of the year and evaluate it on another. However, this would have been relevant only for the physics-inspired model (previously referred to as “physically-based”), since the standard regression one simulates the entire annual curve of stream temperature at once. Calibrating the standard regression model on given months of the year would have introduced a strong bias in its estimation of the annual mean and standard deviation of the monthly mean stream temperature. As a consequence, this procedure would have penalized it because of its structure

rather than reveal its actual performances, and might therefore have erroneously made the physics-inspired model appear better in comparison.

“8. The authors used all data for model selection, including the test data set (p. 4113, line 8). Would different models have been selected had only the calibration set been used in this first step?”

The same question was asked by Anonymous Referee #1 in his specific comments. We actually followed Burnham and Anderson (2002) in using both calibration and validation data sets to select the best model. The citation was already present a few lines above in the manuscript, but we repeated it for the sake of clarity (see line 944 of the marked-up manuscript below).

“9. The authors refer to their model as ‘physically-based.’ Given the high degree of parameterization and the many simplifying assumptions, it is probably more accurate to refer to the model as ‘analytical.’ ”

Anonymous Referee #1 raised the same issue in his general comments. We changed the denomination of the model from “physically-based” to “physics-derived statistical”. The “regression model” was also renamed “standard regression model” so as to clearly make the distinction between the two.

“[10]a. p. 4093, line 3. The authors use sigma to represent the stream surface heat flux, and later for the standard deviation of monthly stream temperatures. I suggest using a different symbol for stream surface heat flux (e.g., phi, to be consistent with later usage).”

We thank the referee for this suggestion, and followed his/her advice.

“[10]b. I found it difficult to follow the description of the model and had to read it several times. For example, the authors refer to measured solar radiation but do not explain where it was measured or how it was processed, except by saying it was assumed to be a linear function of elevation. Also, the authors present Eq. (16) and then discuss parameter estimation before explaining how the variables were quantified. I recommend that the authors attempt to make the model description simpler to follow as they revise the manuscript, perhaps by changing the order of presentation (e.g., describe data sources prior to describing the models).”

We acknowledge some parts of the model description were unclear. We followed the referee’s advice by first describing the data prior to presenting the models. We also added new a section describing the meteorological data used by the models (see lines 375–397 of the marked-up manuscript below). We also completely rewrote the paragraph referred to by the referee when mentioning unclear parts (see lines 602–666).

“[10]c. p. 4410, line 22. The authors refer to groundwater “infiltrating” – “discharging” would be more appropriate.”

We followed the referee’s advice.

References

Burnham, K. P. and Anderson, D. R.: *Model selection and multimodel inference: a practical information theoretic approach*, Springer, 2002.

Johnson S.L.: *Factors influencing stream temperatures in small streams: substrate effects and a shading experiment*, Canadian Journal of Fisheries and Aquatic Sciences, 61, 913–923, 2004.

Webb, B. W., Hannah, D. M., Moore, R. D., Brown, L. E., and Nobilis, F.: *Recent advances in stream and river temperature research*, Hydrol. Process., 22, 902–918, 2008.

Stream temperature prediction in ungauged basins: review of recent approaches and description of a new ~~physically-based analytical~~ physics-derived statistical model

A. Gallice¹, B. Schaefli¹, M. Lehning^{1,2}, M. B. Parlange^{1,3}, and H. Huwald¹

¹School of Architecture, Civil and Environmental Engineering (ENAC), École Polytechnique Fédérale de Lausanne (EPFL), Switzerland

²SLF, WSL Institute for Snow and Avalanche Research, 7260 Davos, Switzerland

³Faculty of Applied Sciences, University of British Columbia, Vancouver, Canada

Correspondence to: A. Gallice (aurelien.gallice@epfl.ch)

Abstract. The development of stream temperature regression models at regional scales has regained some popularity over the past years. These models are used to predict stream temperature in ungauged catchments to assess the impact of human activities or climate change on riverine fauna over large spatial areas. A comprehensive literature review presented in this study shows that the temperature metrics predicted by the majority of models correspond to yearly aggregates, such as the popular annual maximum weekly mean temperature (MWMT). As a consequence, current models are often unable to predict the annual cycle of stream temperature, nor can the majority of them forecast the interannual variation of stream temperature. This study presents a new statistical model to estimate the monthly mean stream temperature of ungauged rivers over multiple years in an Alpine country (Switzerland). Contrary to ~~the similar~~ models developed to date, which ~~mostly rely upon statistical regression to express stream temperature as a function of physiographic and climatic variables~~ are mostly based on standard regression approaches, this one ~~rests upon attempts at incorporating physical aspects into its structure. It is based on~~ the analytical solution to a simplified version of the energy-balance equation over an entire stream network. ~~This physically-based~~ Some terms of this solution cannot be readily evaluated at the regional scale due to the lack of appropriate data, and are therefore approximated using classical statistical techniques. This physics-inspired approach presents some advantages: (1) ~~the functional form linking stream temperature to the predictor variables~~ the main model structure is directly obtained from first principles, (2) ~~the spatial extent over which the predictor variables are averaged naturally arises during model development, and (3) most of~~ the regression coefficients can be interpreted from a physical point of view—their values can therefore be constrained to remain within plausible bounds. The evaluation of the model over a new

freely available data set shows that the monthly mean stream temperature curve can be reproduced with a root mean square error of $\pm 1.3^{\circ}\text{C}$, which is similar in precision to the predictions obtained with a multi-linear regression model. We illustrate through a simple example how the physical ~~basis of the model~~ aspects contained in the model structure can be used to gain more insight into the stream temperature dynamics at regional scales.

1 Introduction

Among the parameters affecting the ecological processes in streams, temperature occupies a predominant role. It influences the concentration of chemicals, such as dissolved oxygen, and may increase the toxicity of dissolved substances (Langford, 1990). It also affects the life cycle of many fish species, particularly the salmonids whose rate of spawning, timing of birth and rate of death are directly influenced by stream temperature (Caissie, 2006; Benyahya et al., 2007). Water temperature is also a relevant factor for many thermal power plants which rely on cooling by river water, and whose electricity production decreases when water temperature exceeds a certain limit (Haag and Luce, 2008).

~~Stream~~ As a result of the raising concern about climate change and water management impacts on aquatic life, stream temperature modeling has regained some interest over the past 10–15 years; ~~which led to~~. This fostered the development of many stochastic and deterministic models (e.g. Mohseni et al., 1998; Segura et al., 2014; Chang and Psaris, 2013; DeWeber and Wagner, 2014; Meier et al., 2003; Westhoff et al., 2007). The former type relies on a statistical analysis to empirically relate stream temperature to climatic and physiographic variables, such as air temperature, discharge, altitude or channel width (see Benyahya et al., 2007, for a complete review of this subject). Deterministic models, on the other hand, rely on a physically-based formulation of the stream energy conservation to compute water temperature (Caissie, 2006). Both model types have usually been applied to a single stream reach or a limited number of catchments (e.g. Sinokrot and Stefan, 1993; Roth et al., 2010; Caissie et al., 2001; Caldwell et al., 2013; Grbić et al., 2013). ~~However~~ As a response to the lack of stream temperature data, some studies have recently attempted at developing regionalized ~~stream temperature~~ models. This effort was certainly ~~fostered~~ encouraged by the incentive of the International Association of Hydrological Sciences (IAHS), which set the focus of the last decade on hydrological prediction in ungauged basins (Sivapalan et al., 2003; Hrachowitz et al., 2013). In the case of stream temperature, ~~authors relied~~ the difficulty to meet the data requirements of the physically-based models led the authors to mostly rely on statistical approaches to make predictions in ungauged catchments; ~~to the best of our knowledge, no study has investigated the regionalization of a deterministic model to date~~. However, the validity of these models for studying climate change impacts or water management techniques is not assessed yet.

In this paper, more than 30 studies describing regionalized statistical models for stream temperature estimation were reviewed to put our work in a larger context (see Table 1). The extensive introduction below discusses several aspects of the reviewed literature which motivated the development of the novel stream temperature model described in the next section.

60 1.1 Predictions with limited precision

One recurring issue described in the reviewed literature is the difficulty to predict stream temperature with a high level of precision. A typical example is the statistical model of Isaak et al. (2010) for the estimation of mean summer stream temperature (15 July–15 September) in the Boise River Basin, Idaho. Despite considering a significant number of predictor variables and two different modeling
65 approaches a priori, Isaak et al. could not reduce the root-mean-square error (RMSE) of their model below 1.5°C. Prediction uncertainties of the same order of magnitude are reported e.g. by Wehrly et al. (2009), Ruesch et al. (2012), Moore et al. (2013) or Hill et al. (2013).

In general, it seems that the model error originates partly from the lack of appropriate field data, such as measures of riparian shading, groundwater infiltration or irrigation withdrawals (Moore et al.,
70 2013). As noted by Hill et al. (2013), “these types of data are not readily available everywhere and will take time to develop”. In the mean time, they can in some circumstances be accounted for through indirect measures. For example, Tague et al. (2007) used the geological aquifer type as a proxy for the presence or absence of groundwater infiltration. Similarly, Hrachowitz et al. (2010) and Scott et al. (2002) estimated riparian shading based on riparian forest coverage, computed over
75 buffer areas of various widths and lengths around the streams. In the absence of such proxies, the model cannot represent some known processes and must concede some increase in its prediction error (Moore et al., 2013). The size of the areas over which stream temperature is modeled—and hereby the diversity of encountered climatic and geomorphologic conditions—constitutes another factor potentially explaining the model uncertainties for some studies.

80 Regarding the impact of the modeling approach, Wehrly et al. (2009) investigated four different statistical model types and showed that their difference in prediction accuracy was relatively small. The same conclusion was reached by Daigle et al. (2010), who compared four other modeling approachestechniques. Isaak et al. (2010) found that networked kriging regression performed better than multi-linear regression over the calibration data set, but this assertion became much less evident
85 over the validation set. Similarly, Pratt and Chang (2012) and Chang and Psaris (2013) concluded that geographically weighted regression is slightly more accurate than multi-linear regression, but they did not validate their results on an independent data set. As such, These studies tend to suggest that no significant decrease in the prediction errors should be awaited from a change in the statistical modeling technique.

90 Further comparisons between the different models reported in the literature is unfortunately hindered by the diversity of temperature metrics and error measures used by the authors. As mentioned

in several studies already, we advocate here the systematic use of the different error measures that are RMSE, bias and coefficient of determination R^2 for the evaluation of the model precision, possibly combined with a benchmark model (Schaeffli and Gupta, 2007). It should be noted that R^2 is also referred to as Nash-Sutcliffe efficiency by the hydrological community (Nash and Sutcliffe, 1970), and is defined as one minus the ratio of the model error variance over the variance of the observed data.

1.2 Few models can predict the stream temperature annual cycle

Inspecting Table 1, it can be seen that most regionalization efforts have concentrated on some particular periods of the year. For example, Jones et al. (2006), Isaak et al. (2010) and Chang and Psaris (2013) focused on the annual maximum of the 7-day moving average of the daily maximum temperature (MWMT). Similarly, both Pratt and Chang (2012) and Hill et al. (2013) aimed at estimating mean stream temperature in summer and winter. Very few studies have actually attempted at deriving regional models to compute the complete annual cycle of stream temperature over several years.

Miyake and Takeuchi (1951) and Stefan and Preud'homme (1993) were probably the first authors to address this issue; they relied on linear regression against air temperature to simultaneously estimate stream temperature at multiple sites. However, their respective works are restricted to a limited number of rivers (20 and 11, respectively) and could probably not be applied to larger areas. In an attempt at generalizing these models, Ozaki et al. (2003) and Kelleher et al. (2012) separately regressed stream temperature against air temperature in each one of the catchments they considered, and subsequently regionalized the slopes of the regression lines. However, both studies were only partly successful in completing the regionalization step, since the modeled regression slopes had large prediction errors. They would additionally have had to model the intercepts of the regression lines to completely regionalize the stream–air temperature relationship. In a similar fashion, Johnson et al. (2014) relied on the logistic equation introduced by Mohseni et al. (1998) to relate stream temperature to air temperature in each catchment. Also they faced difficulties to regress the equation parameters against geomorphological properties of the catchments. The two most complete works on the regionalization of the linear stream–air temperature relationship were recently conducted by Ducharne (2008) and Segura et al. (2014). These two studies attempted at regionalizing both the slopes and intercepts of the regression lines between stream and air temperature. To this end, Ducharne grouped the streams according to their Strahler order and fitted a single line in each group. Segura et al., on the other end, expressed the slopes and intercepts as linear combinations of climatic and physiographic variables. The model of Ducharne had nominally a higher explanatory power ($R^2 = 0.88$ – 0.96 depending on the Strahler order) than Segura et al.'s model ($R^2 = 0.79$), but was effectively based on about 10 times fewer rivers.

Instead of using air temperature as independent variable, Bogan et al. (2003) relied on equilibrium temperature. This variable corresponds to the stream temperature at which the net energy flux at the

air–water interface vanishes (e.g. Edinger et al., 1968). It was shown by Bogan et al. to be a fairly good estimator of stream temperature for almost 600 rivers in the Eastern and Central United States, with a prediction error of about 3°C.

As an alternative to the above-mentioned studies, the annual cycle of stream temperature has been modeled by some authors as a function of time directly, rather than air or equilibrium temperature. Hrachowitz et al. (2010), Imholt et al. (2013) and Rivers-Moore et al. (2012) expressed water temperature as a linear combination of climatic and physiographic variables for each month of the year separately. Their models were derived for a particular year, but can be transferred to other years by estimating stream temperature at a few measurement points using Mohseni’s logistic equation and fitting the multi-linear regression model to the resulting values (Hrachowitz et al., 2010). Based on a similar approach, Macedo et al. (2013) succeeded to derive one single regression model to estimate daily mean stream temperature at 12 different sites in Brazil over 1.5 years. The performance of their model was not tested using data from subsequent years, though.

Johnson (1971) relied on a yet different technique to estimate the thermal regime of 6 rivers in New Zealand. He first fitted the stream temperature annual cycles with sine curves. In a second step, he identified the physiographic properties of the catchments which best correlated with the fit coefficients. The focus of his study being on the investigation of these physiographic properties, he did not evaluate the prediction error of his model. Although not intended for this purpose, the work of Garner et al. (2014) is based on a somewhat similar approach and may be used to get a first estimate of the annual cycle of temperature in UK streams. The authors classified rivers into several groups according to the shape and magnitude of their respective thermal regimes. Then, they investigated the similarities and dissimilarities of some geomorphological properties among and between the groups. This processing could be inverted to infer the thermal regime from the physiographic properties of the catchments.

Finally, some studies have evaluated the possibility of modeling the time evolution of stream temperature using machine learning techniques. For example, DeWeber and Wagner (2014) trained an artificial neural network to reproduce daily mean temperature values from May to October over more than 30 years for 1080 streams in the Eastern United States. Their approach could be easily extended so as to model the complete annual cycle of stream temperature each year.

1.3 Space-averaging of the predictor variables

Some of the reviewed publications on regional stream temperature modeling addressed the question of the spatial scale over which the predictor variables should be averaged. It is common knowledge that stream temperature is not only affected by local environmental conditions, but also by the conditions prevailing upstream. However, the exact extent of the area controlling the stream energy-balance at a given point is not clear (Moore et al., 2005).

Due to this uncertainty, different approaches have been used in the literature to average the predictor variables. Based on studies of the effect of forest harvesting on stream temperature (e.g. Moore et al., 2005), some authors considered riparian buffer zones of various widths and lengths as averaging areas. This approach was usually applied to average the land-cover characteristics only, particularly forest coverage (e.g. Sponseller et al., 2001; Scott et al., 2002; Macedo et al., 2013; Segura et al., 2014), but also in some cases to average most of the predictor variables, including elevation or slope (Hrachowitz et al., 2010; Imholt et al., 2013). Other authors considered larger portions of the catchments as averaging areas, sometimes extending far beyond the riparian zone. For example, Wehrly et al. (2009) used the whole area drained by the stream segment located directly upstream of the temperature measurement point. Whereas most studies relied on simple spatial averaging, a few of them applied a weighting scheme to give more emphasis to the conditions prevailing near the gauging point. As such, Isaak et al. (2010) and Hill et al. (2013) applied a weight w decreasing exponentially with the distance d to the catchment outlet, $w = \exp(-d/L_c)$, where the e-folding distance L_c controls the spatial extent of the averaging area.

In response to this diversity of methods, we could not find a general consensus in the reviewed literature concerning the extent of the spatial area which is relevant for stream temperature prediction. While some studies conclude that this area should have a length of about 1–4 km (Isaak et al., 2010; Hrachowitz et al., 2010; Chang and Psaris, 2013; Macedo et al., 2013), others tend to indicate that the catchment scale is the most appropriate one (Sponseller et al., 2001; Scott et al., 2002). Similarly, values between 30 and 200 m are assumed for the width of the riparian buffer affecting stream temperature at a given point (e.g. Jones et al., 2006; Scott et al., 2002; DeWeber and Wagner, 2014).

1.4 State-of-the-art in ~~Switzerland~~ the European Alps

Of all the regional models reported in Table 1, less than a third were developed for stream temperature prediction outside from North America, and ~~none in Switzerland. The only only one—the model developed by Arscott et al. (2001)—is applied over a European Alpine region. An unpublished attempt at developing such a model over the Swiss territory another model for an Alpine country (Switzerland)~~ was conducted by Rubin et al. (2012), ~~who concentrated on the Canton de Vaud (one of the 26 administrative regions composing Switzerland)~~. They relied on the regionalization of the stream–air temperature relationship, but unfortunately did not evaluate the precision of their model. Other studies have sought to classify the thermal regimes of ~~Swiss Alpine~~ rivers (Jakob, 2010; Müller, 2011), sometimes with minimal success (see Schädler, 2008, for a review of the classification efforts before 2008). These authors grouped the streams according to the physiographic characteristics of their associated watershed, such as mean basin altitude, water origin (lake, artificial reservoir, deep aquifer or shallow subsurface groundwater), channel width or slope. ~~For each group, they~~ They computed the characteristics of the ~~thermal regime that a typical river in this class should have~~ typical thermal regime of each group. However, inter-annual

variations of the thermal regime cannot be accounted for by this method. ~~The impact of climate change on the temperature of some Swiss rivers was addressed by some studies based on linear time extrapolation (Beyeler, 2007; Hari and Zobrist, 2003; Hari and Güttinger, 2004; Jakob, 2010). Finally, other authors relied on physically-based models to simulate water temperature dynamics in single stream reaches or catchments (Meier et al., 2003; Roth et al., 2010; Magnusson et al., 2012; Carlier, 2013), which is not the subject of the present study.~~

205 1.5 Investigation of a new modeling approach

~~The above literature review revealed two subjective choices that are faced by all authors: (a) the selection of the predictor variables, and (b) the choice of the statistical approach to link these~~ All the reviewed models rely on standard statistical techniques to estimate stream temperature. The range of methods encompasses traditional approaches such as multi-linear regression (e.g. Arscott et al., 2001; Mayer, 2012; Imholt et al., 2013), linear mixed modeling (Macedo et al., 2013), but also more advanced techniques such as geographically weighted regression (Pratt and Chang, 2012; Chang and Psaris, 2013), networked kriging models (Gardner and Sullivan, 2004; Isaak et al., 2010), and machine learning techniques (e.g. Westenbroek et al., 2010; Hill et al., 2013; DeWeber and Wagner, 2014).

All these methods are general, in the sense that they can be used to model almost any possible relationship between given input and output variable(s). As a consequence of this generality, the user has to specify the set of predictor variables to ~~stream temperature. The be considered by the model.~~ Although some objective methods can help to perform this selection (e.g. Burnham and Anderson, 2002), the original set of variables on which these methods act must initially be indicated by the user. In the end, the choice of predictor variables ~~varies quite significantly from one model to the other.~~

Some is necessarily affected to some extent by the training and experience of the authors, hereby introducing some diversity in the sets of predictor variables. Thus, in the case of stream temperature modeling in ungauged catchments, some studies consider only physiographic characteristics as predictor variables (e.g. Scott et al., 2002; Jones et al., 2006; Nelson and Palmer, 2007; Hrachowitz et al., 2010), while others also include climatic variables (e.g. Isaak et al., 2010; Ruesch et al., 2012; Moore et al., 2013), stream morphological factors such as channel width or bed gravel size (e.g. Hawkins et al., 1997; Arscott et al., 2001; Daigle et al., 2010), or even markers of anthropogenic activities (e.g. Pratt and Chang, 2012; Hill et al., 2013; Macedo et al., 2013). Similarly, ~~the range of used statistical methods is rather broad. It encompasses traditional methods such as multi-linear regression (e.g. Arscott et al., 2001; Mayer, 2012; Imholt et al., 2013) or linear mixed modeling (Macedo et al., 2013), but also more advanced techniques such as geographically weighted regression (Pratt and Chang, 2012; Chang and Psaris, 2013), networked kriging models (Gardner and Sullivan, 2004; Isaak et al., 2010), and machine learning techniques (e.g. Westenbroek et al., 2010; Hill et al., 2013; DeWeber and Wagner, 2014).~~

It should be mentioned that this diversity also largely results from the varying availability and reliability of data among different geographic areas. This is particularly true for riparian shading,

235 which is never directly measured and can only be estimated based on the data at disposal. For
example, Isaak et al. (2010) approximated riparian shading using a sophisticated combination of
satellite orthoimages and ground hemispherical canopy pictures, whereas DeWeber and Wagner (2014) could
only rely on country-wide landuse data.

~~We advocate here that the statistical modeling approach suffers two drawbacks. First, the selection~~
240 ~~of the predictor variables relies on the training and experience of the author(s); as such, it is not a~~
~~deterministic procedure. This explains the fact that, despite some similarities, all lists of predictor~~
~~variables are different among the reviewed articles. Secondly, Although the generality of the choice~~
~~of the statistical model directly dictates the functional form of the relationship between the predictand~~
~~(stream temperature) and predictor variables, standard statistical methods allows them to be applied~~
245 ~~to many problems, it prevents them to incorporate prior knowledge about the system dynamics into~~
~~their structure.~~ For example, a multi-linear model ~~assumes that the predictand can be expressed~~
~~expresses the predictand~~ as a linear combination of the predictors ~~regardless of the problem at hand.~~
This fact is also true for non-parametric methods such as ~~the~~ artificial neural networks, which im-
plicitly impose some (flexible) functional form ~~. The selection of one type of model over another one~~
250 ~~is generally a heuristic decision.~~

~~We acknowledge that the statistical approach cannot be avoided in many disciplines, such as in~~
~~Social Sciences or Economics, where the interactions between the different variables of interest are~~
~~so complex that no functional form prevails in practice. Yet stream temperature is known to obey~~
~~the energy conservation law. Our aim is to investigate whether the consideration of this physical law~~
255 ~~leads to the development of a more general and justifiable model. This to the model. As advocated by~~
~~Burnham and Anderson (2002), our idea is therefore to attempt at deriving a statistical model whose~~
~~structure includes known dynamics of the predictand variable of interest, namely stream temperature~~
~~in the case at hand.~~

Our approach is strongly inspired from the physically-based models which have been used for
260 decades to predict water temperature along stream reaches (e.g. Brown, 1969; Sinokrot and Stefan,
1993; Westhoff et al., 2007). However, it differs from these models in the sense that we seek a much
simpler expression for stream temperature, expressed as a function of variables which are readily
available at the regional scale.

~~Our method requires analytically solving. To this end, we analytically solve a simplified version~~
265 ~~of the energy-balance equation using some simplifying assumptions (see Sect. over an entire stream~~
~~network (see Section 3.1). The resulting expression for stream temperature is in theory not subject~~
~~to the two above-mentioned drawbacks pertaining to the statistical approach. As a matter of fact, the~~
~~predictor variables do not need to be specified a priori, as they naturally appear during the derivation~~
~~of involves variables whose value cannot be estimated based on the available spatial data sets. Due to~~
270 ~~our lack of knowledge regarding the nature of the relationships between the expression. Moreover,~~
~~the analytical formula directly provides the functional form of the relationship between stream~~

~~temperature and the predictor variables. In practice these two assertions have to be moderated, since some of the predictor variables appearing in the analytical expression are not available in ungauged catchments—such as the long-wave radiation emitted by the stream. So as to keep our~~
275 ~~model as simple as possible~~unknown variables and the available data, we choose to ~~express these~~
~~unknown quantities as linear functions of the available data~~rely on multi-linear regression to estimate
the former as a function of the latter. Although this step involves the subjective selection of predic-
tor variables and assumes a linear relationship~~with the unknown quantities~~, we do not think that it
entirely questions our incentive to ~~base our work on physical grounds. Indeed~~incorporate physical
280 considerations into the model structure. As a matter of fact, only the unknown variables are replaced
in the analytical formula, letting the global form of the relationship unaffected. Assuming that the
major non-linearities are already captured by the global structure of the model, the specific form of
the expressions used to approximate the unknown terms may be considered to have a minor effect.
Moreover, our approach attributes a physical meaning to some of the terms appearing in the for-
285 mula. These terms can be constrained to remain within physical bounds, hereby restricting the range
of values that the calibration parameters can adopt.

The objectives of the present work are three-fold: (1) describe a new ~~analytical-expression~~physics-inspired
statistical model for the prediction of stream temperature in ungauged basins, allowing for the com-
putation of the monthly resolved annual cycle and capturing inter-annual variability; (2) through
290 proper calibration of the model, determine the length of the upstream area which controls stream
temperature at a given point; and (3) compare the ~~analytical-physic-inspired~~ model with a more
standard statistical approach over a set of various Swiss catchments.~~The models, so as to evaluate~~
~~the potential benefits of the incorporation of physical considerations into the model structure. The~~
~~data set used to evaluate the performances of the models is presented in Sect. 2. The models are~~
295 ~~described in Sect. 3. The characteristics of the Swiss catchments are presented in Sect. Results are~~
~~detailed in Sect. 4 and discussed in Sect. 5, followed by the conclusion.~~

2 Data description

2.1 Selected catchments for model evaluation

In order to test the two stream temperature models, catchments are selected in Switzerland such that:
300 (a) the natural regime of the river is as little affected by anthropogenic activities as possible, and
(b) measurements of discharge and stream temperature are available for more than one year. This
results in a set of 29 catchments, whose locations are depicted in Fig. 1 and physiographic properties
are summarized in Table 3.

About half of the selected catchments is situated on the Swiss Plateau—a large area with little
305 altitude variations between Lake Geneva in the south west and Lake Constance in the north east.
The climate in this region is relatively mild, with precipitation mostly falling as rain in winter

and mean daily maximum air temperature hardly exceeding 30°C in summer. The hydrological regimes of the catchments in the Plateau depend on the precipitation patterns and are therefore strongly variable from year to year (Aschwanden and Weingartner, 1985). Discharge does not vary by more than a factor two over the year; it usually reaches its maximum during winter, when evapotranspiration is the lowest. As catchments gain in altitude, the discharge control mechanism changes from evapotranspiration to snowmelt: higher-altitude catchments present a discharge peak during the melt season, in April–May.

Only two catchments are found in the Jura mountains, a relatively low-altitude (< 1700 m) mountainous range with rigorous winters. This region is characteristic for its karstic aquifers with preferential flow paths, generating fast and complex responses to precipitation events. Although more marked, the hydrological regimes of the Jura catchments are relatively similar to those of the watersheds in the Plateau. A clear peak in discharge is noticeable in April for the highest catchments (Aschwanden and Weingartner, 1985).

The Alpine region of Switzerland is typically subdivided into its north and south parts, based on their difference in climate. The Southern Alps are influenced by Mediterranean weather, implying warmer winters and more precipitation in autumn than in the Northern Alps. The hydrological regimes of the catchments in the Northern Alps are strongly related to altitude. The month in which the peak of discharge is observed ranges from May for low-altitude watersheds to July–August for catchments partially covered by glaciers. Moreover, the ratio of annual maximum to annual minimum discharge increases with altitude. Similar hydrological regimes are observed in the Southern Alps, except for a second discharge peak in autumn due to rainfall (Aschwanden and Weingartner, 1985). As seen in Fig. 1, only three unperturbed catchments could be found in the Northern Alps, while five are located in the Southern Alps.

All in all, 10 of the 16 hydrological regimes identified by Aschwanden and Weingartner (1985) in Switzerland are present among the 29 selected catchments (see Table 3). The surface area distribution is quite large, with catchments ranging from 3.31 ~~2.1, and the~~ km² (Rietholzbach at Mosnang) to 392 km² (Broye at Payerne). The mean altitudes of the watersheds are also spanning a wide range of values. Few catchments are partially covered by a glacier, with only two of them having a glacier cover fraction over 10%.

2.2 Stream temperature data

The stream temperature data ~~set in Sect. 2. Results are detailed in~~ which are used in the present study was provided by the Swiss Federal Office for the Environment (FOEN). Advantage is taken of the present publication to describe this new data set, which is freely accessible for research purposes at the following address: <http://www.bafu.admin.ch/hydrologie/01832/01856/index.html?lang=en>. A map displaying the position of all available hydrological stations which measure stream temperature

can also be found on the website of the FOEN (http://www.hydrodaten.admin.ch/en/index.html#temperatur_fliessgewaesser).

345 The FOEN is operating an automatic network of stream gauging stations, continuously measuring
water level and discharge at more than 180 locations in Switzerland. Water level is recorded using
an ultrasonic distance sensor, and converted into discharge values through a rating curve adapted
each year. The water level values are validated against the measurements of a second instrument—a
pressure probe—and rejected in case the difference between the two values is greater than 2 cm. A
350 limited number of gauging stations has been equipped with a thermometer, the earliest starting in
1968. This number has increased greatly since 2002, with now more than 70 stations automatically
probing water temperature every 10 minutes (Jakob, 2010). The measurement values are automatically
uploaded and displayed in real time on the web page of the FOEN (same page as for the map
displaying the positions of the stations).

355 Among the watersheds in which temperature is monitored, 25 have been identified in the present
study as being little affected by anthropogenic activities. In order to complete this data set, the
temperature and discharge measurements of four additional gauging stations were obtained from
the Department for Construction, Transport and Environment of Canton Aargau (see Table 3). The
period in which water temperature was measured by each station is also indicated in Table 3.

360 The temperature data are usually not quality-proofed by the FOEN or Canton Aargau. As a
validation procedure, we performed two different tests on the data at the hourly time step, on top
of visual inspection:

- (a) all temperature measurements lower than 0°C or greater than 30°C were removed, except for
the values between -0.5°C and 0°C which were set to 0°C, and the values between 30°C
and 30.5°C which were set to 30°C. Although the limit of 30°C might be naturally reached
365 in shallow areas, some temperature series showed clear evidence that such temperature was
recorded as a result of the sensor being out of water. As a consequence, it was decided to
remove all data points above 30°C, potentially discarding correct data.
- (b) the temperature variation between consecutive time steps was checked to remain within physical
bounds. In particular, it was verified that temperature varied by more than 0.01°C over 5 hours,
370 but less than 3°C within one hour. Constant temperature values could result from a defect in
the sensor, but also from the fact that the hourly values had been replaced with their daily mean
in some cases. In order to distinguish between the two, the present quality control procedure
was performed semi-manually.

After quality control, the hourly data were aggregated into monthly mean values.

375 2.3 Meteorological data

The two statistical models described in Sect. 4 and discussed in 3 use monthly mean air temperature and incoming solar radiation as predictor variables. Data for these variables were obtained from the Swiss Meteorological Office (MeteoSwiss), which provides free access to them for research purposes. For each one of the selected catchments described in Sect. 5, followed by the conclusion 2.1, the air
380 temperature and incoming solar radiation values measured by all the meteorological stations located at less than 20 km from the catchment outlet were collected. In case less than 3 stations could be found within a 20 km radius, data from the 3 closest meteorological stations were retained. The value of 20 km was chosen so as to ensure that data interpolations would remain representative of the climatic conditions at the catchment outlet, while being based on 3 stations at least. In fact, 27
385 of the 29 selected catchments are entirely contained within the disk of radius 20 km centered on their respective outlet point (not shown). As such, the collected meteorological data can actually be considered as representative for the entire catchments, and not just for their outlet point.

We were provided with hourly mean data, which we aggregated into monthly mean values. We did not perform any quality checks on the data, since MeteoSwiss already follows strict quality control
390 procedures (see <http://www.meteosuisse.admin.ch/home/systemes-de-mesure-et-de-prevision/gestion-des-donnees/preparation-des-donnees.html>; web page only available in German, French or Italian).

Among its network of operated meteorological stations, MeteoSwiss selected a subset of 14 stations which are considered to be representative of the climate diversity in Switzerland (see http://www.meteosuisse.admin.ch/web/fr/climat/climat_aujourd'hui/donnees_mensuelles_homogeneises.html;
395 description only available in German, French or Italian). These stations, referred to as “reference stations” in the following, are used by the standard statistical model to estimate the monthly mean air temperature over the entire Swiss territory (see Sect. 3.2).

2.4 Thermal regime classification

A preliminary study of the selected catchments was performed, with the aim of classifying the rivers
400 according to their thermal behavior. This classification was intended to be used later in order to investigate whether the performance of the models was affected by the river thermal regime.

As a first attempt, we examined whether the catchments could be classified based on the shape of their stream temperature curve. To this end, we z-scored (i.e. standardized) the monthly mean stream temperature values in each watershed similarly to Garner et al. (2014). However, as observed
405 by these authors, we could identify only one single thermal regime (Fig. 2a). Only two catchments among the 29 did not to present the same thermal regime as the others, namely those labeled as 5 and 14 in Table 3.

As an alternative approach, we tested whether the characteristics of the stream–air temperature curve could be used to characterize the thermal regime of the catchments. For this purpose, monthly
410 mean stream temperature was linearly regressed against monthly mean air temperature, excluding the points with negative air temperature values (e.g. Kelleher et al., 2012). Based on the values of

the slope and intercept of this relationship, three groups of catchments could be clearly identified (Fig. 2b). The first group contains the watersheds in which a significant portion of discharge originates from deep aquifer infiltration (watersheds 9 and 14 in Table 3, labeled as “groundwater-fed streams” in Fig. 2b). This group is characterized by low slope and high intercept values, as reported by many studies (e.g. Caissie, 2006; Webb et al., 2008). The second group of watersheds corresponds to the high-altitude basins with more than 50% glacier cover. Both the slope and intercept of the stream–air temperature relationship are small for the members of this group, which is actually composed of only one catchment (watershed 5 in Table 3, denoted as “proglacial” in Fig. 2b). The vast majority of the watersheds do not fall into any of the two aforementioned groups. These catchments, denoted as “thermally climate-driven”, are characterized by relatively low intercept and high slope values, i.e. their stream temperature is strongly correlated to air temperature.

Because of the predominance of the “thermally climate-driven” streams, only the latter will be considered for the testing of the physics-inspired and standard regression models. The inclusion of the groundwater-dominated streams in the test set would require the amount of groundwater discharging into the stream to be estimated. We tested several methods, including the derivation of the baseflow index from discharge measurements (e.g. Eckhardt, 2005; van Dijk, 2010) or from the TOPMODEL topographic index (e.g. Ducharme, 2009). However, none of the investigated techniques succeeded in predicting a larger baseflow index for the catchments labeled as “groundwater-fed” as compared to the others (not shown). Similarly, the consideration of the proglacial streams would imply the glacier cover fraction to be included into the models. This addition of one calibration parameter was not considered justified given that this group contains only one catchment. In total, 26 catchments were used for the calibration and validation of the models, namely all those listed in Table 3 except watersheds 5, 9 and 14.

3 Formulations of the stream temperature models

The new **analytical-expression-physics-inspired statistical model** for stream temperature prediction is derived in the following subsection. The standard statistical model used for comparison is presented in Sect. 3.2.

3.1 **Physically-based stream temperature-Physics-inspired statistical model**

As mentioned above, the **physically-based-physics-inspired** stream temperature model presented in this paper is based on the analytical solution to the stream energy-balance equation. This topic has been investigated extensively in the literature (e.g. Edinger et al., 1968; Theurer et al., 1984; Gosink, 1986; Polehn and Kinsel, 2000; Toffolon et al., 2010), although used only once for stream temperature prediction in ungauged basins (Bogan et al., 2003). In order to analytically solve the energy-balance equation, all studies relied on the linearization of the heat flux $\sigma \phi_a$ at the air–water interface

as a function of stream temperature T : $\sigma = -k(T - T_e)\phi_a = -k(T - T_e)$. Some of them assumed the heat transfer coefficient k to be constant and used prescribed functions of time, space or both to express the equilibrium temperature T_e (e.g. Gosink, 1986; Polehn and Kinsel, 1997; Daly, 2005). Other studies derived analytical formulations for k and T_e based on the physical expressions of the heat fluxes occurring at the stream–air interface (e.g. Edinger et al., 1968; Bogan et al., 2003; Caissie et al., 2005; Bustillo et al., 2014). While a minority of authors considered the temperature distribution to be spatially homogeneous (Edinger et al., 1968; Caissie et al., 2005; Bustillo et al., 2014), most of them assumed the stream to be in a steady-state or, equivalently, the stream celerity to be constant. In addition, they all imposed the river width to remain constant along the stream so as to analytically solve the energy-balance equation. Very few studies accounted for the heat exchange with the stream bed or the heat advected by lateral inflow of water (Bogan et al., 2004; Herb and Stefan, 2011). Bogan et al. (2003) were the only authors to evaluate their analytical expression over ungauged basins. They tested their model in the Central and Eastern United States, since this region has a topography flat enough for a meteorological station located even at more than 100 km from a given point to be still representative of the climate at that point. Their work is therefore hardly transferable to Switzerland, where the mountainous landscape prevents the proper interpolation of variables such as air humidity or wind speed, which are required as input by the model.

3.1.1 Derivation of the analytical solution to the energy-balance equation

Assuming a well-mixed water column and a negligible longitudinal heat dispersion, the mass and energy-balance equations along a stream reach read (adapted from Westhoff et al., 2007):

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = q_\ell, \quad (1)$$

$$\frac{\partial(AT)}{\partial t} + \frac{\partial(QT)}{\partial x} = \frac{w\phi_a + p\phi_b}{\rho c_p} + q_\ell T_\ell - Q \frac{g}{c_p} \frac{\partial z}{\partial x}, \quad (2)$$

where w [m], p [m], A [m²], Q [m³s⁻¹] and T [°C] denote the width, wetted perimeter, cross-sectional area, discharge and temperature of the stream, respectively; t [s] refers to time, x [m] to the downstream distance, z [m] to altitude, and g [ms⁻²] to the gravitational acceleration. The water mass density ρ [kgm⁻³] and the specific heat capacity of water c_p [J°C⁻¹kg⁻¹] are both assumed constant. The quantities ϕ_a [Wm⁻²] and ϕ_b [Wm⁻²] refer to the energy fluxes at the stream–air and stream–bed interfaces, respectively. The lateral heat fluxes due to the inflow of surface, fast subsurface and slow subsurface runoffs into the stream are merged into a single term, $q_\ell T_\ell$, where q_ℓ [m²s⁻¹] denotes the sum of these three runoffs per unit stream length and T_ℓ [°C] stands for their mean temperature. The last term on the right-hand side of Eq. (2) corresponds to friction, which is usually neglected in stream temperature models (e.g. Sinokrot and Stefan, 1993; Westhoff et al., 2007), but has been shown by Hannah et al. (2004) and Leach and Moore (2014) to be an important term in the energy-balance of small streams during winter.

480 The present study builds mainly upon the work of Theurer et al. (1984), which is one of those considering the less restrictive approximations for the derivation of the solution to Eqs. (1)–(2). Our own assumptions are the following:

- (i) At the time scale of the month, the stream temperature is assumed to be in a steady-state.
- (ii) The energy flux at the stream–air interface is expressed as

$$485 \quad \phi_a = \phi_r + k(T_a - T), \quad (3)$$

where ϕ_r [W m^{-2}] denotes the net radiative heat flux, incorporating both the shortwave and longwave components. The second term on the right-hand side accounts for both the latent and sensible heat fluxes (e.g. Polehn and Kinsel, 1997; Toffolon et al., 2010), where the bulk heat transfer coefficient k [$\text{W m}^{-2} \text{ } ^\circ\text{C}^{-1}$] between water and air is assumed to be constant, and T_a [$^\circ\text{C}$] refers to the air temperature.

- (iii) The energy flux at the stream–bed interface is neglected, i.e. $\phi_b = 0$ (e.g. Bogan et al., 2003; Caissie et al., 2005; Bustillo et al., 2014).
- (iv) The lateral inflow of water q_ℓ is assumed to be spatially constant (e.g. Biswal and Marani, 2010; Mutzner et al., 2013).
- 495 (v) The ratio of stream width to discharge w/Q is assumed to be spatially constant. ~~This approximation is different from the one of~~, as opposed to Theurer et al. (1984) and Polehn and Kinsel (2000) ~~who both assumed a constant stream width.~~ This approximation also differs from the typical relationship used in fluvial geomorphology, which expresses stream width as a power-law function of discharge with exponent ~ 0.5 (see e.g. Knighton, 1998). It allows for the defini-
500 tion of a characteristic stream length L_c [m],

$$L_c = \frac{c_p \rho Q}{w k}. \quad (4)$$

- (vi) All sources in the network are supposed to have the same discharge, denoted as Q_s in the following. This approximation is discussed in more details in Sect. 3.1.2.

Using the above assumptions, the mass and energy-balance equations simplify to Eqs. (5)–(6),

$$505 \quad \frac{dQ}{dx} = q_\ell, \quad (5)$$

$$\frac{dT}{dx} = \frac{1}{L_c}(\gamma \phi_r + T_a - T) + \frac{q_\ell}{Q}(T_\ell - T) - \frac{g}{c_p} \frac{dz}{dx}, \quad (6)$$

where $\gamma = 1/k$. The reader is referred to Appendix A for the complete derivation of the analytical solution to these equations. Only the final expressions for discharge Q_{out} and stream temperature

T_{out} at the outlet of a catchment are reported here,

$$510 \quad Q_{\text{out}} = n_s Q_s + L_{\text{tot}} q_\ell, \quad (7)$$

$$T_{\text{out}} = \omega_1 \overline{T_s} + \omega_2 \langle T_\ell \rangle_{\mathcal{L}} + \omega_3 \langle \gamma \phi_r + T_a - L_c \frac{g}{c_p} \frac{dz}{dx} \rangle_{\mathcal{L}}, \quad (8)$$

with

$$\omega_1 = (1 - \eta) \delta_s, \quad (9)$$

$$\omega_2 = \eta \delta_\ell, \quad (10)$$

$$515 \quad \omega_3 = 1 - \omega_1 - \omega_2. \quad (11)$$

In the above equations, L_{tot} and n_s correspond to the total length of the river network and the number of sources in the catchment, respectively. The operator $\langle \cdot \rangle_{\mathcal{L}}$ refers to the distance-weighted average; it computes the average of its operand over the entire stream network using a weight equal to $\exp(-d/L_c)$, where d denotes the distance to the catchment outlet. This operator gives much
520 more emphasis to the points located near to the catchment exit. It should be noted that the spatial extent of the area over which the average is computed is controlled by the characteristic length L_c : the smaller L_c , the smaller the contributing area. The quantity $\overline{T_s}$ appearing in Eq. (8) denotes the weighted average of water temperature at the network sources. The latter are weighted by a factor $\exp(-d_{s,i}/L_c)$, where $d_{s,i}$ is the distance along the stream between the i -th source point and the
525 catchment outlet. The weights ω_1 , ω_2 and ω_3 are all in the interval $[0, 1]$. In Eqs. (9)–(10), the factor η refers to the fraction of discharge at the catchment outlet originating from lateral inflow of water along the stream network—i.e. excluding the fraction coming from the sources,

$$\eta = \frac{q_\ell L_{\text{tot}}}{Q_{\text{out}}} = 1 - \frac{n_s Q_s}{Q_{\text{out}}}. \quad (12)$$

The two factors δ_s and δ_ℓ are defined as

$$530 \quad \delta_s = \frac{1}{n_s} \sum_{i=1}^{n_s} e^{-d_{s,i}/L_c}, \quad (13)$$

$$\delta_\ell = \frac{L_c}{L_{\text{tot}}} \sum_{k=1}^{n_r} e^{-d_k/L_c} (1 - e^{-L_k/L_c}), \quad (14)$$

where n_r denotes the number of reaches in the stream network, d_k the streamwise distance between the downstream point of stream reach k and the catchment outlet, and L_k the length of stream reach k . The factor δ_s corresponds to the average of the weight $\exp(-d(x)/L_c)$ over all the network
535 sources, and the factor δ_ℓ refers to the average of the same weight over the set of all stream reaches in the catchment. It follows that both δ_s and δ_ℓ decrease roughly exponentially as a function of the network length.

Equation (8) expresses stream temperature as a linear function of air temperature, the slope of the regression line between the two being equal to $\omega_3 = 1 - \omega_1 - \omega_2$. Assuming η to vary only slightly

540 along the network, it can be seen from Eqs. (9)–(10) that ω_1 and ω_2 decrease roughly exponentially with the stream network length. As a consequence, the present model predicts ω_3 to tend towards 1 as the catchment size increases, a fact which has been observed at many locations (e.g. Ozaki et al., 2003; Ducharne, 2008; Kelleher et al., 2012; Chang and Pсарis, 2013; Segura et al., 2014).

The present expression for T_{out} differs from those reported previously in the literature in at least 545 two aspects (see Sect. 1 for a review of the analytical solutions to the energy-balance equation published to date). First, the terms on the right-hand side of Eq. (6) were not assumed to be spatially homogeneous when integrating them. This explains the presence of the spatial averaging operator $\langle \cdot \rangle_{\mathcal{L}}$ in Eq. (8), which in turn translates the fact that stream temperature is not impacted by local conditions only. This operator has already been used for the computation of predictor variables 550 in regression-based stream temperature models (Isaak et al., 2010; Hill et al., 2013), but never in association with analytical solutions to the energy-balance equation. Second, the source and lateral inflow terms have not been neglected. These two terms are weighted by the factors ω_1 and ω_2 in Eq. (8), respectively, and tend to decrease exponentially with the stream length (see discussion above). Although negligible in large catchments, they might be of the same order of magnitude as 555 the heat exchange term in small watersheds. Only a few studies relying on an analytical expression for stream temperature modeling have considered the lateral inflow term to date (Bogan et al., 2004; Herb and Stefan, 2011), and none has retained the source term.

As noted above, the extent of the zone over which $\langle \cdot \rangle_{\mathcal{L}}$ averages its operand is controlled by the characteristic length L_c . Given that this length is a function of the river discharge-to-width ratio 560 Q/w (see Eq. (4)) and that the stream celerity is assumed here to be constant, L_c is approximately proportional to the water height. Its value should therefore be expected to change over the course of the year. Based on a formula similar to Eq. (4), Herb and Stefan (2011) have estimated L_c to vary between 3 and 45 km for discharge values between 0.4 and 5.8 m³ s⁻¹ in the case of the Vermillion River in Minnesota. As most of the catchments considered in the present study have discharges 565 contained within this range, we should expect a marked variation in the values of L_c both during the course of the year and across catchments. However, since the characteristic length will be treated as a calibration parameter here (see Sect. 3.1.2), only its seasonal variability will be investigated. A single value will be assumed in each season for all the catchments (see Sect. 4), for otherwise L_c would have to be calibrated independently for each catchment, which would prevent prediction in 570 ungauged basins. We acknowledge this as a limitation of our model.

3.1.2 Parametrization of the unknown terms

Equation (8) contains several unknown quantities. The procedure used to calculate their respective values is detailed below.

The channel slope dz/dx is computed along the center line of each stream. A vector representation of the center lines was extracted from a land cover map at scale 1:25'000 (for more 575

information on this map, see <http://www.swisstopo.admin.ch/internet/swisstopo/en/home/products/maps/national/25.html>) ~~and is freely available on the website of the Swiss Federal Office for the Environment (see ; description available only in German, French or Italian).~~ This map was overlaid with a digital elevation model of Switzerland with 2 m horizontal resolution produced by
 580 the Swiss Federal Office of Topography (see <http://www.swisstopo.admin.ch/internet/swisstopo/en/home/products/height/swissALTI3D.html>) in order to extract the altitude of each point. As an alternative approach, a geomorphological analysis of the stream watersheds could have been performed so as to automatically extract the stream networks. However, it was observed that the results of this analysis did not match with the land cover map in some basins (not shown).

585 The monthly mean air temperature $\overline{T_a}$ along the streams is ~~assumed to be a linear function of altitude. The slope and intercept of this function are computed every month and for each catchment i separately~~ computed based on the values measured by the neighbouring meteorological stations (see Sect. 2.2 for a description of the selected catchments), 2.3). Within each catchment i , air temperature $T_{a,i}$ is assumed to be a linear function of altitude only,

$$590 \quad T_{a,i}(z) = a_{l,i}T_{a,i}(z - z_i) + b_{l,i}T_{a,i}, \quad (15)$$

where z_i [m] refers to the altitude of the gauging station. The lapse rate $\alpha_{l,i}T_{a,i}$ [$^{\circ}\text{C m}^{-1}$] is ~~derived using~~ computed each month separately by regressing the air temperature measurements of ~~all the meteorological stations located at less than 20 km from the catchment outlet—or the 3 closest meteorological stations if less than 3 are located within a 20 km radius~~ neighbouring meteorological
 595 stations against the station altitudes. In case the coefficient of determination R^2 of the regression line ~~between the air temperature values and the altitudes of the meteorological stations~~ is lower than 0.6, $\alpha_{l,i}T_{a,i}$ is set equal to zero. The intercept $b_{l,i}T_{a,i}$ [$^{\circ}\text{C}$] is computed each month as the inverse-distance weighted average of the ~~air temperature values measured by the same neighbouring stations. Before computing their average, these values are~~ same air temperature measurements, which are
 600 first corrected for the altitude effect by virtually transferring them to altitude z_i using the lapse rate $\alpha_{l,i}T_{a,i}$.

The ~~monthly mean quantity~~ $\gamma\phi_r$, which accounts for the effect of the net radiation heat flux at the ~~water surface~~ ϕ_r is mostly dominated by incoming solar radiation, particularly in summer (Caissie, 2006; Webb et al., 2008). In order to keep our model simple, and because ϕ_r eventually has
 605 to be distance-averaged using the operator $\langle \cdot \rangle_L$, we express the unknown quantity $\gamma\phi_r$ as a linear combination of the measured incoming solar radiation $\phi_{\text{sw,inc}}$, an index quantifying the shading by topography f_s , the stream orientation θ and the fraction f_{rc} of the riparian zone covered with forests;

$$\gamma\phi_r = a_{\phi,\text{sw}}\phi_{\text{sw,inc}} + a_{\phi,s}f_s + a_{\phi,\theta}\theta + a_{\phi,\text{f}}f_{\text{f}} + b_{\phi}.$$

610 The above equation requires the calibration of 5 unknown coefficients, namely $\{a_{\phi,x}\}_{x=\text{sw},s,\theta,\text{f}}$ and b_{ϕ} . In order to avoid equifinality issues (see e.g. Beven, 1993), this expression is not directly

used as is. Instead, more parsimonious formulations are evaluated by setting at least two of the coefficients $\{a_{\phi,x}\}_{x=sw,s,\theta,T}$ to zero. Considering all possible combinations of these coefficients, 11 different models are tested for γ_{ϕ_T} in total—including the constant expression with only b_{ϕ} as calibration parameter. In Eq. , the monthly mean incoming solar radiation $\phi_{sw,inc}$ is approximated as a linear air–water interface, cannot be readily computed based on the available data. As a matter of fact, long-wave radiation and reflected short-wave radiation measurements are performed by MeteoSwiss at a few locations only. Incoming short-wave radiation ϕ_{isw} [$W m^{-2}$], on the other hand, is a commonly measured variable which can be interpolated along the stream networks. To this end, it is assumed that the incoming short-wave radiation $\phi_{isw,i}$ in each catchment i is a function of altitude z , whose slope and intercept are computed every month and for every catchment using the same procedure as described above for air temperature (see only,

$$\phi_{isw,i}(z) = a_{\phi,i}(z - z_i) + b_{\phi,i}, \quad (16)$$

where the lapse rate $a_{\phi,i}$ [$W m^{-3}$] and the intercept $b_{\phi,i}$ [$W m^{-2}$] are computed similarly to $a_{T,i}$ and $b_{T,i}$ in Eq. (15)). Shading by topography f_s is expressed at each point along the stream network as a value between 0 and 1, 1 indicating complete shading. It is derived from the above-mentioned 2-m digital elevation map of Switzerland at nine different hours of day time—corresponding to the fractions 0.1–0.9 of the day-time length—on the 15th day of each month of the year. These values are then averaged at each grid cell and in each season to obtain the spatial distribution of f_s . The stream. An attempt is made at correcting the values computed using Eq. 16 in order to account for riparian shading. As discussed in Sect. 1.5 above, very few spatial data sets exist for riparian shading, which in practice often has to be estimated using proxy variables. In the present case, riparian shading at a given stream point is approximated based on the stream orientation θ and riparian forest cover f_r at that point. Using the land cover map at scale 1:25'000 mentioned above, θ is computed as the cosine of the angle between the North–north and the stream direction at the point of interest along the network. It follows that θ –flow direction; it is a measure of northing, i.e. values close to 1 (resp. -1) indicate a catchment that is oriented towards the North (resp. the South). The fraction north and values close to -1 a catchment that is south-oriented. The riparian forest cover f_r is defined here as the fraction of the riparian zone which is covered with forests is extracted from the same national according to the land cover map at scale 1:25'000 which is also used to derive the stream network (see above). As the extent of the riparian zone affecting stream temperature is unclear (Moore et al., 2005), the forest cover fraction is computed over riparian buffers with different widths: 25, 50 and 100 m on each side of the center line of the streams (total buffer widths are 50, 100 and 200 m, respectively). The map does unfortunately not allow for the distinction between coniferous and deciduous forests. In addition to θ and f_r , topographical shading f_s is also computed in order to correct the incoming solar radiation values estimated from Eq. (16). f_s is expressed at each point along the streams as a value between 0 and 1, 1 indicating complete shading. It is

derived from the above-mentioned 2 m digital elevation map of Switzerland at nine different hours of day time—corresponding to the fractions 0.1–0.9 of the day-time period—on the 15th day of each month of the year. These values are then averaged at each grid cell and in each season to obtain the spatial distribution of f_s . Since shading by topography and by the riparian forest do not only affect incoming solar radiation, but also incoming long-wave radiation, it was decided not to use the variables θ , f_f and f_s to directly modify the values of ϕ_{isw} . Instead, it is the unknown term $\gamma\phi_r$ which is approximated as a linear combination of ϕ_{isw} , θ , f_f and f_s :

$$\gamma\phi_r = a_{\phi, isw}\phi_{isw} + a_{\phi, s}f_s + a_{\phi, \theta}\theta + a_{\phi, f}f_f + b_{\phi}. \quad (17)$$

As discussed in Sect. 1.5, the choice of a linear relationship is motivated by our wish to keep the model simple and by our ignorance of the actual form of the function linking $\gamma\phi_r$ to the above-mentioned predictor variables. A linear relationship also significantly simplifies the computation of the distance-average of $\gamma\phi_r$ using the operator $\langle \cdot \rangle_{\mathcal{L}}$. Equation (17) requires the calibration of 5 unknown coefficients, namely $\{a_{\phi, x}\}_{x=isw, s, \theta, f}$ and b_{ϕ} . In order to limit the number of model parameters, this expression is not directly used as is, but more parsimonious formulations are evaluated instead. All possible sub-expressions involving any combination of either one or two of the predictor variables $\{\phi_{isw}, \theta, f_f, f_s\}$ is considered for approximating $\gamma\phi_r$. It should be mentioned that the choice to consider expressions with at most two terms (plus the intercept) is arbitrary and only introduced to avoid equifinality issues (Beven, 2012). In total, 11 different models are tested for $\gamma\phi_r$ —including the constant expression with only b_{ϕ} as calibration parameter.

The two weights ω_1 and ω_2 cannot be readily estimated from Eqs. (9)–(10). While the values of the factors δ_s and δ_ℓ can be easily derived from the vector representation of the stream network described above, the parameter η requires additional assumptions. It should be reminded that this parameter corresponds to the fraction of the outlet discharge which originates from lateral inflow. Assuming a typical power-law relationship between drainage area and discharge (e.g. Mutzner et al., 2013), η could in principle be approximated as the ratio between the area A_{net} drained by the network (excluding the area drained by the sources) and the total catchment area A_{tot} , raised to some power α : $\eta \sim (A_{net}/A_{tot})^\alpha$. However, the computation of A_{net} would require a geomorphological analysis, which was discarded based on the discrepancy between the stream network predicted by this analysis and the observed one (see above). As alternative methods, we consider two different techniques for estimating η . The simplest approach assumes a constant single value for η , calibrated over all catchments. The second approach relies on the analytical expression for η presented in Eq. (12), in which the ratio Q_s/Q_{tot} is replaced with $(A_s/A_{tot})^\alpha$:

$$\eta = 1 - n_s \left(\frac{A_s}{A_{tot}} \right)^\alpha. \quad (18)$$

The calibration parameters of this second method correspond to the area A_s drained by a single source and the exponent α .

In order to compute \bar{T}_s and $\langle T_\ell \rangle_{\mathcal{L}}$ in Eq. (8), two different methods for the estimation of the source and lateral inflow temperatures are considered. In a first approximation, these two temperatures are
685 assumed to be both constant and equal. The second method considers them to be linearly related to air temperature as measured at their respective altitudes. In other words, it expresses the temperature $T_{s,i}$ of each source $i = 1 \dots n_s$ and the lateral inflow temperature $T_\ell(z)$ at any point with altitude z along the network as

$$T_{s,i} = a_w T_a(z_{s,i}) + b_w, \quad (19)$$

690
$$T_\ell(z) = a_w T_a(z) + b_w, \quad (20)$$

where $z_{s,i}$ [m] denotes the altitude of source i , and a_w [$^{\circ}\text{C}^{\circ}\text{C}^{-1}$] and b_w [$^{\circ}\text{C}$] are two parameters to be calibrated over the set of all catchments. Notice that the same slope a_w and intercept b_w are used to derive both $T_{s,i}$ and T_ℓ from air temperature, hereby assuming that the source and lateral inflows originate from the same hydrological processes. Moreover, since these two parameters are the same
695 for all catchments, it is implicitly supposed that the ratio of surface runoff to subsurface runoff is the same in all watersheds. As discussed in Sect. 2.4, this requires catchments to be classified by hydrological regime before a_w and b_w can be calibrated separately for each regime. In Eqs. (19)–(20), the monthly mean air temperature is computed in each catchment using Eq. (15).

The distance-average of variables T_ℓ , $\gamma\phi_r$, T_a and dz/dx are computed by discretizing the operator
700 $\langle \cdot \rangle_{\mathcal{L}}$ over the stream segments,

$$\langle f \rangle_{\mathcal{L}} = \frac{\sum_{k \in \Gamma} e^{-(d_k + L_k/2)/L_c} L_k f_k}{\sum_{k \in \Gamma} e^{-(d_k + L_k/2)/L_c} L_k}, \quad (21)$$

where f_k denotes the unweighted mean value of variable f along stream segment k ; the other quantities have been defined previously in Sect. 3.1.1. Except for the riparian forest cover f_t which is derived over buffers of widths 25, 50 and 100 m, the unweighted mean of all other quantities (namely
705 $\phi_{\text{sw,inc}}$, ϕ_{isw} , f_s , T_a and dz/dx) along each stream segment are computed over a 20 m wide buffer centered around the center line of the segment, as extracted from the vector representation of the stream network at scale 1:25'000 (see above). The value of 20 m is considered as being typical for the width of the streams investigated in the present study; although only this value has been tested, it is expected to have little impact on the computed averages. It should be noted that the expressions
710 for $\langle T_a \rangle_{\mathcal{L}}$ and $\langle \phi_{\text{sw,inc}} \rangle_{\mathcal{L}} \langle \phi_{\text{isw}} \rangle_{\mathcal{L}}$ both reduce to linear functions of the distance-weighted average of altitude along the stream network $\langle z \rangle_{\mathcal{L}}$ as per Eqs. (15) and (17). The length L_k of stream reach k and the distance d_k between the downstream end of reach k and the catchment outlet are derived from the vector map of the stream network.

Replacing the terms in Eq. (8) by their above expressions, the stream temperature model reads

715
$$T_{\text{out}} = (1 - \eta)\delta_s(a_w \bar{T}_a + b_w) + \eta\delta_\ell(a_w \langle T_a \rangle_{\mathcal{L}} + b_w) + (1 - \delta_s - \eta\delta_s - \eta\delta_\ell)\langle T_{\text{eq}} \rangle_{\mathcal{L}}, \quad (22)$$

where

$$T_{\text{eq}} = a_{\phi, \text{sw}} \langle \phi_{\text{sw, inc}} \rangle_{\mathcal{L}} + a_{\phi, \text{s}} \langle f_{\text{s}} \rangle_{\mathcal{L}} + a_{\phi, \theta} \langle \theta \rangle_{\mathcal{L}} + a_{\phi, \text{f}} \langle f_{\text{f}} \rangle_{\mathcal{L}} + \langle T_{\text{a}} \rangle_{\mathcal{L}} - L_{\text{c}} \frac{g}{c_{\text{p}}} \langle \frac{\text{d}z}{\text{d}x} \rangle_{\mathcal{L}} + b_{\phi}. \quad (23)$$

720 The calibration parameters of the model are listed in Table 2). When testing a constant parametrization for the source and lateral inflow temperatures, a_w should be set to zero. Similarly, at least two of the coefficients $\{a_{\phi, x}\}_{x=\text{sw}, \text{s}, \theta, \text{f}}$ are assumed equal to zero as per the parametrization of the radiation term discussed above. **As such**, between 3 and 8 parameters must be calibrated depending on the methods used to approximate the respective unknown variables in Eq. (8). Advantage is taken of the fact that each parameter can be interpreted from a physical point of view to restrict its associated calibration range (see Table 2). For example, η is imposed to adopt a value between 0 and 1 as per Eq. (12), and only positive values are considered for $a_{\phi, \text{sw}}$ based on the fact that solar radiation is contributing positively to the net radiation heat flux. Moreover, six different values are tested for the characteristic length L_{c} used in the definition of $\langle \cdot \rangle_{\mathcal{L}}$: 1, 2, 4, 8, 16 and 32 km (see Sect. 4.2). All possible combinations of the different parametrizations of the model terms are tested for each one of these values of L_{c} . The model associated with the lowest value of the modified Akaike information criterion (AICc) is considered to be the best one among the tested set (e.g. Burnham and Anderson, 2002). As mentioned in Sect. 3.1.1, the model is calibrated in each season separately to account for the fact that the value of the parameter L_{c} varies over the year.

735 3.2 **Regression** Standard regression model

In order to assess its performances, the **physically-based** physics-inspired statistical model described by Eq. (22) is compared with a more classical regression model which we developed based on a combination of some of the standard statistical approaches reviewed in Sect. 1. The regression model takes advantage of the fact that most stream temperature curves have a similar shape (see 740 Sect. 2.4). This shape is first estimated by the model based on air temperature, before being mapped to the respective stream temperature curves of the catchments using a linear transformation.

The model assumes all streams to have the same z-scored (i.e. standardized) temperature \hat{T} [-]. The latter is related to the monthly mean temperature T_i of each individual catchment i through (see e.g. Garner et al., 2014)

$$745 T_i = \sigma_i \hat{T} + \bar{T}_i, \quad (24)$$

where \bar{T}_i [$^{\circ}\text{C}$] and σ_i [$^{\circ}\text{C}$] correspond to the annual mean and standard deviation of monthly mean stream temperature in catchment i , respectively. These two quantities are estimated each year independently using multi-linear regression models (MLR). Although more sophisticated techniques could have been used, Wehrly et al. (2009) and Daigle et al. (2010) showed that MLR performs

750 at least as good as several more complicated statistical methods for stream temperature predic-
 tion. The MLR models were constructed using similar predictor variables as in the [physically-based
 physics-inspired statistical](#) model, namely the annual mean and standard deviation of both air tem-
 perature and incoming short-wave radiation, the riparian forest cover fraction, stream channel slope,
 stream orientation, the difference in topographical shading between summer and winter, the number
 755 of sources in the network and the watershed area. All multi-linear models based on any possible
 subset of these variables were tested, with a maximum number of terms per model arbitrarily fixed
 to ~~6 to prevent 6~~. [This limitation was introduced in order to avoid over-parametrization, but also to
 ensure that the number of parameters in the final standard regression model was about the same as
 in the physics-inspired model, hereby guarantying a more even comparison between the two.](#) Mul-
 760 ticollinearity issues were avoided by discarding [MLR](#) models whose variance inflation factor (VIF)
 exceeded 5. Each predictor variable was distance-averaged over the stream networks using the oper-
 ator $\langle \cdot \rangle_{L_c}$, as in the case of the [physically-based-physics-inspired statistical](#) model. Different values
 of L_c were considered when applying this operator as per Eq. (21): 1, 2, 4, 8, 16 and 32 km. The
 best predicting MLR models for \bar{T}_i and σ_i were selected based on AICc.

765 In Eq. (24), the z-scored stream temperature is computed each month ~~by the model~~ based on a
 non-linear relationship with air temperature,

$$\hat{T} = \mu + \frac{\alpha - \mu}{1 + \exp(-\kappa(\hat{T}_a - \beta))}, \quad (25)$$

where μ [$^{\circ}\text{C}$], α [$^{\circ}\text{C}$], β [$^{\circ}\text{C}$] and κ [$^{\circ}\text{C}^{-1}$] are coefficients obtained through ordinary least squares
 regression, and \hat{T}_a [-] denotes the mean z-scored air temperature over Switzerland. \hat{T}_a is obtained
 770 by averaging the z-scored measurements of [the 14 MeteoSwiss](#) reference meteorological stations ;
~~which are considered by the Swiss Meteorological Office to be representative of the climate diversity
 in Switzerland (see ; description only available in German, French or Italian (see Sect. 2.3),~~

$$\hat{T}_a = \frac{1}{14} \sum_{k=1}^{14} \frac{T_{a,k} - \bar{T}_{a,k}}{\sigma_{a,k}}. \quad (26)$$

In the above equation, $T_{a,k}$ [$^{\circ}\text{C}$] denotes the monthly mean air temperature measured at [reference](#)
 775 station k , and $\bar{T}_{a,k}$ [$^{\circ}\text{C}$] and $\sigma_{a,k}$ [$^{\circ}\text{C}$] refer to the annual average and standard deviation of $T_{a,k}$
 computed each year independently, respectively.

In summary, the [standard](#) regression model proceeds as follows to estimate stream temperature
 in an ungauged basin: (a) it first computes the mean z-scored air temperature over Switzerland ac-
 cording to Eq. (26), based on the measurements of 14 meteorological stations, (b) it then uses \hat{T}_a to
 780 estimate the z-scored stream temperature in any catchment as per Eq. (25), and finally (c) it converts
 \hat{T} to the actual stream temperature using Eq. (24), where the scaling coefficients \bar{T}_i and σ_i are esti-
 mated for the catchment of interest using multi-linear regression models. These different steps will
 be illustrated ~~into in~~ more detail in Sect. 4.4.

4 Selected catchments for model evaluation

785 In order to test the two stream temperature models, catchments are selected in Switzerland such that:
(a) the natural regime of the river is as little affected by anthropogenic activities as possible, and
(b) measurements of discharge and stream temperature are available for more than one year. This
results in a set of 29 catchments, whose locations are depicted in Fig. 1 and physiographic properties
are summarized in Table 3.

790 About half of the selected catchments is situated on the Swiss Plateau—a large area with little
altitude variations between Lake Geneva in the south west and Lake Constance in the north east.
The climate in this region is relatively mild, with precipitation mostly falling as rain in winter
and mean daily maximum air temperature hardly exceeding 30°C in summer. The hydrological
regimes of the catchments in the Plateau depend on the precipitation patterns and are therefore
795 strongly variable from year to year (Aschwanden and Weingartner, 1985). Discharge does not vary
by more than a factor two over the year; it usually reaches its maximum during winter, when
evapotranspiration is the lowest. As catchments gain in altitude, the discharge control mechanism
changes from evapotranspiration to snowmelt: higher-altitude catchments present a discharge peak
during the melt season, in April–May.

800 Only two catchments are found in the Jura mountains, a relatively low-altitude (<1700 m) mountainous
range with rigorous winters. This region is characteristic for its karstic aquifers with preferential
flow paths, generating fast and complex responses to precipitation events. Although more marked,
the hydrological regimes of the Jura catchments are relatively similar to those of the watersheds in the
Plateau. A clear peak in discharge is noticeable in April for the highest catchments (Aschwanden and Weingartner, 1985).

805 The Alpine region of Switzerland is typically subdivided into its north and south parts, based on
their difference in climate. The Southern Alps are influenced by Mediterranean weather, implying
warmer winters and more precipitation in autumn than in the Northern Alps. The hydrological
regimes of the catchments in the Northern Alps are strongly related to altitude. The month in which
810 the peak of discharge is observed ranges from May for low-altitude watersheds to July–August
for catchments partially covered by glaciers. Moreover, the ratio of annual maximum to annual
minimum discharge increases with altitude. Similar hydrological regimes are observed in the Southern
Alps, except for a second discharge peak in autumn due to rainfall (Aschwanden and Weingartner, 1985).
As seen in Fig. 1, only three unperturbed catchments could be found in the Northern Alps, while five
815 are located in the Southern Alps.

All in all, 10 of the 16 hydrological regimes identified by Aschwanden and Weingartner (1985) in
Switzerland are present among the 29 selected catchments (see Table 3). The surface area distribution
is quite large, with catchments ranging from 3.31 km² (Rietholzbach at Mosnang) to 392 km²
(Broye at Payerne). The mean altitudes of the watersheds are also spanning a wide range of values.

820 Few catchments are partially covered by a glacier, with only two of them having a glacier cover fraction over 10%.

4 Stream temperature data

The stream temperature data which is used in the present study was provided by the Swiss Federal Office for the Environment (FOEN). Advantage is taken of the present publication to describe this
825 new data set, which is freely accessible for research purposes at the following address: . A map displaying the position of all available hydrological stations which measure stream temperature can also be found on the website of the FOEN ().

3.1 Data collection and validation

The FOEN is operating an automatic network of stream gauging stations, continuously measuring
830 water level and discharge at more than 180 locations in Switzerland. Water level is recorded using an ultrasonic distance sensor, and converted into discharge values through a rating curve adapted each year. The water level values are validated against the measurements of a second instrument—a pressure probe—and rejected in case the difference between the two values is greater than 2 cm. A limited number of gauging stations has been equipped with a thermometer, the earliest starting in
835 1968. This number has increased greatly since 2002, with now more than 70 stations automatically probing water temperature every 10 minutes (Jakob, 2010). The measurement values are automatically uploaded and displayed in real time on the web page of the FOEN (same page as for the map displaying the positions of the stations).

Among the watersheds in which temperature is monitored, 25 have been identified in the present
840 study as being little affected by anthropogenic activities. In order to complete this data set, the temperature and discharge measurements of four additional gauging stations were obtained from the Department for Construction, Transport and Environment of Canton Aargau (see Table 3). The period in which water temperature was measured by each station is also indicated in Table 3.

The temperature data is usually not quality-proofed by the FOEN or Canton Aargau. As a validation
845 procedure, we performed two different tests on the data at the hourly time step, on top of visual inspection: all temperature measurements lower than 0°C or greater than 30°C were removed, except for the values between -0.5°C and 0°C which were set to 0°C, and the values between 30°C and 30.5°C which were set to 30°C. the temperature variation between consecutive time steps was checked to remain within physical bounds. In particular, it was verified that temperature varied
850 by more than 0.01°C over 5 hours, but less than 3°C within one hour. Constant temperature values could result from a defect in the sensor, but also from the fact that the hourly values had been replaced with their daily mean in some cases. In order to distinguish between the two, the present

quality control procedure was performed semi-manually. After quality control, the hourly data were aggregated into monthly mean values.

855 3.1 Thermal regime classification

A preliminary study of the selected catchments was performed, with the aim of classifying the rivers according to their thermal behavior. This classification was intended to be used later in order to investigate whether the performance of the models was affected by the river thermal regime.

860 As a first attempt, we examined whether the catchments could be classified based on the shape of their stream temperature curve. To this end, we z-scored (i.e. standardized) the monthly mean stream temperature values in each watershed similarly to Garner et al. (2014). However, as observed by these authors, we could identify only one single thermal regime (Fig. 2a). Only two catchments among the 29 did not to present the same thermal regime as the others, namely those labeled as 5 and 14 in Table 3.

865 As an alternative approach, we tested whether the characteristics of the stream-air temperature curve could be used to characterize the thermal regime of the catchments. For this purpose, monthly mean stream temperature was linearly regressed against monthly mean air temperature, excluding the points with negative air temperature values (e.g. Kelleher et al., 2012). Based on the values of the slope and intercept of this relationship, three groups of catchments could be clearly identified (Fig. 2b). The first group contains the watersheds in which a significant portion of discharge originates from deep aquifer infiltration (watersheds 9 and 14 in Table 3, labeled as “groundwater-fed streams” in Fig. 2b). This group is characterized by low slope and high intercept values, as reported by many studies (e.g. Caissie, 2006; Webb et al., 2008). The second group of watersheds corresponds to the high-altitude basins with more than 50% glacier cover. Both the slope and intercept of the stream-air 875 temperature relationship are small for the members of this group, which is actually composed of only one catchment (watershed 5 in Table 3, denoted as “proglacial” in Fig. 2b). The vast majority of the watersheds do not fall into any of the two aforementioned groups. These catchments, denoted as “thermally climate-driven”, are characterized by relatively low intercept and high slope values, i.e. their stream temperature is strongly correlated to air temperature.

880 Because of the predominance of the “thermally climate-driven” streams, only the latter will be considered for the testing of the physically-based and regression models. The inclusion of the groundwater-dominated streams in the test set would require the amount of groundwater infiltrating into the stream to be estimated. We tested several methods, including the derivation of the baseflow index from discharge measurements (e.g. Eckhardt, 2005; van Dijk, 2010) or from the TOPMODEL topographic index (e.g. Ducharme, 2009). However, none of the investigated techniques succeeded in predicting a larger 885 baseflow index for the catchments labeled as “groundwater-fed” as compared to the others (not shown). Similarly, the consideration of the proglacial streams would imply the glacier cover fraction to be included into the models. This addition of one calibration parameter was not considered

justified given that this group contains only one catchment. In total, 26 catchments were used for
890 the calibration and validation of the models, namely all those listed in Table 3 except watersheds 5,
9 and 14.

4 Model evaluation

In order to rigorously evaluate the performance of the two models described in the previous sec-
tion, 5 of the 26 selected catchments were removed from the data set to create an independent
895 validation set (watersheds 3, 6, 11, 13 and 27, displayed in orange in Fig. 1). Caution was given
to single out basins with different size, mean elevation and geographic location. Among the four
climatic regions of Switzerland, only the Jura could not be represented in the validation set, given
that only one station (number 26) among the 26 available was located in this area. A bootstrap on
the validation stations was not possible because of too high computational requirements. Indeed,
900 Burnham and Anderson (2002) recommend using at least 10'000 bootstrap samples, which led to a
prohibitively high number of model evaluations in our case.

The measurement time period is also split into a calibration (2007–2012) and a validation (all dates
before and including 2006) period. Only the measurements performed by the calibration stations—
whose drainage area is marked in green in Fig. 1—during the period 2007–2012 are used to calibrate
905 the models. Four different validation sets can be formed with the remaining station-months:

1. the data set containing the measurements of the validation stations during the calibration pe-
riod. This set can be used to evaluate the ability of the models to make predictions in ungauged
basins.
2. the data set containing the measurements of the calibration stations during the validation pe-
910 riod. This set will be used to evaluate the precision of the models when predicting stream
temperature in past or future years.
3. the data set formed by the measurements of the validation stations during the validation period.
This set serves to evaluate the performance of the models when predicting stream temperature
both in ungauged basins and in ungauged years.
- 915 4. the data set corresponding to the union of all three previous validation sets, which may be used
to obtain a synthetic evaluation of the precision of the models.

The complete data set is almost equally subdivided into its calibration and validation parts, with the
former containing 1223 station-months and the validation sets 1–4 regrouping 360, 705, 204 and
1269 station-months, respectively.

920 As mentioned in Sect. 3.1, the value of the characteristic stream length L_c is expected to change
over the course of the year. In order to ease capturing of this variability, the **physically-based**

[physics-inspired statistical](#) model is calibrated over each season separately. As such, the calibration and validation data sets are each subdivided into four groups, corresponding to winter (January–March), spring (April–June), summer (July–September) and autumn (October–December), respectively. Each one of these subgroups contains approximately one fourth of the station-months originally belonging to the parent group. The [standard](#) regression model is calibrated over all seasons at once, but evaluated in each season separately so as to investigate a potential effect of the period of the year on its precision.

Since the [physically-based physics-inspired](#) model expresses stream temperature as a linear function of air temperature, it cannot reproduce the asymptotic behavior of the former as the latter drops below 0°C. Consequently, data points associated with negative air temperature values are removed from the data set before calibration (Kelleher et al., 2012). When evaluating the model over the validation sets, all stream temperatures predicted to be negative are replaced with 0°C values.

In the following, the best seasonal formulations of the [physically-based physics-inspired](#) model are presented first. The precision of this model is then evaluated, and the influence of the stream network resolution on the model results investigated. Finally, comparison is made with the [regression-based standard regression](#) model. All the results presented in this section will be discussed and analyzed in Sect. 5.

4.1 Model formulations

As mentioned in Sect. 3.1.2, the different possible formulations of the [physically-based physics-inspired statistical](#) model are ranked in each season according to their respective AICc value. [Following Burnham and Anderson \(2002\)](#), AICc is preferred here over the classical definition of the Akaike information criterion (AIC) since it includes a correction term for finite-sized data sets ([Burnham and Anderson, 2002](#)). It should be mentioned that, [following Burnham and Anderson \(2002\)](#), AICc is computed by calibrating the models not over the calibration set only, but rather over the entire data set (i.e. both the calibration set and validation set 4). Only in a second time is each model calibrated over just the calibration set, so as to evaluate its performances in terms of RMSE, R^2 and bias.

Table 4 presents the best model formulations selected in each season, ranked according to their respective Akaike weights w_i . The latter corresponds to the probability of each model to be a better descriptor of the observed data (according to information theory) as compared to the model with the minimum AICc value (Burnham and Anderson, 2002; Wagenmakers and Farrell, 2004). Considering models with $w_i \leq 0.1$ to be statistically insignificant, it can be observed that only a few formulations were identified in each season as being relevant for stream temperature prediction. The characteristic stream length L_c is found to be consistent among these formulations, with a value of 4 km in spring, summer and autumn, and 8 km in winter, regardless of the formulation.

The model selection reveals the radiation term $\gamma\phi_r$ to be preferentially expressed as a function of topographical shading f_s and riparian forest cover fraction f_r , or as a function of f_s alone. Among

the tested buffer widths used to compute f_t , none of the three values 50, 100 or 200 m prevails significantly over the others. The order in which they appear in the ranked models varies depending
960 on the season; for example in winter, forest cover computed over a 100 m wide buffer is expected to be a better predictor of $\gamma\phi_r$ than forest cover over a 50 m wide buffer, whereas the opposite is true in spring. Focusing on each season separately, the linear coefficient associated with any given term is observed to have a fairly constant value among the different expressions tested for $\gamma\phi_r$. For example, the coefficient multiplying f_s remains within a narrow range (at most 1°C large) in each season.

965 This behavior is even more pronounced in the case of the term associated with the source and lateral inflow temperatures (T_s and T_ℓ). This term is expressed as a linear function of air temperature, whose slope a_w and intercept b_w are constant among the various model formulations in a given season (see Table 4). The values of a_w are observed to be rather low independently of the period of the year, which indicates a weak coupling between the stream source (resp. lateral inflow) temperature
970 and air temperature. Moreover, a_w and b_w differ among the seasons in such a way that T_s and T_ℓ are the least coupled to air temperature in winter and the most in summer.

The model ranking based on AICc also identified a single expression for η in each season. This parameter is found equal to one in summer and autumn, and zero in winter. Its expression is slightly more complicated in spring, where the selected formulation is the one based on the source drainage
975 area (see Sect. 3.1.2).

4.2 Model performance

The RMSE, R^2 and bias of the best selected model formulation in each season—i.e. the one with $w_i = 1$ —are reported in Table 5. Based on the results of the evaluation over validation set 4, the model precision is observed to be rather satisfactory. Its RMSE and R^2 are relatively constant over
980 the year (about 1.3°C and 0.87, respectively), except in winter where the value of the coefficient of determination is much lower (0.55). Similarly, the bias is small in all seasons (-0.11 to 0.14°C) apart from winter (-0.47°C).

Regarding the different validation sets, it can be observed in Table 5 that the model performs better when predicting in ungauged catchments as compared to simulating past or future years. Indeed, the
985 RMSE values computed using validation set 1 are smaller than those based on set 2, particularly in winter, autumn and summer. Similarly, the values of R^2 are higher over set 1 than over set 2, despite the fact that the model bias is larger over the former set as compared to the latter. As expected, the weakest model performances are generally associated with validation set 3, which contains the measurements performed by the validation stations during the validation period. The only noticeable
990 exception is in summer, where the model evaluation over set 3 provides satisfactory results (RMSE = 1.13°C , $R^2 = 0.90$, bias = -0.01°C).

4.3 Influence of the stream network resolution

The results reported above are based on the stream network geometries extracted from the land cover map at scale 1:25'000 (see Sect. 3.1.2). These geometries directly affect the values of the distance-averaged predictor variables, since the operator $\langle \cdot \rangle_L$ averages over the entire stream network. As a consequence, modifying the network resolution is expected to impact the model performance.

To test this hypothesis, two additional stream networks with coarser resolution than the original one were investigated. These networks were obtained by removing stream segments with Strahler order 1, and those with Strahler order 1 and 2, respectively. Through this procedure, the mean drainage density of the 26 selected catchments decreased from 2.1 km km^{-2} for the original network to 0.5 km km^{-2} for the coarsest one, passing through 1.0 km km^{-2} for the intermediate resolution network. The different model formulations were evaluated over the two additional networks using the same procedure as described in the previous section. Although the results are not reported here, it was essentially observed that the network resolution had little influence on the ranking of the model formulations based on AICc in each season. Almost all selected models were associated with a characteristic stream length $L_c = 4$ or 8 km , as in the case of the original stream network. The parametrization of the net radiation heat flux $\gamma\phi_r$ was also similar to the one reported in Table 4. Topographical shading and riparian forest cover remained the two most statistically significant predictors for this term, except during winter, where stream orientation appeared as a relevant variable.

The values of the coefficients a_w and b_w were noted to vary little among the selected model formulations in a given season. Finally, the parameter η was preferentially expressed as a constant term. Its value was identified as being zero in all seasons except summer in the case of the intermediate resolution stream network, whereas its parametrization was close to the one described in Table 4 in the case of the coarser network.

As a consequence of the little influence of the network resolution on the model parametrization, few variations in the model precision were observed between the three stream networks. As seen in Fig. 3, no clear tendency can be identified among the residuals. At most can a small increase in the model prediction error be detected for the coarsest network as compared to the first two, especially in autumn. The largest absolute residuals are also observed to be generated by this network. On the other hand, the strong bias previously noted in winter is present in the case of the intermediate resolution network, but less so in the case of the coarsest resolution one.

4.4 Comparison with the standard regression model

This section describes the characteristics of the calibrated standard regression model first, before presenting the results of its evaluation in a second step. Figure 4 pictures the observed monthly mean stream temperature, z-scored and averaged over the 21 calibration stations, as a function of \hat{T}_a over the period 2007–2012. As can be observed, the relationship between these two quantities

displays a small hysteresis effect, which can be explained by stream cooling due to snow-melt in spring (Mohseni et al., 1998). The logistic equation introduced by Mohseni et al. (1998) is fitted to each one of the hysteresis branches separately,

$$1030 \quad \hat{T} = -1.87 + \frac{4.88}{1 + e^{-0.99 \cdot (\hat{T}_a - 0.66)}} \quad (\text{in January–June}), \quad (27)$$

$$\hat{T} = -1.86 + \frac{3.96}{1 + e^{-1.16 \cdot (\hat{T}_a - 0.06)}} \quad (\text{in July–December}). \quad (28)$$

It should be noted that the parameters corresponding to the lower and upper asymptotic values of the logistic curve are particularly sensitive to the data points located at both ends of the hysteresis. To limit inaccuracy errors, the temperatures measured in January and July were used to fit both branches
 1035 of the hysteresis, as they usually correspond to the annual extreme values. Equations (27)–(28) are those used in the model to determine the z-scored stream temperature \hat{T} at any location based on \hat{T}_a (see Sect. 3.2).

The multi-linear regression models which were selected to estimate the annual mean \bar{T}_i and the standard deviation σ_i of stream temperature in a given year are presented in Table 6. They correspond
 1040 to the models associated with the lowest AICc values among the tested formulations (see Sect. 3.2). As observed in the table, the characteristic stream length used by the operator $\langle \cdot \rangle_{\mathcal{L}}$ to average the predictor variables over the stream networks is significantly different in the two cases: $L_c = 4$ km for the \bar{T}_i model, whereas $L_c = 32$ km for the σ_i one.

Table 7 summarizes the prediction errors of the standard stream temperature regression model
 1045 when evaluated over validation set 4 using the original stream network. Comparison with Table 5 reveals that its precision is greater than the one of the ~~physically-based~~ physics-inspired model. Its RMSE is about 0.2°C lower, its R^2 about 0.03 to 0.12 larger, and its absolute bias 0.05 to 0.20°C smaller depending on the season. However, its performance worsens when using the two stream networks with coarser resolution: its yearly average RMSE increases to 1.26°C in the case of the
 1050 intermediate resolution network, and even 1.29°C for the coarsest network, which is close to the value obtained with the ~~physically-based~~ physics-inspired model.

5 Discussion

The formulations of the ~~physically-based~~ physics-inspired model selected by AICc ranking are consistent among the different seasons. In particular, topographical shading systematically appears to
 1055 be the strongest predictor of the net radiation heat flux $\gamma\phi_r$. This observation is not particularly surprising in a mountainous country like Switzerland, where some valleys are steep enough for their bottom not to be illuminated by direct sun light for some period of the year. The basins referred to as numbers 14 and 24 in Table 3 are examples of such watersheds, both having a mean catchment slope larger than 35 degrees. Riparian forest cover fraction corresponds to the second most important
 1060 predictor for the net radiation heat flux term. It was rather unexpected to identify this parameter as

relevant during autumn and winter, especially since more than half of the selected catchments are mainly covered with deciduous forests due to their relatively low mean elevation (< 1000 m). This result has to be balanced with the fact that a given fraction increase in riparian forest cover is predicted by the model to have an effect on $\gamma\phi_r$ about 4 to 6 times smaller in magnitude than the same
1065 fraction increase in topographical shading. It should also be reminded that the precision of the model is rather low in winter, hereby questioning the validity of its parametrization in this season. Certainly more unexpected is the absence of solar radiation among the predictors of $\gamma\phi_r$, which will be explained below. Regarding the parametrization of the discharge fraction due to lateral water inflow η , the model predicts the water in the stream channel to originate principally from surface and sub-
1070 surface runoff during summer and autumn ($\eta = 1$). This partly matches our expectations, since the fraction of discharge originating from the sources is expected to decrease when moving downstream along a given network. The characteristic catchment size defining the transition from source-water-dominated to lateral-inflow-dominated discharge is controlled here by L_c , which is equal to 4 km in summer and autumn. This value is smaller than the main channel length in more than 90% of
1075 the selected watersheds (not shown), hereby strengthening our confidence in the parametrization of η during these two seasons. On the other hand, a value of 1 for η in all catchments may appear as a too simplified approach (see below). The questioning of the parametrization of η is all the more true in winter, where its value is predicted to be zero. Only in spring did the model ranking select the more physically-based formulation for η , expressed as a function of the area drained by each
1080 source. Concerning the parametrization of the source and lateral inflow temperatures, it should be mentioned that the linear expression as a function of air temperature was systematically preferred over the constant term. This certainly results from the large altitudinal range covered by the selected catchments, which does not allow for a constant inflow temperature to reflect the diversity of encountered climatological conditions, and mainly air temperature.

1085 As defined in Eq. (22), the ~~physically-based~~ physics-inspired model linearly relates air temperature to stream temperature through the proportionality coefficient ω_3 . The latter is compared in Fig. 5 with its actual observed value, namely the slope of the regression line between the monthly mean temperature measurements of the stream and air. As seen in the figure, the model systematically overestimates the value of ω_3 , particularly in winter and summer where the mean bias equals
1090 0.2. Referring to Eq. (22), this implies that ω_1 and ω_2 are globally underestimated by the model, hereby indicating that the parametrization of the factor η could possibly be improved. As noted in Sect. 3.1.2, a more physically-based expression could be used to compute η , as long as a geomorphological analysis of the river watersheds can be performed. This approach was not investigated here for the reasons mentioned earlier.

1095 The overestimation of ω_3 is probably at the origin of the fact that solar radiation is unexpectedly missing from the selected expressions for the net radiation heat flux $\gamma\phi_r$ (see Table 4). Contrary to the standard regression model, the ~~physically-based~~ physics-inspired model presents the advantage

that the calibration range of ~~each-one-most~~ of its parameters can be restricted based on physical considerations (see Table. 2). An attempt was made at removing these constraints, which resulted
1100 in incoming short-wave radiation being present in almost all models for $\gamma\phi_r$, but associated with a negative coefficient. It was concluded that the unconstrained model takes advantage of the fact that the annual cycles of air temperature and solar radiation have a similar shape to artificially reduce the value of air temperature by subtracting a fraction of solar radiation, hereby compensating for the too large value of ω_3 . This observation argues once more in favor of a better parametrization of the
1105 factor η .

As mentioned in Sect. 4, the characteristic stream length L_c is found to be of the order of 4–8 km regardless of the season or the stream network resolution in the case of the ~~physically-based physics-inspired~~ model. This range is in agreement with the findings of Isaak et al. (2010) and Macedo et al. (2013). Hrachowitz et al. (2010) and Chang and Psaris (2013) rather concluded that
1110 L_c was around 1 km, however they did not investigate values for L_c larger than 1 km. Contrary to our expectations, we do not observe a marked variation of L_c across seasons, probably due to the fact that we assumed a single value for all the catchments. The annual cycle of L_c may have been better captured by separately calibrating this parameter in each ~~individual-catchment-catchment individually~~, but this would have contradicted our aim to derive a regional model. It should be emphasized that the absence of observed annual cycle for L_c does not question the decision to calibrate the ~~physically-based physics-inspired~~ model on a seasonal basis, since the source temperature parametrization is observed to vary significantly over the year (see Table 4).

Our model is rather equivocal regarding the width of the riparian buffer which is relevant for the determination of stream temperature at a given point. As a matter of fact, none of the tested
1120 buffer widths appears to prevail over the other ones in the retained parametrizations of $\gamma\phi_r$. This ambiguity reflects the range of buffer widths used in the literature, which extends from 30 m (e.g. Jones et al., 2006; Macedo et al., 2013) to 200 m (e.g. Scott et al., 2002; Segura et al., 2014). This also points at the difficulty to adequately account for the effect of riparian vegetation using the available spatial data sets, which often lack important details such as the distinction between deciduous and coniferous forest.
1125

The precision of the ~~physically-based physics-inspired~~ model was reported in the previous section to be rather low in January–March. This can be explained by the fact that the non-linearity of the stream–air temperature relationship at low air temperature values is not captured by the model. The latter rather simulates a sharp transition from the linear regime to a constant one, since the stream
1130 temperature values predicted to be negative are systematically replaced with 0°C. This implies a faster decrease towards 0°C, which is at the origin of the strong negative model bias in winter.

As noticeable in Table 5, the model RMSE is larger in spring as compared to the other seasons. This is attributable to the fact that many of the selected watersheds are impacted by snow melt in spring. Since the snow cover conditions are strongly variable both spatially and temporally, a large

1135 dispersion of the stream temperature values is typically observed in spring. The model performs nonetheless relatively well in this season, for its R^2 value is of the same magnitude in spring as during the rest of the year.

Advantage can be taken of the ~~physical basis of the model~~ physics integrated into the model structure to investigate some aspects of the stream temperature dynamics. For example, Fig. 6 displays the respective values of the factors ω_1 , ω_2 and ω_3 appearing in Eq. (22) as a function of the season. These factors correspond to the weights associated with the mean source temperature \bar{T}_s , average lateral inflow temperature along the network $\langle T_\ell \rangle_{\mathcal{L}}$ and average equilibrium temperature along the network $\langle T_e \rangle_{\mathcal{L}}$, respectively. As seen in the figure, ω_3 is the largest factor of the three in all seasons, with a value of about 0.6–0.8. This results from the fact that stream temperature is primarily impacted by the atmospheric conditions. The other two factors are nonetheless non-negligible, with ω_1 being of the order of 0.4 in winter and ω_2 being approximately equal to 0.2 from April to December. The value of ω_1 has to be put into perspective with respect to the fact that our confidence in the model parametrization is relatively low in winter. Moreover, following the above discussion about the computation of the factor η , the values reported here for ω_2 should be considered as a lower limit. It therefore appears that not only the net total heat flux at the air–water interface is important in determining stream temperature, but also the heat flux associated with the lateral inflow of water. This conclusion is in agreement with the findings of Bogan et al. (2004), who found that the precision of their stream temperature model was improved by including a term accounting for the lateral water inflow. Similarly, Herb and Stefan (2011) mention that the heat input associated with groundwater infiltration may be of the same order of magnitude as the heat flux due to atmospheric forcing in some cases. This effect seems to be largely underestimated in the literature, since the lateral inflow of water has often been neglected in previous stream temperature models (e.g. Edinger et al., 1968; Bogan et al., 2003; Caissie et al., 2005; Bustillo et al., 2014).

The simplifying assumptions (i)–(vi) reported in Sect. 3.1.1 are likely to have limited the performance of the physics-inspired model. In particular, the assumption of a spatially homogeneous lateral inflow rate q_ℓ is expected to fail in most catchments. For example, only the highest regions of low-altitude catchments experience snow-melt in spring. In higher-altitude catchments, snow-melt leads to an increase in q_ℓ only at low altitudes at the beginning of spring, and only at higher altitudes latter in the season. These mechanisms introduce an altitude-dependence in q_ℓ which contradicts our assumption and may partly explain the higher RMSE of the model in spring. Similarly, assumption (v) expresses stream width as a linear function of discharge. As compared to the typical power-law relationship used in fluvial geomorphology (Knighton, 1998), this simplification may lead to an overestimation of stream width at low discharge rates—i.e. in small catchments—and to an underestimation of stream width at high discharge rates—i.e. in large catchments. This may in turn decrease the ability of the model to simulate catchments of various sizes, hereby increasing its prediction error. Assumption (vi) stating that all sources in a given catchment have the same discharge rate is also disputable. This is

particularly true for small catchments, where the short distance to the outlet and the low number of sources do not allow the averaging effect to be significant enough to compensate for the introduced error.

1175 In addition to the simplifying assumptions discussed above, the parametrizations of the unknown terms in the analytical solution might also have impacted the model precision. Indeed, the estimation of the source and lateral inflow temperatures using only air temperature has recently been questioned, particularly for the catchments impacted by snow- or glacier-melt (Leach and Moore, 2015). This simplification may notably have contributed to increase the model RMSE in spring. Regarding the
1180 parametrization of the term accounting for the net radiation heat flux at the air–water interface, the use of a linear expression may appear as limiting. We actually tested a power-law function as well, but did not succeed in calibrating the model due to convergence issues. We also considered an alternative expression based on an estimation of the incoming atmospheric radiation and a first-order approximation of the long-wave radiation emitted by the stream (not shown). Rather than an improvement,
1185 this parametrization actually led to a decrease in the model precision, as a result of its inability to compensate for the overestimation of ω_3 (see above).

As opposed to the ~~physically-based physics-inspired~~ model, the parameter values of the standard regression model could not be constrained using physical considerations. As a result, the sign of some of the linear coefficients relating the predictor variables to \bar{T}_i and σ_i are in contradiction with
1190 our understanding of stream temperature dynamics (see Table 6). For example, the stream orientation θ , measured as the cosine of the angle between north and the channel direction, is positively related to the annual mean stream temperature \bar{T}_i . It was rather expected that north-oriented catchments receive less radiation from the sun, hereby implying lower stream temperatures. The same observation is true for the riparian forest cover fraction, which is positively associated in the model with
1195 the annual standard deviation σ_i of stream temperature. However, experimental observations tend to conclude that riparian shading has a buffering effect on stream temperature, therefore dampening the amplitude of the variations of the latter (see e.g. Moore et al., 2005). Despite these inconsistencies, the standard regression model performs better than the ~~physically-based physics-inspired~~ one in terms of RMSE, R^2 and bias. This fact questions further the validity of the parametrization
1200 of the ~~physically-based physics-inspired~~ model, which could certainly be improved (see Sect. 6). On the other hand, the standard regression model appears to be much more sensitive to the stream network resolution as compared to its counterpart, possibly as a consequence of its lack of physical ~~basiselements into its structure~~. This lack does also not allow for the investigation of the physics governing stream temperature, as can be done with the ~~physically-based physics-inspired~~ model (see
1205 Fig. 6).

6 Conclusions

This study aimed at presenting a new statistical model for the prediction of monthly mean stream temperature in ungauged basins. ~~The model corresponds to an~~ As opposed to the standard statistical methods, this model is devised so as to incorporate physical considerations into its structure. To
1210 this end, it is built upon the analytical solution to a simplified version of the one dimensional heat advection equation, ~~simplified using some approximations. Among the latter, the most important one assumes the ratio of stream width to discharge to remain spatially constant. As opposed to previous studies relying on the same approach, the heat advection equation is solved here.~~ Contrary to previously reported analytical solutions, the present one is obtained by solving the equation
1215 over an entire stream network instead of a single stream ~~reacheach~~. Moreover, the various terms of the equation are ~~also~~ not supposed to be spatially homogeneous, which leads to the apparition of a space averaging operator $\langle \cdot \rangle_L$ applied to most terms of the solution. This operator uses a weight which decreases exponentially with the distance to the catchment outlet, hereby giving more emphasis to the points located near the gauging station. ~~The extent of the spatial area over which the operator acts is controlled by a parameter L_c , referred to as the stream characteristic length. This parameter is typically dynamic, which explains why the model needs to be calibrated separately in each season.~~
1220 Both the source and the lateral inflow terms—which are usually neglected—are retained in the final solution to the heat advection equation. This notably enables the model to be applied in small watersheds, where the influence of source temperature on the value of stream temperature measured at
1225 the catchment outlet cannot be discarded.

While most terms of the ~~model expression could be evaluated based on~~ analytical expression can be evaluated using meteorological observations or topographic maps, some ~~required an additional modeling effort. require data which is not available. These terms are replaced with approximations based on the spatial data sets at hand.~~
1230 was approximated is expressed as a linear combination of several ~~predictor variables, e. g. topographical shading or riparian forest cover. physiographic variables.~~ Similarly, the source and lateral inflow temperatures ~~were expressed are approximated~~ as a linear function of air temperature measured at the source location and along the stream, respectively. Finally, the fraction η of discharge at the catchment outlet originating from lateral water inflow along the network ~~was is~~ estimated
1235 based on the number of sources in the watershed. ~~These parametrizations present the advantage that the information they require is readily available at the regional scale, namely: air temperature measurements, a land cover map and a digital elevation map. They differ from the physically-based parametrizations used in other studies (e.g. Bogan et al., 2004; Herb and Stefan, 2011), which require measurements of meteorological variables being hard to interpolate in mountainous terrain, such as~~
1240 ~~wind speed or relative humidity~~ As a consequence of these approximations, the resulting model is statistical in nature, but nevertheless retains physical aspects due to its global structure being derived from the heat-balance equation.

The performance of the model is quite satisfactory, with a root-mean-square error of about 1.3°C and a coefficient of determination R^2 of 0.87 when used for stream temperature prediction in “thermally climate-driven” catchments. These catchments, which are by far the most abundant ones in Switzerland, correspond to those with a glacier cover lower than 50% and whose stream is not impacted by groundwater infiltration from deep aquifer. Model precision is the lowest in winter, due to the inability of the model to reproduce the fact that stream temperature asymptotically tends towards 0°C for negative air temperature values. ~~Closer investigation of the model error could attribute part of its imprecision to the parametrization of the factor η , which corresponds to the fraction of discharge originating from the lateral inflow of water (in opposition to the discharge fraction originating from the sources).~~

The precision of the model was also assessed by comparing it with a more ~~simple~~ standard regression model. The latter was observed to perform slightly better, with a RMSE about 0.2°C lower. However, its parameters could not be interpreted from a physical point of view, hereby hindering the restriction of their respective calibration ranges based on physical considerations. This led the regression model to simulate some aspects of the stream temperature dynamics wrongly. For example, some physiographic variables known to have a cooling effect on water temperature were modeled as warming up the stream. The standard regression model was also observed to be much more sensitive than its ~~physically-based~~ physics-inspired counterpart with respect to the stream network resolution. When discarding all stream segments with a Strahler order equal to one, the RMSE of the regression model increased from 1.12°C to 1.26°C, whereas the one of the ~~physically-based~~ physics-inspired model remained constant up to 0.01°C.

Despite a few deficiencies, the ~~physically-based~~ physics-inspired statistical model can be used to analyse some aspects of the physics governing stream temperature. As an example, the relative importance of each one of the stream heat sources could be determined from the model. Climatic forcing was found to be the major driver of water temperature, as expected (e.g. Caissie, 2006). More interestingly, the lateral inflow of water was identified as a non-negligible secondary heat flux. This fact is confirmed by other studies (e.g. Bogan et al., 2004; Herb and Stefan, 2011), but nonetheless fails to be accounted for in many stream temperature models (e.g. Caissie et al., 2005; Bustillo et al., 2014). We therefore wish ~~at emphasizing to emphasize~~ the role of lateral water inflow on stream temperature, even in catchments—such as those used in this study—which are not impacted by groundwater infiltration originating from a deep aquifer.

Among the improvements that can be brought to the ~~physically-based~~ physics-inspired model, a more accurate parametrization of the discharge fraction originating from lateral water inflow η appears as a promising enhancement. In particular, η could be estimated from a geomorphological analysis of the catchments. This approach was not retained here due to the discrepancy between the stream networks predicted by the geomorphological analysis and the observed ones. In case it could be implemented, such a revision is expected to improve the predicted slope of the stream–air

1280 temperature curve. A geomorphological analysis could also positively influence the modeling of the source and lateral inflow temperatures. The parametrization of these two terms could be improved by including predictor variables accounting for e.g. the glacier cover fraction or the mean altitude of the area drained by each source (resp. stream reach). ~~As a matter of fact, surface and subsurface runoffs originating from glacier-covered or high-altitude subcatchments are expected to have particularly~~
1285 ~~low temperatures, especially during the snow and ice-melt season.~~ The model could also be substantially improved in case the characteristic stream length L_c , which controls the extent of the spatial area over which the operator $\langle \cdot \rangle_c$ acts, could be computed instead of calibrated. Indeed, L_c does not only present a seasonal variation but also differs across the individual catchments, a fact which was neglected in the present work. ~~The a-priori computation of L_c would imply the ratio of~~
1290 ~~discharge to channel width to be independently estimated in each catchment. Such an estimation would in turn require the stream rating curve to be known, hereby limiting the applicability of the model to gauged watersheds.~~ Finally, one might expect the model precision to improve by using a more physically-based parametrization for the net radiation heat flux—instead of the multi-linear model used here. ~~We actually tested an alternative expression based on an estimation of the incoming~~
1295 ~~atmospheric radiation and a first-order approximation of the long-wave radiation emitted by the stream (not shown). Rather than an improvement, this parametrization led to a decrease of the model precision, as a result of its inability to compensate for the overestimation of another parametrized term (ω_3) in the model—contrary to the multi-linear expression.~~

We expect the physically-based physics-inspired model to be easily transferable to other regions
1300 of the globe. The parametrization of the net radiation heat flux at the air–water interface might need some adaptation in order to correctly reflect the dominant physiographic controls on local stream climate. For example, topographic shading is certainly not a relevant predictor variable over flat regions. Similarly to the approach presented in this work, the most appropriate set of predictor variables for the net radiation heat flux over a particular region can be obtained through AICc ranking.
1305 Once set, the stream temperature model can be used to investigate e.g. the extent of the stream network which is thermally suitable for sensitive fish species at the regional scale (e.g. Isaak et al., 2010). This investigation can in turn serve as a basis for the introduction of regulation policies or protection measures.

Appendix A: Solution to the energy-balance equation over a stream network

1310 The analytical solution to Eqs. (5)–(6) is first derived for the case of a simple stream reach of length L (see Fig. 7a). Let the downstream distance be denoted as x , with x_0 and x_1 referring to the positions of the reach origin and end points, respectively. The discharge $Q(x_1)$ and stream temperature $T(x_1)$

can be easily computed by integrating Eqs. (5)–(6) between x_0 and x_1 ,

$$Q(x_1) = Q(x_0) + q_\ell L, \quad (\text{A1})$$

$$1315 \quad T(x_1) = \frac{Q(x_0)}{Q(x_1)} e^{-d(x_0)/L_c} T(x_0) + \frac{1}{Q(x_1)} \int_{x_0}^{x_1} e^{-d(x)/L_c} \psi(x) dx, \quad (\text{A2})$$

where

$$\psi(x) = q_\ell T_\ell(x) + Q(x) \left(\frac{\gamma \phi_r(x) + T_a(x)}{L_c} - \frac{g}{c_p} \frac{dz}{dx} \right). \quad (\text{A3})$$

In Eq. (A2), $d(x) = x_1 - x$ denotes the distance between any point x and the downstream end point x_1 (see Figure 7a).

1320 The above equations require the values of discharge and temperature at the upstream end of the reach to be known. By applying them iteratively to all the reaches of a network, starting from the most downstream one, the expressions for discharge Q_{out} and water temperature T_{out} at the network outlet can be expressed as a function of the discharges and temperatures of the sources. At the confluences, the discharges $Q_{u,1}$ and $Q_{u,2}$ and the temperatures $T_{u,1}$ and $T_{u,2}$ of the two upstream
1325 reaches can be related to the discharge Q_d and temperature T_d of the downstream reach using the mass and energy-balance equations (Westhoff et al., 2007),

$$Q_d = Q_{u,1} + Q_{u,2}, \quad (\text{A4})$$

$$T_d = \frac{Q_{u,1}}{Q_d} T_{u,1} + \frac{Q_{u,2}}{Q_d} T_{u,2}. \quad (\text{A5})$$

Based on Eqs. (A1)–(A5), the derivation of the expressions for Q_{out} and T_{out} is straightforward and
1330 leads to the following relations,

$$Q_{\text{out}} = \sum_{i=1}^{n_s} Q_{s,i} + q_\ell L_{\text{tot}} \quad (\text{A6})$$

$$T_{\text{out}} = \sum_{i=1}^{n_s} \frac{Q_{s,i}}{Q_{\text{out}}} e^{-d_{s,i}/L_c} T_{s,i} + \frac{1}{Q_{\text{out}}} \int_{\mathcal{L}} e^{-d(x)/L_c} \psi(x) dx. \quad (\text{A7})$$

In the above equations, n_s refers to the number of sources in the network, L_{tot} denotes the total length of the stream network, $d_{s,i}$ corresponds to the downstream distance of source point $x_{s,i}$ to the
1335 network outlet, $d(x)$ refers to the distance between any point x along the network and the network outlet, \mathcal{L} corresponds to the geometrical union of all reaches in the stream network, and $T_{s,i}$ and $Q_{s,i}$ denote the stream temperature and discharge at source point $x_{s,i}$, respectively. The reader is referred to Fig. 7b for a graphical illustration of some of the variables. The integral over \mathcal{L} is a short-hand notation for the sum of the respective integrals over all the reaches in the network.

1340 Equations (A6)–(A7) can be written in a more convenient form using space-averaging operators. Replacing ψ with its expression defined in Eq. (A3), the integral on the right-hand side of Eq. (A7) can be written as

$$\int_{\mathcal{L}} e^{-d(x)/L_c} \psi(x) dx = A_1 q_\ell \langle T_\ell \rangle_{\mathcal{L}} + \frac{A_2}{L_c} \left\langle \gamma \phi_r + T_a - L_c \frac{g}{c_p} \frac{dz}{dx} \right\rangle_{\mathcal{Q}}, \quad (\text{A8})$$

where the two averaging operators $\langle \cdot \rangle_{\mathcal{L}}$ and $\langle \cdot \rangle_{\mathcal{Q}}$ are defined as

$$1345 \quad \langle f \rangle_{\mathcal{L}} = \frac{1}{A_1} \int_{\mathcal{L}} e^{-d(x)/L_c} f(x) \, dx, \quad (\text{A9})$$

$$\langle f \rangle_{\mathcal{Q}} = \frac{1}{A_2} \int_{\mathcal{L}} e^{-d(x)/L_c} Q(x) f(x) \, dx, \quad (\text{A10})$$

for any integrable function f defined on \mathcal{L} , with the normalizing factors A_1 and A_2 being defined as

$$\begin{aligned} A_1 &= \int_{\mathcal{L}} e^{-d(x)/L_c} \, dx, \\ &= \sum_{k=1}^{n_r} \int_{x_{k,0}}^{x_{k,1}} e^{-d(x)/L_c} \, dx \\ 1350 \quad &= \sum_{k=1}^{n_r} L_c e^{-d_k/L_c} (1 - e^{-L_k/L_c}), \end{aligned} \quad (\text{A11})$$

$$\begin{aligned} A_2 &= \int_{\mathcal{L}} e^{-d(x)/L_c} Q(x) \, dx \\ &= \sum_{k=1}^{n_r} \int_{x_{k,0}}^{x_{k,1}} e^{-d(x)/L_c} Q_k(x) \, dx. \end{aligned} \quad (\text{A12})$$

In the above equations, n_r denotes the number of reaches in the network; $x_{k,0}$, $x_{k,1}$ and L_k refer to the upstream point, downstream point and length of reach k respectively; d_k refers to the distance
1355 along the stream network between $x_{k,1}$ and the network outlet; and $Q_k(x)$ denotes the discharge along reach k (see Fig. 7b). Based on Eq. (A1), Q_k may be expressed as

$$Q_k(x) = Q(x_{k,0}) + q_{\ell}(x - x_{k,0}) = \sum_{j \in \Pi_k} Q_{s,j} + q_{\ell} \sum_{r \in \Gamma_k} L_r + q_{\ell}(x - x_{k,0}), \quad (\text{A13})$$

where Π_k and Γ_k denote the set of source points and reaches draining into reach k , respectively, as illustrated in Fig. 7c. Inserting the above equation in Eq. (A12) and re-arranging the terms, A_2 may
1360 be written as

$$\begin{aligned} A_2 &= \sum_{i=1}^{n_s} Q_{s,i} \sum_{j \in \Omega_i} \int_{x_{j,0}}^{x_{j,1}} e^{-d(x)/L_c} \, dx \\ &+ q_{\ell} \sum_{k=1}^{n_r} L_k \sum_{r \in \Delta_k} \int_{x_{r,0}}^{x_{r,1}} e^{-d(x)/L_c} \, dx \\ &+ q_{\ell} \sum_{k=1}^{n_r} \int_{x_{k,0}}^{x_{k,1}} e^{-d(x)/L_c} x \, dx, \end{aligned} \quad (\text{A14})$$

where Ω_i refers to the set of reaches linking the i -th source point to the network outlet, and Δ_k denotes the set of reaches linking reach k to the network outlet, not including reach k itself (see

Fig. 7c). Assuming that all source points have the same discharge Q_s , and replacing the integrals
 1365 along the reaches with their respective values, Eq. (A14) can be written more simply as

$$\begin{aligned}
 A_2 &= L_c Q_s \sum_{i=1}^{n_s} (1 - e^{-d_{s,i}/L_c}) \\
 &\quad + q_\ell L_c \sum_{k=1}^{n_r} \left\{ L_k - L_c e^{-d_k/L_c} (1 - e^{-L_k/L_c}) \right\} \\
 &= L_c (Q_{\text{tot}} - q_\ell L_{\text{tot}}) (1 - \delta_s) + q_\ell L_c L_{\text{tot}} (1 - \delta_\ell),
 \end{aligned} \tag{A15}$$

where Eq. (A6) has been used in the second step to replace $n_s Q_s$ with $Q_{\text{tot}} - q_\ell L_{\text{tot}}$, and the factors
 δ_s and δ_ℓ are defined as

$$\delta_s = \frac{1}{n_s} \sum_{i=1}^{n_s} e^{-d_{s,i}/L_c}, \tag{A16}$$

$$1370 \quad \delta_\ell = \frac{L_c}{L_{\text{tot}}} \sum_{k=1}^{n_r} e^{-d_k/L_c} (1 - e^{-L_k/L_c}) = \frac{A_1}{L_{\text{tot}}}. \tag{A17}$$

Combining Eqs. (A7), (A8), (A11) and (A15), the expression for stream temperature at the network
 outlet can eventually be written in a more convenient form,

$$T_{\text{out}} = (1 - \eta) \delta_s \bar{T}_s + \eta \delta_\ell \langle T_\ell \rangle_{\mathcal{L}} + [1 - (1 - \eta) \delta_s - \eta \delta_\ell] \langle \gamma \phi_r + T_a - L_c \frac{g}{c_p} \frac{dz}{dx} \rangle_{\mathcal{L}}, \tag{A18}$$

where the averaging operator $\langle \cdot \rangle_{\mathcal{Q}}$ has been approximated by $\langle \cdot \rangle_{\mathcal{L}}$, and \bar{T}_s corresponds to the
 1375 distance-weighted source temperature, averaged over all sources and weighted by a factor decreasing
 exponentially with the respective distance of each source to the network outlet,

$$\bar{T}_s = \frac{1}{n_s \delta_s} \sum_{i=1}^{n_s} e^{-d_{s,i}/L_c} T_{s,i}. \tag{A19}$$

The factor η appearing in Eq. (A18) denotes the ratio between the discharge originating from lateral
 inflow and the total discharge at the network outlet,

$$1380 \quad \eta = \frac{q_\ell L_{\text{tot}}}{Q_{\text{tot}}}. \tag{A20}$$

Author contributions. A. Gallice performed the analysis, produced the figures and wrote the manuscript. B. Schae-
 fli gave much appreciated guidance and impulse to the present work. M. P. Parlange commented on the manuscript.
 H. Huwald, M. Lehning and B. Schaepli helped write the manuscript and co-supervised the work.

Acknowledgements. This work was financially supported by the Swiss Federal Office for the Environment
 1385 (FOEN). The Department for Construction, Transport and Environment of the Canton Aargau and FOEN are
 greatly acknowledged for the free access to their hydrological data. All plots have been produced with the
 Matplotlib Python library (Hunter, 2007).

References

- 1390 Arscott, D. B., Tockner, K., and Ward, J.: Thermal heterogeneity along a braided floodplain river (Tagliamento River, northeastern Italy), *Canadian Journal of Fisheries and Aquatic Sciences*, 58, 2359–2373, 2001.
- Aschwanden, H. and Weingartner, R.: *Die Abflussregimes der Schweiz*, Publikation gewässerkunde nr. 65, Geographisches Institut der Universität Bern, 1985.
- Benyahya, L., Caissie, D., St-Hilaire, A., Ouarda, T. B., and Bobée, B.: A Review of Statistical Water Temperature Models, *Canadian Water Resources Journal*, 32, 179–192, doi:10.4296/cwrj3203179, http://dx.doi.org/10.4296/cwrj3203179, 2007.
- 1395 Beven, K.: Prophecy, reality and uncertainty in distributed hydrological modelling, *Advances in Water Resources*, 16, 41–51, doi:http://dx.doi.org/10.1016/0309-1708(93)90028-E, http://www.sciencedirect.com/science/article/pii/030917089390028E, 1993.
- Beven, K. J.: *Rainfall-Runoff Modelling: The Primer*, 2nd Edition, Wiley-Blackwell, 2012.
- 1400 Beyeler, M.: *Wassertemperatur-Trendanalyse, Langzeituntersuchungen der Temperaturen schweizerischer Fliessgewässer*, Master's thesis, Bern University, 2007.
- Biswal, B. and Marani, M.: Geomorphological origin of recession curves, *Geophysical Research Letters*, 37, n/a–n/a, doi:10.1029/2010GL045415, http://dx.doi.org/10.1029/2010GL045415, 2010.
- Bogan, T., Mohseni, O., and Stefan, H. G.: Stream temperature-equilibrium temperature relationship, *Water Resources Research*, 39, n/a–n/a, doi:10.1029/2003WR002034, http://dx.doi.org/10.1029/2003WR002034, 2003.
- 1405 Bogan, T., Stefan, H. G., and Mohseni, O.: Imprints of secondary heat sources on the stream temperature/equilibrium temperature relationship, *Water Resources Research*, 40, n/a–n/a, doi:10.1029/2003WR002733, http://dx.doi.org/10.1029/2003WR002733, 2004.
- 1410 Brown, G. W.: Predicting Temperatures of Small Streams, *Water Resources Research*, 5, 68–75, doi:10.1029/WR005i001p00068, http://dx.doi.org/10.1029/WR005i001p00068, 1969.
- Burnham, K. P. and Anderson, D. R.: *Model selection and multimodel inference: a practical information-theoretic approach*, Springer, New York, 2002.
- Bustillo, V., Moatar, F., Ducharme, A., Thiéry, D., and Poirel, A.: A multimodel comparison for assessing water temperatures under changing climate conditions via the equilibrium temperature concept: case study of the Middle Loire River, France, *Hydrological Processes*, 28, 1507–1524, doi:10.1002/hyp.9683, http://dx.doi.org/10.1002/hyp.9683, 2014.
- 1415 Caissie, D.: The thermal regime of rivers: a review, *Freshwater Biology*, 51, 1389–1406, doi:10.1111/j.1365-2427.2006.01597.x, http://dx.doi.org/10.1111/j.1365-2427.2006.01597.x, 2006.
- 1420 Caissie, D., El-Jabi, N., and Satish, M. G.: Modelling of maximum daily water temperatures in a small stream using air temperatures, *Journal of Hydrology*, 251, 14–28, doi:10.1016/S0022-1694(01)00427-9, http://www.sciencedirect.com/science/article/pii/S0022169401004279, 2001.
- Caissie, D., Satish, M. G., and El-Jabi, N.: Predicting river water temperatures using the equilibrium temperature concept with application on Miramichi River catchments (New Brunswick, Canada), *Hydrological Processes*, 19, 2137–2159, doi:10.1002/hyp.5684, http://dx.doi.org/10.1002/hyp.5684, 2005.
- 1425

- Caldwell, R. J., Gangopadhyay, S., Bountry, J., Lai, Y., and Elsner, M. M.: Statistical modeling of daily and sub-daily stream temperatures: Application to the Methow River Basin, Washington, *Water Resources Research*, 49, 4346–4361, doi:10.1002/wrcr.20353, <http://dx.doi.org/10.1002/wrcr.20353>, 2013.
- Carlier, C.: Water temperature modelling of the Venoge River with meteorological data, Master's thesis, Swiss
1430 Federal Institute of Technology Zurich, 2013.
- Chang, H. and Psaris, M.: Local landscape predictors of maximum stream temperature and thermal sensitivity in the Columbia River Basin, USA, *Science of The Total Environment*, 461–462, 587–600, doi:10.1016/j.scitotenv.2013.05.033, <http://www.sciencedirect.com/science/article/pii/S0048969713005767>, 2013.
- 1435 Daigle, A., St-Hilaire, A., Peters, D., and Baird, D.: Multivariate Modelling of Water Temperature in the Okanagan Watershed, *Canadian Water Resources Journal*, 35, 237–258, doi:10.4296/cwrj3503237, <http://dx.doi.org/10.4296/cwrj3503237>, 2010.
- Daly, S. F.: Anchor ice flooding in Jackson, WY, in: *World Water and Environmental Resources Congress 2005: Impacts of Global Climate Change*, pp. 1–9, American Society of Civil Engineers,
1440 doi:10.1061/40792(173)240, 2005.
- DeWeber, J. T. and Wagner, T.: A regional neural network ensemble for predicting mean daily river water temperature, *Journal of Hydrology*, 517, 187–200, doi:10.1016/j.jhydrol.2014.05.035, <http://www.sciencedirect.com/science/article/pii/S0022169414003990>, 2014.
- Ducharne, A.: Importance of stream temperature to climate change impact on water quality, *Hydrology and
1445 Earth System Sciences*, 12, 13, <http://www.hydrol-earth-syst-sci.net/12/797/2008/>, 2008.
- Ducharne, A.: Reducing scale dependence in TOPMODEL using a dimensionless topographic index, *Hydrology and Earth System Sciences*, 13, 2399–2412, doi:10.5194/hess-13-2399-2009, <http://www.hydrol-earth-syst-sci.net/13/2399/2009/>, 2009.
- Eckhardt, K.: How to construct recursive digital filters for baseflow separation, *Hydrological Processes*, 19,
1450 507–515, doi:10.1002/hyp.5675, <http://dx.doi.org/10.1002/hyp.5675>, 2005.
- Edinger, J. E., Duttweiler, D. W., and Geyer, J. C.: The Response of Water Temperatures to Meteorological Conditions, *Water Resources Research*, 4, 1137–1143, doi:10.1029/WR004i005p01137, <http://dx.doi.org/10.1029/WR004i005p01137>, 1968.
- Gardner, B. and Sullivan, P. J.: Spatial and temporal stream temperature prediction: Modeling nonstationary
1455 temporal covariance structures, *Water Resources Research*, 40, doi:10.1029/2003WR002511, <http://dx.doi.org/10.1029/2003WR002511>, 2004.
- Garner, G., Hannah, D. M., Sadler, J. P., and Orr, H. G.: River temperature regimes of England and Wales: spatial patterns, inter-annual variability and climatic sensitivity, *Hydrological Processes*, 28, 5583–5598, doi:10.1002/hyp.9992, <http://dx.doi.org/10.1002/hyp.9992>, 2014.
- 1460 Gosink, J. P.: Synopsis of Analytic Solutions for the Temperature Distribution in a River Downstream From a Dam or Reservoir, *Water Resources Research*, 22, 979–983, doi:10.1029/WR022i006p00979, <http://dx.doi.org/10.1029/WR022i006p00979>, 1986.
- Grbić, R., Kurtagić, D., and Sliškočić, D.: Stream water temperature prediction based on Gaussian process regression, *Expert Systems with Applications*, 40, 7407–7414, doi:10.1016/j.eswa.2013.06.077, <http://www.sciencedirect.com/science/article/pii/S0957417413004764>, 2013.
1465

- Haag, I. and Luce, A.: The integrated water balance and water temperature model LARSIM-WT, *Hydrological Processes*, 22, 1046–1056, doi:10.1002/hyp.6983, <http://dx.doi.org/10.1002/hyp.6983>, 2008.
- Hannah, D. M., Malcolm, I. A., Soulsby, C., and Youngson, A. F.: Heat exchanges and temperatures within a salmon spawning stream in the Cairngorms, Scotland: seasonal and sub-seasonal dynamics, *River Research and Applications*, 20, 635–652, doi:10.1002/rra.771, <http://dx.doi.org/10.1002/rra.771>, 2004.
- 1470 Hari, R. and Güttinger, H.: *Temperaturverlauf in Schweizer Flüssen 1978–2002, Auswertungen und grafische Darstellungen fischrelevanter Parameter*, Fischnetzpublikation, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland, 2004.
- Hari, R. and Zobrist, J.: *Trendanalyse der NADUF-Messresultate 1974 bis 1998*, Schriftenreihe der EAWAG 17, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland, 2003.
- 1475 Hawkins, C. P., Hogue, J. N., Decker, L. M., and Feminella, J. W.: Channel Morphology, Water Temperature, and Assemblage Structure of Stream Insects, *Journal of the North American Benthological Society*, 16, 728–749, <http://www.jstor.org/stable/1468167>, 1997.
- Herb, W. R. and Stefan, H. G.: Modified equilibrium temperature models for cold-water streams, *Water Resources Research*, 47, n/a–n/a, doi:10.1029/2010WR009586, <http://dx.doi.org/10.1029/2010WR009586>, 2011.
- 1480 Hill, R. A., Hawkins, C. P., and Carlisle, D. M.: Predicting thermal reference conditions for USA streams and rivers, *Freshwater Science*, 32, 39–55, doi:10.1899/12-009.1, 2013.
- Hrachowitz, M., Soulsby, C., Imholt, C., Malcolm, I. A., and Tetzlaff, D.: Thermal regimes in a large upland salmon river: a simple model to identify the influence of landscape controls and climate change on maximum temperatures, *Hydrological Processes*, 24, 3374–3391, doi:10.1002/hyp.7756, <http://dx.doi.org/10.1002/hyp.7756>, 2010.
- 1485 Hrachowitz, M., Savenije, H., Blöschl, G., McDonnell, J., Sivapalan, M., Pomeroy, J., Arheimer, B., Blume, T., Clark, M., Ehret, U., Fenicia, F., Freer, J., Gelfan, A., Gupta, H., Hughes, D., Hut, R., Montanari, A., Pande, S., Tetzlaff, D., Troch, P., Uhlenbrook, S., Wagener, T., Winsemius, H., Woods, R., Zehe, E., and Cudennec, C.: A decade of Predictions in Ungauged Basins (PUB)—a review, *Hydrological Sciences Journal*, 58, 1198–1255, doi:10.1080/02626667.2013.803183, <http://dx.doi.org/10.1080/02626667.2013.803183>, 2013.
- 1490 Hunter, J. D.: Matplotlib: A 2D graphics environment, *Computing In Science & Engineering*, 9, 90–95, 2007.
- Imholt, C., Soulsby, C., Malcolm, I. A., Hrachowitz, M., Gibbins, C. N., Langan, S., and Tetzlaff, D.: Influence of scale on thermal characteristics in a large montane river basin, *River Research and Applications*, 29, 403–419, doi:10.1002/rra.1608, <http://dx.doi.org/10.1002/rra.1608>, 2013.
- 1495 Isaak, D. J. and Hubert, W. A.: A hypothesis about factors that affect maximum summer stream temperatures across montane landscapes, *JAWRA Journal of the American Water Resources Association*, 37, 351–366, doi:10.1111/j.1752-1688.2001.tb00974.x, <http://dx.doi.org/10.1111/j.1752-1688.2001.tb00974.x>, 2001.
- 1500 Isaak, D. J., Luce, C. H., Rieman, B. E., Nagel, D. E., Peterson, E. E., Horan, D. L., Parkes, S., and Chandler, G. L.: Effects of climate change and wildfire on stream temperatures and salmonid thermal habitat in a mountain river network, *Ecological Applications*, 20, 1350–1371, 2010.
- Jakob, A.: Temperaturen in Schweizer Fließgewässern, *GWA*, 3, 221–231, 2010.
- Johnson, F.: Stream temperatures in an Alpine area, *Journal of Hydrology*, 14, 322–336, doi:10.1016/0022-1694(71)90042-4, <http://www.sciencedirect.com/science/article/pii/0022169471900424>, 1971.
- 1505

- Johnson, M. F., Wilby, R. L., and Toone, J. A.: Inferring air–water temperature relationships from river and catchment properties, *Hydrological Processes*, 28, 2912–2928, doi:10.1002/hyp.9842, <http://dx.doi.org/10.1002/hyp.9842>, 2014.
- 1510 Jones, K. L., Poole, G. C., Meyer, J. L., Bumback, W., and Kramer, E. A.: Quantifying expected ecological response to natural resource legislation: a case study of riparian buffers, aquatic habitat, and trout populations, *Ecology and Society*, 11, 15, <http://www.ecologyandsociety.org/vol11/iss2/art15/>, 2006.
- Kelleher, C., Wagener, T., Gooseff, M., McGlynn, B., McGuire, K., and Marshall, L.: Investigating controls on the thermal sensitivity of Pennsylvania streams, *Hydrological Processes*, 26, 771–785, doi:10.1002/hyp.8186, <http://dx.doi.org/10.1002/hyp.8186>, 2012.
- 1515 Knighton, D.: *Fluvial forms and processes: a new perspective*, Arnold, 1998.
- Langford, T.: *Ecological Effects of Thermal Discharges*, Pollution Monitoring Series, Springer, <http://books.google.ch/books?id=f1M6kRZ7MUC>, last access: 17 March 2015, 1990.
- Leach, J. A. and Moore, R. D.: Winter stream temperature in the rain-on-snow zone of the Pacific Northwest: influences of hillslope runoff and transient snow cover, *Hydrology and Earth System Sciences*, 18, 819–838, doi:10.5194/hess-18-819-2014, <http://www.hydrol-earth-syst-sci.net/18/819/2014/>, 2014.
- 1520 Leach, J. A. and Moore, R. D.: Observations and modeling of hillslope throughflow temperatures in a coastal forested catchment, *Water Resources Research*, 51, 3770–3795, doi:10.1002/2014WR016763, <http://dx.doi.org/10.1002/2014WR016763>, 2015.
- Macedo, M. N., Coe, M. T., DeFries, R., Uriarte, M., Brando, P. M., Neill, C., and Walker, W. S.: Land-use-driven stream warming in southeastern Amazonia, *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368, doi:10.1098/rstb.2012.0153, <http://rstb.royalsocietypublishing.org/content/368/1619/20120153.abstract>, 2013.
- Magnusson, J., Jonas, T., and Kirchner, J. W.: Temperature dynamics of a proglacial stream: Identifying dominant energy balance components and inferring spatially integrated hydraulic geometry, *Water Resources Research*, 48, n/a–n/a, doi:10.1029/2011WR011378, <http://dx.doi.org/10.1029/2011WR011378>, 2012.
- 1530 Mayer, T. D.: Controls of summer stream temperature in the Pacific Northwest, *Journal of Hydrology*, 475, 323–335, doi:10.1016/j.jhydrol.2012.10.012, <http://www.sciencedirect.com/science/article/pii/S0022169412008864>, 2012.
- Meier, W., Bonjour, C., Wüest, A., and Reichert, P.: Modeling the Effect of Water Diversion on the Temperature of Mountain Streams, *Journal of Environmental Engineering*, 129, 755–764, doi:10.1061/(ASCE)0733-9372(2003)129:8(755), [http://dx.doi.org/10.1061/\(ASCE\)0733-9372\(2003\)129:8\(755\)](http://dx.doi.org/10.1061/(ASCE)0733-9372(2003)129:8(755)), 2003.
- 1535 Miyake, Y. and Takeuchi, U.: On the temperature of river waters of Japan, *Japanese Journal of Limnology*, 15, 145–151, 1951.
- Mohseni, O., Stefan, H. G., and Erickson, T. R.: A nonlinear regression model for weekly stream temperatures, *Water Resources Research*, 34, 2685–2692, doi:10.1029/98WR01877, <http://dx.doi.org/10.1029/98WR01877>, 1998.
- 1540 Moore, R., Nelitz, M., and Parkinson, E.: Empirical modelling of maximum weekly average stream temperature in British Columbia, Canada, to support assessment of fish habitat suitability, *Canadian Water Resources Journal*, 38, 135–147, doi:10.1080/07011784.2013.794992, <http://dx.doi.org/10.1080/07011784.2013.794992>, 2013.
- 1545

- Moore, R. D., Spittlehouse, D. L., and Story, A.: Riparian microclimate and stream temperature response to forest harvesting: a review, *JAWRA Journal of the American Water Resources Association*, 41, 813–834, doi:10.1111/j.1752-1688.2005.tb03772.x, <http://dx.doi.org/10.1111/j.1752-1688.2005.tb03772.x>, 2005.
- Müller, V.: Erarbeitung eines anthropogen unbeeinflussten, typischen Jahresgangs der Wassertemperatur nach biozönotischen Regionen, Master's thesis, Swiss Federal Institute of Technology Zurich, 2011.
- 1550 Mutzner, R., Bertuzzo, E., Tarolli, P., Weijjs, S. V., Nicotina, L., Ceola, S., Tomasic, N., Rodriguez-Iturbe, I., Parlange, M. B., and Rinaldo, A.: Geomorphic signatures on Brutsaert base flow recession analysis, *Water Resources Research*, 49, 5462–5472, doi:10.1002/wrcr.20417, <http://dx.doi.org/10.1002/wrcr.20417>, 2013.
- Nash, J. and Sutcliffe, J.: River flow forecasting through conceptual models part I – A discussion of principles, *Journal of Hydrology*, 10, 282–290, doi:[http://dx.doi.org/10.1016/0022-1694\(70\)90255-6](http://dx.doi.org/10.1016/0022-1694(70)90255-6), <http://www.sciencedirect.com/science/article/pii/0022169470902556>, 1970.
- 1555 Nelitz, M. A., MacIsaac, E. A., and Peterman, R. M.: A Science-Based Approach for Identifying Temperature-Sensitive Streams for Rainbow Trout, *North American Journal of Fisheries Management*, 27, 405–424, doi:10.1577/M05-146.1, <http://dx.doi.org/10.1577/M05-146.1>, 2007.
- 1560 Nelson, K. C. and Palmer, M. A.: Stream Temperature Surges Under Urbanization and Climate Change: Data, Models, and Responses, *JAWRA Journal of the American Water Resources Association*, 43, 440–452, doi:10.1111/j.1752-1688.2007.00034.x, <http://dx.doi.org/10.1111/j.1752-1688.2007.00034.x>, 2007.
- Ozaki, N., Fukushima, T., Harasawa, H., Kojiri, T., Kawashima, K., and Ono, M.: Statistical analyses on the effects of air temperature fluctuations on river water qualities, *Hydrological Processes*, 17, 2837–2853, 2003.
- 1565 Polehn, R. A. and Kinsel, W. C.: Transient temperature solution for stream flow from a controlled temperature source, *Water Resources Research*, 33, 261–265, doi:10.1029/96WR03016, <http://dx.doi.org/10.1029/96WR03016>, 1997.
- Polehn, R. A. and Kinsel, W. C.: Transient temperature solution for a river with distributed inflows, *Water Resources Research*, 36, 787–791, doi:10.1029/1999WR900350, <http://dx.doi.org/10.1029/1999WR900350>,
- 1570 2000.
- Pratt, B. and Chang, H.: Effects of land cover, topography, and built structure on seasonal water quality at multiple spatial scales, *Journal of Hazardous Materials*, 209–210, 48–58, doi:10.1016/j.jhazmat.2011.12.068, <http://www.sciencedirect.com/science/article/pii/S0304389412000052>, 2012.
- Risley, J. C., Roehl, E. A., and Conrads, P. A.: Estimating Water Temperatures in Small Streams in Western Oregon Using Neural Network Models, *Water-Resources Investigations Report 02-4218*, U.S. Geological Survey, Portland, Oregon, USA, 2003.
- 1575 Rivers-Moore, N., Mantel, A., and Dallas, H.: Prediction of water temperature metrics using spatial modelling in the Eastern and Western Cape, South Africa, *Water SA*, 32, 167–176, doi:10.4314/wsa.v38i2.2, 2012.
- Roth, T. R., Westhoff, M. C., Huwald, H., Huff, J. A., Rubin, J. F., Barrenetxea, G., Vetterli, M., Parriaux, A., Selker, J. S., and Parlange, M. B.: Stream Temperature Response to Three Riparian Vegetation Scenarios by Use of a Distributed Temperature Validated Model, *Environmental Science & Technology*, 44, 2072–2078, doi:10.1021/es902654f, <http://pubs.acs.org/doi/abs/10.1021/es902654f>, 2010.
- 1580 Rubin, J.-F., Richard, A., Chevalley, P.-A., O'Rourke, J., Ingold, S., Rebetez, M., Huwald, H., Barrenetxea, G., Hass, P., Montavon, R., Travaglini, O., Guisan, A., Frossard, P.-A., Prunier, P., Zouaoui, I., Pusterla, C., Cid,

- 1585 M., Gavillet, G., Kummer, N., and Ciotti, V.: Rapport Clim-arbres, Tech. rep., La Maison de la Rivière and Haute École du Paysage, d'Ingénierie et d'Architecture de Genève, 2012.
- Ruesch, A. S., Torgersen, C. E., Lawler, J. J., Olden, J. D., Peterson, E. E., Volk, C. J., and Lawrence, D. J.: Projected Climate-Induced Habitat Loss for Salmonids in the John Day River Network, Oregon, U.S.A., *Conservation Biology*, 26, 873–882, doi:10.1111/j.1523-1739.2012.01897.x, <http://dx.doi.org/10.1111/j.1523-1739.2012.01897.x>, 2012.
- 1590 Schädler, M.: Die Temperatur schweizerischer Fliessgewässer, Tech. rep., Swiss Federal Office for the Environment, Section Hydrology, Bern, Switzerland, 2008.
- Schaefli, B. and Gupta, H. V.: Do Nash values have value?, *Hydrological Processes*, 21, 2075–2080, doi:10.1002/hyp.6825, <http://dx.doi.org/10.1002/hyp.6825>, 2007.
- 1595 Scott, M. C., Helfman, G. S., McTammany, M. E., Benfield, E. F., and Bolstad, P. V.: Multiscale influences on physical and chemical stream conditions across blue ridge landscapes, *JAWRA Journal of the American Water Resources Association*, 38, 1379–1392, doi:10.1111/j.1752-1688.2002.tb04353.x, <http://dx.doi.org/10.1111/j.1752-1688.2002.tb04353.x>, 2002.
- Segura, C., Caldwell, P., Sun, G., McNulty, S., and Zhang, Y.: A model to predict stream water temperature across the conterminous USA, *Hydrological Processes*, doi:10.1002/hyp.10357, <http://dx.doi.org/10.1002/hyp.10357>, 2014.
- 1600 Sinokrot, B. A. and Stefan, H. G.: Stream temperature dynamics: Measurements and modeling, *Water Resources Research*, 29, 2299–2312, doi:10.1029/93WR00540, <http://dx.doi.org/10.1029/93WR00540>, 1993.
- Sivapalan, M., Takeuchi, K., Franks, S. W., Gupta, V. K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J. J., Mendiondo, E. M., O'Connell, P. E., Oki, T., Pomeroy, J. W., Schertzer, D., Uhlenbrook, S., and Zehe, E.: IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences, *Hydrological Sciences Journal*, 48, 857–880, doi:10.1623/hysj.48.6.857.51421, 10.1623/hysj.48.6.857.51421, 2003.
- 1605 Sponseller, R. A., Benfield, E. F., and Valett, H. M.: Relationships between land use, spatial scale and stream macroinvertebrate communities, *Freshwater Biology*, 46, 1409–1424, doi:10.1046/j.1365-2427.2001.00758.x, <http://dx.doi.org/10.1046/j.1365-2427.2001.00758.x>, 2001.
- Stefan, H. G. and Preud'homme, E. B.: Stream temperature estimation from air temperature, *JAWRA Journal of the American Water Resources Association*, 29, 27–45, doi:10.1111/j.1752-1688.1993.tb01502.x, <http://dx.doi.org/10.1111/j.1752-1688.1993.tb01502.x>, 1993.
- 1615 Tague, C., Farrell, M., Grant, G., Lewis, S., and Rey, S.: Hydrogeologic controls on summer stream temperatures in the McKenzie River basin, Oregon, *Hydrological Processes*, 21, 3288–3300, doi:10.1002/hyp.6538, <http://dx.doi.org/10.1002/hyp.6538>, 2007.
- Theurer, F. D., Voos, K. A., and Miller, W. J.: Instream water temperature model. Instream Flow Information Paper 16, Federal Government Series 16, FWS/OBS-84/15, Fort Collins, Colorado, USA, 1984.
- 1620 Toffolon, M., Siviglia, A., and Zolezzi, G.: Thermal wave dynamics in rivers affected by hydropeaking, *Water Resources Research*, 46, n/a–n/a, doi:10.1029/2009WR008234, <http://dx.doi.org/10.1029/2009WR008234>, 2010.

- van Dijk, A. I. J. M.: Climate and terrain factors explaining streamflow response and recession in Australian catchments, *Hydrology and Earth System Sciences*, 14, 159–169, doi:10.5194/hess-14-159-2010, 1625 <http://www.hydrol-earth-syst-sci.net/14/159/2010/>, 2010.
- Wagenmakers, E.-J. and Farrell, S.: AIC model selection using Akaike weights, *Psychonomic Bulletin & Review*, 11, 192–196, doi:10.3758/BF03206482, <http://dx.doi.org/10.3758/BF03206482>, 2004.
- Webb, B. W., Hannah, D. M., Moore, R. D., Brown, L. E., and Nobilis, F.: Recent advances in stream and river temperature research, *Hydrological Processes*, 22, 902–918, doi:10.1002/hyp.6994, <http://dx.doi.org/10.1002/hyp.6994>, 2008. 1630
- Wehrly, K. E., Brenden, T. O., and Wang, L.: A Comparison of Statistical Approaches for Predicting Stream Temperatures Across Heterogeneous Landscapes¹, *JAWRA Journal of the American Water Resources Association*, 45, 986–997, doi:10.1111/j.1752-1688.2009.00341.x, <http://dx.doi.org/10.1111/j.1752-1688.2009.00341.x>, 2009.
- 1635 Westenbroek, S., Stewart, J., Buchwald, C., Mitro, M., Lyons, J., and Greb, S.: A Model for Evaluating Stream Temperature Response to Climate Change Scenarios in Wisconsin, pp. 1–12, *American Society of Civil Engineers*, doi:10.1061/41143(394)1, <http://ascelibrary.org/doi/abs/10.1061/41143%28394%291>, 2010.
- Westhoff, M. C., Savenije, H. H. G., Luxemburg, W. M. J. ., Stelling, G. S., van de Giesen, N. C., Selker, J. S., Pfister, L., and Uhlenbrook, S.: A distributed stream temperature model using high resolution temperature 1640 observations, *Hydrology and Earth System Sciences*, 11, 1469–1480, doi:10.5194/hess-11-1469-2007, <http://www.hydrol-earth-syst-sci.net/11/1469/2007/>, 2007.
- Young, R. G., Quarterman, A. J., Eyles, R. F., Smith, R. A., and Bowden, W. B.: Water quality and thermal regime of the Motueka River: Influences of land cover, geology and position in the catchment, *New Zealand Journal of Marine and Freshwater Research*, 39, 803–825, doi:10.1080/00288330.2005.9517354, <http://dx.doi.org/10.1080/00288330.2005.9517354>, 2005. 1645

Table 1. List of reviewed publications about statistical stream temperature prediction in ungauged basins.

Reference	Geographic Location	Model Type ^a	Number of Sites	Number of Years ^b	Temporal Scale	Model Precision ^{b,c}
Arscott et al. (2001)	Italy	MLR	22	1	season	$R^2 = 0.37-0.8$
Bogan et al. (2003)	Eastern USA	AE	596	30	week	$R^2 = 0.80, \sigma_e = 3.1^\circ\text{C}$
Chang and Psaris (2013)	Western USA	MLR, GWR	74	n/a	week, year	$R^2 = 0.52-0.62, \sigma_e = 2.0-2.3^\circ\text{C}$
Daigle et al. (2010)	Western Canada	Various	16	0.5	month	$\sigma_e = 0.9-2.8^\circ\text{C}$
DeWeber and Wagner (2014)	Eastern USA	ANN	1080	31	day	$\sigma_e = 1.8-1.9^\circ\text{C}$
Ducharne (2008)	France	MLR	88	7	month	$R^2 = 0.88-0.96, \sigma_e = 1.4-1.9$
Gardner and Sullivan (2004)	Eastern USA	NKM	72	1	day	$\sigma_e = 1.4^\circ\text{C}$
Garner et al. (2014)	UK	CA	88	18	month	n/a
Hawkins et al. (1997)	Western USA	MLR	45	≥ 1	year	$R^2 = 0.45-0.64$
Hill et al. (2013)	Conterminous USA	RF	~ 1000	1/site	season, year	$\sigma_e = 1.1-2.0^\circ\text{C}$
Hrachowitz et al. (2010)	UK	MLR	25	1	month, year	$R^2 = 0.50-0.84$
Imholt et al. (2013)	UK	MLR	23	2	month	$R^2 = 0.63-0.87$
Isaak et al. (2010)	Western USA	MLR, NKM	518	14	month, year	$R^2 = 0.50-0.61, \sigma_e = 2.5-2.8^\circ\text{C}$
Isaak and Hubert (2001)	Western USA	PA	26	1/site	season	$R^2 = 0.82$
Johnson (1971)	New Zealand	ULR	6	1	month	n/a
Johnson et al. (2014)	UK	NLR	36	1.5	day	$R^2 = 0.67-0.90, \sigma_e = 1.0-2.4^\circ\text{C}$
Jones et al. (2006)	Eastern USA	MLR	28	3	year	$R^2 = 0.57-0.73$
Kelleher et al. (2012)	Eastern USA	MLR	47	2	day, week	n/a
Macedo et al. (2013)	Brazil	LMM	12	1.5	day	$R^2 = 0.86$

^a AE: Analytical Expression, ANN: Artificial Neural Network, CA: Cluster Analysis, CRT: Classification and Regression Trees, GWR: Geographically Weighted Regression, LMM: Linear Mixed Model, MLR: Multi-Linear Regression, NKM: Networked Kriging Model, NLR: Non-linear Regression, PA: Path Analysis, RF: Random Forest, ULR: Univariate Linear Regression.

^b n/a: not available.

^c σ_e : root-mean-square error, R^2 : coefficient of determination (sometimes referred to as Nash-Sutcliffe index).

Table 1. (continued)

Reference	Geographic Location	Model Type ^a	Number of Sites	Number of Years ^b	Temporal Scale	Model Precision ^{b,c}
Mayer (2012)	Western USA	MLR	104	≥ 2	week, month	$R^2 = 0.72$, $\sigma_e = 1.8^\circ\text{C}$
Miyake and Takeuchi (1951)	Japan	ULR	20	n/a	month	n/a
Moore et al. (2013)	Western Canada	MLR	418	1/site	year	$\sigma_e = 2.1^\circ\text{C}$
Nelitz et al. (2007)	Western Canada	CRT	104	1/site	year	n/a
Nelson and Palmer (2007)	Western USA	MLR	16	3	season	$R^2 = 0.36$ – 0.88
Ozaki et al. (2003)	Japan	ULR	5	8	day	n/a
Pratt and Chang (2012)	Western USA	MLR, GWR	51	1/site	season	$R^2 = 0.48$ – 0.78
Risley et al. (2003)	Western USA	ANN	148	0.25	hour, season	$\sigma_e = 1.6$ – 1.8°C
Rivers-Moore et al. (2012)	South Africa	MLR	90	1/site	month, year	$R^2 = 0.14$ – 0.50
Ruesch et al. (2012)	Western USA	NKM	165	15	year	$R^2 = 0.84$, $\sigma_e = 1.5^\circ\text{C}$
Segura et al. (2014)	Conterminous USA	MLR	171	≥ 1.5	week, month	$R^2 = 0.79$
Sponseller et al. (2001)	Eastern USA	MLR	9	1	year	$R^2 = 0.81$ – 0.93
Scott et al. (2002)	Eastern USA	MLR	36	1/site	season	$R^2 = 0.82$
Stefan and Preud'homme (1993)	Eastern USA	ULR	11	n/a	day, week	$\sigma_e = 2.1$ – 2.7°C
Tague et al. (2007)	Western USA	MLR	43	4	day	$R^2 = 0.49$ – 0.65
Wehrly et al. (2009)	Eastern USA	Various	1131	1/site	month	$\sigma_e = 2.0$ – 3.0°C
Westenbroek et al. (2010)	Eastern USA	ANN	254	1/site	day	$R^2 = 0.70$, $\sigma_e = 1.8^\circ\text{C}$
Young et al. (2005)	New Zealand	MLR	23	1	season	$R^2 = 0.75$ – 0.93

^a AE: Analytical Expression, ANN: Artificial Neural Network, CA: Cluster Analysis, CRT: Classification and Regression Trees, GWR: Geographically Weighted Regression, LMM: Linear Mixed Model, MLR: Multi-Linear Regression, NKM: Networked Kriging Model, NLR: Non-linear Regression, PA: Path Analysis, RF: Random Forest, ULR: Univariate Linear Regression.

^b n/a: not available.

^c σ_e : root-mean-square error, R^2 : coefficient of determination (sometimes referred to as Nash-Sutcliffe index).

Table 2. Calibration parameters of the ~~physically-based~~ physics-inspired statistical model.

Parameter	Defined in	Units	Calibration Range	Physical Constraints
a_w	Eqs. (19)–(20)	$[\text{°C °C}^{-1}]$	Chosen so as to constrain $T_{s,i}$ and T_ℓ to the range $0\text{--}25\text{°C}$	Must be positive
b_w		$[\text{°C}]$		None
$a_{\phi,sw}$ $a_{\phi,sw}$	Eq. (17)	$[\text{°C m}^2 \text{W}^{-1}]$	Chosen so as to constrain $\gamma\phi_r$ to the range $-20\text{--}20\text{°C}$	Must be positive
$a_{\phi,s}$		$[\text{°C}]$		Must be negative
$a_{\phi,\theta}$		$[\text{°C}]$		Must be negative
$a_{\phi,f}$		$[\text{°C}]$		None
b_ϕ		$[\text{°C}]$		None
η	Eqs. (9)–(10)	$[-]$	$0\text{--}1$	None
A_s	Eq. (18)	$[\text{m}^2]$	Chosen so as to constrain η in the range $0\text{--}1$	Must be positive
α		$[-]$	$0\text{--}3$	None

Table 3. Physiographic properties of the 29 selected hydrological catchments in Switzerland. The three watersheds indicated in bold are not used for the model evaluation.

Basin Number	Name	Basin Area [km ²]	Gauging Station Altitude [m]	Mean Basin Altitude [m]	Glacier Cover [%]	Hydrological Regime ^a	Temperature Measurement Period	Data Provider ^b
1	Broye at Payerne	392.0	441	710	0	pluvial inferior	1976–2012	FOEN
2	Sitter at Appenzell	74.2	769	1252	0.1	transition nival	2006–2012	FOEN
3	Murg at Wängi	78.0	466	650	0	pluvial inferior	2002–2012	FOEN
4	Gürbe at Belp, Mülimatt	117.0	522	849	0	transition pluvial	2007–2012	FOEN
5	Massa at Blatten, Naters	195.0	1446	2945	65.9	glacial	2003–2012	FOEN
6	Sense at Thörishaus, Sensematt	352.0	553	1068	0	pre-alpine nivo-pluvial	2004–2012	FOEN
7	Allenbach at Adelboden	28.8	1297	1856	0	alpine nival	2002–2012	FOEN
8	Rosegbach at Pontresina	66.5	1766	2716	30.1	glacial	2004–2012	FOEN
9	Grosstalbach at Isenthal	43.9	767	1820	9.3	alpine nival	2005–2012	FOEN
10	Goldach at Goldach, Bleiche	49.8	399	833	0	pluvial superior	2005–2012	FOEN
11	Dischmabach at Davos, Kriegsmatte	43.3	1668	2372	2.1	glacio-nival	2004–2012	FOEN
12	Langeten at Huttwil, Häberenbad	59.9	597	766	0	pluvial inferior	2002–2012	FOEN
13	Riale di Roggiasca at Roveredo	8.1	980	1710	0	meridional nivo-pluvial	2003–2012	FOEN
14	Riale di Calneggia at Caverigno, Pontit	24	890	1996	3.0	meridional nival	2002–2012	FOEN
15	Poschiavino at La Rösa	14.1	1860	2283	0.4	meridional nival	2004–2012	FOEN
16	Mentue at Yvonand, La Mauguettaz	105.0	449	679	0	jurassian pluvial	2003–2012	FOEN
17	Necker at Mogelsberg, Aachsäge	88.2	606	959	0	pluvial superior	2007–2012	FOEN
18	Grossbach at Gross, Säge	9.1	940	1276	0	pre-alpine nivo-pluvial	2003–2012	FOEN
19	Rietholzbach at Mosnang, Rietholz	3.3	682	795	0	pluvial superior	2002–2012	FOEN
20	Gürbe at Burgistein, Pfandersmatt	53.7	569	1044	0	pre-alpine nivo-pluvial	2007–2008	FOEN
21	Biber at Biberbrugg	31.9	825	1009	0	pluvial superior	2003–2012	FOEN
22	Sellenbodenbach at Neuenkirch	10.5	515	615	0	pluvial superior	2003–2012	FOEN
23	Alp at Einsiedeln	46.4	840	1155	0	transition pluvial	2003–2012	FOEN
24	Riale di Pincascia at Lavertezzo	44.4	536	1708	0	meridional nivo-pluvial	2004–2012	FOEN
25	Rom at Müstair	129.7	1236	2187	0.1	meridional nival	2003–2012	FOEN
26	Sissle at Eiken	123.0	314	529	0	jurassian pluvial	2004–2012	Aargau
27	Bünz at Othmarsingen	110.6	390	526	0	pluvial inferior	2005–2012	Aargau
28	Wyna at Unterkulm	92.1	455	643	0	pluvial inferior	2005–2012	Aargau
29	Talbach at Schinznach-Dorf	14.5	358	559	0	jurassian pluvial	2009–2012	Aargau

^a According to the classification by Aschwanden and Weingartner (1985).^b FOEN: Swiss Federal Office for the Environment, Aargau: Department for Construction, Transport and Environment of Canton Aargau.

Table 4. Formulations of the ~~physically-based~~ physics-inspired statistical model selected in each season based on their corresponding AICc value. The Akaike weights are denoted as w_i . Only the model formulations with $w_i \geq 0.1$ are presented here. The widths of the buffers used to compute the riparian vegetation cover are indicated as subscripts of the variable f_r (the indicated values correspond to the total buffer widths, i.e. accounting for both sides of the stream center line).

Season	w_i	L_c [km]	Formulation of $\gamma\phi_r$	Formulation of T_s and T_ℓ	Formulation of η
Winter	1	8	$\gamma\phi_r = a_{\phi,s}f_s + a_{\phi,f}f_{r,100m} + b_\phi$, with $a_{\phi,s} = 18.2^\circ\text{C}$, $a_{\phi,f} = -2.9^\circ\text{C}$ and $b_\phi = -12.1^\circ\text{C}$	$T_s = a_w T_a + b_w$, with $a_w = 0.15$ and $b_w = 3.1^\circ\text{C}$	$\eta = 0$ (constant)
	0.25		$\gamma\phi_r = \alpha_{\phi,s}f_s + b_\phi$, with $a_{\phi,s} = 19.4^\circ\text{C}$ and $b_\phi = -14.1^\circ\text{C}$		
	0.15		$\gamma\phi_r = a_{\phi,s}f_s + a_{\phi,f}f_{r,50m} + b_\phi$, with $a_{\phi,s} = 19.3^\circ\text{C}$, $a_{\phi,f} = -0.4^\circ\text{C}$ and $b_\phi = -13.9^\circ\text{C}$		
Spring	1	4	$\gamma\phi_r = a_{\phi,s}f_s + a_{\phi,f}f_{r,50m} + b_\phi$, with $a_{\phi,s} = 12.9^\circ\text{C}$, $a_{\phi,f} = -3.7^\circ\text{C}$ and $b_\phi = -11.3^\circ\text{C}$	$T_s = a_w T_a + b_w$, with $a_w = 0.27$ and $b_w = 5.3^\circ\text{C}$	$\eta = 1 - n_s(A_s/A_{\text{tot}})^\alpha$, with $A_s = 0.13 \text{ km}^2$ and $\alpha = 1$
	0.86		$\gamma\phi_r = a_{\phi,s}f_s + a_{\phi,f}f_{r,200m} + b_\phi$, with $a_{\phi,s} = 13.1^\circ\text{C}$, $a_{\phi,f} = -3.4^\circ\text{C}$ and $b_\phi = -11.6^\circ\text{C}$		
	0.53		$\gamma\phi_r = a_{\phi,s}f_s + b_\phi$, with $a_{\phi,s} = 13.1^\circ\text{C}$ and $b_\phi = -12.7^\circ\text{C}$		
	0.31		$\gamma\phi_r = a_{\phi,s}f_s + a_{\phi,f}f_{r,100m} + b_\phi$, with $a_{\phi,s} = 13.4^\circ\text{C}$, $a_{\phi,f} = -2.2^\circ\text{C}$ and $b_\phi = -12.3^\circ\text{C}$		
Summer	1	4	$\gamma\phi_r = a_{\phi,s}f_s + a_{\phi,f}f_{r,100m} + b_\phi$, with $a_{\phi,s} = 13.4^\circ\text{C}$, $a_{\phi,f} = -0.3^\circ\text{C}$ and $b_\phi = -13.0^\circ\text{C}$	$T_s = a_w T_a + b_w$, with $a_w = 0.33$ and $b_w = 6.6^\circ\text{C}$	$\eta = 1$ (constant)
	0.15		$\gamma\phi_r = a_{\phi,s}f_s + b_\phi$, with $a_{\phi,s} = 13.4^\circ\text{C}$ and $b_\phi = -13.1^\circ\text{C}$		
Autumn	1	4	$\gamma\phi_r = a_{\phi,s}f_s + b_\phi$, with $a_{\phi,s} = 10.4^\circ\text{C}$ and $b_\phi = -5.9^\circ\text{C}$	$T_s = a_w T_a + b_w$, with $a_w = 0.25$ and $b_w = 5.1^\circ\text{C}$	$\eta = 1$ (constant)
	0.54		$\gamma\phi_r = a_{\phi,s}f_s + a_{\phi,f}f_{r,100m} + b_\phi$, with $a_{\phi,s} = 10.0^\circ\text{C}$, $a_{\phi,f} = -2.8^\circ\text{C}$ and $b_\phi = -4.7^\circ\text{C}$		
	0.48		$\gamma\phi_r = a_{\phi,s}f_s + a_{\phi,f}f_{r,200m} + b_\phi$, with $a_{\phi,s} = 10.1^\circ\text{C}$, $a_{\phi,f} = -2.9^\circ\text{C}$ and $b_\phi = -4.7^\circ\text{C}$		
	0.4		$\gamma\phi_r = a_{\phi,s}f_s + a_{\phi,f}f_{r,50m} + b_\phi$, with $a_{\phi,s} = 10.2^\circ\text{C}$, $a_{\phi,f} = -2.2^\circ\text{C}$ and $b_\phi = -5.1^\circ\text{C}$		

Table 5. Performance of the best ~~physically-based~~ physics-inspired statistical model in each season ($w_i = 1$) in terms of RMSE, R^2 and bias, depending on the validation set. The numbers between brackets refer to the different validation sets (see beginning of Sect. 4).

Season	RMSE [$^{\circ}\text{C}$]				R^2 [-]				Bias [$^{\circ}\text{C}$]			
	1	2	3	4	1	2	3	4	1	2	3	4
Winter	1.34	1.34	1.58	1.38	0.68	0.40	0.52	0.55	-0.58	-0.30	-0.84	-0.47
Spring	1.51	1.29	1.57	1.40	0.87	0.90	0.86	0.88	0.24	0.02	0.36	0.14
Summer	1.07	1.47	1.13	1.31	0.91	0.84	0.90	0.87	-0.04	0.01	-0.01	-0.01
Autumn	1.16	1.22	1.47	1.25	0.89	0.87	0.84	0.87	-0.49	0.15	-0.34	-0.11
All Year	1.28	1.33	1.45	1.34	0.94	0.94	0.93	0.94	-0.22	-0.03	-0.21	-0.11

Table 6. Best multi-linear regression models for the prediction of annual mean \bar{T}_i and standard deviation σ_i of the monthly mean stream temperature in a given year. All predictor variables are averaged over the stream networks using the operator $\langle \cdot \rangle_{\mathcal{L}}$.

Predictand	Predictors ^a (with coefficients)	L_c [km]
\bar{T}_i [$^{\circ}\text{C}$]	$f_{i,25m}$ (-1.86°C), θ (0.60°C), A_{tot} ($1.6 \times 10^{-3} \text{C km}^{-2}$), Δf_s (-4.90°C), $\bar{T}_{a,i}$ ($0.75^{\circ}\text{C } ^{\circ}\text{C}^{-1}$), intercept (3.88°C)	4
σ_i [$^{\circ}\text{C}$]	$ dz/dx $ (6.03°C), $f_{i,25m}$ (6.70°C), θ (0.93°C), n_s ($1.6 \times 10^{-3} \text{C}$), Δf_s (-12.8°C), $\sigma_{a,i}$ ($0.39^{\circ}\text{C } ^{\circ}\text{C}^{-1}$), intercept (0.34°C)	32

^a Δf_s denotes the difference in topographical shading between summer and winter, $\bar{T}_{a,i}$ and $\sigma_{a,i}$ refer to the annual mean value and standard deviation of air temperature in the year of interest, respectively, A_{tot} denotes the watershed area and $|dz/dx|$ the channel slope. The other variables have been defined in the text.

Table 7. Performance of the standard regression model in terms of RMSE, R^2 and bias computed over the validation set 4 in each season. The stream network used to evaluate the model corresponds to the original one derived from the map at scale 1:25'000.

Season	RMSE [$^{\circ}\text{C}$]	R^2	Bias [$^{\circ}\text{C}$]
Winter	1.18	0.67	-0.27
Spring	1.06	0.93	-0.09
Summer	1.18	0.90	0.11
Autumn	1.03	0.91	0.02
All Year	1.12	0.96	-0.06

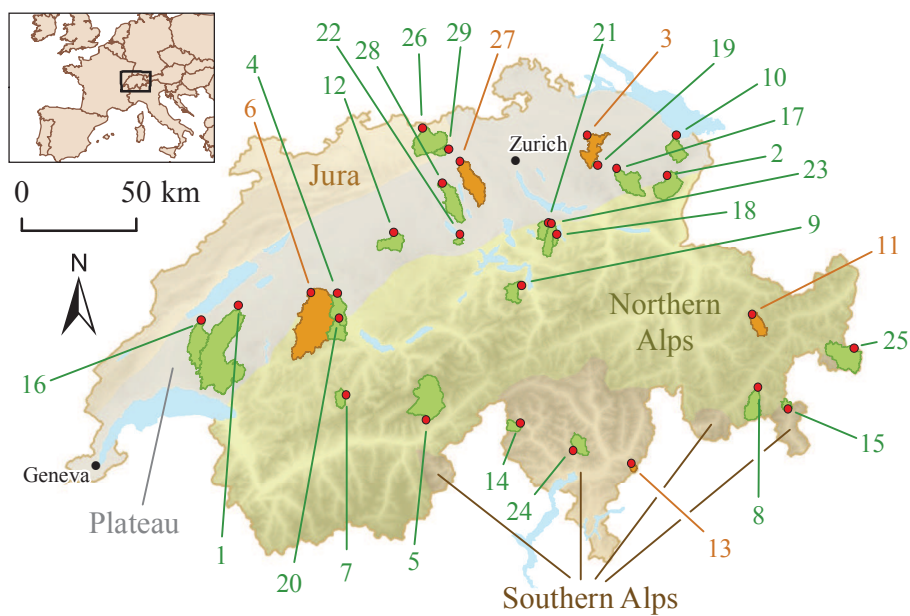


Figure 1. Locations of the gauging stations selected for the evaluation of the physically-based physics-inspired and regression-standard statistical models. The stations are displayed as red points and their associated catchments as green or orange areas, depending on whether they are used to calibrate or validate the model. The four main climatic regions of Switzerland—Jura mountains, Plateau, Northern Alps and Southern Alps—are displayed in different colors. The numbering corresponds to Table 3.

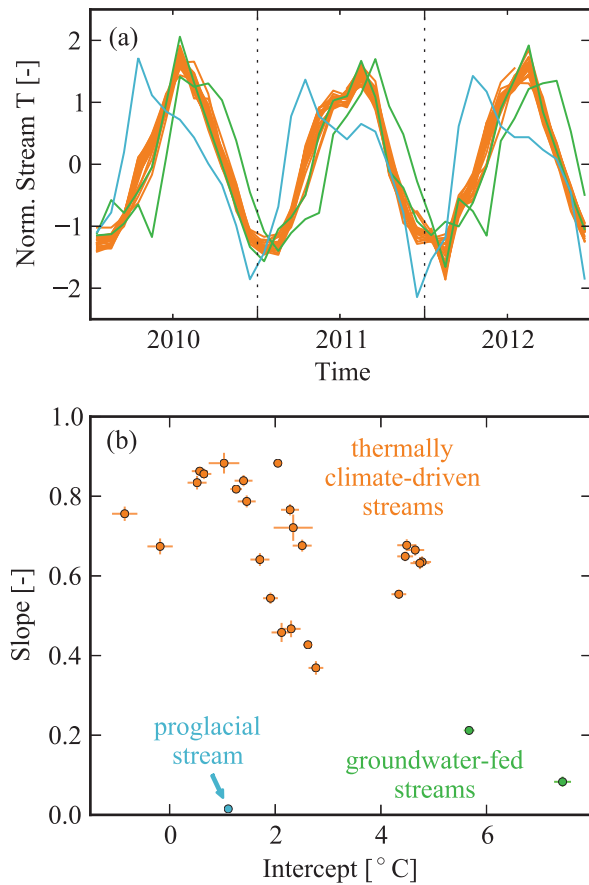


Figure 2. Classification of the thermal regimes of the selected catchments. Streams impacted by groundwater infiltration are shown in green, the proglacial stream in blue and the thermally climate-driven streams in orange. (a) Normalized monthly mean stream temperature curves over 3 consecutive years (2010–2012); all curves are z -scored independently each year. (b) Slopes and intercepts of the regression lines fitted to the stream–air temperature points of the respective catchments. All points with negative air temperature values have been discarded prior to fitting. The bars indicate the standard error estimates.

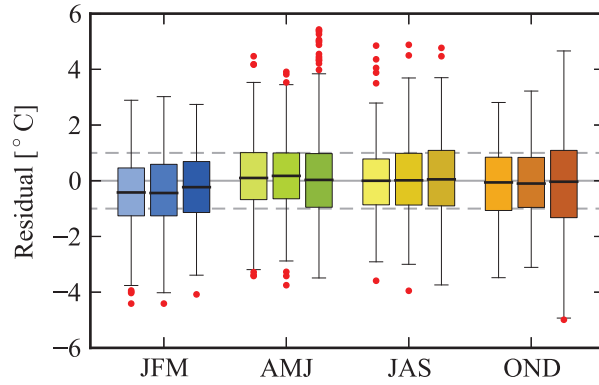


Figure 3. Prediction error of the physically-based-~~physics-inspired~~ statistical model for different resolutions of the stream network. The boxes extend from the first to the third quartile of the error distribution. Outliers are displayed as red dots. In each season, the network resolution decreases from left to right: the left box corresponds to the network with all stream reaches, whereas the central and right boxes contain only the stream segments whose Strahler order is greater or equal to 2 and 3, respectively. The error values 0, -1 and $+1^{\circ}\text{C}$ are displayed as a solid grey line and two dashed grey lines.

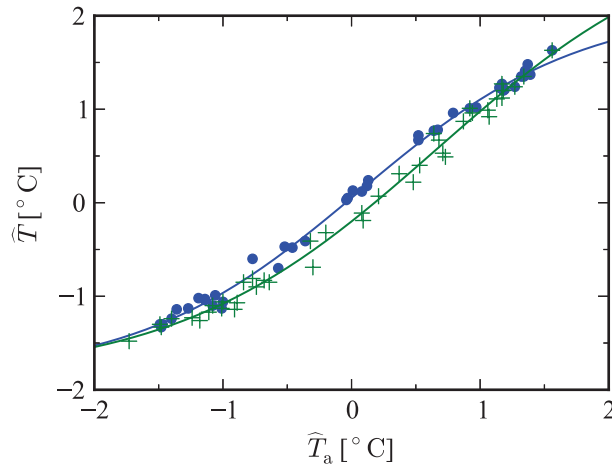


Figure 4. Non-linear relationship between the z-scored stream temperature \hat{T} and the z-scored air temperature \hat{T}_a averaged over 14 reference meteorological stations. The values of \hat{T} are obtained by averaging in each month the z-scored stream temperatures measured at the 21 calibration stations. Each point corresponds to a single month of the calibration period 2007–2012. Months from January to July are displayed as green crosses, and those from July to December as blue dots. The two solid lines correspond to the respective fits of the data points in the two year halves (see Eqs. (27)–(28)).

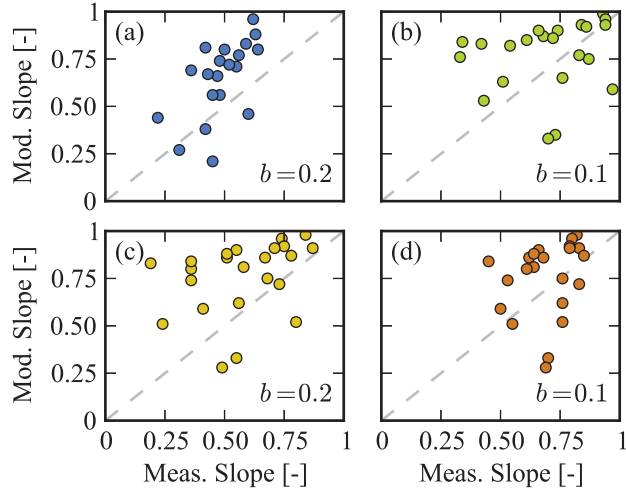


Figure 5. Comparison of modeled against measured slopes of the regression line between stream and air temperatures. The panels correspond to the different seasons: (a) January–March, (b) April–June, (c) July–September, and (d) October–December. The bias b corresponds to the average, in each season, of the difference between the modeled and measured regression slopes over all the selected stations and years (i.e. belonging to both the calibration set and validation set 4). The 1:1 line is indicated as a dashed grey curve.

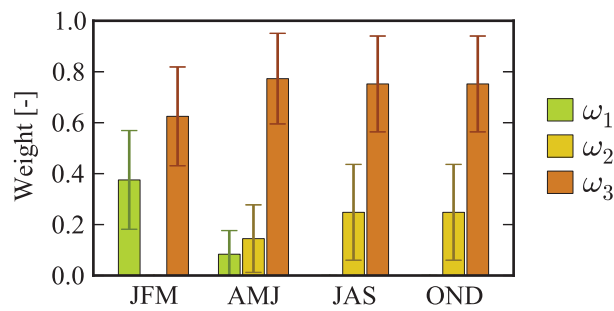


Figure 6. Seasonal values of the factors ω_1 , ω_2 and ω_3 weighting the different terms in Eq. (8). The values of these weights are evaluated over the entire data set, i.e. both the calibration set and validation set 3. The error bars indicate the confidence interval centered around the mean and extending over one standard deviation on each side.

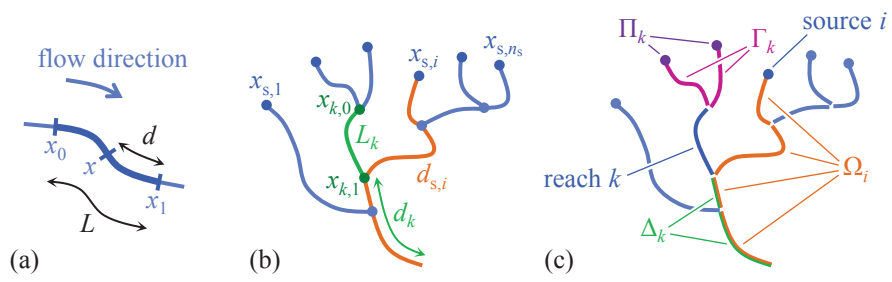


Figure 7. Schematic representations of (a) a stream reach, and (b)–(c) a stream network, illustrating the notation used in Appendix A to derive the analytical solution to the stream energy-balance.