ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/302482139

Prediction of river water temperature: a comparison between a new family of hybrid models and statistical approaches...

Article *in* Hydrological Processes · May 2016

DOI: 10.1002/hyp.10913

CITATION
1

READS
28

6 authors, including:


Sebastiano Piccolroaz
Utrecht University
21 PUBLICATIONS **61** CITATIONS

SEE PROFILE


Bruno Majone
Università degli Studi di Trento
38 PUBLICATIONS **255** CITATIONS

SEE PROFILE


Annunziato Siviglia
ETH Zurich
49 PUBLICATIONS **525** CITATIONS

SEE PROFILE


Marco Toffolon
Università degli Studi di Trento
75 PUBLICATIONS **471** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project

Managing the effects of multiple stressors on aquatic ecosystems under water scarcity - GLOBAQUA
View project

All content following this page was uploaded by Bruno Majone on 27 October 2016.

The user has requested enhancement of the downloaded file. All in-text references underlined in blue are added to the original document and are linked to publications on ResearchGate, letting you access and read them immediately.

HYDROLOGICAL PROCESSES
Hydrol. Process. **30**, 3901–3917 (2016)
Published online 21 June 2016 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/hyp.10913

Prediction of river water temperature: a comparison between a new family of hybrid models and statistical approaches

Sebastiano Piccolroaz,[1]* Elisa Calamita,[1] Bruno Majone,[1] Aurélien Gallice,[2] Annunziato Siviglia[3] and Marco Toffolon[1]

[1] *Department of Civil, Environmental and Mechanical Engineering, University of Trento, via Mesiano 77, I-38123 Trento, Italy*
[2] *School of Architecture, Civil and Environmental Engineering, École Polytechnique Fédérale de Lausanne, Station 2, CH-1015 Lausanne, Switzerland*
[3] *Laboratory of Hydraulics, Hydrology and Glaciology (VAW), ETH-Zürich, HIA C 58 Hönggerbergring 26, CH-8093 Zürich, Switzerland*

Abstract:

River water temperature is a key physical variable controlling several chemical, biological and ecological processes. Its reliable prediction is a main issue in many environmental applications, which however is hampered by data scarcity, when using data-demanding deterministic models, and modelling limitations, when using simpler statistical models. In this work we test a suite of models belonging to *air2stream* family, which are characterized by a hybrid formulation that combines a physical derivation of the key equation with a stochastic calibration of parameters. The *air2stream* models rely solely on air temperature and streamflow, and are of similar complexity as standard statistical models.

The performances of the different versions of *air2stream* in predicting river water temperature are compared with those of the most common statistical models typically used in the literature. To this aim, a dataset of 38 Swiss rivers is used, which includes rivers classified into four different categories according to their hydrological characteristics: low-land natural rivers, lake outlets, snow-fed rivers and regulated rivers. The results of the analysis provide practical indications regarding the type of model that is most suitable to simulate river water temperature across different time scales (from daily to seasonal) and for different hydrological regimes. A model intercomparison exercise suggests that the family of *air2stream* hybrid models generally outperforms statistical models, while cross-validation conducted over a 30-year period indicates that they can be suitably adopted for long-term analyses. Copyright © 2016 John Wiley & Sons, Ltd.

KEY WORDS thermal regime; heat budget; hybrid models; thermal response; hydrological regime; air temperature; stream temperature; statistical models

Received 11 November 2015; Accepted 4 May 2016

INTRODUCTION

Stream ecosystem functioning is mainly controlled by River Water Temperature (RWT) and streamflow (Smakhtin, 2001; Caissie, 2006). These variables exert a strong control on biological and chemical quality of water (Ducharne, 2008), chemical reaction rates and oxygen solubility (Webb *et al.*, 2008), and instream habitats (Benda *et al.*, 2004). The complexity of river hydrological and thermal response depends on both anthropogenic (e.g. pollution, water withdrawals, deforestation, etc.) and climate changes as well as on their feedback mechanisms (Arismendi *et al.*, 2012; Destouni *et al.*, 2013). Joint simulation of these processes is, however, a challenging task, because of the complexity and space-time variability of input meteorological forcing, heat flux exchanges and transport processes in soils and channel networks.

The existing modelling approaches for RWT prediction fall into two main categories: deterministic models and statistical models. Deterministic, or process-based, models generally require a mathematical representation of both hydrological processes occurring in the catchment, and heat exchange processes between the river and the surrounding environment. Most of these models are limited to the processes in the river's surface waters (e.g. Sinokrot and Stefan, 1993; Kim and Chapra, 1997; Younus *et al.*, 2000; Siviglia and Toro, 2009) and rely on suitable boundary conditions for the hydrological inputs. However there are also some recent attempts to jointly simulate hydrologic and thermal regimes (e.g. Comola *et al.*, 2015). In general, deterministic models have the advantage of potentially being applied at different spatial scales and offer the possibility to predict RWT at the locations where observed data are not available (Caissie, 2006). On the other hand, they generally require a large

*Correspondence to: Sebastiano Piccolroaz, Department of Civil, Environmental and Mechanical Engineering, University of Trento, via Mesiano 77, I-38123 Trento, Italy.
E-mail: s.piccolroaz@unitn.it

Copyright © 2016 John Wiley & Sons, Ltd.

amount of observational data (e.g. catchment characteristics and geometry, full set of meteorological variables, hydraulic properties of the river), which makes them relatively complex. Their effective use is further hampered by data scarcity, which still represents one of the main concerns for many environmental applications.

These limitations have favoured the wide use of simpler statistical regression-based models, to an extent that these models are often preferred in both practical applications and scientific research. The use of this kind of models is generally simpler, with most of the regression models considering air temperature as the only predictor in linear or logistic relationships (e.g. Kothandaraman, 1972; Mohseni and Stefan, 1999; Morrill *et al*., 2005, among the others; see also Benyahya *et al*., 2007 for a review). More complex statistical models fall in the class of stochastic models, which account also for the autocorrelation structure of RWT time series (e.g. Caissie *et al*., 2001). The application of statistical models can be controversial, especially when the physical processes controlling the heat exchanges are complex. In these cases, RWT is unlikely to be reproduced appropriately by simple regression relationships forced only by air temperature (Webb *et al*., 2003, 2008) and without explicitly accounting for the thermal inertia of water (Stefan and Preud'homme, 1993). In addition, the inherent disconnection between model's structure and thermo-physical dynamics does not allow for reliable extrapolations beyond the range of calibration conditions, making their use questionable in studies addressing the likely impacts of climate change (e.g. Benyahya *et al*., 2007). This is particularly relevant given the great interest that has been devoted in the last years to study the effects of climate change on RWT (e.g. Morrill *et al*., 2005; Matulla *et al*., 2007; Kurylyk *et al*., 2013), especially in the Alpine region (e.g. Brown *et al*., 2005; Brown and Hannah, 2007).

In light of the practical and technical limitations discussed above, it appears that it is not simple to find models able to simultaneously conjugate simplicity and parsimony, limited computational demand and the ability to provide reliable long-term estimates. Here we test different versions of *air2stream* (Toffolon and Piccolroaz, 2015), a simple lumped model that can be classified into a third alternative family of hybrid, statistical-based semi-empirical models. The model has been developed with the aim to retain the limited data requirements of statistical models, while preserving the intimate structure of process-based models. Starting from first principles of physics and introducing opportune simplifications, the model combines a physically based structure with a stochastic calibration of the parameters. In this way, the information contained in the observational data is directly transferred, through calibration, to model parameters (for a similar approach, see Gallice *et al*., 2015). Parameter values can then provide significant hints on the functioning of the system under investigation.

In this work we investigate and compare the performance of different versions of *air2stream* and of a selection of the most common regression-based and stochastic RWT models of similar complexity. For this purpose, a long-term dataset of 38 Swiss gauging stations is used, which includes rivers belonging to four different hydrological classes: low-land natural rivers, lake outlets, snow-fed rivers and regulated (i.e. affected by significant hydropower releases) rivers. The main objectives of the analysis can be summarized in the following three points:

- to introduce an alternative classification that distinguishes between thermally reactive and thermally resilient rivers on the basis of the slope of the hysteresis cycle between observed water and air temperatures, revisiting the existing classifications (e.g. Kelleher *et al*., 2012 and Mayer, 2012);
- to evaluate the performance of the different models used to predict RWT of rivers with different hydrological and thermal regimes, in particular at different time scales;
- to assess the robustness of the different models when adopted for projections of RWT in periods different from that of calibration.

As a final aim of the work, the existing lack of proper guidelines about RWT modelling is addressed by providing best practices for the choice of the most appropriate models (in terms of reliability and simplicity/parsimony) depending on the hydrological characteristics of the river, the time scale of interest and the scope of the analysis. These guidelines mainly refer to those cases in which data availability is the main constraint, while in other contexts deterministic models maintain their undeniable relevance. We believe that this analysis is a timely and useful contribution to the literature, in that it may help end-users in preventing from the use of simplistic and often inappropriate models for RWT prediction. This is even more significant considering the growing interest on themes like climate change, rising temperatures and extreme events.

MATERIALS AND METHODS

Swiss rivers database

This study is based on streamflow and temperature measurements collected by the Swiss Federal Office for the Environment (FOEN). The available database regroups data from a network of more than 180 water level and discharge gauging stations in Switzerland, among which more than 70 stations provide continuous stream temper-

Copyright © 2016 John Wiley & Sons, Ltd.

ature records, in some cases extending back to the early ’70s (see, e.g. Gallice *et al.*, 2015). In order to test models performances, 38 gauging stations have been selected in the present study (see Table I and Figure 1), which presented the longest and almost uninterrupted daily records of water temperature and streamflow. For each river, air temperature measurements have been retrieved from the closest meteorological station (see Table S1 in the Supporting Information).

The gauging stations examined in this work have been classified into four groups on the basis of available GIS information (https://map.geo.admin.ch) and expert judgement. The classification allowed identifying four hydrological categories, depending on different landscape and anthropic characteristics of the rivers’ catchments:

- *Low-land natural* rivers flowing at low altitudes (approximately in the range 400–800 m a.s.l., corresponding to the Swiss plateau) through catchments affected by low to moderate anthropic pressures (17 rivers).
- *Lake outlets*, which include all rivers that are downstream of a lake (irrespective of the distance between lake outlet and measurement station) and are characterized by low anthropic thermal alterations (eight rivers).
- *Snow-fed* rivers located in the mountainous part of Switzerland and originating from the melting of glaciers and snowfields (eight rivers).
- *Regulated* rivers characterized by the presence of significant upstream releases from high-elevation

Table I. List of river gauging stations with length of time series and the adopted river classification (see Section on ‘Swiss Rivers Database’).

	ID	River	Reach	Time series	Hydrological classification
1	2369	Mentue	Yvonand, La Mauguettaz	2002–2012	Natural low-land
2	2232	Allenbach	Adelboden	2002–2012	Snow-fed
3	2179	Sense	Thörishaus, Sensematt	2003–2013	Natural low-land
4	2159	Gürbe	Belp, Mülimatt	2006–2012	Natural low-land
5	2608	Sellenbodenbach	Neuenkirch	2002–2012	Natural low-land
6	2308	Goldach	Goldach, Bleiche, (H)	2005–2012	Natural low-land
7	2327	Dischmabach	Davos, Kriegsmatte	2003–2012	Snow-fed
8	2034	Broye	Payerne, Caserne D’aviation	1981–2012	Natural low-land
9	2112	Sitter	Appenzell	2005–2012	Natural low-land
10	2126	Murg	Wängi	2002–2012	Natural low-land
11	2161	Massa	Blatten Bei Naters	2004–2012	Snow-fed
12	2256	Rosegbach	Pontresina	2003–2012	Snow-fed
13	2343	Langeten	Huttwil, Häberenbad	2002–2012	Natural low-land
14	2366	Poschiavino	La Rösa	2004–2012	Snow-fed
15	2374	Necker	Mogelsberg, Aachsäge	2007–2012	Natural low-land
16	2414	Rietholzbach	Mosnang, Rietholz	2002–2012	Natural low-land
17	2609	Alp	Einsiedeln	2002–2012	Natural low-land
18	2612	Riale Di Pincascia	Lavertezzo	2003–2012	Natural low-land
19	2617	Rom	Müstair	2002–2012	Snow-fed
20	2276	Grosstalbach	Isenthal	2004–2012	Snow-fed
21	2347	Riale di Roggiasca	Roveredo	2009–2012	Natural low-land
22	2009	Rhone	Porte Du Scex	1984–2013	Regulated
23	2011	Rhone	Sion	1984–2013	Regulated
24	2016	Aare	Brugg, Aegerten	1984–2013	Outlet
25	2019	Aare	Brienzwiler	1984–2013	Regulated
26	2029	Aare	Brugg, Aegerten	1984–2013	Outlet
27	2030	Aare	Thun	1984–2013	Outlet
28	2044	Thur	Andelfinger	1984–2013	Natural low-land
29	2056	Reuss	Seedorf	1984–2013	Regulated
30	2070	Emme	Emmenmatt	1984–2013	Natural low-land
31	2085	Aare	Hagneck	1984–2013	Outlet
32	2091	Rhein	Rheinfelden	1984–2013	Outlet
33	2135	Aare	Bern, Schonau	1984–2013	Outlet
34	2143	Rhein	Rekingen	1984–2013	Outlet
35	2152	Reuss	Luzern, Geissmattbrucke	1984–2013	Outlet
36	2372	Linth	Mollis, Linthbrucke	1984–2013	Regulated
37	2415	Glatt	Rheinsfelden	1984–2013	Natural low-land
38	2457	Aare	Ringgenberg, Goldswil	1984–2013	Snow-fed

Copyright © 2016 John Wiley & Sons, Ltd.

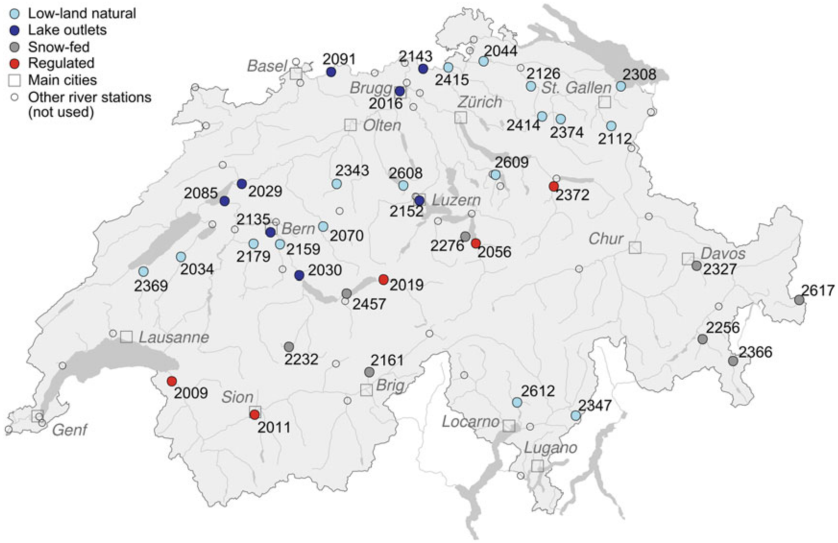

Figure 1. Map of Switzerland and location of river water temperature and streamflow measurement stations. Colours refer to the hydrological classification introduced in the Section on ‘Swiss Rivers Database’

storage hydropower reservoirs, thus affected by strong hydro- and thermo-peaking at various time scales (five rivers).

Models

Hybrid model air2stream. The first set of models is constituted by different versions of *air2stream*, a recently proposed semi-empirical, hybrid model that combines a physically based structure with a stochastic calibration of parameters (Toffolon and Piccolroaz, 2015; source code available at https://github.com/marcotoffolon/air2stream). The fundamental equation is the heat budget applied to an undefined volume V of the river reach (Figure 2), which includes the net heat exchange with the atmosphere, H, and the contribution of upstream tributaries and, possibly, groundwater:

$$\rho c_p V \frac{dT_w}{dt} = AH + \rho c_p \left(\textstyle\sum_i Q_i T_{w,i} - QT_w \right), \qquad (1)$$

where T_w is the daily water temperature (i.e. RWT), t is time, ρ and c_p are water density and specific heat at constant pressure, A is the surface area of the river reach, Q is the streamflow at the closure section and Q_i and $T_{w,i}$ are streamflow and water temperature of the i-th contributing water flux. Subsurface water and energy fluxes (e.g. hyporheic exchanges) are implicitly accounted for, and the net heat flux at the water–atmosphere interface, H, integrates the contributions of the main heat flux components (short- and long-wave radiation, latent and sensible heat fluxes). These, as a first approximation, are evaluated using air temperature, T_a, as the only meteorological forcing (e.g. Caissie, 2006). Heat generated through friction with the

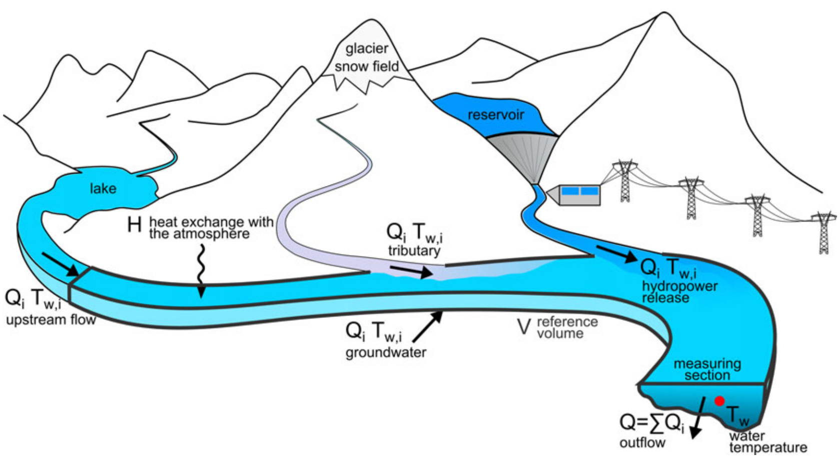

Figure 2. Schematic representation of a river reach with the relevant inflows and heat fluxes

Copyright © 2016 John Wiley & Sons, Ltd.

stream bed (Sinokrot and Stefan, 1993) is neglected. Under these assumptions, Equation 1 can be rewritten as (Toffolon and Piccolroaz, 2015):

$$\frac{dT_w}{dt} = \frac{1}{\delta}\left\{a_1 + a_2 T_a - a_3 T_w + \vartheta\left[a_5 + a_6 \cos\left(2\pi\left(\frac{t}{t_y} - a_7\right)\right) - a_8 T_w\right]\right\}, \quad (2)$$

where the time t is expressed in days, t_y is the duration of one year, and $\vartheta = Q/\overline{Q}$ is the dimensionless streamflow, with $\overline{Q} = (t_2 - t_1)^{-1}\int_{t_1}^{t_2} Q(t)dt$ being a reference value averaged over the entire time series extending from t_1 to t_2. The parameter $\delta = \vartheta^{a_4}$ represents the dimensionless depth according to a power law rating curve.

Equation 2 represents the full version of the *air2stream* model, described by eight parameters ($a_1 - a_8$) in which the main thermo-physical, geometrical and hydraulic properties of the stream are lumped. The time series of air temperature and streamflow are used as the sole input of the model, and the parameters are calibrated against observed RWT, without requiring additional explicit information concerning the characteristics of the river reach. Equation 2 is solved with a daily time step using the second-order accurate Crank–Nicolson numerical scheme, which is implicit in time and unconditionally stable. Moreover, RWT is restricted to be lower-bounded by 0 °C.

The full eight-parameter version (*a2s-8*) of the model can be simplified in four alternative versions, in which some of the parameters are neglected as summarized in Table II (see Toffolon and Piccolroaz, 2015, for details):

- *a2s-7*, obtained by disregarding the dependence on the flow depth (i.e. $\delta = 1$, thus $a_4 = 0$);
- *a2s-5*, neglecting the effect of varying streamflow ($\vartheta = 1$) and the dependence on flow depth ($\delta = 1$), and rearranging the remaining parameters ($a_1 + a_5 \rightarrow a_1$, $a_3 + a_8 \rightarrow a_3$);
- *a2s-4*, obtained by discarding the explicit contribution of water fluxes in the second term of the right hand side of equation 1, hence $\vartheta = 0$;
- *a2s-3*, derived from *a2s-4* with the additional assumption that $\delta = 1$.

The versions with four and three parameters are appropriate when tributaries are absent or their temperature is similar to that of the outlet section.

In addition to the differential versions of *air2stream*, the instantaneous equilibrium water temperature, $T_{w,eq}$, can be defined by neglecting the time derivatives ($dT_w/dt = 0$, Toffolon and Piccolroaz, 2015; see also Edinger *et al*., 1968; Caissie *et al*., 2005). This approximation is valid when the adaptation time scale of RWT to the external forcing is shorter than the averaging time window of the observations, which in this case is 1 day. With this assumption, *air2stream* in both its eight- and seven-parameter versions reduces to a six-parameter ($e_1 - e_6$) equation (*Teq-6* model):

$$T_w = 1(1 + \theta e_4)\left\{e_1 + e_2 T_a + \vartheta\left[e_5 + e_6 \cos\left(2\pi\left(\frac{t}{t_y} - e_3\right)\right)\right]\right\}. \quad (3)$$

Table II. Model types and associated parameters.

Model family	Type	Acronym	Parameters: Number	Parameters: List
Regression based	Linear regression	*lin*	2	$l_1 - l_2$
	Seasonal linear regression	*s-lin*	4	$s_1 - s_4$
	Logistic function	*log*	4	$f_1 - f_4$
Stochastic	Caissie *et al.* (2001)	*CES*	6	$c_1 - c_6$
air2stream equilibrium temperature	Two-parameter	*Teq-2*	2	$e_1 - e_2$
	Four-parameter	*Teq-4*	4	$e_1 - e_3$, e_6
	Six-parameter	*Teq-6*	6	$e_1 - e_6$
air2stream (differential equation)	Three-parameter	*a2s-3*	3	$a_1 - a_3$
	Four-parameter	*a2s-4*	4	$a_1 - a_4$
	Five-parameter	*a2s-5*	5	$a_1 - a_3$, a_6, a_7
	Seven-parameter	*a2s-7*	7	$a_1 - a_3$, $a_5 - a_8$
	Eight-parameter	*a2s-8*	8	$a_1 - a_8$
*air2wate*r (lake model)	Four-parameter	*a2w-4*	4	$b_1 - b_4$
	Six-parameter	*a2w-6*	6	$b_1 - b_6$

Copyright © 2016 John Wiley & Sons, Ltd.

Analogously, the five-parameter version reduces to a four-parameter equation, *Teq-4*, and the four- and three-parameter versions to a two-parameter equation, *Teq-2* (see Table II). In all equilibrium solutions, water temperature at day i is evaluated using the average values of air temperature at day i and $i-1$, which was shown to provide the best results.

We note that the five-parameter and the three-parameter versions of *air2stream*, and the corresponding equilibrium temperature solutions, do not include any information about the streamflow, and air temperature is used as the sole predictor.

Lake surface temperature model air2water. Water temperature in river stations located downstream of lakes (classified as lake outlets) is strongly affected by the surface temperature of the lake. In lakes the thermal response is regulated by different processes than in rivers and thermal inertia is larger, producing a wider hysteretic cycle (i.e. temporal phase lag) between air and water temperature (Toffolon *et al*., 2014). Hence, we tested the *air2water* model (Piccolroaz *et al*., 2013; Toffolon *et al*., 2014; Piccolroaz *et al*., 2015; Piccolroaz, 2016, source code available at https://github.com/spiccolroaz/air2water), which is conceptually analogous to *air2stream*, but has been developed for lentic waters such as lakes. In particular, *air2water* was applied in its four- (*a2w-4*) and six-parameter (*a2w-6*) versions (see Table II, and Piccolroaz *et al*., 2013 for further details).

The version *a2w-4* is analogous to *a2s-4* except for the definition of δ. For lakes, δ represents the dimensionless volume of the surface well-mixed layer participating to the heat exchanges with the atmosphere, and is not a function of streamflow but rather of the strength of lake thermal stratification, assessed on the basis of T_w (see Piccolroaz *et al*., 2013). In *a2w-6* an additional sinusoidal term is added to account for seasonal variations of the forcing parameters:

$$\frac{dT_w}{dt} = \frac{1}{\delta}\left\{b_1 + b_2T_a - b_3T_w + b_5\cos\left[2\pi\left(\frac{t}{t_y} - b_6\right)\right]\right\}, \tag{4}$$

keeping the same definition of δ as in *a2w-4*.

Regression-based models. Statistical regression-based models are widely used in climate studies because of their simplicity and limited amount of data required for implementation. Indeed, they are generally based on air temperature as the sole predictor. In this analysis, we used the standard linear regression model (e.g. Stefan and Preud'homme, 1993; Erickson and Stefan, 2000; Kelleher *et al*., 2012)

$$T_w = l_1 + l_2T_a \tag{5}$$

considering two different versions: an overall regression of the available data over the whole year, and two distinct regressions for periods January–June (in the rising limb of RWT, starting from the cold season) and for July–December (in the falling limb, starting from the warm season). In the latter case, four parameters identify the model (s_1, s_2 for the first period, and s_3, s_4 for the second period). We note that the two-parameter version of the equilibrium temperature model, *Teq-2*, corresponds to a linear regression with lower bound at 0 °C.

As a third regression-based model, we tested the widely used logistic regression function (e.g. Mohseni *et al*., 1998; Arismendi *et al*., 2014)

$$T_w = f_1 + \frac{f_2 - f_1}{1 + \exp[f_3(f_4 - T_a)]} \tag{6}$$

which uses only air temperature as predictor and relies on 4 calibration parameters (i.e. f_1 and f_2 are the minimum and maximum RWT of the analysed period of data, respectively, f_3 is a measure of the steepest slope of the function, and f_4 is the air temperature at the inflection point). The logistic function is characterized by a typical S-shape with decreasing slope at extreme (both maxima and minima) temperatures.

Stochastic models. As a well-known example of the class of stochastic models we considered the model proposed by Caissie *et al*. (2001). The model was originally developed for daily maximum RWT, but can be extended to daily averaged values. The model is based on the decomposition of RWT into two terms: $T_w = A_{Tw} + R_{Tw}$. The first term, A_{Tw}, represents the annual component and can be calculated as the first term of a Fourier expansion applied to the mean year time series, $\overline{T}_w$ (defined, for each day of the year, as the average value of all RWT measurements available in the data set for that same specific day):

$$A_{Tw} = c_1 + c_2\cos\left[2\pi\left(\frac{t}{t_y} - c_3\right)\right]. \tag{7}$$

The second term, R_{Tw}, is the residual, which is calculated by subtracting the actual water temperatures from the annual component 7, and is subsequently used to calibrate a stochastic second-order Markov model

$$R_{Tw}(t) = c_4R_{Tw}(t-1) + c_5R_{Tw}(t-2) + c_6R_{Ta}(t), \tag{8}$$

where R_{Ta} is the residual of air temperature with respect to its corresponding annual component evaluated as in equation 7. The model is characterized by six parameters. The parameters $c_1 - c_3$ are directly computed from

Copyright © 2016 John Wiley & Sons, Ltd.

measured RWT; c_4 and c_5 represent the autocorrelation coefficients of RWT for a lag of 1 and 2 days, respectively; and c_6 is calibrated by minimizing the root mean square error in Equation 8. For further details on the method we refer to the original paper of Caissie *et al.* (2001).

Model calibration and error estimation

Model calibration is obtained by means of the Particle Swarm Optimization algorithm (Kennedy and Eberhart, 1995), an evolutionary, self-adapting computation technique aimed at optimizing a metric of model performance. In this analysis, the adopted metric was the Root Mean Square Error (*RMSE*) between simulated, T_w, and observed, $\hat{T}_w$, RWT:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(T_{w,i} - \hat{T}_{w,i}\right)^2}, \tag{9}$$

where n is the length of the observational time series.

In order to obtain a robust identification of optimal values of parameters, the series of observational data should be long enough to include significant inter-annual variability and, possibly, also the occurrence of extreme events (e.g. heatwaves, droughts, floods). Based on our experience, parameters can be satisfactorily calibrated relying on time series covering at least 3–5 years, although the optimal length of observational series is inherently case specific (see also Piccolroaz, 2016 for the case of *air2water*). As a standard setup, we used 2/3 of the observational time series for model calibration, and the remaining 1/3 for model validation.

In order to allow for a fair comparison among different cases, we also calculated the Nash–Sutcliffe Efficiency index (Nash and Sutcliffe, 1970), which provides a normalized measure ($-\infty$,1) of model performance by weighting the average variability of model residuals with respect to the variability of the target series. Given the strong seasonality typical of RWT series, we considered a modified definition of the NSE index that uses the inter-annual mean value for each day of the year (i.e. the mean year, $\overline{T}_w$) as a benchmark model, instead of the overall mean of the target series:

$$NSE^* = 1 - \frac{RMSE^2}{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{T}_{w,i} - \overline{T}_w\right)^2}. \tag{10}$$

*NSE** is to be preferred to *NSE* for appreciating how well the model captures dynamics not contained in the seasonal component (Schaefli and Gupta, 2007). We note that $NSE^* = 1$ indicates perfect model fitting, while $NSE^* = 0$ indicates that the model performs as good as assuming the mean year of measurements, $\overline{T}_w$, as a predictor.

We opted to use both *RMSE* and *NSE** in order to provide a quantitative, simple to understand and practical evaluation of the error in the first case, and a fair, normalized measure of model performance in the latter. This is in line with the recommendation, suggested by Bennett *et al.* (2013), of choosing indices of model performance that are appropriate for the expert modeller (*NSE**) but also comfortable for the end-user (*RMSE*). Finally, we note that in our case these two indices are suitable to rank competing models on the basis of goodness of fit and number of parameters. Indeed, the number of available data in the selected rivers is much larger (tens of years, see Table I, i.e. $n \sim 10^3 - 10^4$ at daily time scale) than the maximum number of parameters (i.e. 8); thus, indices based on information theory (e.g. Akaike or Bayesian information criterions, Burnham and Anderson, 1998) provide almost equivalent rankings (see also Section on 'Guidelines for the Selection of the Optimal Model').

RESULTS AND DISCUSSION

Thermal classification

In this section we examine the GIS-based classification presented in the Section on 'Swiss Rivers Database' by complementing it with the analysis of the hydrological and thermal regimes of the investigated rivers. Figure 3 shows as an example the typical seasonal patterns of RWT, air temperature and streamflow in four selected cases. Substantial differences among the river categories can be noted, providing a direct verification of the adopted classification. This is further confirmed by the markedly different hysteresis cycles between observed air and water temperature time series registered for all the investigated rivers (Figure 4). The hysteresis cycle results from the delay of RWT with respect to the external forcing, thus is a good descriptor of the thermal response of the water body. Distinctive clustering patterns and shape features of the four river categories are clearly distinguishable in Figure 4.

The air–water temperature hysteresis cycles are wider for lake outlets, in agreement with the general understanding that RWT of these rivers is strongly affected by the large thermal inertia of lakes. Low-land natural rivers are characterized by narrower hysteretic loops and generally colder water temperatures with respect to lake outlets (see also Figure 3a and 3b). The hysteresis cycles further shrink in regulated and snow-fed rivers, which are the two categories characterized by the coldest water temperatures. We note that all stations located along low-land natural rivers, lake outlets and regulated rivers present approximately the same range of variability for air temperature, suggesting that the different air–water temperature relationships are ascribable to their specific

Copyright © 2016 John Wiley & Sons, Ltd.

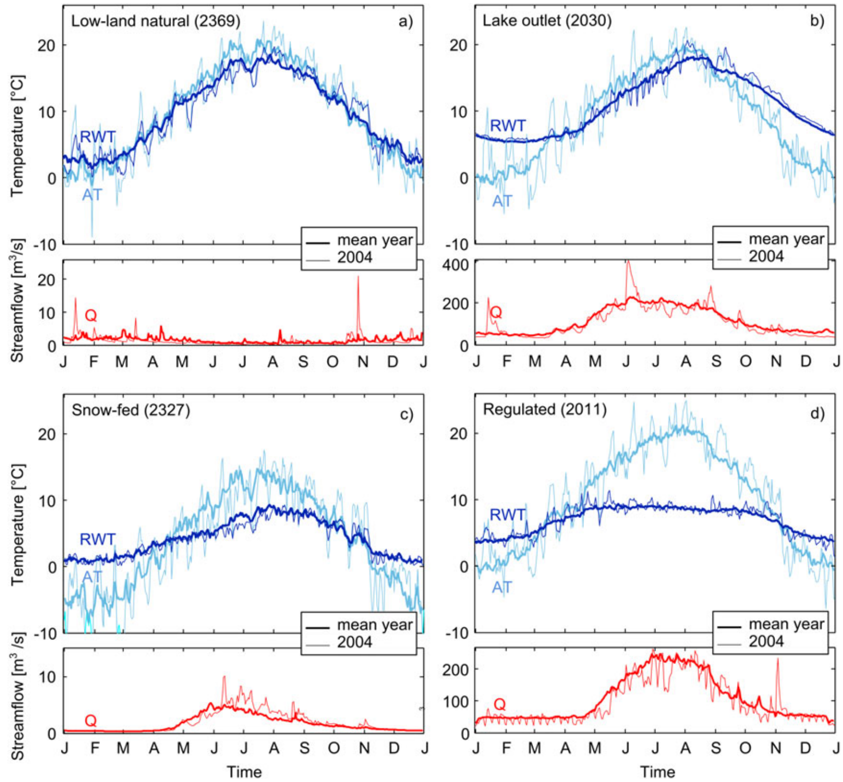

Figure 3. Typical annual variation of RWT, air temperature (AT) and streamflow (Q) for four examples of the different river categories (a–d). For each category, both the mean year (see Section on 'Model Calibration and Error Estimation') and a single year (2004) are presented. In the lower panel of subplot d, weekly fluctuations of streamflow typical of regulated rivers are clearly recognizable

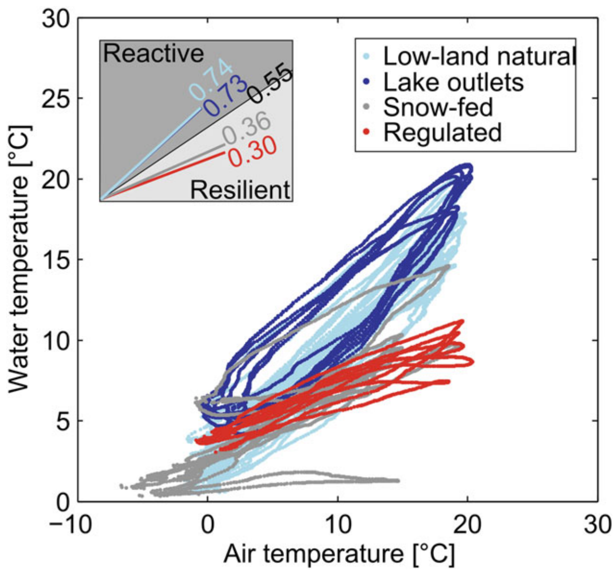

Figure 4. Air–water mean annual hysteresis cycles of the 38 examined rivers, following the classification introduced in the Section on 'Swiss Rivers Database'. For the sake of clarity, the hysteresis cycles have been smoothed with a 30-day moving average filter. The top left inset shows the mean slopes of the hysteretic loops (ΔTw / ΔTa) of each river category, and the threshold slope (ΔTw / ΔTa = 0.55) separating thermally resilient and thermally reactive rivers

hydrological regimes and catchment characteristics. Being generally located at higher altitudes, snow-fed rivers experience colder air temperatures, which are reflected into lower water temperature throughout the whole year. This effect is particularly evident during the winter season when RWT is generally close to freezing temperature (0 °C, see Figure 3c).

In addition, Figure 4 clearly shows the existence of two distinct thermal patterns, which are characterized by different slopes of the hysteretic loops and, interestingly, are fully coherent with the hydrological regime classification described above. A threshold slope $\Delta T_w/\Delta T_a \cong 0.55$ separates low-land natural rivers and lake outlets on the one hand, and snow-fed and regulated rivers on the other hand (see the inset of Figure 4). The clear separation into these two groups supports the adoption of an alternative classification that, consistently with the concept of thermal sensitivity introduced by Kelleher *et al*., 2012 and successively developed by Mayer, 2012, allows to distinguish between thermally *reactive* (above threshold) and thermally *resilient* (below threshold) rivers. Low-land natural rivers and lake outlets, which belong to the first group, are characterized by the steepest hysteresis cycles, suggesting that for a given variation in air temperature, the thermal response will be more intense than for snow-fed and regulated rivers. Conversely, the thermal reactivity of the second group of rivers is inhibited, especially during the warm season when large volumes of cold waters exert a strong control on river thermal

Copyright © 2016 John Wiley & Sons, Ltd.

response (Toffolon *et al*., 2010). This water originates from either snow or glacier melting or releases from high-altitude storage reservoirs for hydropower production (Zolezzi *et al*., 2009, 2011; Vanzo *et al*., 2015). The overall result is a marked flattening of the thermal seasonal pattern (Figure 3c,d; see also Meier and Wüest, 2004; Toffolon and Piccolroaz, 2015) with a consequent lower slope of the air–water temperature hysteretic loop, and a damped response to extreme events as for example heatwaves.

We note that, among snow-fed rivers, two cases exhibit peculiar thermal behaviours. The Massa creek (station 2161) is characterized by extremely small water temperatures changes throughout the year (i.e. nearly horizontal air–water hysteresis cycle), which is symptomatic of its high-altitude position (1446 m a.s.l) and its location immediately downstream of a moraine. River Aare immediately downstream of Lake Brienz (station 2457) shows a wide hysteresis cycle typical of lakes but with a mean slope characteristic of thermally resilient rivers. The moderate slope is explained by the low temperatures of the two major inflows of Lake Brienz, river Aare and river Lütschine, which are characterized by intense hydropower activity and nival/glacial hydrological regime, respectively (Finger *et al*., 2007). Because the RWT gauging station is located in close proximity of the Lütschine entrance into the lake (see http://www.hydrodaten.admin.ch/en/2457.html), we classified station 2457 as snow-fed.

Analysis of performances

Comparison among models. The performances of the different models introduced in the Section on 'Models' are compared in Figure 5, which shows the box plots of *NSE** values obtained for all the gauging stations during both calibration and validation periods (see also Figure S1 in Supporting Information for the same analysis in terms of *RMSE*). Average values of *RMSE* and *NSE** for the different models are also reported in Table III.

All versions of *air2stream*, with the exception of *Teq-2* (which coincides with the linear regression model, except for the lower bound at 0 °C) largely outperform the regression-based models, suggesting a greater predictive capability. The mean values of *NSE** and *RMSE* for regression-based models are −0.48 and 1.50 °C in calibration, and −0.53 and 1.53 °C in validation. Among this family of models, the seasonal linear regression provides the overall best performances (*NSE** = −0.31; *RMSE* = 1.43 °C; values averaged over calibration and validation periods). Nonetheless, all regression-based models present negative mean values of *NSE**, indicating that their predictive capability is worse than the mean year. The average performance of *air2stream* models (i.e. the eight model versions listed in Table II with the exclusion of *Teq-2*, which can be assimilated to the regressive model category) is significantly better, with *NSE** increasing to 0.41 and 0.36, and *RMSE* reducing to 0.92 °C and 0.95 °C, respectively for calibration and validation. The best performances are obtained with *a2s-8* (*a2s-7* performs similarly) with mean values of *NSE** and *RMSE* of 0.54 and 0.81 °C (0.49 and 0.84 °C in validation). The stochastic model by Caissie *et al*. (2001) has better performances than regression-based models (*NSE** and *RMSE* of 0.25 and 1.09 °C, and 0.15 and 1.15 °C in calibration and validation, respectively), but is less satisfactory than the *air2stream* versions with a comparable number of parameters (i.e. *Teq-4*, *Teq-6*, *a2s-5*, *a2s-7* and *a2s-8*). Furthermore, when compared with

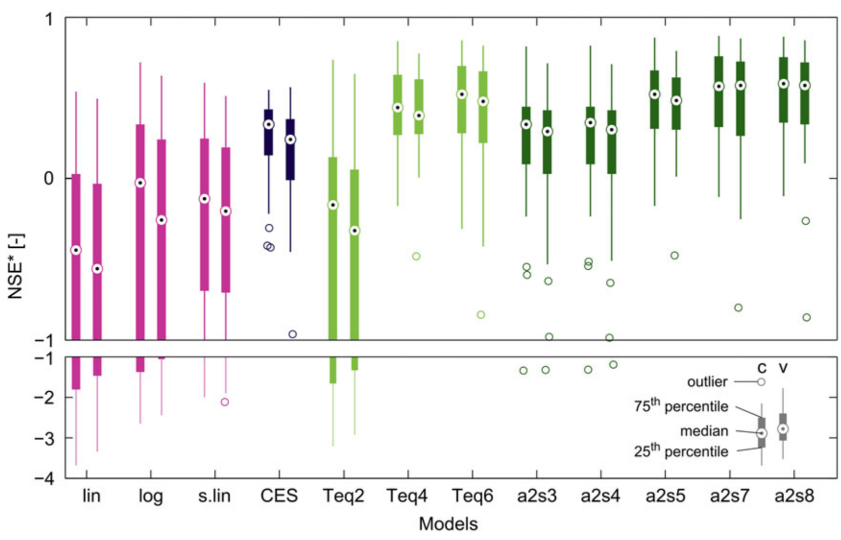

Figure 5. Box plots of ⟦NSE⟧^* obtained for all statistical models and all versions of air2stream at daily time scale distinguishing calibration ('c') and validation ('v') periods. The acronyms used to identify the models are defined in Table 2

Table III. Mean values of *NSE** and *RMSE* obtained in calibration and validation at daily time scale by applying all models to the entire dataset (Table 1), and separating the dataset according to the rivers classification introduced in the Section on 'Swiss Rivers Database'.

Model acronym	*NSE** [-] (*RMSE* [°C])									
	Entire dataset		Low-land natural		Lake outlets		Snow-fed		Regulated	
	cal	val	cal	val	cal	val	cal	val	cal	val
lin	−0.80 (1.65)	−0.83 (1.68)	−0.15 (1.72)	−0.35 (1.74)	−2.61 (2.55)	−2.09 (2.59)	−0.91 (1.11)	−1.13 (1.11)	0.03 (0.85)	−0.01 (0.94)
log	−0.37 (1.43)	−0.41 (1.46)	0.17 (1.46)	0.04 (1.46)	−1.79 (2.24)	−1.42 (2.30)	−0.50 (0.98)	−0.71 (0.99)	0.26 (0.74)	0.17 (0.84)
s-lin	−0.28 (1.41)	−0.34 (1.44)	0.09 (1.55)	−0.07 (1.57)	−1.11 (1.95)	−0.83 (2.00)	−0.54 (0.98)	−0.74 (0.98)	0.21 (0.76)	0.18 (0.85)
CES	0.25 (1.09)	0.16 (1.15)	0.40 (1.31)	0.22 (1.40)	0.34 (1.09)	0.43 (1.11)	−0.07 (0.80)	−0.26 (0.82)	0.12 (0.80)	0.16 (0.86)
Teq-2	−0.58 (1.52)	−0.58 (1.54)	0.09 (1.52)	−0.05 (1.53)	−2.22 (2.40)	−1.75 (2.45)	−0.76 (1.06)	−0.93 (1.05)	0.09 (0.82)	0.05 (0.91)
Teq-4	0.42 (0.92)	0.40 (0.95)	0.64 (1.00)	0.60 (1.00)	0.29 (1.13)	0.32 (1.22)	0.15 (0.71)	0.11 (0.69)	0.35 (0.68)	0.31 (0.78)
Teq-6	0.44 (0.90)	0.39 (0.93)	0.67 (0.93)	0.63 (0.93)	0.13 (1.25)	0.19 (1.32)	0.17 (0.71)	0.03 (0.71)	0.57 (0.56)	0.50 (0.66)
a2s-3	0.22 (1.06)	0.17 (1.09)	0.48 (1.19)	0.41 (1.20)	0.31 (1.11)	0.38 (1.15)	−0.34 (0.89)	−0.50 (0.90)	0.10 (0.82)	0.06 (0.91)
a2s-4	0.24 (1.04)	0.18 (1.08)	0.49 (1.17)	0.41 (1.20)	0.35 (1.08)	0.40 (1.14)	−0.30 (0.87)	−0.46 (0.88)	0.10 (0.82)	0.06 (0.91)
a2s-5	0.46 (0.88)	0.44 (0.91)	0.66 (0.96)	0.63 (0.97)	0.41 (1.03)	0.43 (1.11)	0.16 (0.70)	0.12 (0.68)	0.36 (0.68)	0.32 (0.77)
a2s-7	0.53 (0.82)	0.48 (0.85)	0.73 (0.85)	0.71 (0.84)	0.35 (1.07)	0.39 (1.14)	0.23 (0.68)	0.07 (0.69)	0.59 (0.55)	0.51 (0.65)
a2s-8	0.54 (0.81)	0.49 (0.84)	0.74 (0.85)	0.71 (0.85)	0.39 (1.04)	0.43 (1.10)	0.24 (0.68)	0.07 (0.69)	0.59 (0.55)	0.51 (0.65)
a2w-4	—	—	—	—	0.48 (0.96)	0.52 (1.01)	—	—	—	—
a2w-6	—	—	—	—	0.54 (0.90)	0.56 (0.97)	—	—	—	—

Copyright © 2016 John Wiley & Sons, Ltd.

the simplest versions of *air2stream* (i.e. *a2s-3*, *a2s-4*), we note that the better performances of *CES* in calibration are subjected to a larger deterioration in validation compared to the family of hybrid models.

A general improvement of *air2stream* performances is obtained for models with higher number of parameters. However, differences are negligible when including or not the effect of the water volume (thus thermal inertia) by means of the parameter δ (i.e. *a2s-3* vs *a2s-4*, and *a2s-7* vs *a2s-8*). This was already evidenced by Toffolon and Piccolroaz (2015), who showed that the time scale for the adaptation of RWT to the external forcing in the examined rivers is typically smaller than (or comparable to) the time averaging window of the data (1 day). The relatively fast adaptation of RWT to external forcing is also the reason for performances of equilibrium versions (*Teq*) being comparable to their corresponding differential versions. We also note that, despite the number of degrees of freedom is the same (four parameters), the equilibrium temperature model *Teq-4* allows for better results than the differential model *a2s-4*, with the mean *NSE** and *RMSE* being respectively 0.18 higher and 0.12 °C smaller. This is because of the effect of considering the phase adjustment of RWT with respect to air temperature by means of the additional sinusoidal term. Figure 5 also suggests that *air2stream* models (with the exception of *Teq-2*) are more robust than those with purely regression-based models as indicated by the lower variability of *NSE** across all considered sites (i.e. narrower boxplots), similarly to *CES*.

Comparison among hydrological categories of rivers. Comparison of model performances for each gauging station during the calibration period is presented in Figure 6 (in terms of *NSE**; see also Figure S2 in Supporting Information for the same comparison but in terms of *RMSE*), where rivers are grouped according to their hydrological and thermal regimes. Visual inspection of the figure allows for an immediate understanding of the suitability of different models for different categories of rivers.

Regression-based models fail to predict RWT dynamics of lake outlets (always negative values of *NSE**) as a consequence to their inability to describe the existing hysteresis cycle between air and water temperature (see Figures 3b and 4, and the discussion in the Section on 'Thermal Classification'). Good results for lake outlets are obtained by the differential and equilibrium versions of *air2stream*: the average *NSE** and *RMSE* calculated excluding *Teq-2* are equal to 0.32 and 1.10 °C in calibration, similar to *CES* (Table III). The best version is *a2s-5* while the versions explicitly including the presence of contributing water fluxes (through the parameter ϑ, i.e. *a2s-7*, *a2s-8* and *Teq-6*) do not necessarily provide higher performances with respect to simpler versions, indicating that streamflow is not always a significant predictor variable for this category of rivers. As expected, higher performances are obtained by employing *air2water* models (*a2w-4* and *a2w-6*) with average *NSE** increasing to about 0.51 and average *RMSE* reducing to about 0.94 °C in calibration. Interestingly, *air2water* models are matched or even outperformed by *a2s-7* and *a2s-8* when applied to river stations located at a large distance from the lake outlet (station 2016 at 96 km downstream Lake Biel; station 2085 at 57 km downstream Lake Thun; station 2091 at 117 km downstream Lake Constance). This evidence suggests a progressive decline of the influence of lake's thermal conditions in favour of a gradual larger control exerted by local exchanges.

For low-land natural rivers, the discrepancy between regression-based models and *air2stream* versions is not as

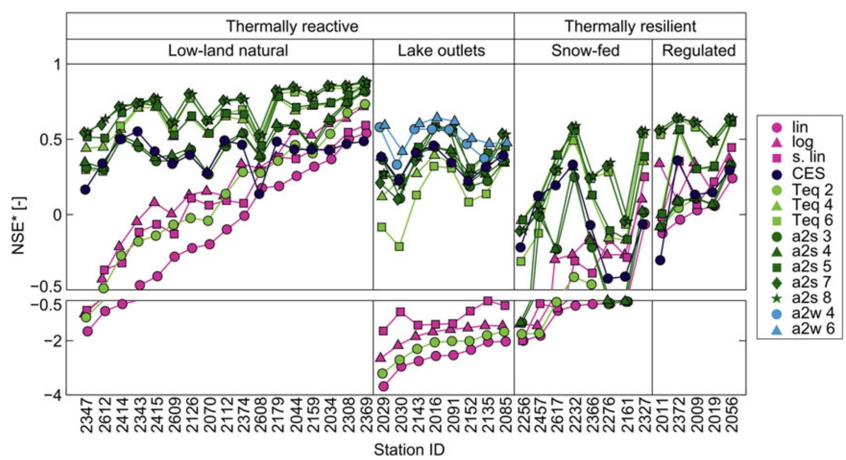

Figure 6. Comparison of ⟦NSE⟧^*values obtained for each river station using different models (different symbols) at daily scale. Rivers are grouped according to their hydrological and thermal classification. ⟦NSE ⟧^* values refer to the calibration period

abrupt as for lake outlets (at least not for all rivers), although the latter family of models always outperforms the first. Among the different model versions, *a2s-7* and *a2s-8* provide overall better performances (*NSE** and *RMSE* of about 0.74 and 0.85 °C in calibration, see Table III) indicating that, in general, the effect of streamflow cannot be neglected. The stochastic model *CES*, which does not include any information about streamflow, is in this case less accurate than the whole family of *air2stream* models (again excluding *Teq-2*), with *NSE** and *RMSE* equal to 0.39 and 1.31 °C in calibration.

The dependence of performances on streamflow is accentuated in thermally resilient rivers as evidenced by the larger *NSE** indices obtained for *a2s-7*, *a2s-8* and (for regulated rivers) *Teq-6* models compared to the simpler versions (Figure 5 and Table III). This is consistent with the actual physical processes controlling the thermal response, whereby the colder water coming from hydropower reservoirs or snowfields is the first reason for the mitigation of warmer RWT in summer (a more detailed analysis of the role of streamflow is provided in Toffolon and Piccolroaz, 2015). Notably, *a2s-5*, *Teq-4* and (for snow-fed rivers) *Teq-6* models provide relatively high *NSE** indices, comparable to those obtained with model versions including the effect of streamflow. This is because of the presence of the sinusoidal term in Equations 2 and 3 which introduces a phase shift between RWT and air temperature mimicking the effect of cold contributions because of snow/ice melting water and hydropower releases. We also note that in snow-fed rivers the prediction of RWT provided by all regression-based models, *CES*, and the remaining versions of *air2stream* is certainly not adequate, being associated to negative values of *NSE** (for the sake of completeness, we refer also to Table S2, where *NSE** and *RMSE* values are calculated excluding the peculiar case of station 2161, see Section on 'Thermal Classification'). Concerning regulated rivers, despite most of the models provide positive values of *NSE** (in particular the seasonal linear regression and the logistic function models), still they are clearly suboptimal compared to the more advanced versions of *air2stream*.

Performances at longer time scales

For many applications or in the case of coarser data availability RWT is often predicted at time scales longer than one day. To this end, all models have been calibrated also at weekly, monthly and seasonal (i.e. 3-month period) time scales. The calibration of the regression-based models and of *Teq* versions of *air2stream* has been conducted optimizing the fitting between air temperature averaged over the different time scales as predictor, and the corresponding averaged RWT as the response variable. Conversely, the differential versions of *air2stream* (whose computational time step is 1 day) have been calibrated using as external forcing the daily air temperature (and discharge when required) linearly reconstructed from the mean values at a given time scale, and evaluating the RMSE on the timescale-averaged RWT values. The same has been done for the *CES* model, but in this case also RWT has been linearly reconstructed, as it is required in Equation 8. In this way, the definition of *RMSE* is equivalent for all models and the input information is the same, making the comparison fair.

Figure 7 presents the simulation results as box plots, considering the entire dataset and all the models listed in Table II. The relationships between air and water temperatures change at different time scales, producing a systematic increase of prediction performances with increasing time window. In particular, *RMSE* is approximately halved on

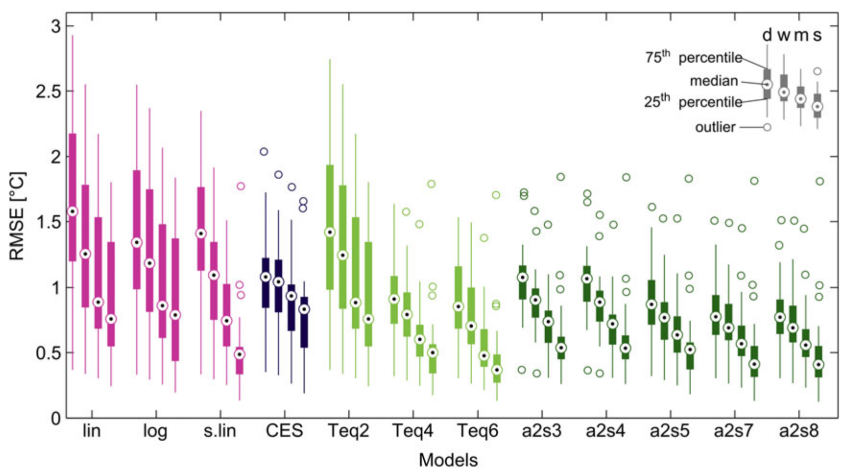

Figure 7. Box plots of RMSE obtained for all models calibrated at daily ('d'), weekly ('w'), monthly ('m') and seasonal ('s') time scale, considering the entire dataset for the calibration period

Copyright © 2016 John Wiley & Sons, Ltd.

Table IV. Averaged *RMSE* obtained for all models calibrated at daily ('d'), weekly ('w'), monthly ('m') and seasonal ('s') time scale, considering the entire dataset for the calibration period.

Model acronym	d	w	m	s
lin	1.65	1.36	1.08	0.90
log	1.43	1.25	1.01	0.90
s-lin	1.41	1.10	0.78	**0.49**†
CES	1.09	1.04	0.89	0.78
Teq-2	1.52	1.35	1.08	0.90
Teq-4	0.92	0.81	**0.62**†	0.51
Teq-6	0.90	0.78	**0.56***	**0.42***
a2s-3	1.06	0.90	0.72	0.59
a2s-4	1.04	0.88	0.70	0.58
a2s-5	**0.88**†	**0.77**†	0.66	0.54
a2s-7	0.82	0.74	0.62	0.47
a2s-8	**0.81***	**0.73***	0.60	0.47

*Indicates the overall best model.
† Indicates the best model among those not using streamflow.

average, moving from daily to seasonal time scale (see also Table IV). Note that results are shown in terms of *RMSE* because the analysis in terms of *NSE** would be misleading (see Figure S3 in Supporting Information). In fact, the denominator of the second term on the right hand side of equation 10 depends on the variance of the observations with respect to the benchmark model (the mean year $\overline{T}_w$), both evaluated at the selected time scale. This variance significantly decreases for longer time scales because of the increased performances of the benchmark model. Thus, the relative improvement of the benchmark model compared to the one under analysis may produce an unexpected decrease of *NSE** with longer aggregation periods.

Regression-based models and *Teq-2* are accompanied by larger reductions of *RMSE*, with seasonal linear regression presenting the largest improvement moving from daily to seasonal time scales (*RMSE* from 1.41 to 0.49 °C, i.e. a decrease of up to 65%). In particular, at monthly and seasonal time scales, *s-lin* reaches, and in some cases overcomes, the performances of *air2stream* models, with the only exception of *a2s-7*, *a2s-8* and the corresponding equilibrium temperature model *Teq-6*, which are the models with the highest performances. This is consistent with the general understanding that at large time scales (e.g. monthly to seasonal) water temperature tends to be less correlate within the time series and regression-based models become quite effective (Caissie, 2006). Finally, with a reduction of *RMSE* of less than 30%, the *CES* model is the one with the smallest improvement at seasonal time scale, while overall *Teq-6* replaces *a2s-8* as the best model for time scales greater than or equal to one month, again because of the lower autocorrelation of RWT at larger time scales. The general considerations outlined above apply also when analysing the four river categories individually (Table S3).

It is worth noting that only few studies have proposed a quantitative estimate of the increased performances of different models across time scales, and in any case focusing on linear and logistic models (e.g. Stefan and Preud'homme, 1993; Pilgrim *et al*., 1998; Webb *et al*., 2003; Morrill *et al*., 2005; Kelleher *et al*., 2012). Here we are able to offer a more systematic view on the effect of considering longer time windows, an element that is crucial when assessing the performance of RWT models.

Cross validation

The robustness of model predictions for all versions of *air2stream* and all regression-based models is tested by means of a cross-validation procedure applied at daily time scale. The objective is to evaluate the errors of the various models in predicting RWT in periods different from that adopted during the calibration. The analysis is restricted to 17 selected stations presenting a 30-year uninterrupted time series of observed RWT data (river stations 22–38 of Table I: three low-land rivers, eight lake outlets, one snow-fed river, five regulated rivers) during the period 1984–2013. This 30-year period presents large variations in the thermal regime with two extremely hot summers (2003 and 2006) and an abrupt warming of Swiss inland water bodies occurred in the late 1980s, as a consequence of a sudden climate regime shift throughout the Northern Hemisphere (North *et al*., 2013). Therefore, this can be considered a severe test of the model's performance.

Calibration was performed separately for five different 6-year time windows (i.e. 1984–1989, 1990–1995, 1996–2001, 2002–2007 and 2008–2013), thereby obtaining five sets of parameters for each model. Subsequently, the validation was undertaken by running all the models on the remaining four time windows of observation by using, in turn, the above five sets of parameters and computing NSE separately for each time window. This procedure, commonly known as cross validations (e.g. Merz *et al*., 2011; Majone *et al*., 2012), is still relatively uncommon in the hydrological modelling literature (Seibert, 2003) and, to the best of our knowledge, has been applied to RWT modelling only by Arismendi *et al*. (2014).

The results of the analysis are presented in terms of *NSE** by considering separately the four different hydrological categories. For the sake of brevity, Figure 8 reports as an example the model performances when the first time window (1984–1989) is used for calibration, but similar results have been obtained when calibration was performed in the other time windows (Figures S4–S7). A more synthetic picture of the cross validation exercise is provided in Table V, where *NSE** indices averaged over of the various time windows are

Copyright © 2016 John Wiley & Sons, Ltd.

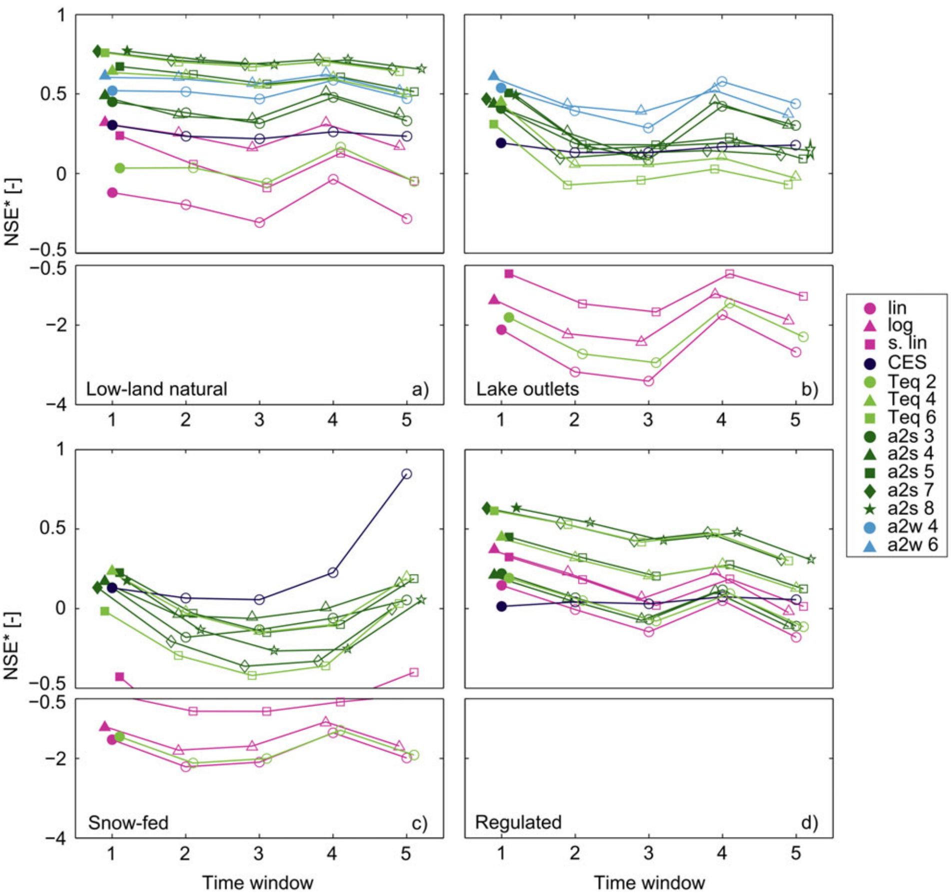

Figure 8. Cross validation of all models considering five time windows: ⟦NSE⟧^* for all models obtained with calibration in time window 1 (1984–1989, filled symbols) and validation in the remaining period (time windows 2 to 5). Average values are presented for the four hydrological categories

Table V. Averaged *NSE** resulting from cross validation for the four hydrological categories. For each group, values are averaged over the five time windows used for calibration, and over the remaining 20 (5 × 4) time windows used for validation.

Model acronym	Low-land natural			Lake outlets			Snow-fed			Regulated		
	cal	val	cal-val	cal	val	cal-val	cal	val	cal-val	cal	val	cal-val
lin	−0.16	−0.20	0.04	−2.58	−2.64	0.06	−1.81	−1.83	0.03	0.05	−0.04	0.09
log	0.27	0.24	0.03	−1.73	−1.80	0.07	−1.44	−1.48	0.04	0.29	0.18	0.11
s-lin	0.13	0.09	0.05	−1.06	−1.14	0.08	−0.55	−0.63	0.08	0.26	0.14	0.12
CES	0.30	0.30	0.00	0.19	0.19	0.00	0.29	0.28	0.01	0.08	0.08	0.00
Teq-2	0.05	0.01	0.03	−2.19	−2.24	0.05	−1.73	−1.75	0.03	0.12	0.03	0.09
Teq-4	0.66	0.60	0.06	0.37	0.22	0.15	0.19	0.06	0.13	0.40	0.28	0.13
Teq-6	0.76	0.70	0.05	0.23	0.09	0.13	0.00	−0.18	0.17	0.62	0.48	0.14
a2s-3	0.42	0.39	0.03	0.34	0.29	0.05	0.00	−0.04	0.04	0.14	0.04	0.10
a2s-4	0.44	0.40	0.04	0.38	0.30	0.09	0.09	0.01	0.07	0.13	0.04	0.10
a2s-5	0.68	0.62	0.06	0.46	0.33	0.13	0.19	0.06	0.13	0.41	0.28	0.13
a2s-7	0.77	0.71	0.06	0.43	0.28	0.15	0.10	−0.11	0.21	0.63	0.49	0.15
a2s-8	0.77	0.71	0.06	0.46	0.31	0.14	0.13	−0.07	0.20	0.64	0.49	0.15
a2w-4	—	—	—	0.50	0.44	0.06	—	—	—	—	—	—
a2w-6	—	—	—	0.57	0.49	0.09	—	—	—	—	—	—

presented for both calibration and validation time windows, with reference to the four hydrological categories. The performances of the different models confirm the outcomes discussed in the previous sections (but with a dataset restricted to 17 rivers), and allow for specific remarks in terms of model robustness.

Copyright © 2016 John Wiley & Sons, Ltd.

Overall, the results confirm that the sets of parameters obtained by the 5-year calibration do not lead, in general, to an excessive deterioration of the performances when applied to longer time windows, indicating that model predictions can be transferred well to other time periods. This is valid for all models and all regimes, although with different relative performances. In general, the stochastic model *CES* is the one with the highest stability within time windows, also when the other families of models show some deterioration (e.g. in low-land natural and regulated rivers). The deterioration is higher with increasing complexity of the model (Table V). Nonetheless, in absolute terms, the stochastic model *CES* is always outperformed by nearly all the *a2s* and the corresponding *Teq* models in low-land natural rivers and lake outlets during both calibration and validation (on average), and in regulated rivers even by the seasonal linear regression and by the logistic function (Table V). The only exception is for snow-fed rivers, where *CES* provides the best performance in both calibration and validation. However, only one river is considered in this case, thus making the analysis not fully representative. We stress that the apparent larger deterioration of performance typical of the more complex versions of *air2stream* is inherently associated to their generally higher performances. In other words, it is difficult that a low performance model largely worsens, while it is more likely that high performance models undergo a certain worsening, albeit maintaining their superiority.

Concerning lake outlets, the best models are, as expected, *a2w-4* and *a2w-6*. Contrarily to their *air2-stream* counterparts, these models are also associated to a relatively smaller worsening of performance moving from calibration to validation, thus suggesting *air2water* as particularly robust for long-term analyses.

Guidelines for the selection of the optimal model

The performances of the more advanced versions of *air2stream* (*a2s-5*, *a2s-7*, *a2s-8* and the corresponding equilibrium temperature models) are generally higher than regression-based and stochastic models, even in those cases where the number of parameters is the same (see also Table S5 for model ranking based on the Akaike information criterion). Despite some deterioration in model performance, this rule holds also when these models are applied to different time windows, a feature that is relevant in the case of questions related to climate change.

The best models at daily and weekly time scales are always *a2s-7* or *a2s-8*, while at monthly and seasonal time scales is the *Teq-6* model, except for lake outlets where the *air2water* family of models prevails. Among models that do not require discharge as input information, *a2s-5* and the corresponding equilibrium temperature models (*Teq-4*) provide the best performances, respectively at short and longer time scales, even if in some cases *Teq-4* is outperformed by the regression *s-lin* at seasonal time scale. We remark that the equilibrium temperature models are suitable to be used at larger time scales because of the progressive loss of autocorrelation of RWT.

In general, purely regression models are inadequate in the case of regulated rivers, although the seasonal linear regression and the logistic function may provide acceptable results in some cases. Analogously, the stochastic model *CES* performed rather poorly when tested with low-land natural rivers, and performed better for the other river categories, but still with performances lower than the hybrid models, even for those versions that do not consider streamflow.

The inclusion of streamflow is required to obtain more accurate results. The only exception is represented by lake outlets where discharge plays a secondary role compared to the thermal dynamics of the upstream lake, and the RWT model has to account for thermal inertia and stratification seasonality. This is true only until a certain distance from the lake outlet, which has been evaluated equal to about 50 km based on present analysis and for the rivers of the present database. For longer distances the effect of the lake is likely to become negligible.

Furthermore, for the examined dataset (much larger rivers may behave differently) the primary value of considering streamflow is not related to the inclusion of the thermal inertia of the river through the explicit incorporation of water depth (indeed performances of *a2s-7* and *a2s-8* are fully comparable). Rather, the increased performances are related to the possibility to account for the effect of upstream or lateral contributing water/heat fluxes (e.g. regulated and snow-fed rivers, significant hyporheic exchanges, presence of several tributaries etc.).

CONCLUSIONS

In this study we address a key question that is often posed in the literature concerning the thermal response of rivers (e.g. Arismendi *et al.*, 2014): is air temperature enough to predict RWT? As often happens, the answer is not univocal, but depends on the hydrological regime of the river, the type of model and the time scale of the analysis. We used different types of models (regression-based, stochastic and hybrid with different assumptions) to predict RWT considering a large database of Swiss rivers. The rivers were classified into four hydrological categories (low-land natural, lake outlets, regulated and snow-fed rivers), which neatly correspond to the two thermal categories (reactive and resilient rivers) that we introduced on the basis of the slope of the air–water temperature hysteresis curve.

We demonstrated that the *air2stream* family of models, which is based on hybrid model conceptualization,

Copyright © 2016 John Wiley & Sons, Ltd.

provides very satisfying results improving the current capability to predict RWT with a small amount of input data and information. In fact, by exploiting a physically based model structure, relatively good predictions of RWT can be achieved also when using air temperature as the only external forcing (e.g. *a2s-5* and *Teq-4*). Conversely, performances of purely regression-based or stochastic models are lower, with the only exception of seasonal regression at seasonal time scale.

The analysis highlighted the value of the hybrid *air2stream* approach as a tool to improve the current capability to predict RWT without using more complex deterministic models and to indirectly learn more about the main processes controlling the river thermal response, depending on the model's conceptual structure. This emerged clearly when comparing the different versions of *air2stream* to identify the best model. Indeed, the comparative analysis of the performances within the *air2stream* family allows to identify those cases where streamflow plays a predominant role and to characterize the typical time scale of the thermal response to external forcing. Finally, an exercise of cross-validation allowed us to evaluate the robustness of the different models for long-term projections as required by climate change studies. The analysis consistently indicated that the *air2stream* models represent the optimal choice.

ACKNOWLEDGEMENTS

Water temperature and discharge data were kindly provided by the Swiss Federal Office of the Environment (FOEN), and meteorological data by the Swiss Meteorological Institute (MeteoSchweiz). We thank the anonymous Reviewers for their constructive comments, which helped us to improve the manuscript. The source codes of the *air2stream* and *air2water* models are freely available at https://github.com/marcotoffolon/air2stream and https://github.com/spiccolroaz/air2water, respectively.

REFERENCES

Arismendi I, Johnson SL, Dunham JB, Haggerty R, Hockman-Wert D. 2012. The paradox of cooling streams in a warming world: regional climate trends do not parallel variable local trends in stream temperature in the pacific continental United States. *Geophysical Research Letters* **39**: L10401. DOI:10.1029/2012GL051448

Arismendi I, Safeeq M, Dunham JB, Johnson SL. 2014. Can air temperature be used to project influences of climate change on stream temperature? *Environmental Research Letters* **9**: 084015. DOI:10.1088/1748-9326/9/8/084015

Benda L, Poff NL, Miller D, Dunne T, Reeves G, Pess G, Pollock M. 2004. The network dynamics hypothesis: how channel networks structure riverine habitats. *Bioscience* **54**: 413–427. DOI:10.1641/0006-3568(2004)054[0413:TNDHHC]2.0.CO;2

Bennett ND, Croke BFW, Guariso G, Guillaume JHA, Hamilton SH, Jakeman AJ, Marsili-Libelli S, Newham LTH, Norton JP, Perrin C, Pierce SA, Robson B, Seppelt R, Voinov AA, Fath BD, Andreassian V. 2013. Characterising performance of environmental models. *Environmental Modelling & Software* **40**: 1–20. DOI:10.1016/j.envsoft.2012.09.011

Benyahya L, Caissie D, St-Hilaire A, Ouarda TBMJ, Bobe B. 2007. A review of statistical water temperature models. *Canadian Water Resources Journal* **32**: 179–192. DOI:10.4296/cwrj3203179

Brown L, Hannah D. 2007. Alpine stream temperature response to storm events. *Journal of Hydrometeorology* **8**: 952–967. DOI:10.1175/JHM597.1

Brown L, Hannah D, Milner A. 2005. Spatial and temporal water column and streambed temperature dynamics within an alpine catchment: implications for benthic communities. *Hydrological Processes* **19**: 1585–1610. DOI:10.1002/hyp.5590

Burnham KP, Anderson DR. 1998. *Model selection and multimodel inference: a practical information-theoretic approach*, 2nd edition. Springer-Verlag: New York.

Caissie D, El-Jabi N, Satish MG. 2001. Modelling of maximum daily water temperatures in a small stream using air temperature. *Journal of Hydrology* **251**: 14–28. DOI:10.1016/S0022-1694(01)00427-9

Caissie D, Satish MG, El-Jabi N. 2005. 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

Caissie D. 2006. The thermal regime of rivers: a review. *Freshwater Biology* **51**: 1389–1406. DOI:10.1111/j.1365-2427.2006.01597.x

Comola F, Schaefli B, Rinaldo A, Lehning M. 2015. Thermodynamics in the hydrologic response: travel time formulation and application to Alpine catchments. *Water Resources Research* **51**: 1671–1687. DOI:10.1002/2014WR016228

Ducharne A. 2008. Importance of stream temperature to climate change impact on water quality. *Hydrology and Earth System Sciences* **12**: 797–810. DOI:10.5194/hess-12-797-2008

Destouni G, Jaramillo F, Prieto C. 2013. Hydroclimatic shifts driven by human water use for food and energy production. *Nature Climate Change* **959**: 213–217. DOI:10.1038/nclimate1719

Edinger JE, Duttweiler DW, Geyer JC. 1968. The response of water temperatures to meteorological conditions. *Water Resources Research* **4**: 1137–1143. DOI:10.1029/WR004i005p01137

Erickson TR, Stefan HG. 2000. Linear air/water temperature correlations for streams during openwater periods. *Journal of Hydrologic Engineering, ASCE* **5**: 317–321. DOI:10.1061/(ASCE)1084-0699(2000)5:3(317)

Finger D, Schmid M, Wüest A. 2007. Comparing effects of oligotrophication and upstream hydropower dams on plankton and productivity in perialpine lakes. *Water Resources Research* **43**: W12404. DOI:10.1029/2007WR005868

Gallice A, Schaefli B, Lehning M, Parlange MB, Huwald H. 2015. Stream temperature prediction in ungauged basins: review of recent approaches and description of a new physics-derived statistical model. *Hydrology and Earth System Sciences* **19**: 3727–3753. DOI:10.5194/hess-19-3727-2015

Kelleher C, Wagener T, Gooseff M, McGlynn B, McGuire K, Marshall L. 2012. Investigating controls on the thermal sensitivity of Pennsylvania streams. *Hydrological Processes* **26**: 771–785. DOI:10.1002/hyp.8186

Kennedy J, Eberhart R. 1995. Particle swarm optimization, in Proceedings of IEEE International Conference on Neural Networks, Institute of Electrical & Electronics Engineering, University of Western Australia, Perth, Western Australia, pp. 1942–1948, DOI:10.1109/ICNN.1995.488968.

Kim KS, Chapra SC. 1997. Temperature model for highly transient shallow streams. *Journal of Hydraulic Engineering ASCE* **123**: 30–40. DOI:10.1061/(ASCE)0733-9429(1997)123:1(30)

Kothandaraman V. 1972. Air-water temperature relationship in Illinois river. *Journal of the American Water Research Association* **8**: 38–45. DOI:10.1111/j.1752-1688.1972.tb05091

Kurylyk B, Bourque CA, MacQuarrie K. 2013. Potential surface temperature and shallow groundwater temperature response to climate change: an example from a small forested catchment in east-central New Brunswick (Canada). *Hydrology and Earth System Sciences* **17**: 2701–2716. DOI:10.5194/hess-17-2701-2013

Copyright © 2016 John Wiley & Sons, Ltd.

Majone B, Bovolo CI, Bellin A, Blenkinsop S, Fowler HJ. 2012. Modeling the impacts of future climate change on water resources for the Gàllego river basin (Spain). *Water Resources Research* **48**: W01512. DOI:10.1029/2011WR010985

Matulla C, Schmutz S, Melcher A, Gerersdorfer T, Haas P. 2007. Assessing the impact of a downscaled climate change simulation on the fish fauna in an inner-alpine river. *International Journal of Biometeorology* **52**: 127–137. DOI:10.1007/s00484-007-0107-6

Mayer T. 2012. Controls of summer stream temperature in the Pacific Northwest. *Journal of Hydrology* **475**: 323–335. DOI:10.1016/j.jhydrol.2012.10.012

Meier W, Wüest A. 2004. Wie verändert die hydroelektrische Nutzung die Wassertemperatur der Rhone?. *Wasser Energie Luft* **96**: 305–309 (www.rhone-thur.eawag.ch/wel_rhone.pdf)

Merz R, Parajka J, Blöschl G. 2011. Time stability of catchment model parameters: implications for climate impact analyses. *Water Resources Research* **47**: W02531. DOI:10.1029/2010WR009505

Mohseni O, Stefan HG, Erickson TR. 1998. A nonlinear regression model for weekly stream temperatures. *Water Resources Research* **34**: 2685–2692. DOI:10.1029/98WR01877

Mohseni O, Stefan HG. 1999. Stream temperature/air temperature relationship: a physical interpretation. *Journal of Hydrology* **218**: 128–141. DOI:10.1016/S0022-1694(99)00034-7

Morrill JC, Bales RC, Conklin MH. 2005. Estimating stream temperature from air temperature: implications for future water quality. *Journal of Hydraulic Engineering ASCE* **131**: 139–146. DOI:10.1061/(ASCE)0733-9372(2005)131:1(139)

Nash JE, Sutcliffe JV. 1970. River flow forecasting through conceptual models 1. A discussion of principles. *Journal of Hydrology* **10**: 282–290. DOI:10.1016/0022-1694(70)90255-6

North RP, Livingstone DM, Hari RE, Köster O, Niederhauser P, Kipfer R. 2013. The physical impact of the late 1980s climate regime shift on Swiss rivers and lakes. *Inland Waters* **3**: 341–350. DOI:10.5268/IW-3.3.560

Piccolroaz S, Toffolon M, Majone B. 2013. A simple lumped model to convert air temperature into surface water temperature in lakes. *Hydrology and Earth System Sciences* **17**: 3323–3338. DOI:10.5194/hess-17-3323-2013

Piccolroaz S, Toffolon M, Majone B. 2015. The role of stratification on lakes' thermal response: the case of Lake Superior. *Water Resources Research* **51**: 7878–7894. DOI:10.1002/2014WR016555

Piccolroaz S. 2016. Prediction of lake surface temperature using the air2water model: guidelines, challenges, and future perspectives. *Advances in Oceanography and Limnology* **7**: 36–50. DOI:10.4081/aiol.2016.5791

Pilgrim JM, Fang X, Stefan HG. 1998. Stream temperature correlations with air temperatures in Minnesota: implications for climate warming. *Journal of the American Water Resources Association* **34**: 1109–1121. DOI:10.1111/j.1752-1688.1998.tb04158.x

Schaefli B, Gupta HV. 2007. Do Nash values have value? *Hydrological Processes* **21**: 2075–2080. DOI:10.1002/hyp.6825

Seibert J. 2003. Reliability of model predictions outside calibration conditions. *Nordic Hydrology* **34**: 477–492. DOI:10.2166/nh.2003.028

Sinokrot BA, Stefan HG. 1993. Stream temperature dynamics: measurements and modelling. *Water Resources Research* **29**: 2299–2312. DOI:10.1029/93WR00540

Siviglia A, Toro EF. 2009. The WAF method and splitting procedure for simulating hydro and thermal peaking waves in open channel flows. *Journal of Hydraulic Engineering ASCE* **135**: 651–662. DOI:10.1061/(ASCE)HY.1943-7900.0000054

Smakhtin VU. 2001. Low flow hydrology: a review. *Journal of Hydrology* **240**: 147–186. DOI:10.1016/S0022-1694(00)00340-1

Stefan HG, Preud'homme EB. 1993. Stream temperature estimation from air temperature. *Journal of the American Water Resources Association* **29**: 27–45. DOI:10.1111/j.1752-1688.1993.tb01502.x

Toffolon M, Siviglia A, Zolezzi G. 2010. Thermal wave dynamics in rivers affected by hydropeaking. *Water Resources Research* **46**: W08536. DOI:10.1029/2009WR008234

Toffolon M, Piccolroaz S, Majone B, Soja AM, Peeters F, Schmid M, Wüest A. 2014. Prediction of surface temperature in lakes with different morphology using air temperature. *Limnology and Oceanography* **59**: 2185–2202. DOI:10.4319/lo.2014.59.6.2185

Toffolon M, Piccolroaz S. 2015. A hybrid model for river water temperature as a function of air temperature and discharge. *Environmental Research Letters* **10**: 114011. DOI:10.1088/1748-9326/10/11/114011

Webb BW, Clark PD, Walling DE. 2003. Water–air temperature relationships in a Devon river system and the role of flow. *Hydrological Processes* **17**: 3069–3084. DOI:10.1002/hyp.1280

Webb BW, Hannah DM, Moore RD, Brown LE, Nobilis F. 2008. Recent advances in stream and river temperature research. *Hydrological Processes* **22**: 902–918. DOI:10.1002/hyp.6994

Younus M, Hondzo M, Engel BA. 2000. Stream temperature dynamics in upland agricultural watersheds. *Journal of Environmental Engineering, ASCE* **126**: 518–526. DOI:10.1061/(ASCE)0733-9372(2000)126:6(518)

Vanzo D, Siviglia A, Carolli M, Zolezzi G. 2015. Characterization of sub-daily thermal regime in alpine rivers: quantification of alterations induced by hydropeaking. *Hydrological Processes* (in press) DOI:10.1002/hyp.10682

Zolezzi G, Bellin A, Bruno MC, Maiolini B, Siviglia A. 2009. Integrating standard and novel approaches to assess hydrological alterations due to hydropower production at multiple time scale: Adige River, Italy. *Water Resources Research* **45**: W12421. DOI:10.1029/2008WR007266

Zolezzi G, Siviglia A, Toffolon M, Maiolini B. 2011. Thermopeaking in Alpine streams: event characterization and time scales. *Ecohydrology* **4**: 564–576. DOI:10.1002/eco.132

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site.

Copyright © 2016 John Wiley & Sons, Ltd.