*Water Resources Research*

Supporting Information for

# Modeling Seasonal Effects of River Flow on Water Temperatures in an Agriculturally Dominated California River

J. Eli Asarian[1], Crystal Robinson[2†], Laurel Genzoli[3]

[1]Riverbend Sciences, Eureka, CA, USA, [2]Quartz Valley Indian Reservation, Fort Jones, CA, USA, [3]Flathead Lake Biological Station and the University of Montana in Missoula, MT, USA, [†]current affiliation: California Department of Fish and Wildlife, Yreka, CA, USA

## Contents of this file



## Introduction

The supporting text contains methodological details on quality control procedures and combining of stream temperature data from multiple entities (Text S1), choosing autocorrelation coefficients in the GAMs and why we included a random effect for year (Text S3, Figures S12, S13, S14), and additional discussion of the 22 °C salmonid temperature threshold (Text S4) that were excluded from the manuscript for the sake of brevity. A sensitivity analysis on the effects of using different methods for summarizing air temperatures is provided in Text S2 and Figure S11. Text S6 discusses the modal relationship between flow and stream temperature at some sites during the October–November period. Supporting figures include additional outputs from the stream temperature models, including time series plots comparing modeled data to observed data (Figures S5, S6, and S7), a GAM smoother plot (Figure S3), Bayesian information criteria scores (Figure S4 and accompanying Text S5), daily outputs from model scenarios (Figure S8); and standardized flow coefficients from previous regional studies (Figure S9, S10). Table S1 provides site characteristics and data sources for stream temperature modeling sites. Table S2 lists mean ranks for each model.

**Text S1.**

Primary quality control was conducted by the entities who collected the stream temperature data. These entities check probe calibration before and after every deployment, and data not meeting calibration criteria are discarded. In addition, we reviewed the data and removed any suspicious values (e.g., when there were calibration issues or probes appear to have been exposed to air). The Quartz Valley Indian Reservation (QVIR) Environmental Department uses YSI (Yellow Springs, Ohio) 6600 multi-parameter datasondes to monitor Scott River water temperatures at the U.S. Geological Survey (USGS) gage 11519500 near the outlet of Scott Valley (QVIR, 2016; Asarian et al., 2020), recording temperature measurements every 30 minutes with a reported accuracy of ±0.15 °C. The YSI 6600 multi-parameter datasondes do not require calibration but are compared to a reference sonde every two weeks and serviced by the manufacturer annually (QVIR, 2016). KNF's stream temperature monitoring equipment has changed over time, but calibration and deployment protocol has remained similar with pre- and post-deployment testing against a National Institute of Standards and Technology (NIST) traceable thermometer (KNF, 2010, 2011). Since 2010, KNF has used ONSET Pro v2 data logger u22-001 for all temperature monitoring (KNF, 2011). Prior to 2010 KNF used a combination of ONSET Pro v2 u22-001, Optic StowAway, and other ONSET temperature logger models. USFWS protocols are described by Romberger & Gwozdz (2018). USBR data were subjected to a detailed quality control review by USGS prior to inclusion in the database from which we accessed them (Smith et al., 2018).

For days on which Scott River daily stream temperatures were available from multiple entities, we averaged the values together. For the 1216 days with both QVIR and USFS records, root mean standard error (RMSE) was 0.31 °C and 0.18 °C for $T_{max}$ and $T_{mean}$, and respectively. Only one entity collected data at each of the other nine sites, so averaging values was unnecessary there.

**Text S2.**

At the beginning of this project, we only modeled stream temperatures at the Scott River site. Our final analyses at all 10 Klamath Basin sites use a 2-day weighted average air temperatures ($A_{2w}$) from the gridded PRISM air temperature dataset (Daly et al., 2008); however, for the initial Scott River analyses, we used daily mean air temperature data from USFS' Quartz Hill weather station (Global Historical Climatology Network - Daily [GHCND] station USR0000CQUA; Menne et al., 2012a, 2012b) located approximately 8 km southeast of the stream temperature gage, with missing values infilled by linear regression with nearby weather stations or PRISM. In initial explorations of Scott River stream temperature models, we explored many air temperature metrics including multi-day averages (Webb et al., 2003; Siegel et al. 2022), exponential weights (Koch & Grünewald, 2010; Piotrowski & Napiorkowski, 2019; Soto, 2016), and including the day of interest and preceding days separately (Siegel & Volk 2019). These explorations tested

five categories of air temperature metrics, where $A_i$ is the mean air temperature on the day *i*, using Equations (1), (2), (3), (4), and (5):

Single-day average $A_1$:

$$A_1 = A_i \quad (1)$$

Multi-day averages $A_2 \ldots A_7$:

$$A_2 = \frac{(A_i + A_{i-1})}{2}, \ldots, A_7 = \frac{(A_i + A_{i-2} \ldots A_{i-6})}{7} \quad (2)$$

Multi-day weighted averages $A_{2w}$ and $A_{3w}$, with preceding days discounted by 50% per day:

$$A_{2w} = \frac{A_i + (0.5 \times A_{i-1})}{1.5} \text{ and } A_{3w} = \frac{(A_i + 0.5A_{i-1} + 0.25A_{i-2})}{1.75} \quad (3)$$

Lagged averages $A_{L3}$ and $A_{L5}$:

$$A_{L3} = \frac{(A_{i-1} + A_{i-2} + A_{i-3})}{3} \quad \text{and} \quad A_{L5} = \frac{(A_{i-1} + A_{i-2} + A_{i-3} + A_{i-4} + A_{i-5})}{5} \quad (4)$$

Differences between lagged average and day *i*:

$$A_{\Delta 3} = (A_i - A_{L3}) \quad \text{and} \quad A_{\Delta 5} = (A_i - A_{L5}) \quad (5)$$

These initial Scott River explorations, using a model structure similar to GAM4 (tensors for Q-D and A2w-D), indicated that the 2-day weighted air temperature ($A_{2w}$) had excellent performance for predicting both $T_{max}$ and $T_{mean}$, so we proceeded to use $A_{2w}$ for all subsequent stream temperature models except one that uses a seven-day average ($A_7$) (Section 3.2).

After completing our final modeling at all 10 sites using PRISM $A_{2w}$ (or $A_7$) and selecting our final model GAM7, we did a sensitivity analysis comparing performance of variants of Scott River GAM7 using the same air temperature summaries that were initially tested, except this time using data from PRISM instead of the local GHCND weather station measurements. Interestingly, the results of this GAM7 PRISM sensitivity analysis (Figure S11) differed from the initial GAM4 GHCND sensitivity analysis (not shown here), with the single-day average $A_1$ performing better (i.e., lower RMSE and BIC) than $A_{2w}$. Surprised, we explored further (i.e., ran a similar sensitivity analysis on GAM4 PRISM, results not shown here) and determined that which air temperature summary worked the best (i.e., $A_1$ or $A_{2w}$) was not due to differences between the modeling structure of GAM4 and GAM7 (i.e., tensors or non-linear smoothers, etc.) but rather between the PRISM data and

GHCND data. We speculate, but did not confirm, that this may be due to differences in how days are defined between PRISM and GHCND. In summarizing daily stream temperatures, we defined days as midnight-to-midnight local time, but PRISM days are defined at 1200–1200 UTC (e.g., 0700–0700 EST) and stations with reporting times (i.e., day definition) within four hours are used as inputs to PRISM (Daly et al. 2021). We could not readily ascertain the reporting time for the GHCND station we used.

**Text S3.**

The bam in mgcv function cannot automatically derive the AR-1 coefficient (rho), so it must be manually assigned. Following Baayen et al. (2018) and van Rij et al. (2019, 2020), we initially fit each model without an autocorrelation term, and then re-ran the model with an autocorrelation term, assigning a rho value based on the lag 1 autocorrelation from the residuals of the initial model. Comparing models fit using fast restricted maximum likelihood (fREML) with a range of rho values, as recommended by Baayen et al. (2018), van Rij et al. (2019), and Wood (2017), confirmed these initial values were reasonable. These tests indicated that rho values that minimized fREML scores were 0.02–0.16 higher than the initial rho values (Figure S12 shows example of Scott River GAM7, Figure S13 shows all models for all sites). However, autocorrelation function (ACF) plots indicated that these higher rho values often had the undesirable side effect of exacerbating the negative autocorrelation at lag 1 or lag 2 (e.g., Figure S14 shows example of Scott River GAM7), leading to our decision to use the initial rho values instead. BIC scores, included as a supplementary measure of model fit, show the same pattern as fREML scores regarding optimal rho values (Figure S12). Using BIC scores to assess optimal rho values in fREML-fit models is acceptable because the models compared had the same fixed effects and differed only in their rho values.

A random effect for year was included to account for year-to-year variability in other factors not included in the models such as changes in channel morphology or riparian vegetation. From a statistical perspective, including a random effect for year is beneficial because it helps reduce temporal autocorrelation within years that arises from a combination of the natural hierarchical structure of both the physical system (i.e., see previous sentence) and how the data were collected. For example, some sites and years have data for summer only (or other periods that do not span across multiple years), so for those years the random effect would account for differences in the exact placement of the temperature probe and/or any bias in the probe itself. However, we acknowledge for those sites and years when data were collected year-round and the probes were visited multiple times per year, year would be less of a natural break.

**Text S4.**

We chose 22 °C as an indicator of biological effects on juvenile salmonids that rear in the mainstem Scott River or outmigrate downstream using the river as a migratory corridor.

Given the potential for local genetic adaptation to thermal regimes (Zillig et al., 2021), we prioritized geographically proximal studies in selecting thresholds. When the Klamath River exceeds 22–23 °C, juvenile salmonids move to tributary confluences (Brewitt & Danner, 2014; Sutton & Soto, 2012; Sutton et al., 2007). Similar behavior was observed in the Shasta River (Nichols et al., 2014) and 22 °C was also used by McGrath et al. (2017). In recognition of our study site's location on a mainstem river where temperatures would naturally be higher than a small well-shaded or spring-fed tributary, we chose 22 °C over colder thresholds that would more fully protect coho salmon like Stenhouse et al.'s (2012) recommendation of 15.5 °C for spring-fed tributaries to the Shasta River or Welsh et al.'s (2001) 18 °C maximum weekly maximum temperature (MWMT) derived from coastal streams. In addition, juvenile coho salmon grew fast in experimental cages in the food-rich Shasta River with MWMT as high as 24.0 °C, although survival was higher at cooler sites (Lusardi et al., 2019). Our data and code are public, so future researchers could choose a different threshold. Recognizing the drawbacks of any single statistic or threshold (Steel et al., 2013), we also examine annual maximum temperature.

**Text S5.**

BIC scores (Figure S4) largely corroborate the extrapolation CV results identifying the importance of seasonally varying flow effects. Of eight models with seasonally varying flow effects, the most complex model (three-way tensor GAM1) had the worst overall (averaged across all sites) BIC rank, but intermediate extrapolation CV RMSE. Averaging BIC ranks across sites, our extrapolation CV-selected model, GAM7, had the best BIC ranks for both Tmax and Tmean (Figure S4); however, at many individual sites including Scott River, other models had better BIC scores (Figure S4).

**Text S6.**

At Scott River (Figure 6) and two other sites (Figure S15), the modeled flow-temperature relationship is modal (i.e., highest water temperatures at moderate flows) instead of monotonic in October–November, a period of hydrologic transition when precipitation ends seasonal baseflow recession, increases flow, and refills the valley aquifer (Figure 1). The reasons for this non-monotonic behavior are unclear, but could reflect processes such as groundwater-surface water dynamics, variation in timing of fall precipitation, or other seasonal variables; regardless, these departures from monotonic are of low consequence because they are <1 °C and occur when temperatures are not a biological concern.

**Table S1.** Site characteristics and data sources for stream temperature modeling sites. Drainage areas are from NHDPlus version 2.1 (Moore & Dewald, 2016). Key to abbreviations: CDWR = California Department of Water Resources, QVIR = Quartz Valley Indian Reservation, USBR = U.S. Bureau of Reclamation, USFWS = U.S. Fish and Wildlife Service, USFS KNF = United States Forest Service Klamath National Forest, and USGS = U.S. Geological Survey.

| | | Stream temperature data | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Site number and name of USGS flow gage | Drainage area ($km^2$) | Data source | Original site code | N. of days | N. of years | Date range | Latitude | Longitude | Notes |
| 11530500 Klamath R Nr Klamath CA | 34550 | USFWS | KRTG2 | 5002 | 16 | 2004–2019 | 41.51118 | -123.97844 | |
| 11523000 Klamath R A Orleans | 25159 | USFWS | KROR1 | 4138 | 17 | 2001–2018 | 41.30358 | -123.53439 | |
| 11520500 Klamath R Nr Seiad Valley CA | 21171 | USFWS | KRSV1 | 5684 | 19 | 2001–2019 | 41.85409 | -123.23147 | |
| 11528700 SF Trinity R Bl Hyampom | 2414 | USFWS | SFTR1 | 4627 | 19 | 2001–2019 | 40.88943 | -123.60221 | Temperature monitoring site located at confluence with Trinity River, 42.5 km downstream of the USGS gage |
| 11522500 Salmon R A Somes Bar CA | 1946 | CDWR | F3410000 | 5200 | 18 | 2002–2019 | 41.37695 | -123.47736 | |
| 11517500 Shasta R Nr Yreka CA | 1934 | USFWS | SHKR1 | 5172 | 18 | 2001–2019 | 41.82476 | -122.59392 | |
| 11519500 Scott R Nr Fort Jones CA | 1716 | QVIR | SRGA | 3180 | 13 | 2007–2020 | 41.64000 | -123.01380 | |
| | | USFS KNF | H2O_Temp_LOCID103 | 977 | 8 | 2006–2016 | | | |
| | | USFS KNF | H2O_Temp_ScottNearFtJones | 1048 | 3 | 2009–2011 | | | |
| | | USBR | 11519500 | 682 | 3 | 1998–2000 | | | |
| | | USFS KNF | Scott River at USGS Gage | 341 | 3 | 2003–2019 | | | |
| | | USFS KNF | H2O_Temp_LOCID224 | 118 | 1 | 2004–2004 | | | |
| 11521500 Indian C Nr Happy Camp | 310 | USFS KNF | H2O_Temp_LOCID056 | 3540 | 17 | 2000–2016 | 41.83525 | -123.38291 | |
| 11525670 Indian C Nr Douglas City CA | 87 | USFWS | ICTR1 | 5197 | 18 | 2002–2019 | 40.65645 | -122.91388 | |
| 11525530 Rush C Nr Lewiston CA | 58 | USBR/USGS | RCL | 5679 | 18 | 2001–2019 | 40.72500 | -122.83400 | |

**Table S2.** Overall model ranks from extrapolation cross-validation tests, each calculated as mean RMSE rank of all 10 sites and both temperature response variables ($T_{max}$ and $T_{mean}$). See Figure 4 for model formulas and a key to abbreviations.

| Model number and name | Mean rank RMSE |
|---|---|
| GAM7: vary Q & A2w (final) | 3.60 |
| GAM2: tensors Q-D & A2w-D | 3.65 |
| GAM4: tensors Q-D & A2w-Q | 3.65 |
| GAM3: tensor Q-D & vary A2w | 3.80 |
| GAM1: tensor Q-A2w-D | 4.10 |
| GAM5: tensor Q-D no vary A2w | 5.10 |
| GAM8: vary Q & no vary A2w | 5.70 |
| GAM6: vary Q & A2w linear | 6.55 |
| GAM10: A2w no Q or vary | 9.40 |
| GAM9: A2w no vary | 9.45 |
| GAM11: A7 only no AR1 | 11.00 |

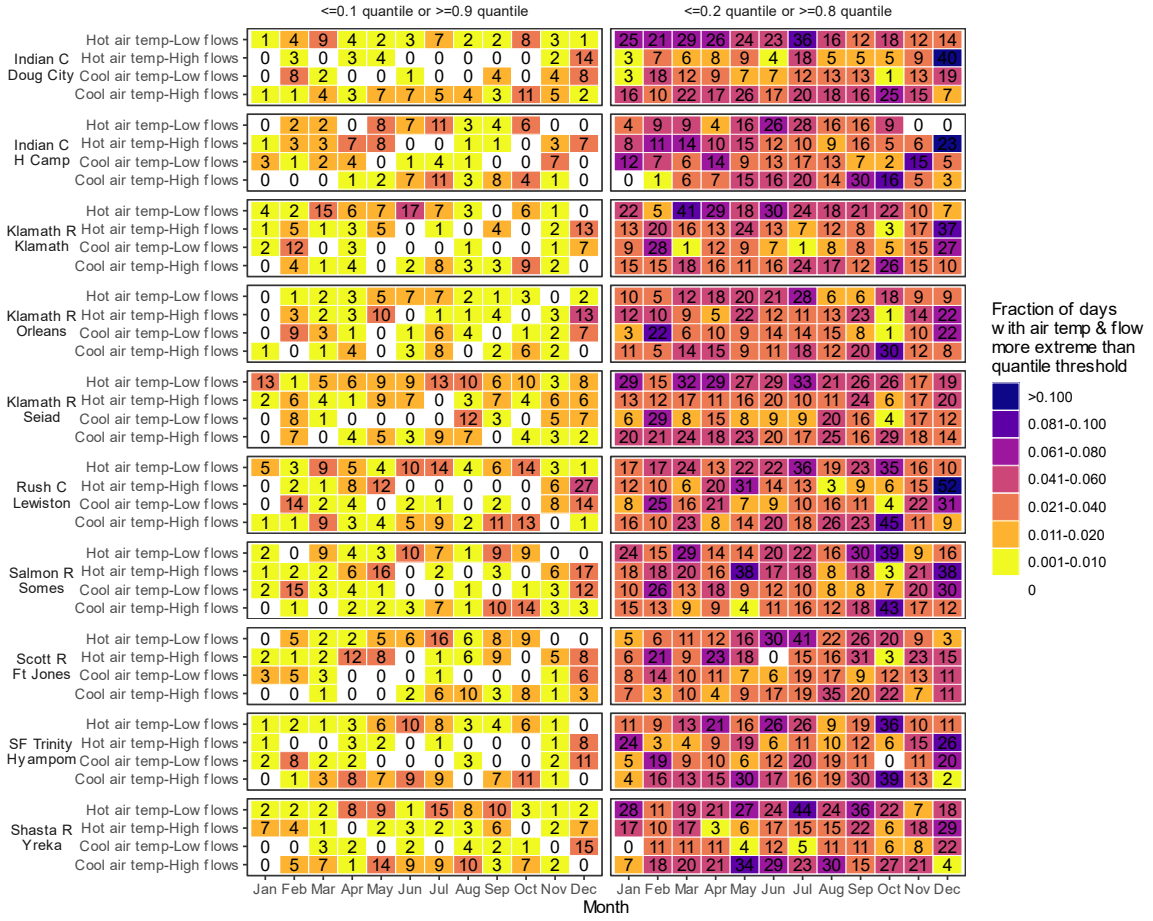


**Figure S1.** Availability of measured water temperature data for days when extreme quantiles of air temperature and flow co-occur. Shading indicates the fraction of days for each site and month when air temperatures and flow were more extreme than the quantile threshold (≤0.1 and ≥0.9 for left panels, ≤0.2 and ≥0.8 for right panels). Data labels inside each square indicate the total number of days exceeding the quantile threshold. For example, in July at Scott River there were 16 days (2.8% of the 572 days when water temperature data were available for that site and month) when air temperatures were ≥0.9 quantile and flows were ≤0.1 quantile.

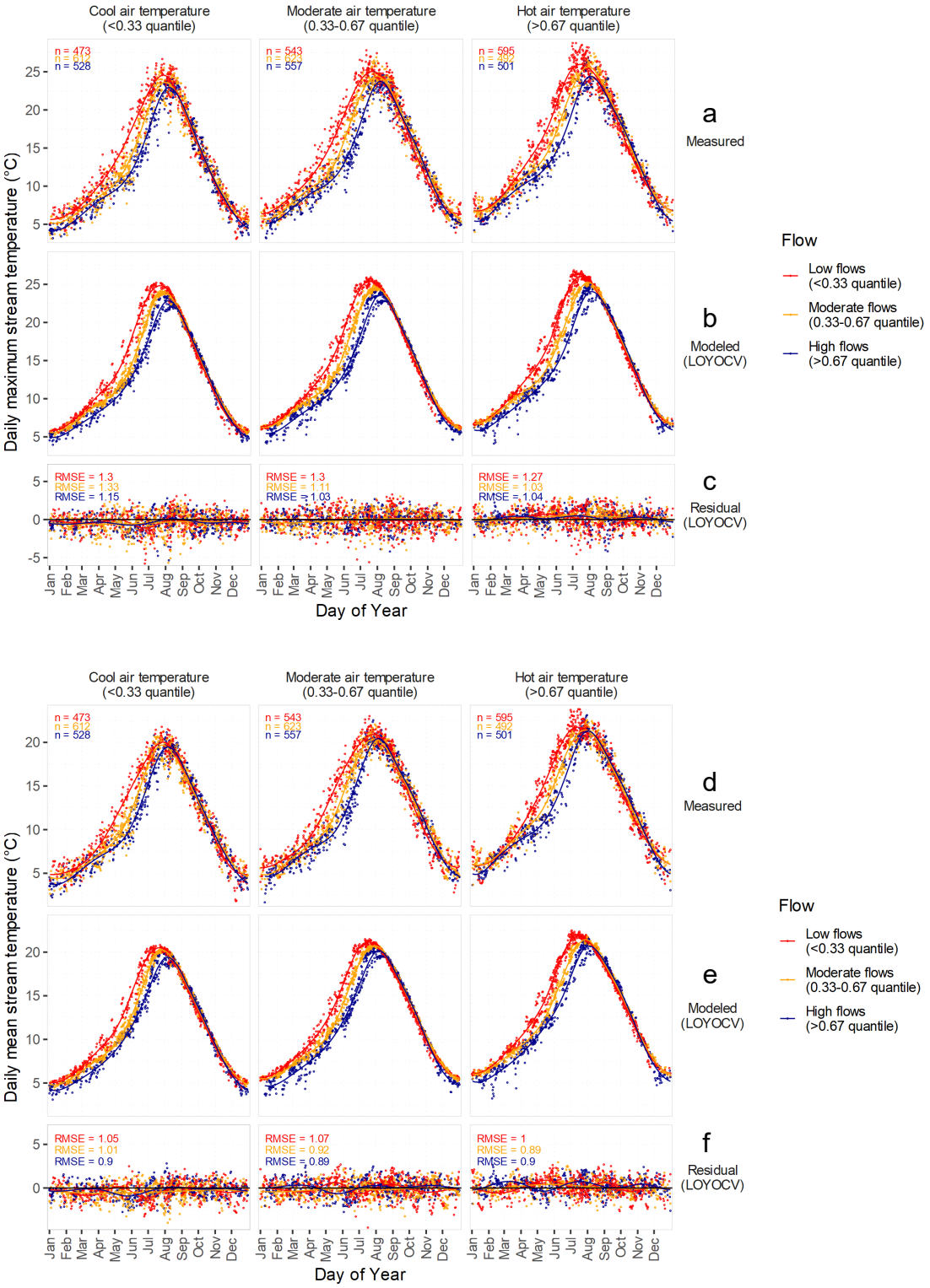


**Figure S2.** Measured (a) $T_{max}$ and (d) $T_{mean}$ at Scott River for dates with combinations of cool, typical, or hot air temperatures (arranged in columns) and low, typical, or high flows (shown by color). (b,e) Modeled LOYO CV temperatures predicted by selected model GAM7 for the same dates, and (c,f) LOYO CV residuals, calculated as measured minus modeled. Lines are GAM smoothers fit to points, shown as visual aids.

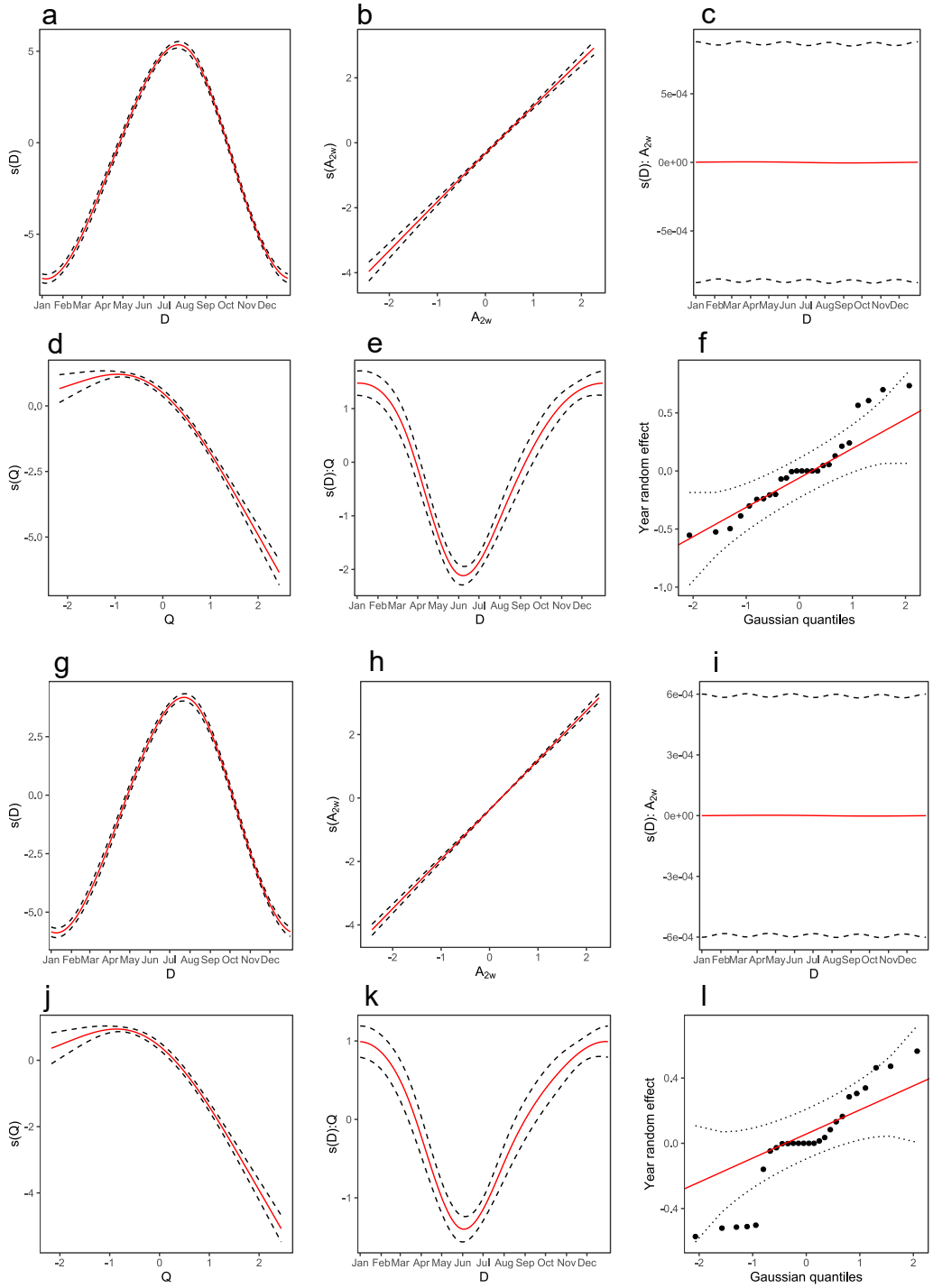


**Figure S3.** GAM smooths (i.e., covariate responses and interactions) from Scott River model GAM7 for $T_{max}$ (top six panels) and $T_{mean}$ (bottom six panels) showing partial effects of smooth functions of: (a,g) day of year D, (b,h) two-day air temperature A2w, (c,i) interaction of $A_{2w}$ and D (i.e, slope of $A_{2w}$ varying as non-linear function of D), (d,j) flow Q, and (e,k) interaction of Q and D. Dashed lines are 95% confidence intervals. (f,l) shows random effects for year.

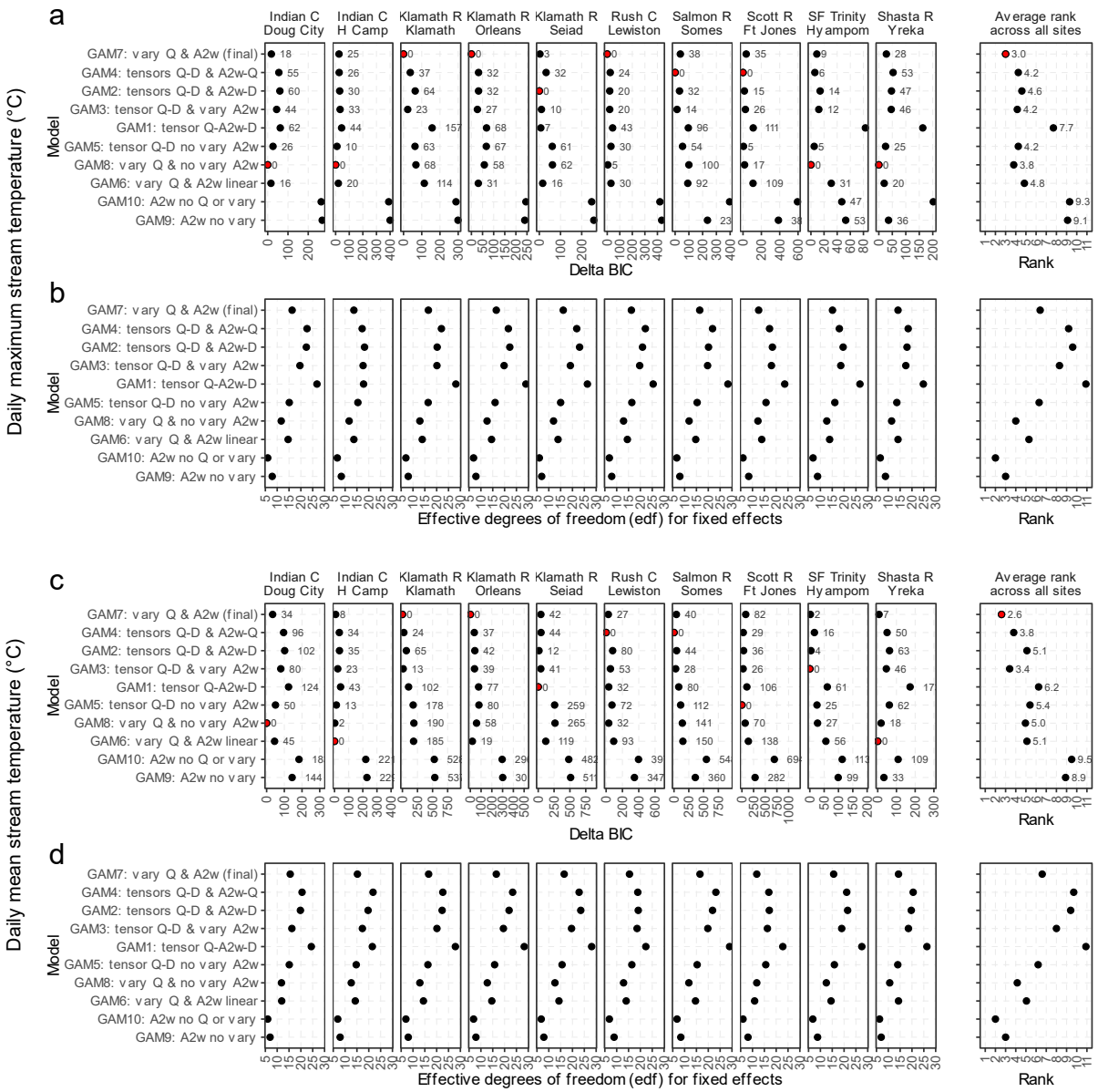


**Figure S4.** Comparison of (a,c) delta BIC, and (b,d) effective degrees of freedom (edf) for models of (a,b) $T_{max}$ and (c,d) $T_{mean}$ at 10 sites in the Klamath Basin. Symbols for models with lowest delta BIC are colored red. Models are sorted in same order as in Figure 4 (i.e., by overall RMSE rank). Average ranks in right column were calculated by first ranking model scores within each site (i.e., 1=best, 11=worst), then averaging those model ranks across sites. Model GAM11 was excluded from this figure because its model fit was so poor it would expand the axes making it difficult to see differences between the other models.

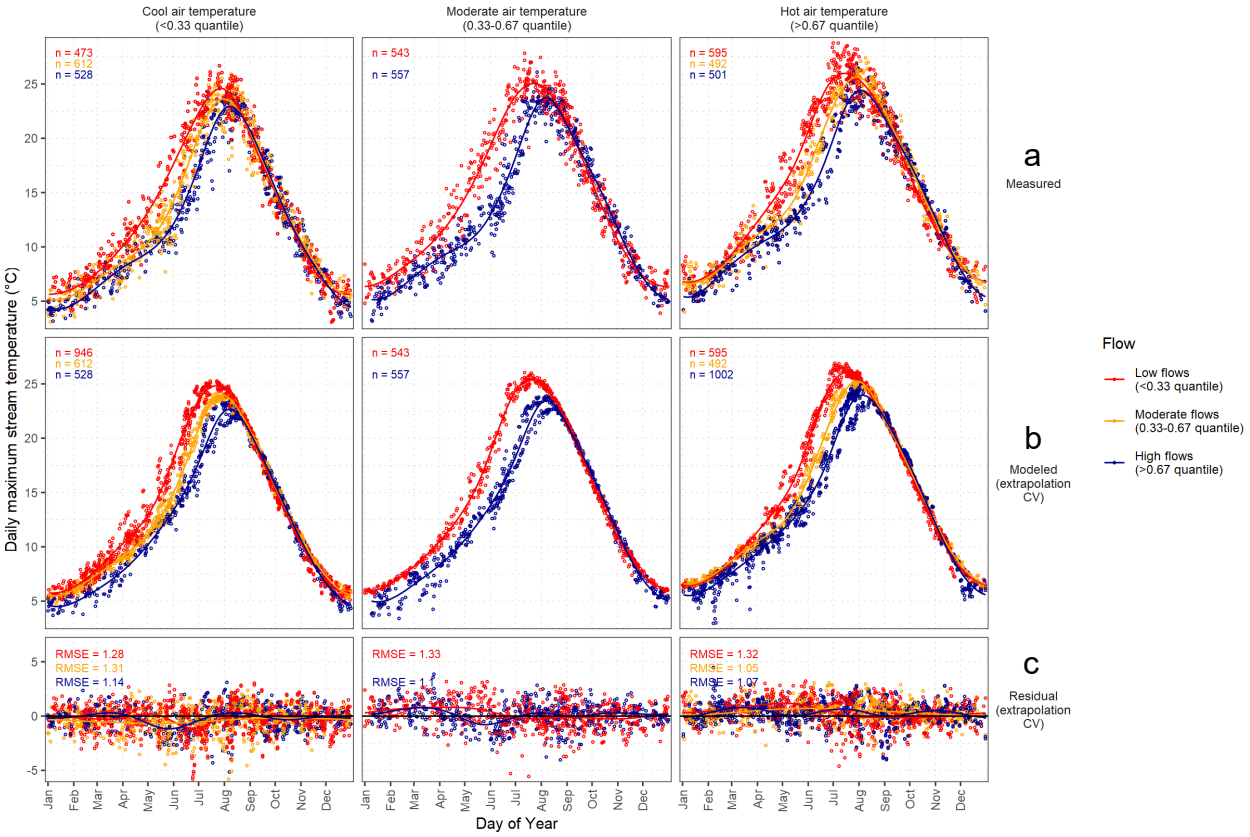


**Figure S5.** (a) Measured $T_{max}$ at Scott River for dates with combinations of cool, moderate, or hot air temperatures (arranged in columns) and low, moderate, or high flows (shown by color). (b) Modeled extrapolation CV temperatures predicted by the selected model 'GAM7' for the same dates, and (c) extrapolation CV residuals, calculated as measured minus modeled. Lines are GAM smoothers fit to the points, shown as visual aids.

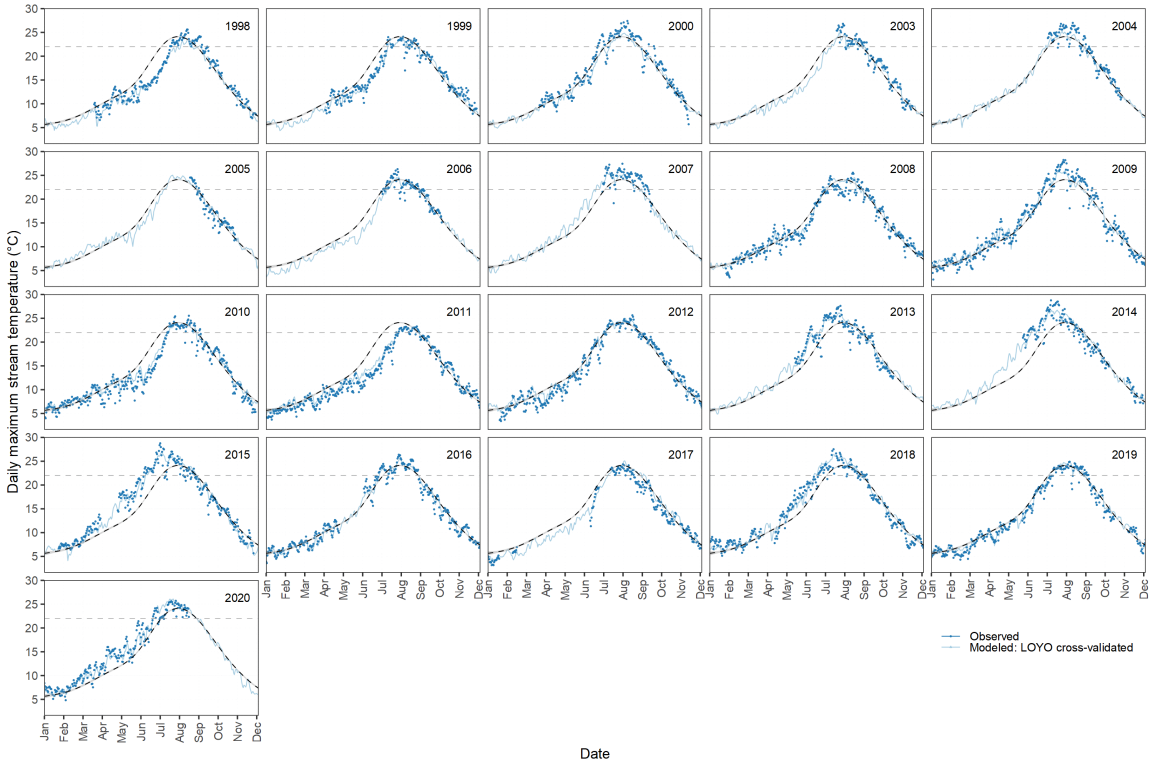


**Figure S6.** Daily time series of measured (dots) and modeled (solid lines, from leave-one-year-out [LOYO] cross-validation) $T_{max}$ in the Scott River at the USGS gage for the years 1998–2020 (no data 2001-2002). Horizontal dashed gray line at 22 °C indicates a temperature threshold for juvenile salmonids. Curved black dashed line is GAM smoother of all measured $T_{max}$ for all years 1998-2020, indicating typical conditions for each day of year.

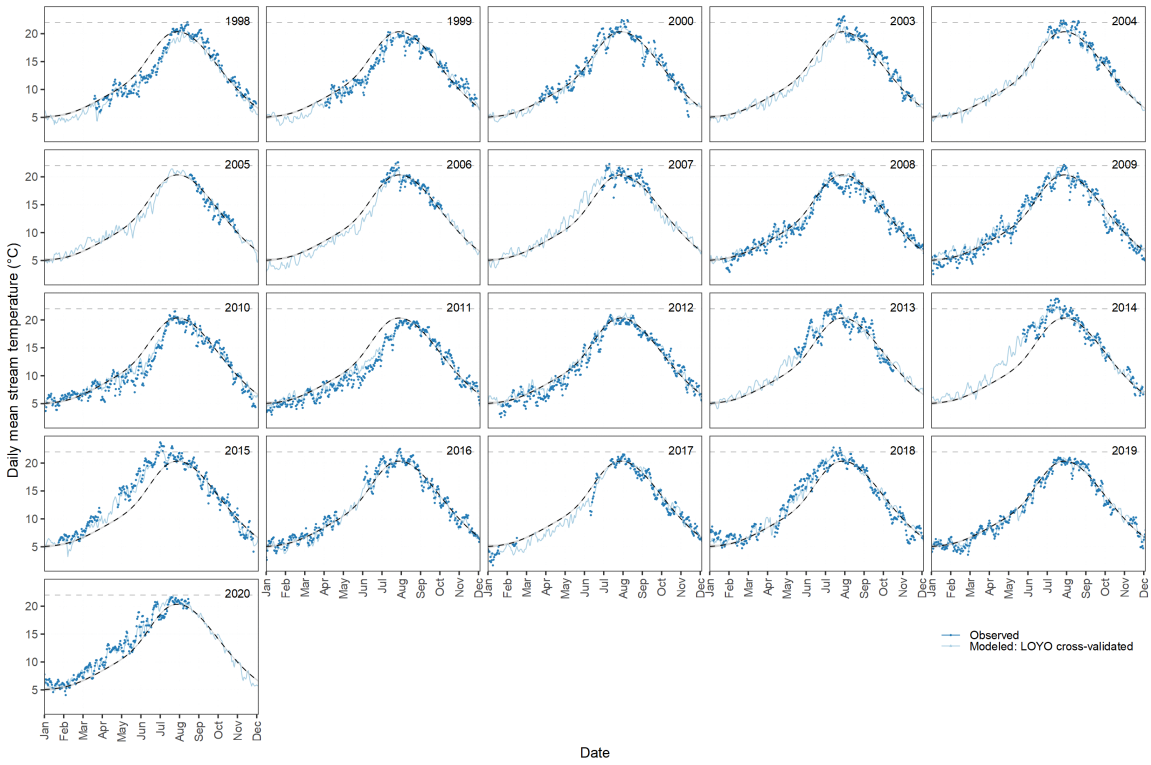


**Figure S7.** Daily time series of measured (dots) and modeled (solid lines, from leave-one-out [LOYO] cross-validation) $T_{mean}$ in the Scott River at the USGS gage for the years 1998–2020 (no data 2001-2002). Horizontal dashed gray line at 22 °C indicates a temperature threshold for juvenile salmonids. Curved black dashed line is GAM smoother of all measured $T_{mean}$ for all years 1998-2020, indicating typical conditions for each day of year.

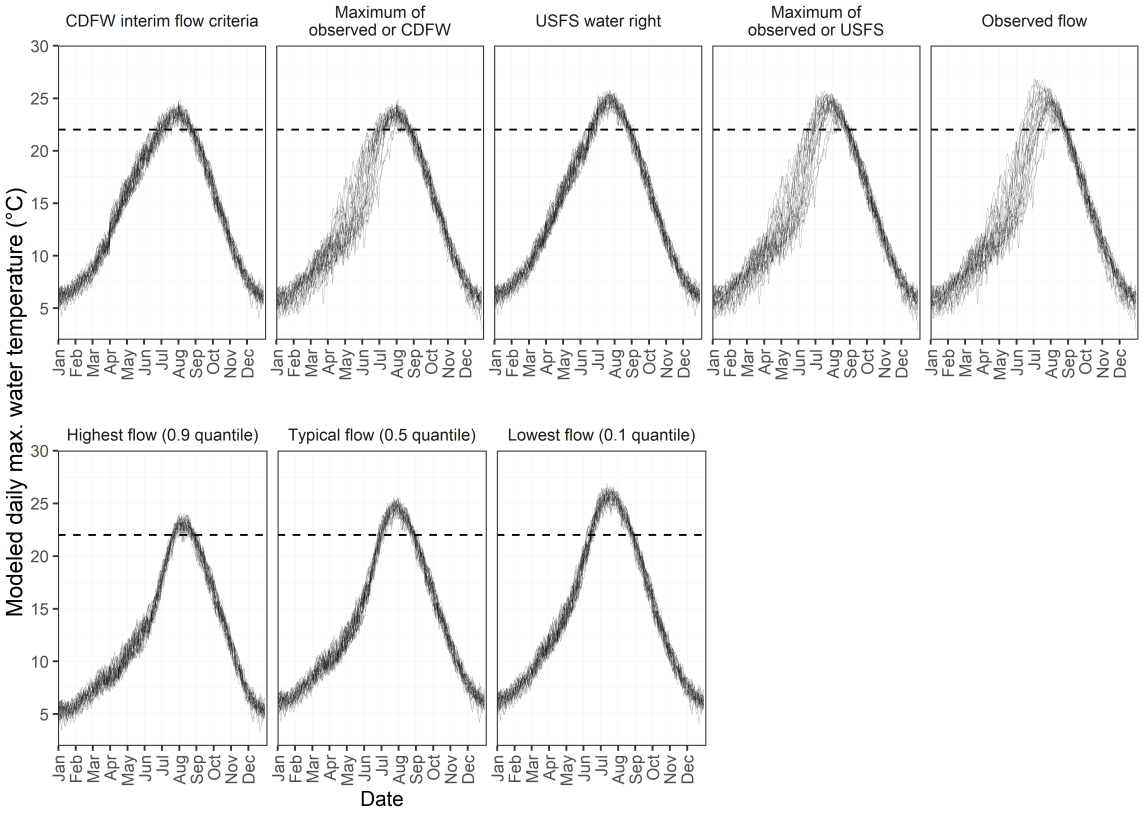


**Figure S8.** Scott River $T_{max}$ predicted with a statistical model under the group of scenarios that pair observed air temperatures for 1998–2020 with eight different flow conditions (Table 1): observed time series of USGS measured flows, three quantile flow scenarios, and four flow scenarios based on the CDFW interim instream flow criteria and USFS water right. Two scenarios use the exact flows (based on month and day) specified in the CDFW flow criteria and USFS water right, while in the other two the CDFW and USFS flows were replaced by observed USGS flows on dates when the observed flows were higher than the management flows (Table 1). Horizontal dashed gray line at 22 °C indicates a temperature threshold for juvenile salmonids.

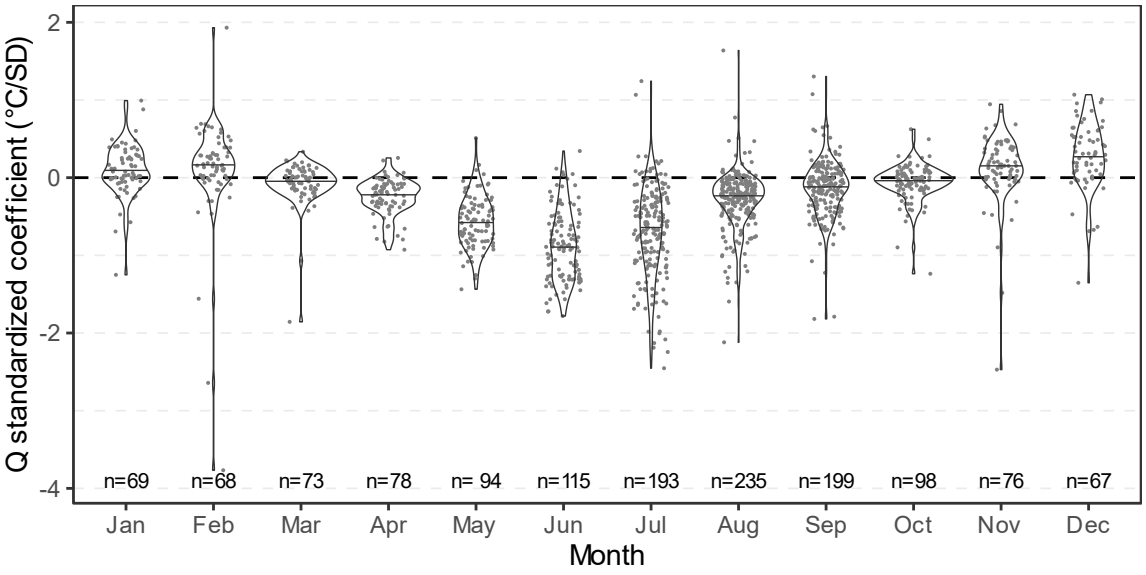


**Figure S9.** Violin plot (i.e., combination of box plot and density plot) of standardized coefficients for flow (Q) from multiple regression models of monthly stream temperatures at 239 river sites in the Northwestern U.S. where flow is not regulated by dams, from Isaak et al.'s (2018) analysis. Within each month, horizontal lines are median values, gray points are coefficients for individual sites (jittered for legibility), and labels are the number of sites. Isaak et al. (2018) developed these models in the original units of $m^3/s$. We obtained the coefficients from the study authors, converted the coefficients to standardized units by multiplying each coefficient by the standard deviation of Q for each month and site, and then created this figure.

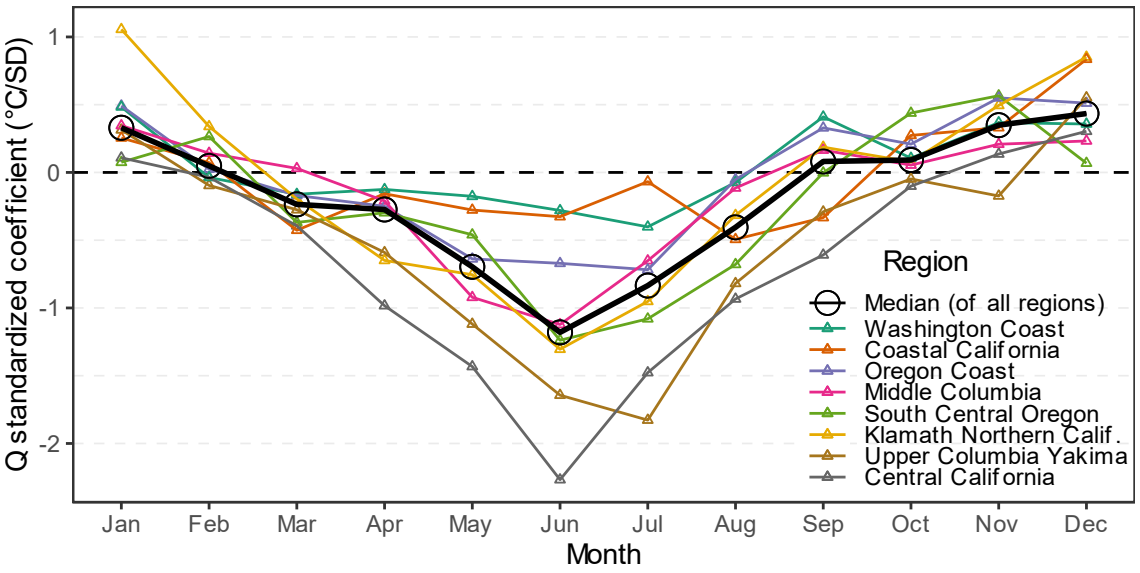


**Figure S10.** Standardized coefficients for flow (Q) from monthly spatial stream network models of stream temperature in eight Western U.S. regions, from FitzGerald et al.'s (2021) analysis. We created this figure using coefficients provided by the study authors.

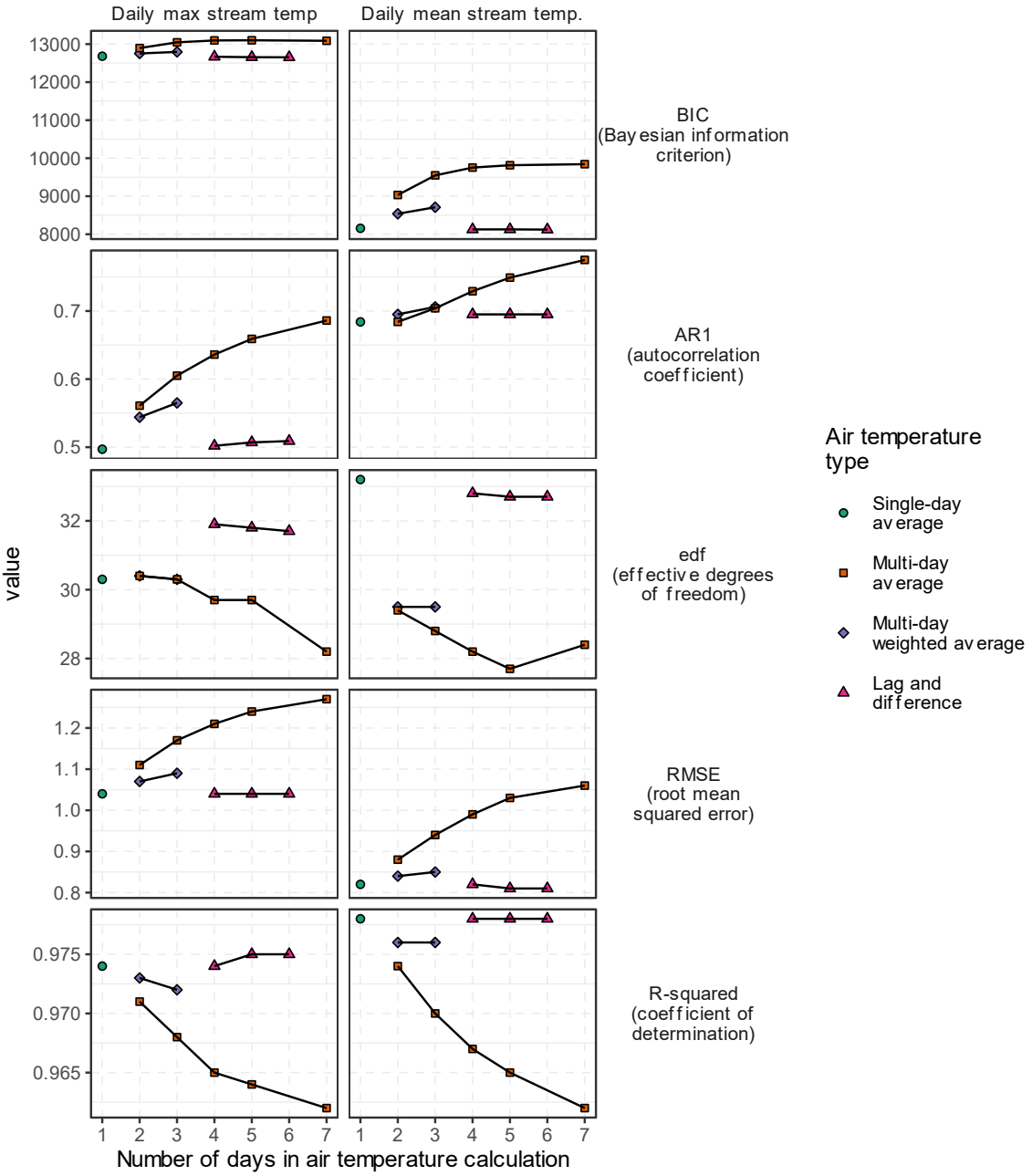


**Figure S11.** Effect of choice of air temperature metric on model training statistics, comparing 11 models of $T_{max}$ (left panels) and $T_{mean}$ (right panels). Models are alternative versions of the final model "GAM7", differing only in the choice of the air temperature metric.

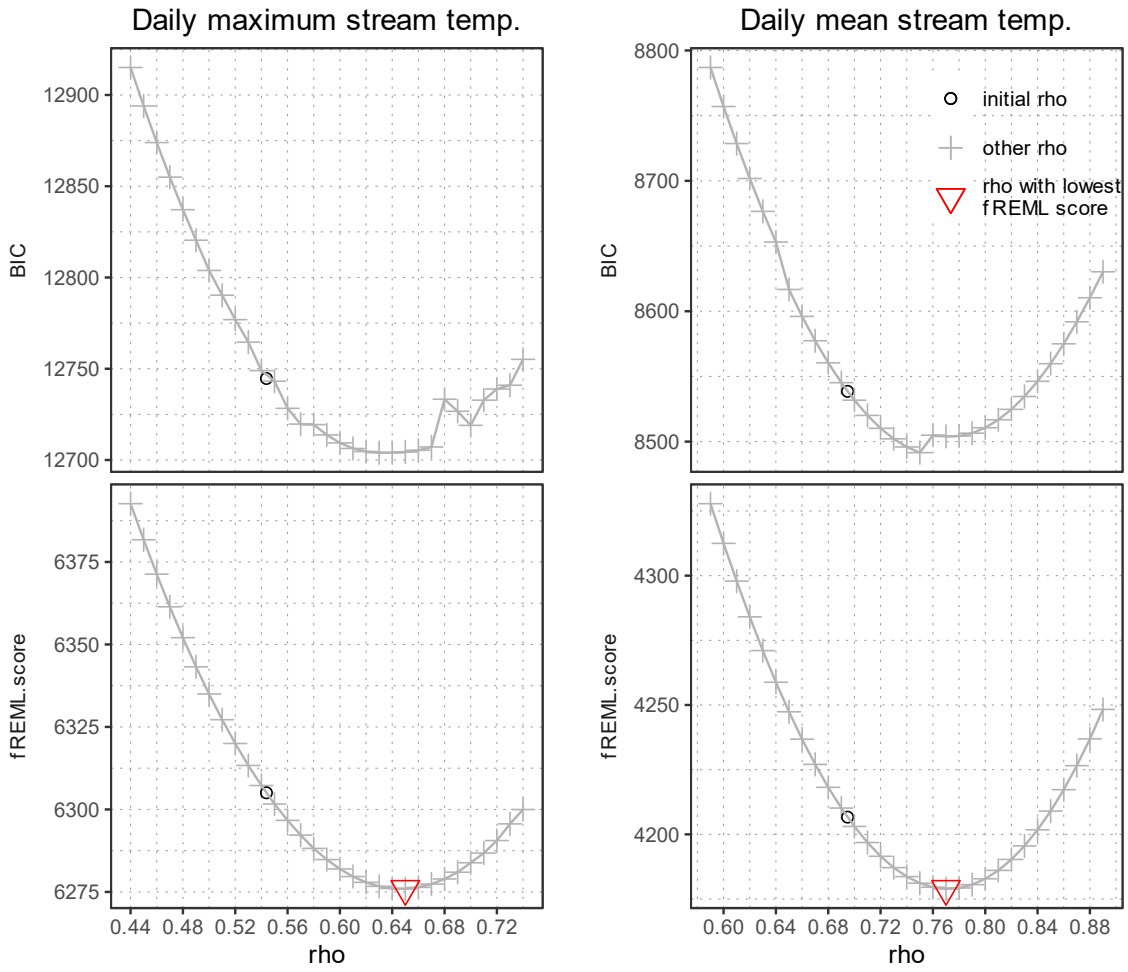


**Figure S12.** Plots comparing BIC (top panels), and fREML scores (bottom panels) for alternative versions of the Scott River final "GAM7" model for $T_{max}$ (left panels) and $T_{mean}$ (right panels) that use different autocorrelation values (i.e., rho, on x-axis). The "initial" rho value is the lag 1 autocorrelation value of the residuals from an initial model without autocorrelation. "Other" rho values range from 0.1 below the initial value to 0.2 above the initial value, in 0.01 increments.

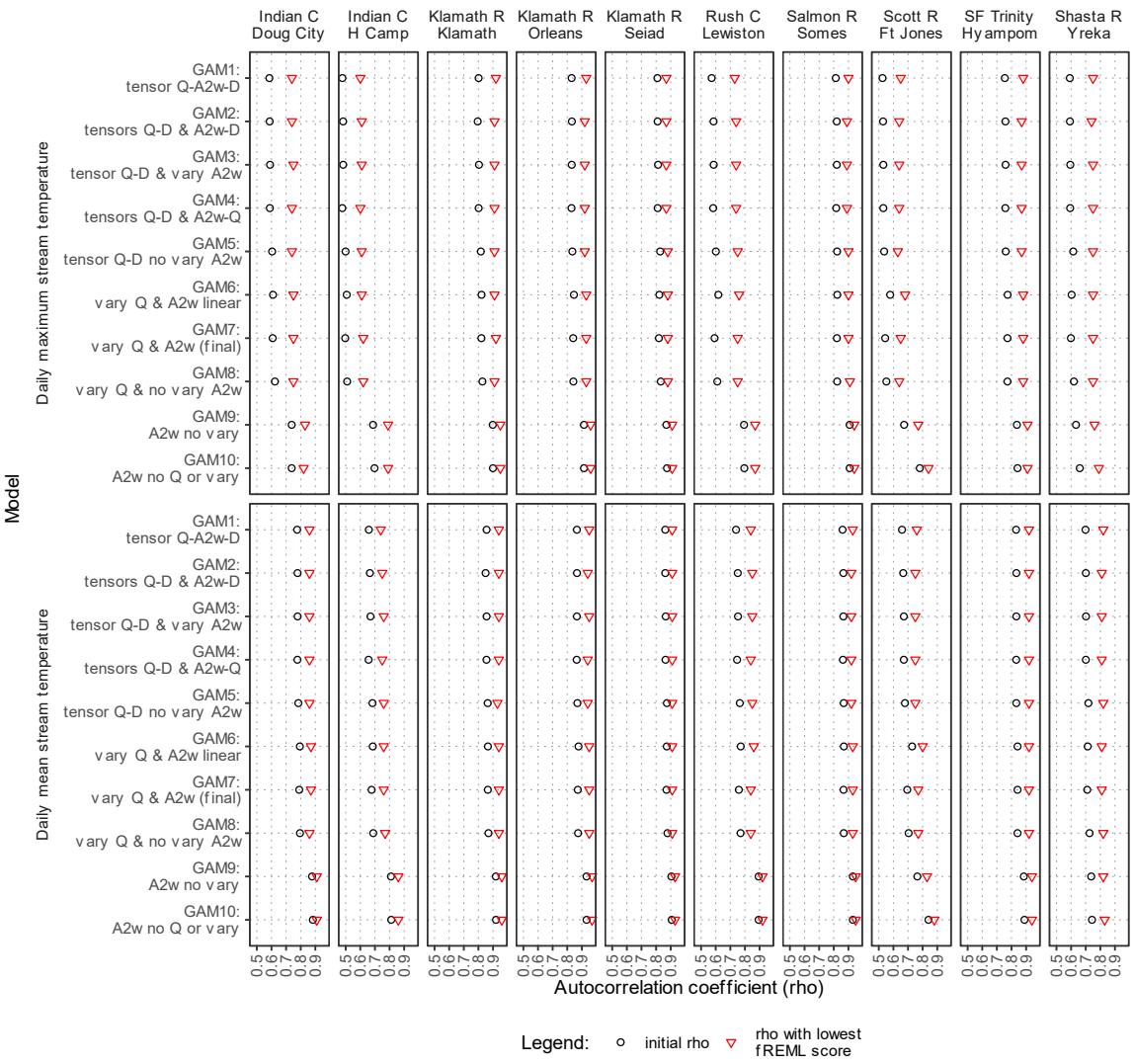


**Figure S13.** Comparison, for each site and model, of initial rho values and rho values that minimizes the fast restricted maximum likelihood (fREML) score for $T_{max}$ (top panels) and $T_{mean}$ (bottom panels). The "initial" rho value is the lag 1 autocorrelation value of the residuals from an initial model without autocorrelation. GAM11 does not have an autocorrelation coefficient so is not included here.

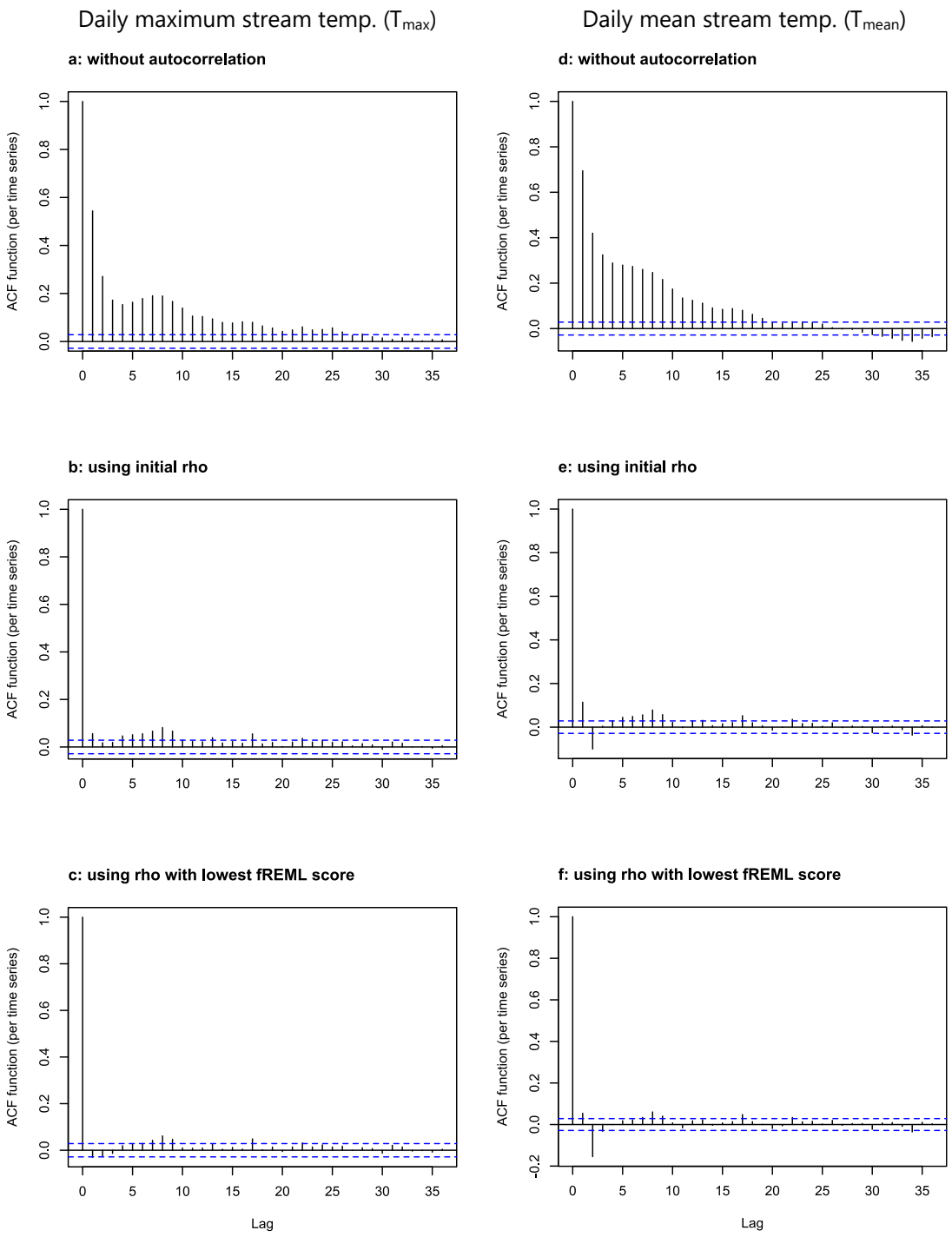


**Figure S14.** Autocorrelation function (ACF) plots for alternative versions of the Scott River final "GAM7" models for $T_{max}$ and $T_{mean}$ with (a,d) no autocorrelation structure, or autocorrelation values (i.e., rho) set as either (b,e) the lag 1 autocorrelation value of the residuals from an initial model without autocorrelation, or (c,f) the rho value that minimizes the fast restricted maximum likelihood (fREML) score (i.e., red triangle in Figure S12).

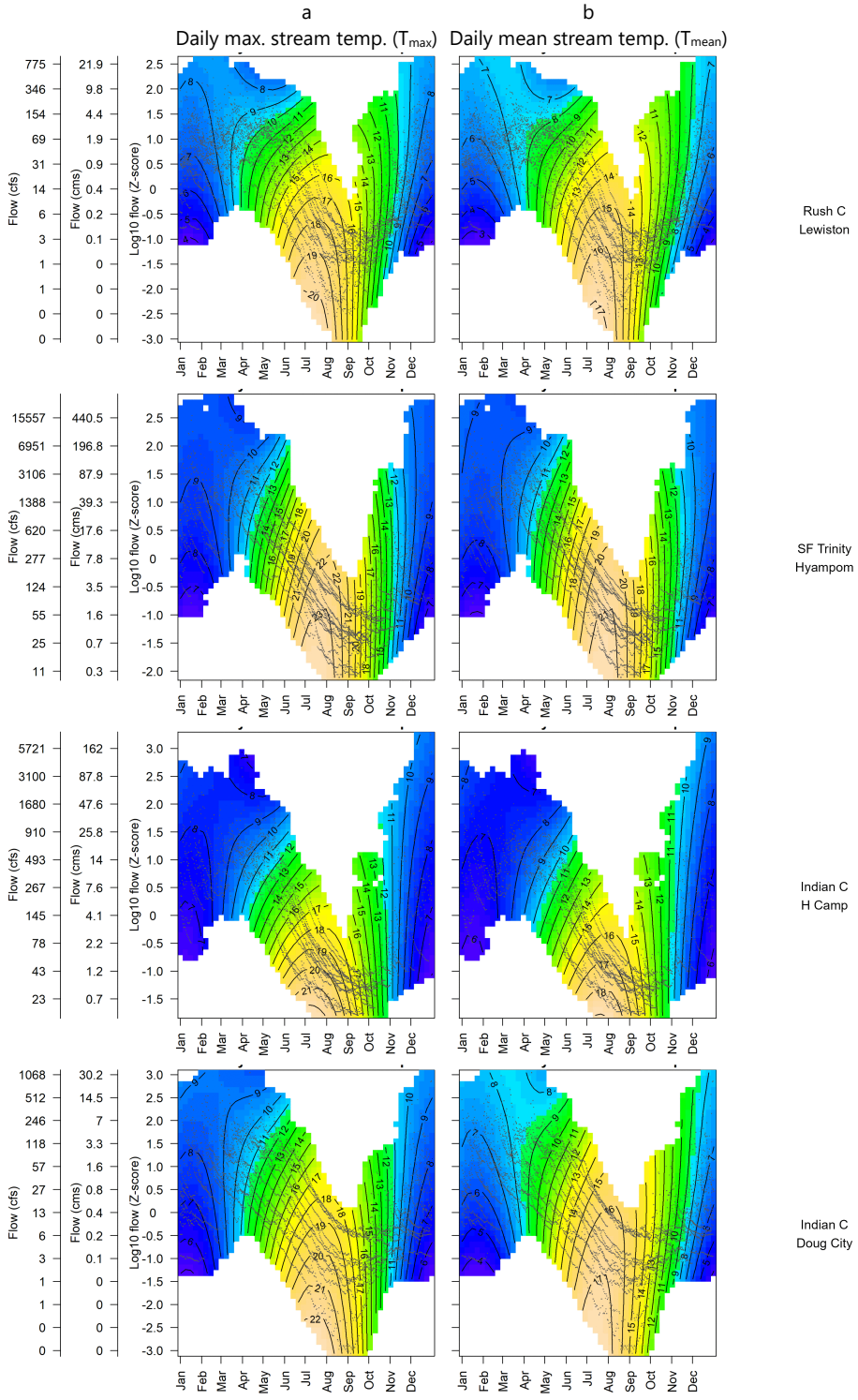


**Figure S15.** Effects of flow (Q) and day of year (D) on predicted values of (a) $T_{max}$ and (b) $T_{mean}$ in selected model GAM7 at four example sites. Sites in top two rows have non-monotonic relationships in Oct–Nov (Section 5.4) while sites in the bottom two rows do not). Colors and labeled contour lines show predicted temperatures (°C). Underlying gray dots show calibration data. Y-axis labels provide multiple units to facilitate interpretation.