Upper Delaware River Temperature Modeling: Phase I Final Report

By J.M. Bartholow and J. Heasley

April 28, 2005

Report to the Delaware Basin's Flow Management Technical Advisory Committee

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
Gale A. Norton, Secretary

U.S. Geological Survey
Charles G. Groat, Director

U.S. Geological Survey, Reston, Virginia 2005

For product and ordering information:
World Wide Web: http://www.usgs.gov/pubprod
Telephone: 1-888-ASK-USGS

For more information on the USGS—the Federal source for science about the Earth,
its natural and living resources, natural hazards, and the environment:
World Wide Web: http://www.usgs.gov
Telephone: 1-888-ASK-USGS

Contents

Figures

Tables

Conversion Factors

Inch/Pound to SI

Multiply	By	To obtain
	Length	
mile (mi)	1.609	kilometer (km)
	Area	
square mile (mi^2)	2.590	square kilometer (km^2)
	Flow rate	
cubic foot per second (ft^3/s)	0.02832	cubic meter per second (m^3/s)

SI to Inch/Pound

Multiply	By	To obtain
	Length	
kilometer (km)	0.6214	mile (mi)
	Area	
square kilometer (km^2)	0.3861	square mile (mi^2)
	Flow rate	
cubic meter per second (m^3/s)	35.31	cubic foot per second (ft^3/s)

Temperature in degrees Celsius (°C) may be converted to degrees Fahrenheit (°F) as follows:

$$°F=(1.8\times°C)+32$$

Temperature in degrees Fahrenheit (°F) may be converted to degrees Celsius (°C) as follows:

$$°C=(°F-32)/1.8$$

Upper Delaware River Temperature Modeling: Phase I Final Report

By J.M. Bartholow (USGS/BRD) and J. Heasley (IAP World Services)

Executive Summary

The objective of this study (Phase I) was to gather data for and test the applicability of physically-based water temperature models for the East and West Branches of the Delaware River and Neversink River below their respective water supply reservoirs to see if these models could predict water temperatures as a function of streamflow better than the existing nomograph-based approach. Hydrologic, meteorologic, and stream geometry data were collected or synthesized from existing records and used as input for models depicting the summers of a 3-year period, 1997-1999. Considerable historic and other data were found to be available, albeit with some missing elements, and incorporated into each of the two models.

The models preformed adequately, though not as well as had been expected. Both models had r-values greater than 0.8, tended to over-predict temperatures slightly, and had an expected prediction error of approximately ±1.2-1.5°C (±2.2-2.7°F) for both mean and maximum daily water temperatures, with the Neversink model performing slightly better than the East/West Branch model. Error analysis showed that model biases were moderately correlated with model input variables, notably air temperature, but attempts at model improvement through calibration proved unsuccessful.

Because the mechanistic model did not perform quite as well as we would have liked, we developed some preliminary statistical models driven by meteorology and reservoir discharge. These statistical models compared favorably to the mechanistic models, but were not clearly superior.

Since we were not able to objectively compare any of the models developed with the nomograph technique currently being used to forecast water temperatures and adjust reservoir releases, we offered our comparison of the attributes of each of the modeling techniques and six recommendations as to potentially profitable next steps, including better testing, further development, and possibilities for improving the method used to “spend” the reservoirs’ established conservation pools. The principle recommendation is to exercise both the SNTEMP and nomograph techniques with a common data set and see which does a better job of predicting in-river temperatures.

Background

USGS involvement in the Upper Delaware River Basin is the result of Congressional funding directed towards the study of instream habitat needs in the basin. This study was proposed for federal funding by a coalition of non-profit groups (including The Nature Conservancy, Trout Unlimited, and the Delaware River Foundation) and supported by the Delaware River Basin Commission. A USGS study plan was developed in conjunction with the Subcommittee on Ecological Flows (SEF) for the Delaware Basin's Flow Management Technical Advisory Committee. The SEF's goal is to "to develop ecological flow requirements for the maintenance for restoration of healthy, self-sustaining and managed aquatic ecosystems in the Delaware Basin." This goal must be accomplished given the legal requirements for export of water from the rivers and downstream water delivery requirements for municipal water supplies. The complete USGS study plan includes consideration of water temperature and flow variability as they influence ecological communities, with particular attention to brown and rainbow trout habitat.

The goal of the water temperature component of the overall study is to substantially improve the ability to forecast longitudinal water temperatures as a function of reservoir releases, network hydrology, and ambient meteorological predictions on something close to a "real time" basis. If successful, this capability will ultimately become part of the overall basin planning and operations processes for the New York-Delaware River reservoir system operation.

Objectives

As mentioned, the goal is to advance the state-of-the-art of water temperature prediction in the Upper Delaware River Basin. Existing methods for doing this rely on a set of nomograms that have a tendency to overestimate (but occasionally underestimate) the volume of water necessary to be released from the three reservoirs under consideration to support existing thermal requirements at specific downstream locations. The three existing reservoirs are Cannonsville Reservoir on the West Branch of the Delaware River, Pepacton Reservoir on the East Branch, and the Neversink Reservoir on the Neversink River. If a more mechanistic model can be successfully developed that allows input of daily network flows and ambient meteorological predictions (e.g., air temperature, wind speed, relative humidity, cloud cover), it may be possible to better manage the volumes of water available to conserve habitats with suitable water temperatures.

Specific objectives address this goal in two phases. Phase I has involved gathering data for and testing the applicability of an existing network water temperature model. The test included determining historical data sufficiency, model calibration and validation to objective standards, and technology transfer to Basin participants (to the degree warranted and desired). Model validation normally would include comparison of model predictions with real-world conditions to see if the model can perform better than the existing nomogram approach, i.e., lead to achievement of temperature and habitat targets while conserving as much storage as practical. However, as will be shown, this was not possible given existing data and budgetary constraints.

If and only if Phase I were judged successful, Phase II may follow. This phase would make enhancements to the modeling software that facilitates turning its use into a "real-time" process. By real-time, we mean the ability to incorporate predictions of "tomorrow's" meteorology and hydrology into the model to forecast reservoir releases necessary to meet downstream temperature targets. Phase II will address factors such as data harvest from the Internet (automated to the degree possible) and incorporation into, or linking with, the Basin model OASIS that "optimizes" reservoir releases and water supply withdrawals under a given set of constraints to achieve specified goals. Phase II will also include technology transfer to institutionalize this process.

This document reports on the progress for Phase I but will further discuss Phase II.

Study Area and Study Period

The Upper Delaware stream network delimits the boundaries of the system modeled (Figure 1), but was divided into two parts. The first portion includes the West and East Branches of the Delaware from their respective reservoirs to their confluence (18 miles on the West Branch and 33 miles on the East Branch, and approximately 25 miles down the mainstem to Callicoon, NY. It also includes the Beaver Kill upstream to the flow and temperature gaging station at Cooks Falls. The second portion includes the Neversink River from the reservoir approximately 17 miles to Bridgeville. Consideration was given to linking these two study areas, but because of travel time constraints (see below), it was decided not to do so at this time. It may be possible at a later date to address a single integrated network if it proves desirable to do so. Water temperature predictions at multiple intermediate locations along each river above their respective termini are available from the models.

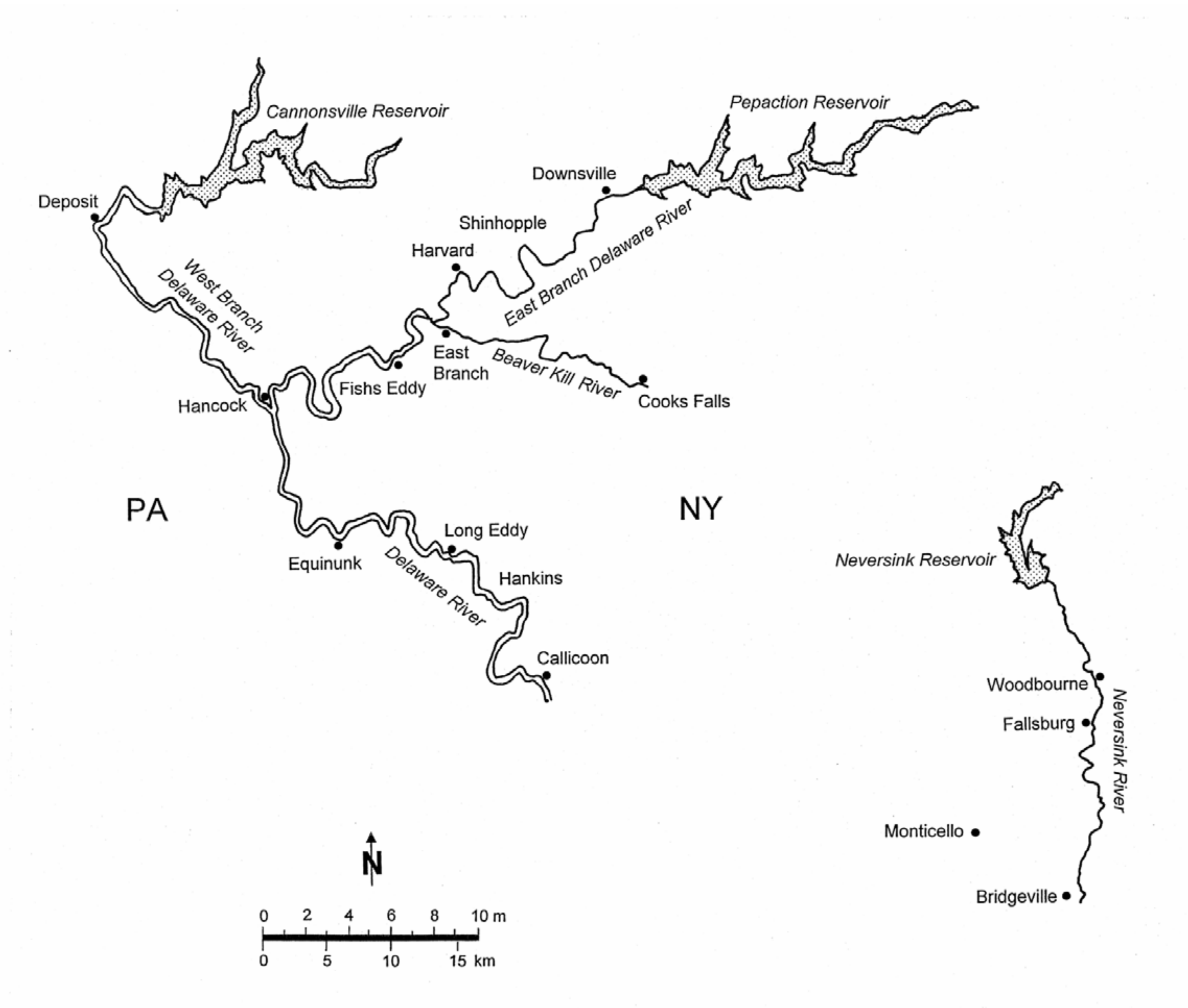

Figure 1. Upper Delaware River Basin showing the two study areas and approximate river segmentation (discussed below).

The study period for Phase I was dependent on the overlap determined in historical gaging data (both flows and water temperatures) and meteorological data, and was chosen for the test that covers the range from dry and hot to wet and cool. As will be described below, we chose 1997-1999 for model calibration. Hot summers were the focus of the study, so the chosen time period extended from 1-May through 31-October of each year. The time step was one day, the shortest time step allowed by the chosen model (see below), with provision for estimating the maximum daily water temperature. Though there was some discussion of the issue of anchor ice, winter conditions were not covered by this work.

Model Selection

The SNTEMP model (Theurer et al. 1984) was chosen for Phase I. This is a well-tested model, though most use has been in the western US. The model has proven especially robust in predicting mean daily water temperatures. Estimating maximum daily temperatures requires additional manual parameter adjustment and goodness-of-fit calibration over and above what is typically done for mean daily water temperature alone. The SNTEMP model is normally capable of predicting mean daily water temperatures ± 0.5°C (0.9°F), and almost always to within 1°C (1.8°F), depending on the quality of the input data. In addition, SNTEMP is far less demanding than many other models in terms of data requirements.

SNTEMP is an appropriate model to test for this application because of its public domain status and support. This means that not only is the model available, but the source code may be modified as necessary for potential Phase II "real-time" work. In addition, there is a considerable body of material available for the technology transfer portion of these tasks, including documentation (Theurer et al. 1984), self-paced learning material (Bartholow 2000) and background on data collection techniques (Bartholow 1989). The model runs on readily available PC platforms and executes rapidly, thus facilitating calibration and gaming.

In spite of this model's advantages, there are also some potential disadvantages. One data input item, "percent possible sun" or cloud cover, is no longer regularly collected by National Climatic Data Center stations and often requires additional effort to estimate. Also, the model assumes steady state hydrologic conditions, which might signal problems when abrupt changes to reservoir releases or short term rainfall-driven runoff events occur. Though none of the existing reservoirs has a peaking power release, they can and do spill. The SNTEMP model is not a reservoir water temperature model, and requires reservoir release temperature estimates as a boundary condition. The USGS supported version of the model runs as a DOS program. This can be cumbersome, but would be addressed under Phase II as appropriate. (Note that we also tested a consultant-developed version of SNTEMP for this project that uses a Windows-based interface and has potentially useful model enhancements. Unfortunately, we found that this new version did not prove suitable for either Phase at this time.)

As mentioned, the model operates on a daily time step under steady-state conditions. A consequence of this is that the maximum extent of the study area should typically be no more than one day's travel time from the furthest upstream point to the furthest downstream point. This constraint can be compromised, but with some degradation in predictive power. Given that Hankins is approximately 20 hours below Cannonsville and Callicoon is approximately 30 hours below Cannonsville at low flows (Robert Klosowski, NY-DEC, personal communication; White and Kratzer, 1994), we would expect that the model will perform slightly less well under dynamic conditions at Callicoon than at Hankins. However, there may be no noticeable degradation in predictive ability under conditions of relatively stable hydrology and meteorology.

Data Gathering and Synthesis

Data gathering generally followed guidelines presented in Bartholow (1989). There are three broad categories of data required by all temperature models, including SNTEMP: meteorological data, hydrologic data, and stream geometry data. Measured water temperature data are also required to perform a competent model calibration and validation (discussed below).

Representative meteorological data includes air temperature, wind speed, relative humidity, percent possible sun (cloud cover), and solar radiation. In addition, the elevation of the meteorological station must be known. On occasion, it is advantageous to have more than one met station to be able to cross check outlying data values, fill missing values, or create composite sets that might better represent the whole watershed. It is our understanding that NY-DEC has chosen two stations, Liberty and Walton NY, to use for gleaning daily weather forecasts (NY-DEC No Date). Though our original plan was to use data from these stations, we found that they did not have a suitable complement of data. Table 1 lists the major meteorological stations we evaluated for this project.

Table 1. Measured meteorological data summary.

Location	Source	Years Included	Period Covered	Comments
Binghamton	NCDC	1994-2004	1 May - 30 Sept	1996 missing 61 %-sun values 1999 missing 5 %-sun values 2003-2004 has no %-sun values
Monticello	NCDC	1994-2004	1 May - 30 Sept	Data has no %-sun values 1996 missing 7 data records 1997 missing 22 data records 2000 missing 2 data records 2001 missing 8 data records 2002 missing 10 data records 2003 missing 2 data records
Stonykill	MesoWest	2003-2004	1 May - 30 Sept	Has solar radiation instead of %-sun
Sherburne	MesoWest	2003-2004	1 May - 30 Sept	Has solar radiation instead of %-sun

Hydrologic data includes the best estimates of streamflow throughout the basin. There appeared to be 14 gages with a useful complement of data (Table 2). Many of these stations also have long-term water temperature data available.

Table 2. USGS discharge gages in the Upper Delaware River study sites having at least four years of record. Period of record is given along with the count of daily samples. Inclusion of water temperature records is also indicated.

Site Number	Site Name	From	To	Count	Includes Water Temperature
1417000	EAST BRANCH DELAWARE RIVER AT DOWNSVILLE NY	7/1/1941	9/30/2003	22737	
1417500	EAST BR DELAWARE RIVER AT HARVARD NY	10/1/1934	9/30/2003	21426	Yes
1420500	BEAVER KILL AT COOKS FALLS NY	7/25/1913	9/30/2003	32857	Yes
1420980	E BR DELAWARE RIVER ABV READ CR AT FISHS EDDY NY	11/19/1912	9/30/2003	33136	Yes
1421000	EAST BR DELAWARE R AT FISHS EDDY NY	11/19/1912	9/30/2001	32406	
1425000	WEST BR DELAWARE RIVER AT STILESVILLE NY	7/1/1952	9/30/2003	18719	Yes
1426000	OQUAGA CREEK AT DEPOSIT NY	10/1/1940	9/30/1973	12053	Yes, but not used
1426500	WEST BRANCH DELAWARE RIVER AT HALE EDDY NY	11/15/1912	9/30/2003	33176	Yes
1427405	DELAWARE R NR CALLICOON NY	8/25/1967	7/8/1975	2875	Not used
1427500	CALLICOON CREEK AT CALLICOON NY	10/1/1940	9/30/1982	15340	Not used
1427510	DELAWARE RIVER AT CALLICOON NY	6/27/1975	9/30/2003	10323	Yes
1436000	NEVERSINK RIVER AT NEVERSINK NY	10/1/1941	9/30/2003	22645	
1436500	NEVERSINK RIVER AT WOODBOURNE NY	10/21/1937	9/30/1993	18973	
1436690	NEVERSINK RIVER AT BRIDGEVILLE NY	10/1/1992	9/30/2003	4017	Yes

Stream geometry data included delineation of the rivers into discrete segments and tagging them with reach length, aspect (direction of flow from the N-S axis), latitude, elevation, channel width as a function of discharge, and Manning's n. Manning's n, a measure of "friction" was estimated to be 0.035, a common "default". Stream widths were generally characterized through a preliminary inventory and may be subject to improvement, along with Manning's n, when results of the more detailed micro-habitat modeling come on-line. The model also requires estimates of streamside shade, whether from riparian vegetation or the surrounding topography. Data collection for these data is described below.

Measured water temperature data were derived from existing USGS water quality gaging stations as well as New York DEC measurements. According to NY-DEC (No date, Figure 8) there are about 16-20 DEC stations scattered throughout the two study areas to be modeled, but at present, historical data is only available for the summers of 1997-99 at selected stations (Wayne Elliot, NY-DEC, personal communication). Data for 2004 will soon be available. Reservoir release temperature data were taken from the most upstream site available on each of the three rivers. Unmodeled accretion temperatures were approximated by mean annual air temperature adjusted for elevation.

Historical Data Gathering

Data invariably suffer from two flaws: there is always more than you can ever use and there is never exactly what you are looking for. The Delaware is no exception. We found that full complements of meteorologic data were more scarce than we had initially imagined, necessitating a sort of "patchwork" approach to creating a data set that is as directly representative of the basin as possible, has a long enough record to be meaningful, and contains few missing or suspect values. In other words, meteorologic data had to be synthesized from multiple sources into a single, consistent data set.

The historical data effort was one of electronic gathering and synthesis. All available on-line USGS gaging station data, both for flow and water temperature, were downloaded and converted to formats that allowed better scrutiny for any missing data. These data were supplemented by additional off-line data supplied by the USGS office in Troy, NY. The most appropriate data were converted to SNTEMP input files covering the period 1997 to 1999 for both the East and West Branches and the Neversink.

NY Department of Conservation supplied a considerable amount of additional thermister data covering both networks. These data were processed as above and precisely located using UTM coordinates. A summary of the water temperature data we used is provided in Table 3.

Table 3. **Measured water temperature data summary.**

Location	Source	Years Included	Period Covered	Comments
Harvard	USGS	1994-2004	1 May-30 Sept	Complete data set
Cooks Falls	USGS	1994-2004	1 May-30 Sept	Complete data set
Fishs Eddy	USGS	2001-2004	1 May-30 Sept	Complete data set
Hale Eddy	USGS	1996-2004	1 May-30 Sept	Complete data set
Hancock	USGS	1994-2004	1 May-30 Sept	Complete data set
Hankins	USGS	1994-2004	1 May-30 Sept	Complete data set
Callicoon	USGS	1994-2004	1 May-30 Sept	Complete data set
Bridgeville	USGS	1994-2004	1 May-30 Sept	Complete data set
Stilesville	USGS	1994-2004	1 May-30 Sept	1996 missing 46 days 1998 missing 48 days 2003 missing 93 days 2004 missing 8 days
Abe Lord Crk	NYDEC	1997-1999	1 May-30 Sept	27-79 missing days/year depending on the year
Balls Eddy	NYDEC	1997-1999	1 May-30 Sept	27-79 missing days/year depending on the year
Deutch's Flats	NYDEC	1997-1999	1 May-30 Sept	25-69 missing days/year depending on the year
Fireman's Park	NYDEC	1997-1999	1 May-30 Sept	33-63 missing days/year depending on the year
Hankins	NYDEC	1997-1998	1 May-30 Sept	44-62 missing days/year depending on the year
Harvard	NYDEC	1997-1999	1 May-30 Sept	26-63 missing days/year depending on the year
Kellams	NYDEC	1998-1999	1 May-30 Sept	37-44 missing days/year depending on the year
Leonard's	NYDEC	1997-1999	1 May-30 Sept	27-79 missing days/year depending on the year
Long Eddy	NYDEC	1997-1999	1 May-30 Sept	37-62 missing days/year depending on the year
Men's Club	NYDEC	1997-1999	1 May-30 Sept	25-61 missing days/year depending on the year
Roods Crk	NYDEC	1997-1999	1 May-30 Sept	25-61 missing days/year depending on the year
Shehawken Crk	NYDEC	1997-1999	1 May-30 Sept	25-61 missing days/year depending on the year
Shinhopple Bridge	NYDEC	1997-1999	1 May-30 Sept	25-81 missing days/year depending on the year
Terry's Campground	NYDEC	1997-1999	1 May-30 Sept	25-69 missing days/year depending on the year

Additional data were also supplied by the NY Department of Environmental Protection (DEP), but we were not able to obtain historic reservoir discharge and temperature data from DEP, so the models were initialized using data collected at the first monitoring station downstream. These stations are USGS gages on the Neversink and West Branch, and a DEC thermister on the East Branch. Substantial numbers of "missing" data at these locations from 1 May to the end of June, and/or toward the end of September, degrade the quality of the simulations.

Meteorology data from Binghamton and Monticello were downloaded and similarly processed. With the exception of the DEP meteorological data from the three reservoirs, Table 4 summarizes the historical meteorological data gathered.

Table 4. Data gathering summary for meteorologic and hydrologic data for the Delaware River Basin water temperature modeling, focusing on the 1994-2004 period.

Data Type	Location	Dates	Status	Comments
Air Temperature	Monticello	1994 - 2004	Downloaded	
	Stonykill, NY	2003 - 2004	Requested	Received
	Sherberne, NY	2003 - 2004	Requested	Received
	USGS Water Quality Samples	Scattered dates	Need to be downloaded	
	Binghamton, NY	1948 - 2004	Downloaded and unpacked	Minor purchase
Wind Speed	Monticello	1994 - 2004	Downloaded	
	Stonykill, NY	2003 - 2004	Requested	Received
	Sherberne, NY	2003 - 2004	Requested	Received
	Binghamton, NY	1983 - 2004	Downloaded and unpacked	Minor purchase
Relative humidity/dew point	Monticello	1994 - 2004	Downloaded	
	Stonykill, NY	2003 - 2004	Requested	Received
	Sherberne, NY	2003 - 2004	Requested	Received
	Binghamton, NY	1983 - 2004	Downloaded and unpacked	Minor purchase
Percent Sunshine	Binghamton, NY	1965 - 2001	Downloaded and unpacked	Minor purchase; some missing days
Solar Radiation	Stonykill, NY	2003 - 2004	Requested	Received
	Sherberne, NY	2003 - 2004	Requested	Received
Discharge	USGS gages	1994 - 2004	Downloaded	
Water Temperature	USGS gages	1994 - 2004	Received	
	Tailwater & thermograph data	1997 - 1999	Received	

As will be described later, we ultimately decided that the Monticello station provided the best, most representative data set, though it too had some missing data. A summary of this data set may be found in Table 5.

Table 5. List of days without meteorological data in the Monticello database. Note that Monticello did not have percent sun (cloud cover) data at all, so these data were taken directly from Binghamton for all years.

Year	Comments
1994	no missing days
1995	no missing days
1996	July 2-8
1997	June 21 - July 1, August 8 - 13, August 15 - 18
1998	no missing days
1999	June 17, June 25 - 38, September 29-30
2000	September 23 - 24
2001	May 15, May 21, May 23, August 18 - 20
2002	May 15 - 19, June 18 - 21, September 7
2003	May 31, August 15
2004	no missing days

Field Data Gathering

In addition to the historical hydrology and meteorology data, we visited the study area and collected a variety of representative stream geometry data, including stream width, topographic, and riparian shading measurements. We took numerous sightings (and photographs) to the topographic horizon on both sides of the river for pre-determined river segments. (Segments, in this context, refer to river reaches subdivided based on general stream aspect with respect to the N-S axis.) The horizon measurements used to compute segment shading were not meant to be final determinations, but rather used as confirmation of what a representative range of values would be along the rivers. There is such a high degree of topographic heterogeneity in these basins, and some segments were not accessible by car, that it would be extremely costly to collect more detailed horizon information, especially in cases like the Delaware River where shading is likely best used as a calibration parameter.

We used the MapTech Terrain Navigator software and data base for New York. This software is composed of scanned 1:24000 topographic maps that "sit on top of" a 10 meter Digital Elevation Model grid. Using this software, one can carefully examine each reach and construct a "profile line" perpendicular to, and leading away from, both sides of the river. From these profile lines, one can determine where the visual horizon would be as if you were standing in the middle of the river. You are making the assumption that the line you draw captures the basic topography of each segment -- something that is not always easy or objective. But in areas of extreme heterogeneity, multiple measurements can confirm the reliability of one's subjective approach. From these profile lines, it is possible to then measure the distance from the river and the elevation change to the horizon, allowing computation of the topographic altitude angle required by SNTEMP. The software was also used to calculate river segment lengths, interpolate elevations (somewhat problematic due to 10 meter grid resolution, but adjusted to provide reasonable gradients), and estimate azimuths (i.e., direction of the river from N-S), in addition to latitude and longitude of each segment boundary. This process was completed for the East and West Branches and the mainstem down to Callicoon, as well as the Neversink to Bridgeville.

Riparian vegetative shading was estimated for the same river segments. Unlike topography, estimated vegetative characteristics of tree height, diameter and leaf density were relatively uniform throughout the various river basins. Differences were manifest principally in the relative continuity of trees along each bank and, to some degree, their offset from the river's edge. Field measurements, supplemented by the digital 1:24,000 topographic maps, aided the development of segment-by-segment riparian shading estimates. Table 6 summarizes attributes for the various river components and Figure 2 illustrates the general stream gradient for the two study areas.

Table 6. Delaware system components and their attributes, including shade estimates (comprised of both topography and vegetation) for mid-August.

Area	# Segments	Total length (km)	Mean Shade (%)	Shade Range (%)
Neversink	15	26	16	2 – 28
West Branch	12	28	19	1 – 39
East Branch	23	54	25	0 – 40
Beaver Kill	6	16	20	2 – 41
Mainstem	21	45	19	1 – 34

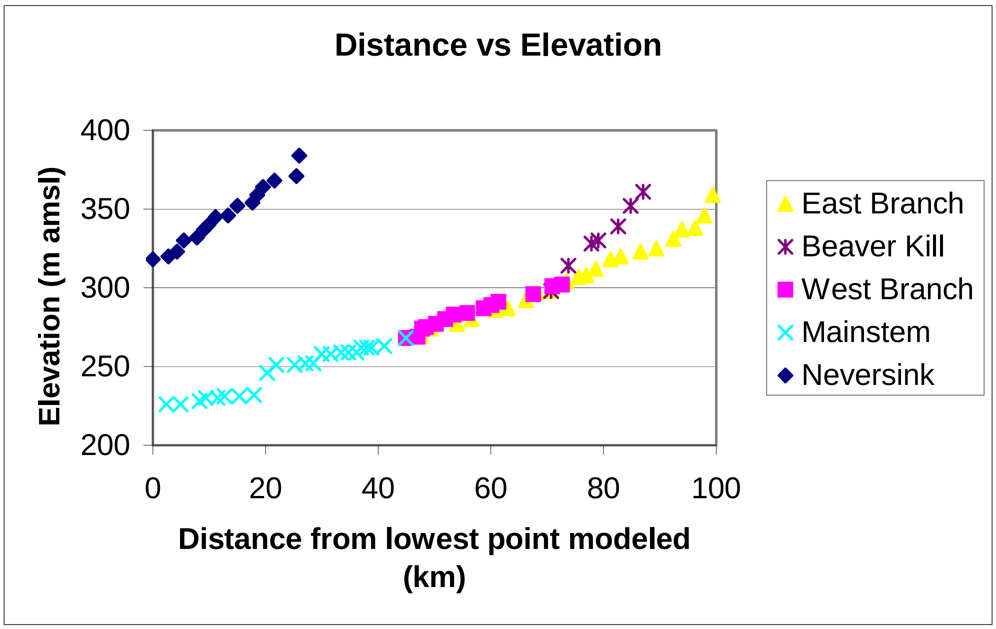

Figure 2. Longitudinal elevations along the courses of the two study area streams.

Computing Accretions between Points of Known Discharge

As mentioned, we gathered flow data from the network of USGS flow gages throughout the basins. However, SNTEMP requires that flows also be specified at many other locations, specifically at all places where temperatures were measured and at major tributary junctions. Flows at locations where discharge data were not available were calculated by simply apportioning the difference in flows between points of known discharge by the ratio of the distance between the upstream known discharge point and the unknown point to the total distance between the known points. We used this method regardless of whether the location was a river junction or simply a location along the river.

We considered using drainage area ratios to prorate unmeasured accretions, but we did not have access to the drainage area at each location we needed. Though we could have sought additional help in getting this information, we did not feel that the benefit would be worth the cost. Experiments in other basins we have worked in have shown that when unmeasured accretions make up such a relatively small proportion of the flow, the influence on "mainstem" water temperatures is minimal. We also considered using relationships developed by others for the Delaware River (Thatcher and Mendoza, 1990), but these proved to be at too gross a scale for our purposes.

Quality Assurance/Quality Control

Large compilations of data must be scrutinized for data quality. It is not uncommon for water temperature or other data to have spurious values that must be weeded out and that was the case here. Few strictly objective measures exist for examining every data value, but obvious outliers were eliminated from each data set. Missing data were generated for meteorological or hydrological data, or the upstream-most water temperatures using the best surrogate available by relying on station-to-station regressions.

One item warrants additional discussion. It is rare that we have the opportunity to compare measured water temperature data collected at ostensibly the same location from two different sources. We compared data collected by USGS and DEC for the Harvard site for 1997-1999 and they agreed very well. Median absolute differences between the two were 0.3°C (0.5°F) for mean daily temperatures and 0.2°C (0.4°F) for the maximum daily temperatures (n = 316). Some of the difference may be explained by the minimum resolution of the data; USGS data is reported at the 0.5°C level where the DEC data is reported at the 0.1°C level. There were some notable exceptions however. The maximum absolute differences were 4.0°C (7.2°F) for the mean and 4.8°C (8.6°F) for the maximum temperatures. We did not make similar comparisons at other “joint” data collection sites, but this would be interesting to do.

Initial Model Simulations

SNTEMP models for both the Neversink, and East and West Branch through the mainstem were run with the data available for the summers of 1997 through 1999, 1-May through 30-September. We initially determined that the models both performed best given meteorological data from Monticello, NY, rather than either Binghamton, NY, or an average of the two, though differences were slight. Monticello is at a somewhat higher elevation than most of the two basins modeled, but it is relatively close, roughly 11 km from Fallsburg, NY, but 54 km from Harvard, NY.

With current data limitations, but without calibration, the Neversink model performed passably, with an overall r-value of 0.84, mean error of 0.12°C (.22°F), and a probable error of 1.16°C (2.09°F). Probable error is a measure of the median dispersion around any given temperature prediction; in other words, each daily temperature prediction would be expected, on average, to be within ± the probable error. As expected, the model did not do well during some high flow events.

Initial model runs for the East and West Branch Delaware also showed that the model was having difficulty with the large amounts of missing observed data at some river locations. For this model, the r-value was 0.89 and the mean error was 0.55°C (.99°F) with a probable error of 1.23°C (2.21°F).

Maximum errors were -7.23°C (-13°F) on the Delaware and -5.88°C (10.6°F) on the Neversink. Both of these last two metrics appear to be directly attributable to the missing Monticello meteorological data and do not likely reflect significantly on the model’s overall predictive ability.

Example graphs showing time series and longitudinal comparisons appear in Figures 3-5. As can be seen, some fits were good and some not so good. A complete set of goodness-of-fit statistics for both river systems may be found in the Appendix, Tables A1 and 2 for mean daily water temperatures.

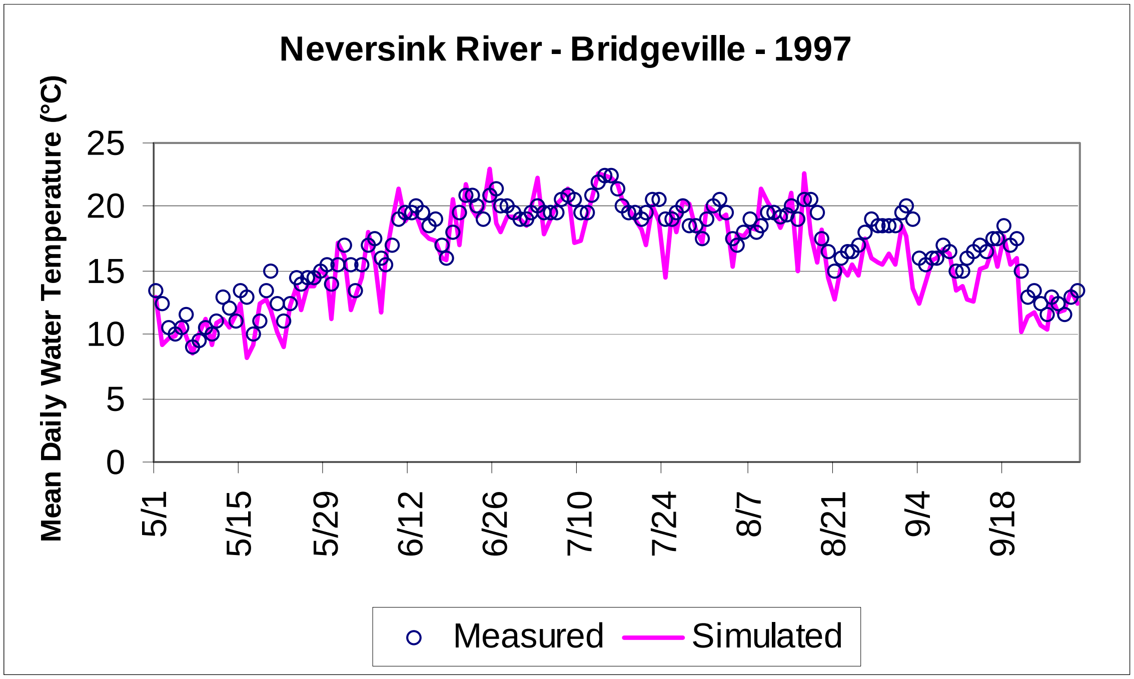

Figure 3. Example goodness-of-fit at the Bridgeville gage for the summer of 1997.

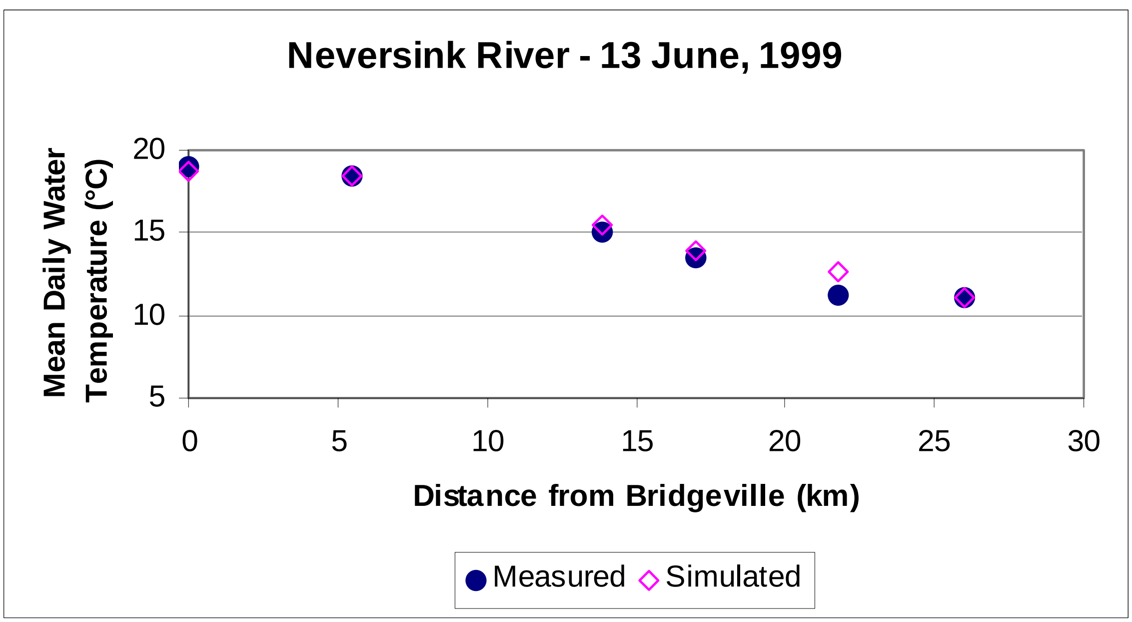

Figure 4. Example longitudinal fit for mean daily water temperature along the Neversink River for a single day.

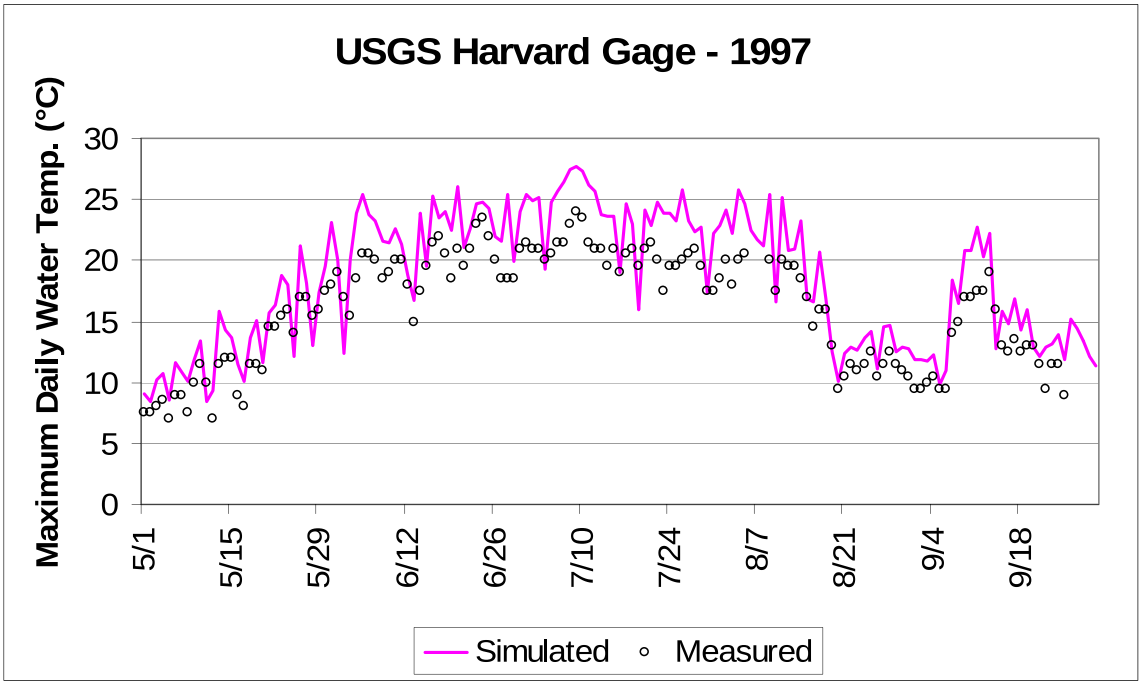

Figure 5. Example maximum daily water temperature goodness-of-fit at the Harvard gage for the summer of 1997.

Overall Fit to Measured Mean Daily Data

In addition to the statistics reported above, it is often useful to look at the overall model fit by simply comparing the measured and simulated data. Figures 6 and 7 do just that. This adds a visual feel to the fidelity and scatter in the models. Clearly, though the overall trend is well captured, there is considerable scatter and there are some days when the models do not correspond well to the observations. Either the model predictions or the measured data, or both, could be in error.

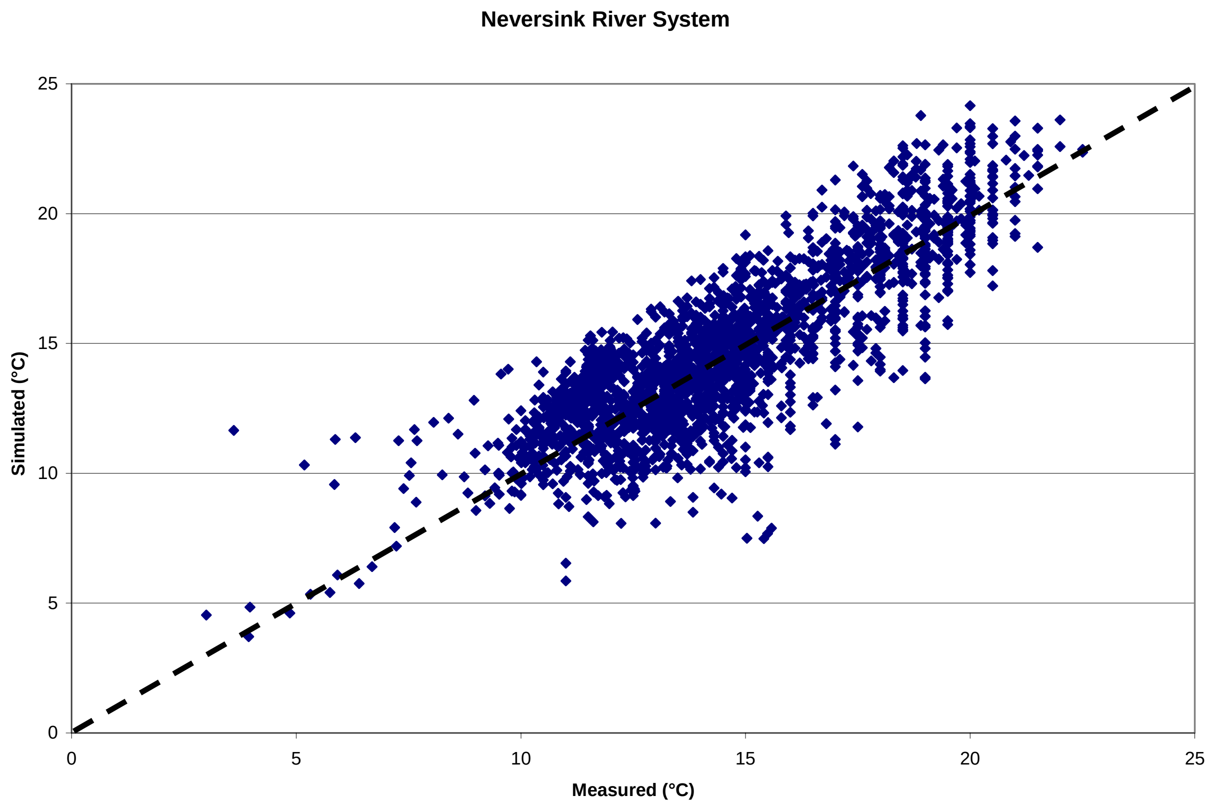

Figure 6. Visual correlation between measured and simulated mean daily water temperatures on the Neversink River across all measurement locations. Obvious outliers were removed.

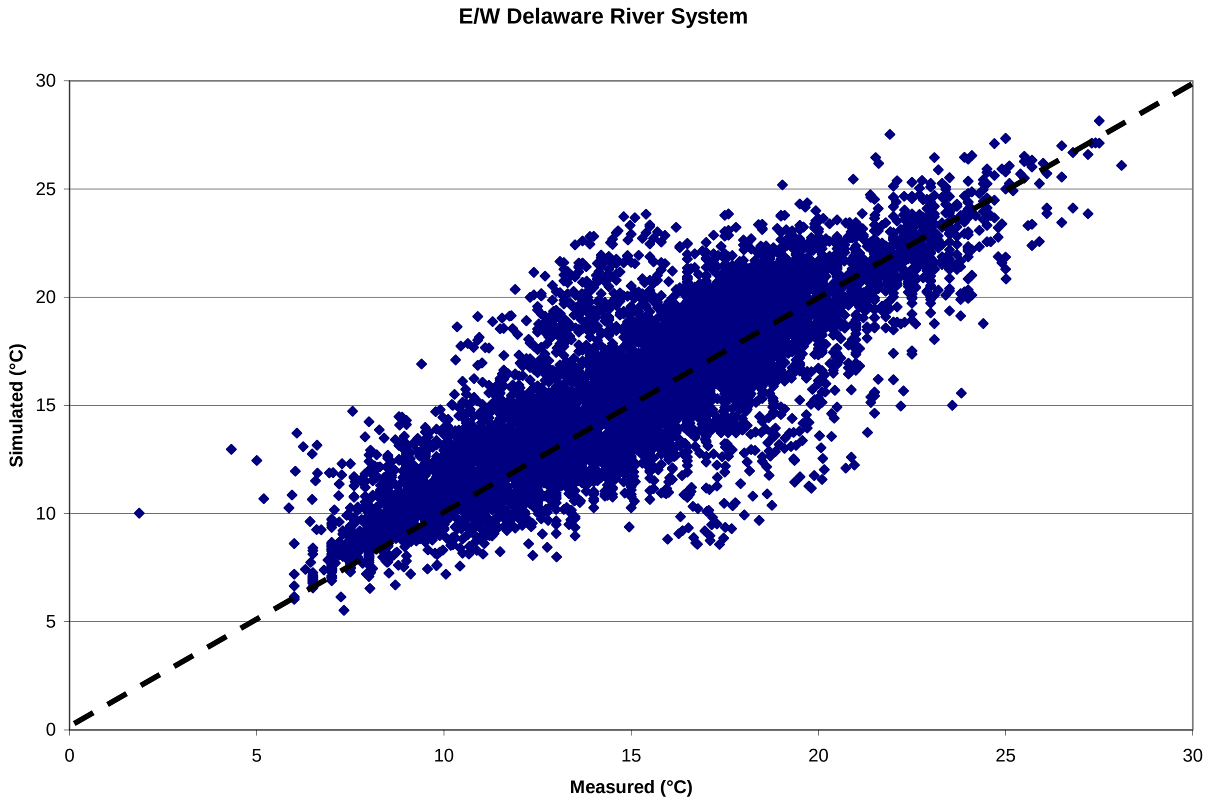

Figure 7. Visual correlation between measured and simulated mean daily water temperatures on the Delaware River across all measurement locations. Obvious outliers were removed.

Evaluation of the SNTEMP model with Meteorological Data Collected by the NY-DEP

Though it took some time to obtain meteorological data collected by NY-DEP at the three reservoir sites (Pepacton, Cannonsville, and Neversink), we did synthesize this data and used it in trials to see if the models were capable of improved predictions. Though we held high hopes that the meteorological data collected at the three reservoir sites would significantly improve the models' performance for the two networks (East and West Branches and the Neversink), this did not prove to be true. Running each of the two networks with a data set derived from each reservoir showed that none of the combinations did quite as well as using the Monticello, NY, meteorological data alone. Correlations between model-predictions and field measurements declined, mean and maximum errors increased, and the probable errors worsened, for both mean and maximum daily water temperature predictions. Interestingly, this was true even with the addition of solar radiation which had been collected at these reservoir sites. Differences were not large, and in one or two cases some individual metrics were slightly better, but Monticello remains the best meteorological station to use.

We can only speculate as to why meteorological data from the three reservoirs did not perform as well as the Monticello-derived data. We carefully compared the data collected at the Neversink Reservoir with data collected at Monticello and found generally very high correlations. Air temperatures agreed quite well, but tended to be a bit cooler at the reservoir, especially at higher Monticello temperatures. Wind speeds at the reservoir were almost all higher than at Monticello, perhaps reflecting the greater exposure of an open site. This was especially true at low wind speeds. Relative humidity showed a very similar trend, but a much higher scatter between the two locations. In other words, neither site would predict the relative humidity at the other site very well on a day to day basis. The same was true for solar radiation, only in this case we were comparing measured solar against that predicted by SNTEMP since we had no measured values at Monticello. Though we cannot rule out that we may have made an error in synthesizing the reservoir met data, one might suppose that a weather station located at the "upper" end of each network may not be quite as representative of a station that might better represent the "interior" of the drainage basin. This would not be true for a *reservoir* temperature model like CE-QUAL-W2, but certainly proved true for the rivers. The Monticello site is also likely to be carefully controlled since it is a National Weather Service station.

In any event, this detour proved to be somewhat costly in terms of the overall effort. At least we are now assured that the Monticello meteorological data can serve as the foundation for whatever temperature prediction models are employed, and, fortunately, may be the more readily accessible station for "real-time" forecasting.

Model Calibration

Well formulated models with high quality input data require little or no calibration, but data are always limited to some degree, particularly in the ability of meteorological data to truly represent conditions at and along long stretches of a river. Model calibration examined the model's bias, correlation, and error from statistics that compare model predictions with water temperature measurements. Bias is simply the average performance: does the model tend to over-predict or under-predict. We also want to know if there are recognizable tendencies to over-predict or under-predict either through time, through space (longitudinally), or with different hydrologic conditions. Correlation (or really correlation squared) is a measure of how well variation in the model's predictions "explains" variation in the measured data. Error is a measure of the overall "closeness" of model predictions to measurements.

The goal of model calibration is to simultaneously minimize bias and error while maximizing correlation. Typical criteria are to have very close to zero bias, 50% of the model's average temperature predictions differ from observations by less than 0.5°C (0.9°F), absolute maximum errors under 4°C (7°F), and overall model correlation (r) greater than 0.9. Criteria for maximum daily temperatures would be similar. The general philosophy in model calibration is to vary the least well-known input values within a representative range to maximize the model's goodness-of-fit.

Mean daily water temperatures were the initial focus of model calibration. Once mean daily temperatures are as close as we can get them, the focus usually shifts to maximum daily water temperatures, accomplished via several empirical coefficients that basically predict the additional heat gained over and above the daily average depending on hydrologic and meteorologic conditions.

Main Sources of Mean Daily Model Error

Because neither model performed as well as we had expected, an extensive analysis was made of the potential sources of model error (bias) by correlating many of the model inputs or calculated values with the model's residuals (model predictions minus observed measurements). These correlations are presented in Tables A3 and A4 in the Appendix. Because of the observed preponderance of wide, shallow pools on these rivers, we expected that air temperature and relative humidity might tend to dominate the thermal response rate when discharge was low. Several visual trends were apparent in examining the residuals as shown in Figures A1 and A2, however, only air temperature was marginally "statistically significant" on both models (Figure 8). Flow was also a "statistically significant" contributor to model error on the Neversink River, but it was felt that this was attributable solely to outlying points that represented spills or rainstorms rather than more "normal" reservoir release conditions.

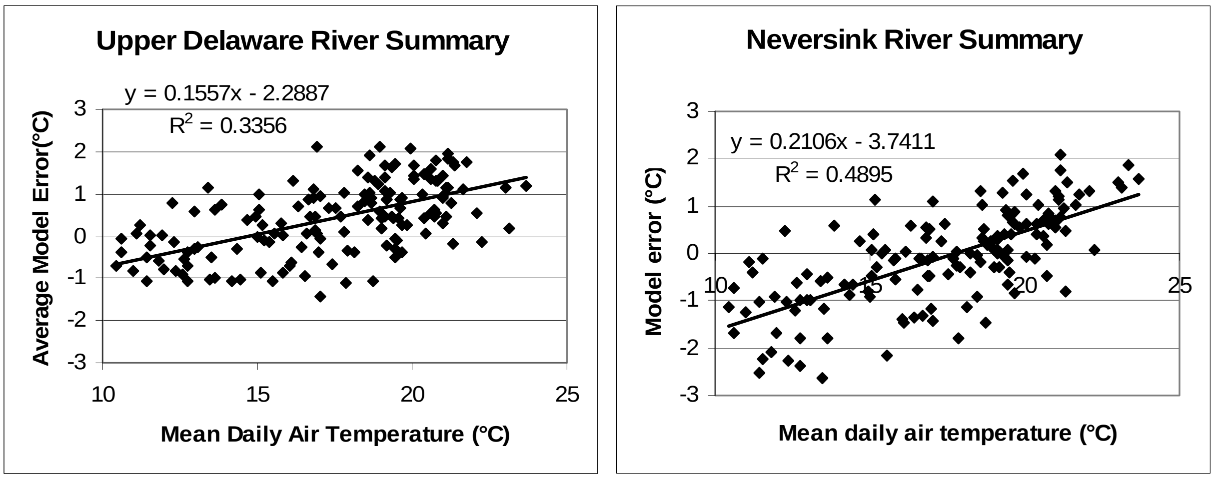

Figure 8. Correlation between mean daily air temperature and SNTEMP mean daily model error across all verification nodes and (averaged) time periods for the Upper Delaware and Neversink River models.

Attempts at Model Calibration

Because we were unsure of the exact causes of error correlated with air temperatures from Monticello (potentially including both elevational and other climatic differences), we elected to try to adjust air temperatures using SNTEMP's "global" correction capability. SNTEMP makes this easy through a formula designed to adjust all input air temperatures:

$$AT_n = a + b\,(AT_g)$$

Where AT_n = new air temperature
AT_g = given air temperature
a and b are coefficients

We tried to adjust the air temperature inputs using this equation, but were unsuccessful in improving the overall goodness-of-fit statistics. In fact, we applied this and similar techniques with several other model inputs, both "globally" and through time, without significant success. This is a good-news/bad-news situation. The good news is that even though the model error is correlated with some of the model inputs, the model does not seem to improve given systematic adjustment of those inputs. This presumably means that even though the Monticello meteorology is "off-site", it is not *systematically* unrepresentative of on-site conditions. Instead the error correlated with inputs reflects more random processes.

The bad news is that this means that the model still contains one or more systematic internal biases for unknown reasons. This is best illustrated in Figure 9 which shows model error through time for the Neversink; the situation is very similar for the Delaware River model. We tried adjusting the model's "time", i.e., fooling the model to think that the time of year had changed to attempt to remove the temporal error – again to no avail. As we have observed this sort of temporal error in other model applications we were not surprised, and we suspect an issue with the model's calculation of solar radiation, though this has never been proven. The model contains scant facility to correct the radiation calculation without good quality ground-level solar radiation measurements. In any event, this systematic bias is difficult to tease apart given the number of model inputs that are themselves non-linearly correlated with time.

It would certainly be possible to apply an empirical adjustment to all water temperature predictions after the model has been run using an interface that may be developed in Phase Two. Such an empirical adjustment could incorporate time as well as any other model inputs that remain significant after temporal error was corrected and offer the promise of improving management predictions. However, this does beg the question of whether a purely statistical model may be more appropriate for these Delaware and Neversink applications.

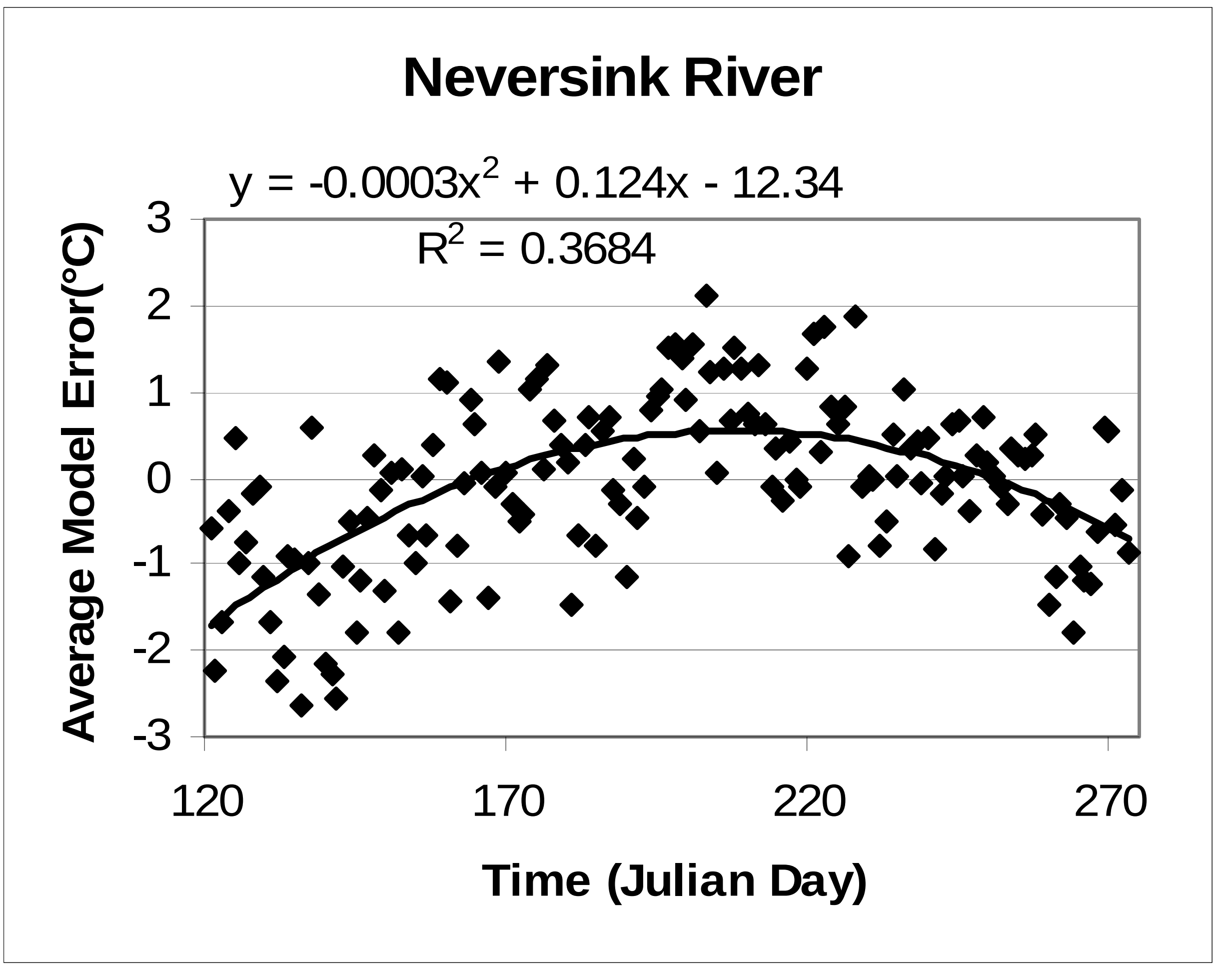

Figure 9. Illustration of the trend in the SNTEMP model's mean daily error through time for the Neversink River.

Maximum Daily Temperature Model Performance

We also examined both uncalibrated models for their ability to predict maximum daily water temperatures. These results are presented in Tables A5 and A6 in the Appendix. Results are very comparable to the mean daily goodness-of-fit statistics, just slightly poorer, which we expected. For what it is worth, these maximum daily statistics were actually better than we expected from an "uncalibrated" model. Because of this, and time constraints, we did not attempt further calibration of maximum daily water temperatures.

General Observations on Model Calibration

For reasons that are not clear at the moment, the SNTEMP model appears to be performing somewhat more poorly on the East Branch than on either the West Branch or Mainstem Delaware River in terms of model bias. This may be due to the characteristically low flows on the East Branch, yet if flows were the "culprit" we'd expect the model's probable error to be higher on the East Branch as well as the bias, which is not the case. In fact, the probable error is consistently in the range of 1.0 to 1.2°C -- on both river systems -- leading us to the conclusion that much of the error is associated roughly in the order of (1) off-site meteorology; (2) rainstorms or spills; and (3) the internal model error noted above, especially given the sinusoidal relationship of error through the summer.

Single-day maximum errors are larger than we would like to see, both for mean daily and maximum daily predictions. As mentioned, many of the days that are generating these maximum errors are those when at least some meteorological data were missing from the Monticello record. However, even if we remove those days from the goodness-of-fit calculations, maximum errors remain virtually unchanged. We are unsure what to conclude from this except to hark back to the three points listed above.

The Neversink SNTEMP implementation is a slightly better model, at least as measured by the mean error. This should not be surprising. The Neversink is a simpler system, more stable hydrologically, and more homogeneous geometrically. Tributaries are fewer and less diverse. At the same time, the overall correlations are generally slightly poorer for reasons that we cannot fully explain – perhaps sample size.

Model Validation

Model "calibration" was conducted for the summers of 1997 to 1999. Though it was always our intention to validate the model with a different set of summers, 2001-2003, we did not do this for two reasons. First, as previously stated, we found that the model performed best without any calibration at all. Second, we ran out of time and dollars to synthesize the post-2000 data. In lieu of a more formal model validation, we developed some preliminary statistical models with which we could compare SNTEMP results, described below.

Development and Testing of the Statistical Models

We were not entirely happy with the "tightness" of the SNTEMP models' predictions for the Neversink and East-West Delaware Rivers because the goodness-of-fit metrics did not meet the criteria we initially laid out in our Scope of Work. For this reason, we wanted to see what would happen if we developed purely statistical models for several important locations throughout the two networks as an alternative to SNTEMP, at least in a preliminary sense. This technique would be somewhat less flexible in predicting temperatures at unmeasured locations, but may offer the opportunity to correct for systematic biases in the two SNTEMP models that we may not be able to eliminate otherwise.

According to Theurer et al. (1984), there are several forms of regression models that appear to provide a high degree of correlation in predicting stream temperatures, at least for "natural" conditions. They range from simple harmonic models:

$$T_w = T_{avg} + T_0 \bullet \cos[(2\bullet/365)(D_i - P)] \tag{1}$$

where T_w = estimated water temperature (either mean or maximum)
T_{avg} = average water temperature over all observations
T_0 = half the initial temperature range over all observations
D_i = Julian day number for day i, January 1 = 1, etc.
P = Phase delay in timing of the maximum seasonal temperature

to models that are straight polynomial:

$$T_w = a_o + a_1\bullet T_a + a_2\bullet W_s + a_3\bullet R_h + a_4\bullet S_s + a_5\bullet H_{sx} + a_6\bullet Q + a_7\bullet T_a^2 + a_8\bullet W_s^2 + a_9\bullet R_h^2 + a_{10}\bullet S_s^2 + a_{11}\bullet H_{sx}^2 + a_{12}\bullet Q^2 \tag{2}$$

where T_a = air temperature (maximum or mean, depending on the situation)
W_s = wind speed
R_h = relative humidity
S_s = percent sunshine (cloud cover)
H_{sx} = maximum possible solar radiation for the latitude and time of year
Q = discharge

to models that incorporate, at least to some degree, the physics of heat flux and heat transport (functional form not presented here). However, water temperatures at various locations on these rivers are not "natural" in that they are influenced to varying degrees by relatively constant reservoir release temperatures depending on the downstream location in question. Further, it is not straightforward to compute the required solar radiation data to effectively use Equation 2 as it stands.

Recall that we identified a temporal bias inherent in the SNTEMP models that likely could be made to fit the functional form in Equation 1. In addition, experience on other modeling projects has shown that using $Log_e(Q)$ often performs better than using raw discharge values in predictive equations, a "trick" that gets at what is known about in-river thermal processes. For these reasons, we chose to combine the various approaches and include the reservoir release (and other known boundary conditions) as independent variables into:

$$T_w = a_o + a_1\bullet T_a + a_2\bullet W_s + a_3\bullet R_h + a_4\bullet S_s + a_5\bullet \log_e(Q) + a_6\bullet T_R + a_7\bullet T_T + a_8\bullet \cos[(2\bullet/365)(D_i - P)] \tag{3}$$

where T_R = reservoir release temperature (actually used both reservoirs, if appropriate)
Tt = tributary initiation temperature (Cooks Falls temperature, if appropriate)
Q = would use both Pepacton and Cannonsville discharge, if appropriate

other parameters defined as above

as a starting point. Because this equation, and its variants, are unpleasant to work with when trying to develop a normal linear regression, data were set up to use Excel's "Solver" function to calculate best-fit parameters for selected sites during the 1997-1999 period of record. Days with any missing data, whether meteorologic or hydrologic, were not included in the regressions. In some cases, this cut the number of observations by a large fraction and often limited the number of measurements representing the months of May and September.

"Solver" can be structured with many objective functions. We chose to minimize the median of the differences between the absolute value of observed and predicted water temperatures. This is useful because it is an all-round measure of closeness-of-fit while simultaneously giving the user a tangible idea about how good predictions may be. Note that this metric is similar to what is known as a Least Absolute Deviation (LAD) regression, except that the LAD method minimizes the mean rather than the median of the absolute differences. The mean would be a useful metric, but the median is more comparable to SNTEMP's probable error, i.e., 50% of our predictions will be within X degrees of the "truth". Note that minimizing the median error, at least conceptually, does little or nothing for the maximum errors. Minimizing the maximum error might also provide a useful approach, but from experience, minimizing the maximum error may come at a cost of lowering the overall fit. It could be that any single metric sacrifices some degree of goodness-of-fit. As one reviewer stated, the objective function is critical to the results and should be carefully chosen to reflect the most critical management needs.

It quickly became apparent that the harmonic function (the cosine portion of Equations 1 and 3) was not required by Solver in finding a best fit, so we eliminated this component. This is good news because it means that a model based strictly on the meteorologic and hydrologic variables that are easy to attain does not seem to require any additional time-dependent component.

Beyond this good news, interpretation of the results becomes clouded. Results of the various trials (and we have not tried to be exhaustive here) are summarized in Tables 7 and 8 for mean daily and maximum daily predictions, respectively. The SNTEMP model performed best at half of the sites tested and some variant of the statistical model performed best at the other half. However, the "winners" were not very clear-cut; differences were generally not large, although some individual goodness-of-fit metrics could be better or worse. Statistical results for predicting maximum daily water temperatures often had respectable median errors, but extremely ill-behaved maximum errors. We did not have the time to develop a credible explanation for this except that the maximum errors occurred on days with very high flows. This may well mean that spills and tributaries would need to be handled in a more robust way than we have done to this point.

Table 7. Goodness-of-fit statistics comparing SNTEMP and statistical models at selected locations throughout the two modeled rivers, East and West Branch Delaware and the Neversink, in their ability to predict mean daily water temperatures. The "best" model for each location is in bold print.

Location	Attribute	R	Mean Error(°C)	Probable or Median Error (°C)	Maximum Error (°C)	Number of Observations
WB Hancock USGS gage	**SNTEMP**	**0.85**	**0.55**	**1.03**	**5.11**	**456**
	Regression (normal Q)	0.73	0.25	1.38	9.33	377
	Regression [Log(Q)]	0.56	0.06	1.61	7.33	377
EB Harvard USGS gage	SNTEMP	0.90	1.21	1.20	6.28	456
	Regression (normal Q)	0.86	0.42	0.83	5.40	308
	Regression [Log(Q)]	**0.91**	**-0.04**	**0.70**	**3.92**	**308**
MS Hankins USGS gage	**SNTEMP**	**0.85**	**0.01**	**1.29**	**-6.09**	**452**
	Regression (normal Q)	0.54	-0.78	1.75	14.11	253
	Regression [Log(Q)]	0.47	0.27	1.59	8.43	253
NV Bridgeville USGS gage	SNTEMP	0.88	-0.54	1.26	-5.88	456
	Regression (normal Q)	0.81	0.41	0.81	6.01	178
	Regression [Log(Q)]	**0.81**	**0.41**	**0.81**	**5.99**	**178**

Table 8. Goodness-of-fit statistics comparing SNTEMP and statistical models at selected locations throughout the two modeled rivers, East and West Branch Delaware and the Neversink, in their ability to predict maximum daily water temperatures. The "best" model for each location is in bold print.

Location	Attribute	R	Mean Error(°C)	Probable or Median Error (°C)	Maximum Error (°C)	Number of Observations
WB Hancock USGS gage	**SNTEMP**	**0.84**	**1.95**	**1.34**	**6.16**	**456**
	Regression (normal Q)	0.02	.88	2.84	19.9	385
	Regression [Log(Q)]	0.05	0.04	2.78	18.84	385
EB Harvard USGS gage	SNTEMP	0.89	2.34	1.56	7.96	456
	Regression (normal Q)	**0.87**	**0.16**	**0.75**	**5.68**	**315**
	Regression [Log(Q)]	0.79	-0.19	1.38	5.23	315
MS Hankins USGS gage	**SNTEMP**	**0.86**	**-.03**	**1.33**	**-6.84**	**452**
	Regression (normal Q)	0.30	-.34	1.38	43.96	258
	Regression [Log(Q)]	0.21	2.12	1.72	57.98	258
NV Bridgeville USGS gage	SNTEMP	0.88	-0.29	1.43	-6.56	456
	Regression (normal Q)	0.69	0.23	0.94	7.08	178
	Regression [Log(Q)]	**0.69**	**0.23**	**0.93**	**7.09**	**178**

Discussion of Model Comparisons

Certain attributes of these statistical models should be explicitly pointed out. First, *as we implemented them*, the best-fit statistical models often had two large faults: (a) they do not always indicate a negative coefficient for the upstream discharge variable, and (b) they tend to be very "conservative" in describing high temperature events. The problem with predicting positive discharge coefficients, of course, means that the model cannot be used to calculate reservoir releases since they would predict higher temperatures with increased discharge. This is likely due to the inclusion of spill events in the training set, i.e., days when the reservoirs are spilling may indeed increase downstream temperatures if the spill were greater, even though these days represented but a small fraction of the days used to develop the regression. The "conservative" issue is best illustrated in Figure 10, which shows that the statistical model may occasionally predict water temperatures near 23°C (73.4°F) when the measured stream temperatures are only 18-19°C (64.4-66.2°F). Using this relationship would lead to the continuation of a problem noted with today's nomogram procedure, namely wasting water from the conservation account.

Both of these problems might be improved, either by including higher order terms in the regression equations and/or by limiting the training sets to exclude spills or only include days with high temperatures, say above 15°C (59°F). It may also be possible to handle spills in a more robust fashion than was done here. In addition, it might prove fruitful to include estimates of the maximum daily air temperature and "yesterday's" stream temperatures as is done to some degree ("qualitatively and subjectively" according to Rob Klosowski) in the current nomogram approach. Including "yesterday's" stream temperatures has proven valuable in predicting water temperatures for "today" in some applications that we are familiar with, but to our knowledge only in free flowing systems without controlled releases. Nonetheless, one reviewer stressed that such an "autoregressive" technique may prove valuable.

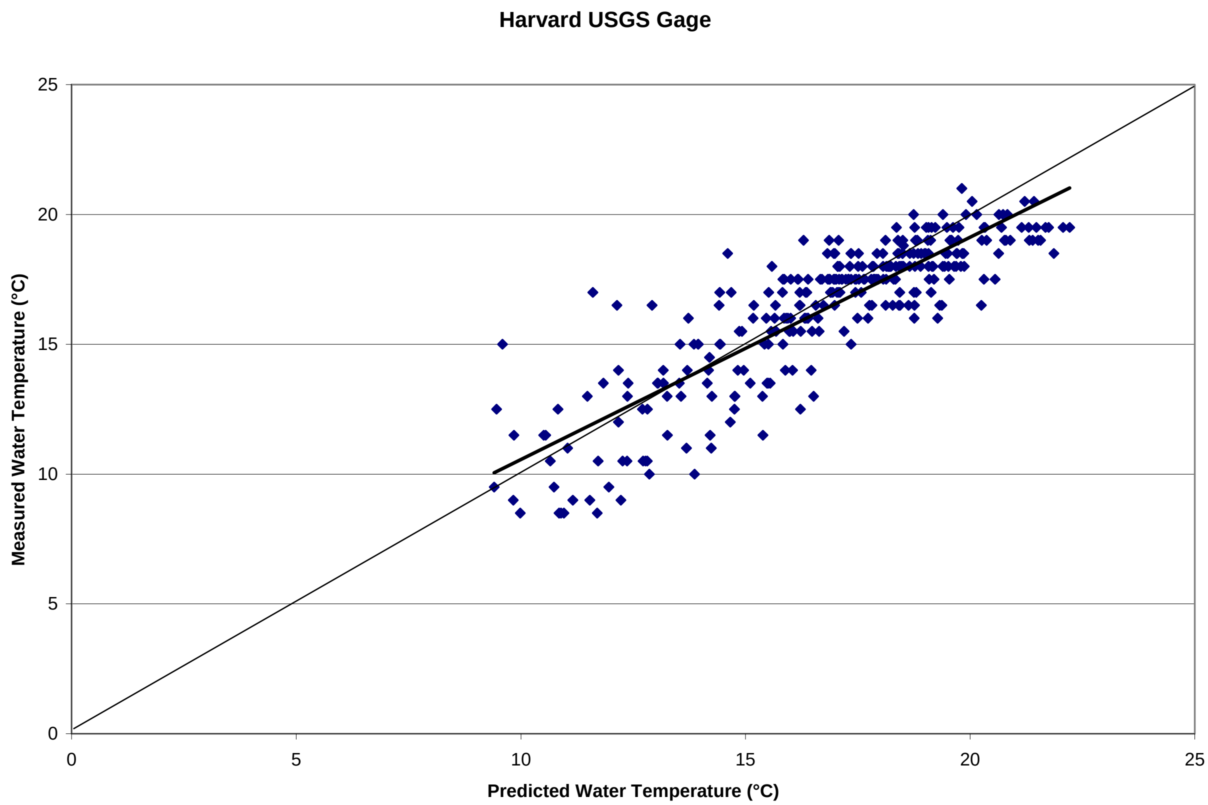

Figure 10. Plot of how one of the statistical models might be too "conservative" in managing reservoir releases. The model itself is generally good as can be seen by the fit of the individual daily points with the thin 45° line. But the points themselves are best described by the darker line, which tilts toward over-prediction at high temperatures.

It should be noted that the SNTEMP models are not free from errors of these sorts, but SNTEMP tended to be far more robust than the statistical models *as we have implemented them.*

Evaluation of SNTEMP with Existing Nomogram Approach

We had anticipated that it would be an easy task to compare SNTEMP's predictive ability with the existing nomogram-derived prediction approach, but did not find this to be the case. We know that the SNTEMP models predict daily water temperatures within about 1.2°C (2.16°F) for mean daily values and within about 1.5°C (2.7°F) for maximum daily values. After speaking with Rob Klosowski, however, it is not apparent that there is currently any historical record comparing the nomogram method with what actually happened in the rivers that would facilitate a comparable evaluation of both methods. Note that even if such a past record were available, this might not be a perfect comparison because the statistics we have for SNTEMP were developed with "perfect" knowledge of the reported meteorology and hydrology whereas the existing nomogram procedure uses forecasts of the weather- and flow-related variables.

It is certainly possible to use SNTEMP to construct a "nomogram" for any given set of meteorologic or hydrologic conditions, and those curves could be compared, but we don't think this would really answer the question. Instead, an SNTEMP modeling technique should be applied for each unique set of circumstances to solve for the discharge required to mitigate high temperature events.

For example, we took the Neversink data set and scanned it for the days with the highest water temperature events (Table 9). We created a scaled-down version of SNTEMP solely for the single worst day (14-July, 1997). On this day, reservoir releases were approximately 80 cfs (2.265 m^3s) at 51.8°F (11°C), the mean daily air temperature was 73.5°F (23.06°C), the relative humidity was 56.6%, the wind speed was 4.0 miles per hour (1.79 meters per second), and it was mostly sunny, about 83% possible sun. Flow in the river at Bridgeville was about 86 cfs (2.435 m^3s).

[For what it is worth, the uncalibrated SNTEMP model performed very well on this day at Bridgeville. The model predicted a mean daily water temperature of 22.47°C (72.45°F) compared to the observed temperature of 22.5°C (72.5°F), and predicted a maximum daily temperature of 26.14°C (79.05°F) compared to 26.0°C (78.8°F). The fidelity of the model on this day at this location could have been happenstance, and, regardless, we would not know this if using the model in a truly forecasting mode. But again there is a "bad news" side to this. Even though the model did well, we would have *expected* some biases and, had we corrected for them, we would have introduced, rather than corrected, a bias.]

Then we ran an SNTEMP utility named TDELTAQ. This program assists in modifying the model's hydrology input file so that you can adjust the flow in the river below reservoirs (or anywhere) and see what the downstream thermal consequences are. We exercised TDELTAQ for this single day by varying the discharge below the Neversink Reservoir from 0.9 to 1.6 times the original 80 cfs flow for that day until temperatures declined to below the target 75°F (23.9°C) maximum. The resulting discharge was 130 cfs (Figure 11).

Table 9. Worst case observed water temperatures at Bridgeville on the Neversink River over the period modeled using SNTEMP.

Year	Day	Mean Daily Temperature (°C)	Maximum Daily Temperature (°C)
1997	14-Jul	22.5	26.0
1997	13-Jul	22.0	26.0
1999	3-Jul	22.0	25.5
1999	28-Jul	21.5	25.5
1997	12-Jul	21.0	25.0
1999	7-Aug	20.5	25.0
1999	12-Aug	20.5	25.0
1998	26-Aug	21.5	24.5
1998	27-Aug	21.5	24.5
1999	16-Jul	21.5	24.5
1997	25-Jun	21.0	24.5
1998	18-Jul	21.0	24.5
1999	6-Aug	21.0	24.5
1999	27-Jul	20.0	24.5
1997	8-Jul	21.0	24.0
1998	4-Aug	21.0	24.0
1997	7-Jul	20.5	24.0
1998	3-Aug	20.5	24.0
1999	29-Aug	20.5	24.0
1997	27-Jun	20.0	24.0
1999	23-Jun	20.0	24.0
1999	15-Jul	20.0	24.0
1999	23-Jul	20.0	24.0

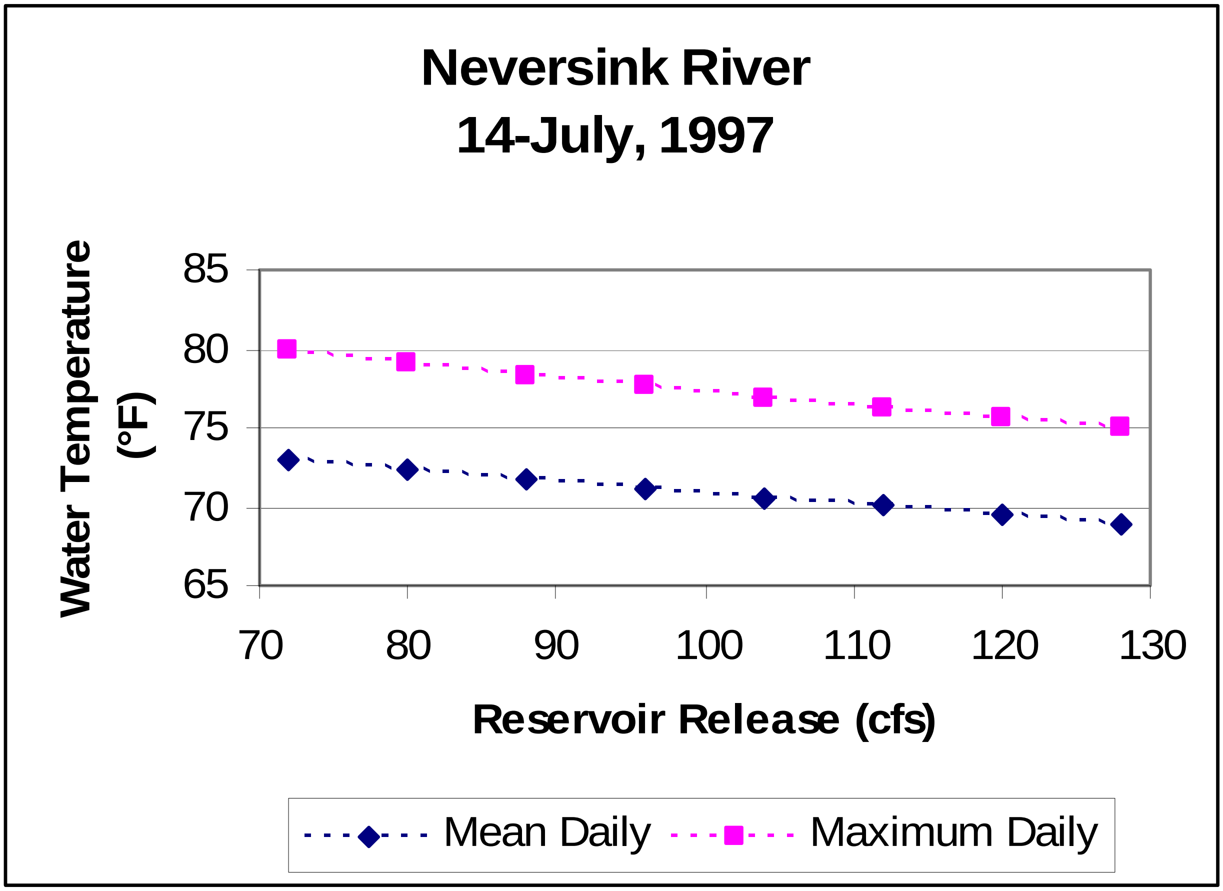

Figure 11. Mean and maximum daily water temperatures predicted at Bridgeville for 14-July, 1997, as a function of reservoir release, showing that the actual discharge near 80 cfs would need to be increased to about 130 cfs to reduce the maximum daily water temperature to the target 75°F.

Because it was not possible to make direct comparisons between methods, perhaps it is best to list some strengths and weaknesses of the various approaches for the Technical Team's use in considering possible next steps. Table 10, including regression models for completeness, captures our interpretations. We are not saying that what we have here is correct; these are just our opinions.

Table 10. Comparison of SNTEMP, Nomogram, and Regression model approaches to deriving reservoir release recommendations.

Attribute	SNTEMP Models	Nomogram Technique	Regression Models
Accuracy	±1.2 to 1.5°C; has some potentially correctable biases	Unquantified	Has potential to be less than 1°C; biases may be correctable
Ease of use (data entry, etc.)	More complex; requires iterative approach to fine-tune release recommendation; uncertain how the two-reservoir system would be handled	Relatively simple	Relatively simple once developed; may require less data entry
Ease of integrating into OASIS or similar model	More difficult and will require more $$	Relatively simple	Relatively simple, but would still require more $$
Can handle a range of hydrologic and meteorologic conditions (basin accretions, etc.)	Yes	Only considers low and high air temperatures	Maybe
Has capacity to “extrapolate” to hydrologic and meteorologic conditions beyond those used to calibrate the model	Yes – but would need to be better tested	To some degree, but is noted to not handle the Beaver Kill thermal inputs well	Maybe – but would need to be better tested
Has the capacity to “interpolate” between locations with measured data	Yes, including handling the interaction of reservoirs on the E-W system	To some degree	To some degree
Has the ability to deal with spills	Yes, to some degree	Unsure	Would likely need two sets of models, one with and one without spills
“Conservatism” in predicting need for too much water	Appears to be less of a problem	Known problem	Needs more work to determine the answer
Can be used to forecast multiple days	Yes	Yes	Yes
Can be improved over time	Yes	Yes	Yes
Sensitivity to errors in the input data	Unquantified	Unquantified	Unquantified
Has ability to estimate thermal benefit from non-flow alternatives	Yes	No	No

Recommendations

Phase I of the temperature model evaluation is now essentially complete. “Complete” in this context means that we accomplished most of what we had set out to do. Delays in getting some of the meteorological data slowed down progress, and we did not perform a full SNTEMP model validation for independent years (previously selected to be 2001-2003) since we found that parameter adjustment did not improve SNTEMP’s predictive ability (the fact that it might be argued that testing the model with five different meteorological stations and choosing the best could be considered a form of calibration, notwithstanding). It is true that our conclusions might have been more solidly based had we used the full six years of data rather than three, but we frankly ran out of time and money. Nonetheless, we can offer the following recommendations to the SEF (besides using the Monticello meteorology):

1. Because evaluation of the SNTEMP results remains to some degree subjective, our principal recommendation is that the SEF carefully review the material in this report to determine whether they believe further SNTEMP model work is warranted. Because the current nomograph approach itself was built to cover worst case scenarios and has not been formally evaluated, we have no objective body of data characterizing its goodness-of-fit to be able to directly compare the two techniques to reality. A valuable step suggested by Rob Klosowski would be to exercise both the existing nomograph approach and SNTEMP using a common historical data set to better define the relative ability of both techniques to predict water temperatures at comparable downstream locations as functions of reservoir release and expected meteorological conditions. Whether the "historical" data set should use "forecast" meteorology or "perfect" meteorology remains an open question.

2. Depending on the outcome of the previous step, further develop either SNTEMP or a statistical approach. Both of these techniques would benefit by accounting for spills as a separate upstream input for both flow volume and temperature. This would allow a much more exact quantification of the effects of reservoir release on downstream temperatures. We have not done this to date because the available upstream data were for the gages at well-mixed locations below the reservoirs. But it might be possible to "back calculate" the two components, resulting in improvements to predictive ability as a function of reservoir releases. Both techniques would also benefit from expanding the data set to include 2001-2003 (or 2004) data.

The regression technique would, in addition, benefit from (a) tuning the model to best predict only water temperatures above some predetermined (biologically relevant) threshold to reduce the effect of over-predicting high temperatures, possibly through using higher-order regression terms; (b) carefully choosing the correct performance measure (e.g., minimize mean, median, or maximum errors), or potentially weighting deviations at high temperatures more heavily than deviations at low temperatures. Finally, (c) accounting for tributary inflows, specifically the Beaver Kill. Any statistical approach would certainly benefit from professional expertise the authors do not have. In a related vein, one could also consider a neural network model that *may* have advantages over more normal regression techniques (Risley et al., 2002).

3. Assuming that either SNTEMP or a statistical approach continued to prove valuable, this would open the door to proceeding with Phase II. In this phase, the chosen model would be prepared for "real time" (or for three-day forecasts) use as a decision support tool. If the statistical approach were chosen, it should be easy to incorporat in the existing spreadsheet.

In the case of SNTEMP, the principal benefit would be to add a user interface that would substitute for the spreadsheet that Rob Klosowski is currently using to specify what "tomorrow's" network hydrology and meteorology (air temperature, wind speed, relative humidity, cloud cover) will be, and develop a procedure that iteratively adjusts reservoir discharges to achieve the specified downstream target temperatures in a fashion similar to that portrayed in Figure 11. Rather than function in degrees Celsius and cubic meters per second like SNTEMP, the interface would also handle temperature and flow units that are the most preferred (or both) and help correct for known model biases.

Phase II should also be able to integrate East Branch flow recommendations (and resulting downstream temperatures) into West Branch recommendations. Perhaps the best way to do this is to have the routine work in that order, i.e., solve for East Branch "constraints" first and then solve for the West Branch. This may turn out to be more complicated than it sounds because there may well be times that high flows in the Beaver Kill make it too "costly" to mitigate water temperatures below its confluence with the East Branch. Adequately determining these situations would take some experimentation. A question unanswered at this time revolves around the microhabitat consequences of varying flows to achieve thermal benefits.

Incorporation of the "real time" temperature model into or linking with a basin model such as OASIS has not been scoped at this time. It well could be that it would be best to use SNTEMP in an "operational" day-to-day forecasting mode and a statistical approach in a "planning" mode.

4. We would envision that Phase II would also more formally address the relationship between the uncertainties inherent in modeling the rivers' thermal responses and conserving the volume of the reservoirs' conservation pools. This is a complex task that depends on two interrelated factors. First, from a fish's perspective, both thermal stress and mortality increase non-linearly with increasing temperatures. Therefore slightly elevated temperatures over a long time result in lower stress and mortality than large temperature increases that may occur if the cold water pools were exhausted. Second, we wish to favor water management strategies that are robust to model uncertainties, yet hedge our bets such that we don't "spend" our thermal account unwisely, essentially putting tomorrow at risk for today's decision. Thoughts on how one might approach this interesting problem follow.

Large volumes of water are necessary to meet temperature targets with a high degree of certainty, but at the risk of exhausting the cold pool. A lower degree of certainty uses less water but with a higher probability that the targets will be periodically exceeded. It may be possible to use information about the distribution of the uncertainty in model predictions to our advantage by calculating the volume of reservoir releases required to achieve desired water temperatures with varying degrees of certainty from our existing data record. Then, for example, one might be willing to live with a ±1.5°C (2.7°F) variance on day one of a hot period. If we found that the calculated flow increment did not succeed in producing the desired temperature and the next day was predicted to be equally hot, one would choose a flow increment that had, say, only a ±1.0°C (1.8°F) expected variance, and so forth. The goal remains to not waste water, but adds to that goal the certainty that you are willing, or need, to attain to reach your target. Though this idea has been widely mentioned in the literature, we are only aware of one paper that has seriously tried to tackle the issue in the context we are addressing here (Neumann 2001).

Then, again from a fish's point of view, we could put a biologically-relevant evaluation metric in place to help decide the certainty level needed. One such metric might be degree-days over an established temperature threshold as a way to track cumulative, detrimental exposures. (Actually we would recommend a degree-squared metric as this better captures cumulative stress related to thermal exposure.) As the cumulative exposure metric increases, our need for ever greater reliability in achieving the temperature standard also increases, reducing the uncertainty you are willing to live with, and presumably increasing the water that you are willing to "spend" to guarantee adherence to the standard. Undoubtedly this sort of strategy is employed today; we could just work on formalizing it. We would want to test the application of such a technique to make sure that it satisfactorily spreads cold releases over the warmest months. However, we must guard against putting a straightjacket on discretion since modeling of any sort will remain an art.

5. More for completeness than anything, we recommend that the SEF consider whether to use SNTEMP to estimate the potential thermal benefits of non-flow rehabilitation alternatives. This last recommendation relates to the last row in Table 10, estimation of thermal benefits through non-flow means. Initial observations of the Upper Delaware River system indicate relatively wide, shallow stream segments with few instream obstacles (e.g., fallen trees) that contribute to channel complexity and may reduce effective stream width. Relatively even-aged tree cover would seem to indicate nearly synchronized timber harvest that (1) reduces the probability of downed trees when there is uniform and continuous riparian cover, and (2) has resulted in some stream reaches that remain lightly forested today. Though we have not yet exercised the SNTEMP model to address non-flow scenarios, it is certainly possible to estimate the thermal benefit of reducing effective stream widths, increasing depths, and/or increasing riparian shade (e.g., Bartholow 1991) if the SEF group felt that actions aimed at river restoration or rehabilitation were in order. On the other hand, from conversations with Ken Bovee, opportunities for "adjusting" the existing channel may be limited due to the current main channels' armored, stable and down-cut configuration (as indicated by Hurricane Ivan's inability to cause much channel change), and the lack of sand and gravel input from the tributaries to build banks, bars, or additional islands. Nonetheless, we wanted to mention non-flow alternatives in case it triggered additional thoughts on the part of the SEF.

6. Consider the provision of getting reliable meteorology data from a more representative station such as Liberty, NY, or installation of one or more on-river stations. Though we have no definitive proof that local meteorological data will improve the predictive ability of either modeling approach, and attempts at SNTEMP model calibration proved fruitless, it remains clear that there are systematic biases between model residuals and several of the meteorological inputs. These biases argue that using more local data may offer a "controllable" way to improve the accuracy of these models. However, weather forecasts will likely be derived from conventional weather stations. For this reason, this recommendation is last on our list and represents a low priority.

A compact disk with the SNTEMP input data files for the two network models, and associated data files used to derive them, is available on request. We would be happy to come and brief the SEF group on our findings, or offer a training session on model use, if you believe that would be worthwhile and travel funds are available.

Acknowledgements

We'd like to thank the many members of the SEF, NY-DEC, NY-DEP, USGS and others who contributed their time and expertise to this effort. Robert Angyal, Alan Ellsworth, Colin Apse, Jerry Butch, Wayne Elliot, Glenn Horton, Robert Klosowski, Bill Lellis, Rick Naro, Charles Olson, Margaret Phillips, Don Pierson, Hernan Quinodoz, and many others we may be forgetting at the moment. All contributed very positively to this analysis, especially Rob Klosowski.

References

Bartholow, J.M. 1991. A modeling assessment of the thermal regime for an urban sport fishery. *Environmental Management.* 15(6):833-845. Available on the Internet at *http://www.fort.usgs.gov/products/pubs/4061/4061.asp*

Bartholow, J.M. 2002. Modeling Uncertainty: Quicksand for Water Temperature Modeling. Available on the Internet at *http://www.fort.usgs.gov/products/Publications/11043/11043.asp*

Bartholow, J.M. 2000. The Stream Segment and Stream Network Temperature Models: A Self-Study Course, Version 2.0. U.S. Geological Survey Open File Report 99-112. 276 pp. Available on the Internet at *http://www.fort.usgs.gov/products/publications/4090/4090.asp*

Bartholow, J.M. 1989. Stream temperature investigations: field and analytic methods. Instream Flow Information Paper No. 13. U.S. Fish Wildl. Serv. Biol. Rep. 89(17). 139 pp. Available on request.

Neumann, D.W. R. Balaji, E. Zagona, M. Bender, and T. Scott. 2001. An operations model for temperature management of the Truckee River . M.S. Thesis, University of Colorado. Available on the Internet at *http://cadswes.colorado.edu/PDF/RiverWare/NeumannLV2002.pdf.*

New York Department of Conservation. No Date. Monitoring Plan for Delaware River Tailwaters, 2004-2006. 14 pages plus tables and graphs. (Sent to me by Colin Apse, TNC; reportedly written by Wayne Elliott, NY-DEC)

Risley, J.C., E.A. Roehl, Jr., and P.A. Conrads. 2002. Estimating Water Temperatures in Small Streams in Western Oregon Using Neural Network Models. U.S. Geological Survey Water-Resources Investigations Report 02-4218. Available on the Internet at http://water.usgs.gov/pubs/wri/wri024218/

Thatcher, M.L. and C. Mendoza. 1990. Extension and testing of the daily flow model. Columbia University Dept. of Civil Engineering and Engineering Mechanics. Report prepared for the Delaware River Basin Commission. v.p.

Theurer, F.D., K.A. Voos, and W.J. Miller. 1984. Instream Water Temperature Model. Instream Flow Inf. Pap. 16. U.S. Fish and Wildl. Serv. FWS/OBS-84/15. v.p. Available on request.

White, K.E. and T.W. Kratzer. 1994. Determination of travel time in the Delaware River, Hancock, New York, to the Delaware Water Gap by use of a conservative dye tracer. USGS Water-Resources Investigations Report 93-4203. 54 pages.

Note: Additional references relevant to the SNTEMP model may be found at *http://www.fort.usgs.gov/products/software/SNTEMP/SNTEMP_refs.asp*

Appendix

Table A 1. Summary of verification location annual mean daily goodness-of-fit statistics for the East/West/Mainstem Delaware River for initial model run. Stream distance is measured from the downstream terminus of the study area. USGS gage numbers or DEC UTM coordinates are given in parentheses.

STREAM NAME	NODE TYPE	(KM) STREAM DISTANCE	(R) CORR. COEF.	(C) MEAN ERROR	(+-C) PROB. ERROR	(C) MAX. ERROR	NO. TERMS	REMARKS
MS Delaware R	V	66.5	0.9145	0.74	0.82	4.16	292	WB DEC Record at Men's Club (465204E/4654777N)
MS Delaware R	V	60.1	0.8684	0.40	1.05	-5.76	427	WB Hale Eddy Gage (01426500)
MS Delaware R	V	57.8	0.8706	0.91	1.04	4.71	292	WB DEC Record at Roods Creek (470129E/4649820N)
MS Delaware R	V	52.4	0.8160	1.23	1.22	6.04	292	WB DEC Record at Balls Eddy (472525E/4646402N)
MS Delaware R	V	47.0	0.8544	0.55	1.03	5.11	456	WB Hancock Gage (01427000)
MS Delaware R	V	45.6	0.8244	0.74	1.13	5.30	347	WB DEC Record at Shehawken Creek (476332E/46433
EB Delaware R	V	81.3	0.9037	1.49	0.85	5.50	339	DEC Record at Terry's Campground (494290E/46509
EB Delaware R	V	78.6	0.8843	1.46	1.04	6.05	339	DEC Record at Deutch's Flats (491976E/4652531N)
EB Delaware R	V	75.7	0.8965	1.21	1.20	6.28	456	Harvard Gage (01417500)
EB Delaware R	V	75.6	0.8987	1.42	1.07	6.24	319	DEC Record at Harvard Gage (490073E/4652531N)
EB Delaware R	V	46.5	0.8890	-1.07	1.14	-7.23	337	DEC Record at Hancock Fireman's Park (476601E/4
MS Delaware R	V	40.7	0.8067	0.66	1.24	5.61	292	DEC Record at Leonard's (477667E/4638781N)
MS Delaware R	V	28.3	0.7891	-0.10	1.21	-4.98	292	DEC Record at Abe Lord Creek (483879E/4635067N)
MS Delaware R	V	19.5	0.7699	0.18	1.29	-6.07	316	DEC Record at Long Eddy (489001/4632656)
MS Delaware R	V	15.4	0.7912	0.53	1.24	-5.89	225	DEC Record at Kellams (490593/4630211)
MS Delaware R	V	15.4	0.8480	0.01	1.29	-6.09	452	Hankins Gage (01427301)
MS Delaware R	V	12.4	0.7125	0.20	1.37	5.99	200	DEC Record at Hankins (492890/4628883)
MS Delaware R	V	0.1	0.8796	-0.44	1.24	-5.88	441	Callicoon Gage (01427510)
SUMMARY: VALIDATION TYPE NODES			0.8916	0.55	1.23	-7.23	6114	ALL VALIDATION & CALIBRATION NODES

Table A 2. Summary of verification location annual mean daily goodness-of-fit statistics for the East/West/Mainstem Delaware River for initial model run. Stream distance is measured from the downstream terminus of the study area. USGS gage numbers or DEC UTM coordinates are given in parentheses.

STREAM NAME	NODE TYPE	(KM) STREAM DISTANCE	(R) CORR. COEF.	(C) MEAN ERROR	(+-C) PROB. ERROR	(C) MAX. ERROR	NO. TERMS	REMARKS
Neversink R	V	17.0	0.6407	0.43	0.76	3.40	313	DEC Record at Woodbourne (533455/4622831)
Neversink R	V	13.9	0.6561	0.56	1.01	-4.18	313	DEC Record at Fallsburg (532927/4620177)
Neversink R	V	5.5	0.7860	0.34	1.28	-5.70	313	DEC Record near Thompsonville/Ranch Road (53334
Neversink R	V	0.0	0.8751	-0.54	1.26	-5.88	456	Bridgeville (01436690)
SUMMARY: VALIDATION TYPE NODES			0.8379	0.12	1.16	-5.88	1395	ALL VALIDATION & CALIBRATION NODES

Table A 3. Correlation of temperature model input parameters with mean daily model error for the Upper Delaware River Network. The R value measures the degree and sign of the linear association of each input parameter with model error (simulated - observed) over time. The R2 value measures the ratio of explained variance with total variance. The PROBability value is the significance level at which the null hypothesis of zero correlation is rejected (the probability whose small value indicates a significant correlation). A flag (*) marks any correlation whose R2 value is greater than 0.20 AND whose PROB value is less than 0.25.

VARIABLE	LOCATION			R	R2	PROB	FLAG
AIR TEMP	MS Delaware R	V	66.5	0.41	0.17	0.00	
AIR TEMP	MS Delaware R	V	60.1	0.70	0.49	******	*
AIR TEMP	MS Delaware R	V	57.8	0.49	0.24	0.00	*
AIR TEMP	MS Delaware R	V	52.4	0.59	0.35	0.00	*
AIR TEMP	MS Delaware R	V	47.0	0.52	0.27	0.00	*
AIR TEMP	MS Delaware R	V	45.6	0.57	0.33	0.00	*
AIR TEMP	EB Delaware R	V	81.3	0.46	0.21	0.00	*
AIR TEMP	EB Delaware R	V	78.6	0.44	0.20	0.00	
AIR TEMP	EB Delaware R	V	75.7	0.59	0.34	0.00	*
AIR TEMP	EB Delaware R	V	75.6	0.42	0.17	0.00	
AIR TEMP	EB Delaware R	V	46.5	0.19	0.04	0.01	
AIR TEMP	MS Delaware R	V	40.7	0.45	0.20	0.00	*
AIR TEMP	MS Delaware R	V	28.3	0.39	0.15	0.00	
AIR TEMP	MS Delaware R	V	19.5	0.42	0.18	0.00	
AIR TEMP	MS Delaware R	V	15.4	0.30	0.09	0.00	
AIR TEMP	MS Delaware R	V	15.4	0.41	0.17	0.00	
AIR TEMP	MS Delaware R	V	12.4	0.25	0.06	0.00	
AIR TEMP	MS Delaware R	V	0.1	0.34	0.12	0.00	
AIR TEMP	SUMMARY: VALIDATION TYPE NODES			0.58	0.34	0.00	*
WIND SPEED	MS Delaware R	V	66.5	-0.18	0.03	0.09	
WIND SPEED	MS Delaware R	V	60.1	-0.27	0.08	0.01	
WIND SPEED	MS Delaware R	V	57.8	-0.23	0.06	0.04	
WIND SPEED	MS Delaware R	V	52.4	-0.23	0.05	0.05	
WIND SPEED	MS Delaware R	V	47.0	-0.20	0.04	0.05	
WIND SPEED	MS Delaware R	V	45.6	-0.26	0.07	0.03	
WIND SPEED	EB Delaware R	V	81.3	-0.25	0.06	0.03	
WIND SPEED	EB Delaware R	V	78.6	-0.28	0.08	0.02	
WIND SPEED	EB Delaware R	V	75.7	-0.34	0.12	0.00	
WIND SPEED	EB Delaware R	V	75.6	-0.29	0.09	0.02	
WIND SPEED	EB Delaware R	V	46.5	-0.29	0.08	0.02	
WIND SPEED	MS Delaware R	V	40.7	-0.27	0.07	0.02	
WIND SPEED	MS Delaware R	V	28.3	-0.23	0.05	0.04	
WIND SPEED	MS Delaware R	V	19.5	-0.18	0.03	0.11	
WIND SPEED	MS Delaware R	V	15.4	-0.16	0.03	0.15	
WIND SPEED	MS Delaware R	V	15.4	-0.30	0.09	0.01	
WIND SPEED	MS Delaware R	V	12.4	-0.22	0.05	0.07	
WIND SPEED	MS Delaware R	V	0.1	-0.32	0.10	0.01	
WIND SPEED	SUMMARY: VALIDATION TYPE NODES			-0.37	0.14	0.00	
RELATIVE HUMIDITY	MS Delaware R	V	66.5	0.10	0.01	0.20	
RELATIVE HUMIDITY	MS Delaware R	V	60.1	0.12	0.01	0.10	
RELATIVE HUMIDITY	MS Delaware R	V	57.8	0.15	0.02	0.05	
RELATIVE HUMIDITY	MS Delaware R	V	52.4	0.12	0.01	0.12	
RELATIVE HUMIDITY	MS Delaware R	V	47.0	0.03	0.00	0.73	
RELATIVE HUMIDITY	MS Delaware R	V	45.6	0.01	0.00	0.88	
RELATIVE HUMIDITY	EB Delaware R	V	81.3	0.05	0.00	0.54	
RELATIVE HUMIDITY	EB Delaware R	V	78.6	0.10	0.01	0.22	
RELATIVE HUMIDITY	EB Delaware R	V	75.7	0.13	0.02	0.08	

RELATIVE HUMIDITY	EB Delaware R	V	75.6	0.02	0.00	0.80
RELATIVE HUMIDITY	EB Delaware R	V	46.5	-0.08	0.01	0.41
RELATIVE HUMIDITY	MS Delaware R	V	40.7	0.10	0.01	0.21
RELATIVE HUMIDITY	MS Delaware R	V	28.3	0.14	0.02	0.08
RELATIVE HUMIDITY	MS Delaware R	V	19.5	-0.01	0.00	0.88
RELATIVE HUMIDITY	MS Delaware R	V	15.4	0.02	0.00	0.86
RELATIVE HUMIDITY	MS Delaware R	V	15.4	-0.01	0.00	0.92
RELATIVE HUMIDITY	MS Delaware R	V	12.4	0.01	0.00	0.91
RELATIVE HUMIDITY	MS Delaware R	V	0.1	-0.01	0.00	0.90
RELATIVE HUMIDITY	SUMMARY: VALIDATION TYPE NODES			0.12	0.01	0.10
PERCENT SUN	MS Delaware R	V	66.5	0.05	0.00	0.56
PERCENT SUN	MS Delaware R	V	60.1	0.14	0.02	0.05
PERCENT SUN	MS Delaware R	V	57.8	0.03	0.00	0.76
PERCENT SUN	MS Delaware R	V	52.4	0.06	0.00	0.51
PERCENT SUN	MS Delaware R	V	47.0	0.18	0.03	0.01
PERCENT SUN	MS Delaware R	V	45.6	0.14	0.02	0.07
PERCENT SUN	EB Delaware R	V	81.3	0.10	0.01	0.22
PERCENT SUN	EB Delaware R	V	78.6	0.07	0.01	0.38
PERCENT SUN	EB Delaware R	V	75.7	0.18	0.03	0.01
PERCENT SUN	EB Delaware R	V	75.6	0.14	0.02	0.07
PERCENT SUN	EB Delaware R	V	46.5	0.28	0.08	0.00
PERCENT SUN	MS Delaware R	V	40.7	0.08	0.01	0.33
PERCENT SUN	MS Delaware R	V	28.3	0.08	0.01	0.31
PERCENT SUN	MS Delaware R	V	19.5	0.17	0.03	0.04
PERCENT SUN	MS Delaware R	V	15.4	0.08	0.01	0.38
PERCENT SUN	MS Delaware R	V	15.4	0.30	0.09	0.00
PERCENT SUN	MS Delaware R	V	12.4	0.15	0.02	0.06
PERCENT SUN	MS Delaware R	V	0.1	0.31	0.10	0.00
PERCENT SUN	SUMMARY: VALIDATION TYPE NODES			0.22	0.05	0.00
OBSERVED SOLAR	MS Delaware R	V	66.5	-----NOT PRESENT-----		
OBSERVED SOLAR	MS Delaware R	V	60.1	-----NOT PRESENT-----		
OBSERVED SOLAR	MS Delaware R	V	57.8	-----NOT PRESENT-----		
OBSERVED SOLAR	MS Delaware R	V	52.4	-----NOT PRESENT-----		
OBSERVED SOLAR	MS Delaware R	V	47.0	-----NOT PRESENT-----		
OBSERVED SOLAR	MS Delaware R	V	45.6	-----NOT PRESENT-----		
OBSERVED SOLAR	EB Delaware R	V	81.3	-----NOT PRESENT-----		
OBSERVED SOLAR	EB Delaware R	V	78.6	-----NOT PRESENT-----		
OBSERVED SOLAR	EB Delaware R	V	75.7	-----NOT PRESENT-----		
OBSERVED SOLAR	EB Delaware R	V	75.6	-----NOT PRESENT-----		
OBSERVED SOLAR	EB Delaware R	V	46.5	-----NOT PRESENT-----		
OBSERVED SOLAR	MS Delaware R	V	40.7	-----NOT PRESENT-----		
OBSERVED SOLAR	MS Delaware R	V	28.3	-----NOT PRESENT-----		
OBSERVED SOLAR	MS Delaware R	V	19.5	-----NOT PRESENT-----		
OBSERVED SOLAR	MS Delaware R	V	15.4	-----NOT PRESENT-----		
OBSERVED SOLAR	MS Delaware R	V	15.4	-----NOT PRESENT-----		
OBSERVED SOLAR	MS Delaware R	V	12.4	-----NOT PRESENT-----		
OBSERVED SOLAR	MS Delaware R	V	0.1	-----NOT PRESENT-----		
OBSERVED SOLAR	SUMMARY: VALIDATION TYPE NODES			-----NOT PRESENT-----		
DUST COEFF	MS Delaware R	V	66.5	0.00	0.00	1.00
DUST COEFF	MS Delaware R	V	60.1	0.00	0.00	1.00
DUST COEFF	MS Delaware R	V	57.8	0.00	0.00	1.00
DUST COEFF	MS Delaware R	V	52.4	0.00	0.00	1.00
DUST COEFF	MS Delaware R	V	47.0	0.00	0.00	1.00
DUST COEFF	MS Delaware R	V	45.6	0.00	0.00	1.00
DUST COEFF	EB Delaware R	V	81.3	0.00	0.00	1.00
DUST COEFF	EB Delaware R	V	78.6	0.00	0.00	1.00
DUST COEFF	EB Delaware R	V	75.7	0.00	0.00	1.00
DUST COEFF	EB Delaware R	V	75.6	0.00	0.00	1.00
DUST COEFF	EB Delaware R	V	46.5	0.00	0.00	1.00
DUST COEFF	MS Delaware R	V	40.7	0.00	0.00	1.00

DUST COEFF	MS Delaware R	V	28.3	0.00	0.00	1.00
DUST COEFF	MS Delaware R	V	19.5	0.00	0.00	1.00
DUST COEFF	MS Delaware R	V	15.4	0.00	0.00	1.00
DUST COEFF	MS Delaware R	V	15.4	0.00	0.00	1.00
DUST COEFF	MS Delaware R	V	12.4	0.00	0.00	1.00
DUST COEFF	MS Delaware R	V	0.1	0.00	0.00	1.00
DUST COEFF	SUMMARY: VALIDATION TYPE NODES			0.00	0.00	1.00
REFLECTIVITY	MS Delaware R	V	66.5	0.00	0.00	1.00
REFLECTIVITY	MS Delaware R	V	60.1	0.00	0.00	1.00
REFLECTIVITY	MS Delaware R	V	57.8	0.00	0.00	1.00
REFLECTIVITY	MS Delaware R	V	52.4	0.00	0.00	1.00
REFLECTIVITY	MS Delaware R	V	47.0	0.00	0.00	1.00
REFLECTIVITY	MS Delaware R	V	45.6	0.00	0.00	1.00
REFLECTIVITY	EB Delaware R	V	81.3	0.00	0.00	1.00
REFLECTIVITY	EB Delaware R	V	78.6	0.00	0.00	1.00
REFLECTIVITY	EB Delaware R	V	75.7	0.00	0.00	1.00
REFLECTIVITY	EB Delaware R	V	75.6	0.00	0.00	1.00
REFLECTIVITY	EB Delaware R	V	46.5	0.00	0.00	1.00
REFLECTIVITY	MS Delaware R	V	40.7	0.00	0.00	1.00
REFLECTIVITY	MS Delaware R	V	28.3	0.00	0.00	1.00
REFLECTIVITY	MS Delaware R	V	19.5	0.00	0.00	1.00
REFLECTIVITY	MS Delaware R	V	15.4	0.00	0.00	1.00
REFLECTIVITY	MS Delaware R	V	15.4	0.00	0.00	1.00
REFLECTIVITY	MS Delaware R	V	12.4	0.00	0.00	1.00
REFLECTIVITY	MS Delaware R	V	0.1	0.00	0.00	1.00
REFLECTIVITY	SUMMARY: VALIDATION TYPE NODES			0.00	0.00	1.00
SHADE	MS Delaware R	V	66.5	-0.15	0.02	0.15
SHADE	MS Delaware R	V	60.1	-0.30	0.09	0.01
SHADE	MS Delaware R	V	57.8	0.00	0.00	0.99
SHADE	MS Delaware R	V	52.4	-0.13	0.02	0.20
SHADE	MS Delaware R	V	47.0	-0.28	0.08	0.01
SHADE	MS Delaware R	V	45.6	-0.10	0.01	0.29
SHADE	EB Delaware R	V	81.3	0.26	0.07	0.00
SHADE	EB Delaware R	V	78.6	0.29	0.09	0.00
SHADE	EB Delaware R	V	75.7	0.19	0.04	0.01
SHADE	EB Delaware R	V	75.6	0.23	0.05	0.00
SHADE	EB Delaware R	V	46.5	0.22	0.05	0.00
SHADE	MS Delaware R	V	40.7	0.23	0.05	0.00
SHADE	MS Delaware R	V	28.3	0.23	0.05	0.00
SHADE	MS Delaware R	V	19.5	-0.05	0.00	0.63
SHADE	MS Delaware R	V	15.4	0.27	0.07	0.00
SHADE	MS Delaware R	V	15.4	0.11	0.01	0.12
SHADE	MS Delaware R	V	12.4	-0.04	0.00	0.71
SHADE	MS Delaware R	V	0.1	0.17	0.03	0.01
SHADE	SUMMARY: VALIDATION TYPE NODES			0.18	0.03	0.01
SOLAR RADIATION	MS Delaware R	V	66.5	0.09	0.01	0.28
SOLAR RADIATION	MS Delaware R	V	60.1	0.17	0.03	0.02
SOLAR RADIATION	MS Delaware R	V	57.8	-0.02	0.00	0.83
SOLAR RADIATION	MS Delaware R	V	52.4	0.13	0.02	0.11
SOLAR RADIATION	MS Delaware R	V	47.0	0.22	0.05	0.00
SOLAR RADIATION	MS Delaware R	V	45.6	0.13	0.02	0.10
SOLAR RADIATION	EB Delaware R	V	81.3	-0.09	0.01	0.36
SOLAR RADIATION	EB Delaware R	V	78.6	-0.13	0.02	0.19
SOLAR RADIATION	EB Delaware R	V	75.7	0.03	0.00	0.69
SOLAR RADIATION	EB Delaware R	V	75.6	-0.03	0.00	0.76
SOLAR RADIATION	EB Delaware R	V	46.5	0.09	0.01	0.28
SOLAR RADIATION	MS Delaware R	V	40.7	-0.07	0.01	0.45
SOLAR RADIATION	MS Delaware R	V	28.3	-0.08	0.01	0.42
SOLAR RADIATION	MS Delaware R	V	19.5	0.10	0.01	0.25
SOLAR RADIATION	MS Delaware R	V	15.4	-0.09	0.01	0.36

SOLAR RADIATION	MS Delaware R	V	15.4	0.18	0.03	0.01	
SOLAR RADIATION	MS Delaware R	V	12.4	0.09	0.01	0.30	
SOLAR RADIATION	MS Delaware R	V	0.1	0.15	0.02	0.04	
SOLAR RADIATION	SUMMARY: VALIDATION TYPE NODES			0.08	0.01	0.27	
TIME	MS Delaware R	V	66.5	0.12	0.01	0.14	
TIME	MS Delaware R	V	60.1	0.27	0.08	0.00	
TIME	MS Delaware R	V	57.8	0.31	0.09	0.00	
TIME	MS Delaware R	V	52.4	0.12	0.01	0.14	
TIME	MS Delaware R	V	47.0	0.10	0.01	0.16	
TIME	MS Delaware R	V	45.6	0.20	0.04	0.01	
TIME	EB Delaware R	V	81.3	0.45	0.20	0.00	
TIME	EB Delaware R	V	78.6	0.50	0.25	0.00	*
TIME	EB Delaware R	V	75.7	0.44	0.19	0.00	
TIME	EB Delaware R	V	75.6	0.39	0.16	0.00	
TIME	EB Delaware R	V	46.5	0.21	0.04	0.00	
TIME	MS Delaware R	V	40.7	0.46	0.21	0.00	*
TIME	MS Delaware R	V	28.3	0.44	0.19	0.00	
TIME	MS Delaware R	V	19.5	0.21	0.04	0.01	
TIME	MS Delaware R	V	15.4	0.39	0.15	0.00	
TIME	MS Delaware R	V	15.4	0.24	0.06	0.00	
TIME	MS Delaware R	V	12.4	0.16	0.03	0.05	
TIME	MS Delaware R	V	0.1	0.27	0.07	0.00	
TIME	SUMMARY: VALIDATION TYPE NODES			0.41	0.17	0.00	
FLOW	MS Delaware R	V	66.5	0.58	0.34	0.00	*
FLOW	MS Delaware R	V	60.1	0.24	0.06	0.00	
FLOW	MS Delaware R	V	57.8	0.42	0.17	0.00	
FLOW	MS Delaware R	V	52.4	0.45	0.20	0.00	*
FLOW	MS Delaware R	V	47.0	0.23	0.05	0.00	
FLOW	MS Delaware R	V	45.6	0.39	0.15	0.00	
FLOW	EB Delaware R	V	81.3	-0.38	0.14	0.01	
FLOW	EB Delaware R	V	78.6	-0.36	0.13	0.01	
FLOW	EB Delaware R	V	75.7	-0.44	0.20	0.00	
FLOW	EB Delaware R	V	75.6	-0.32	0.10	0.01	
FLOW	EB Delaware R	V	46.5	-0.14	0.02	0.18	
FLOW	MS Delaware R	V	40.7	-0.19	0.03	0.09	
FLOW	MS Delaware R	V	28.3	-0.14	0.02	0.19	
FLOW	MS Delaware R	V	19.5	-0.17	0.03	0.13	
FLOW	MS Delaware R	V	15.4	-0.27	0.07	0.03	
FLOW	MS Delaware R	V	15.4	-0.20	0.04	0.05	
FLOW	MS Delaware R	V	12.4	-0.16	0.02	0.17	
FLOW	MS Delaware R	V	0.1	-0.18	0.03	0.07	
FLOW	SUMMARY: VALIDATION TYPE NODES			-0.35	0.12	0.00	

NUMBER OF OBSERVATIONS IS --> 459

DISTANCE	FOR ALL V NODES			0.63	0.40	0.00	*

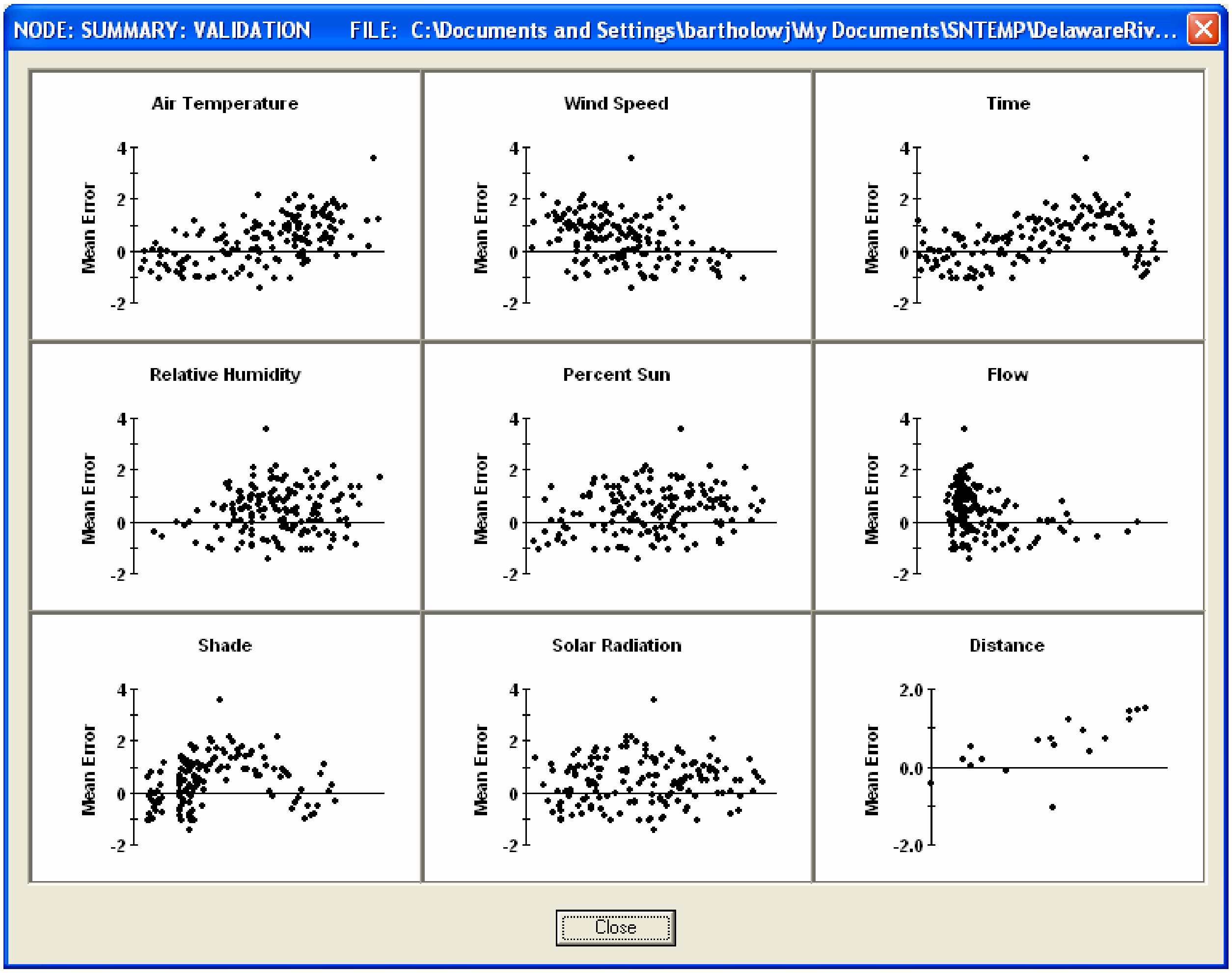

Figure A1. Residual plots for each of the significant model inputs, in this case for all verification nodes on the East/West Branch Delaware River temperature model. X-axis represents the breadth of the input data for each variable.

Table A 4. Correlation of temperature model input parameters with mean daily model error for the Neversink River below Neversink Reservoir. The R value measures the degree and sign of the linear association of each input parameter with model error (simulated - observed) over time. The R2 value measures the ratio of explained variance with total variance. The PROBability value is the significance level at which the null hypothesis of zero correlation is rejected (the probability whose small value indicates a significant correlation). A flag (*) marks any correlation whose R2 value is greater than 0.20 AND whose PROB value is less than 0.25.

VARIABLE	LOCATION			R	R2	PROB	FLAG
AIR TEMP	Neversink R	V	17.0	0.46	0.21	0.00	*
AIR TEMP	Neversink R	V	13.9	0.48	0.23	0.00	*
AIR TEMP	Neversink R	V	5.5	0.49	0.24	0.00	*
AIR TEMP	Neversink R	V	0.0	0.55	0.30	0.00	*
AIR TEMP	SUMMARY: VALIDATION TYPE NODES			0.70	0.49	******	*
WIND SPEED	Neversink R	V	17.0	0.06	0.00	0.49	
WIND SPEED	Neversink R	V	13.9	0.01	0.00	0.93	
WIND SPEED	Neversink R	V	5.5	-0.12	0.02	0.23	
WIND SPEED	Neversink R	V	0.0	-0.41	0.17	0.00	
WIND SPEED	SUMMARY: VALIDATION TYPE NODES			-0.41	0.17	0.00	
RELATIVE HUMIDITY	Neversink R	V	17.0	0.07	0.00	0.45	
RELATIVE HUMIDITY	Neversink R	V	13.9	0.01	0.00	0.90	
RELATIVE HUMIDITY	Neversink R	V	5.5	-0.08	0.01	0.42	
RELATIVE HUMIDITY	Neversink R	V	0.0	0.04	0.00	0.63	
RELATIVE HUMIDITY	SUMMARY: VALIDATION TYPE NODES			0.20	0.04	0.00	
PERCENT SUN	Neversink R	V	17.0	-0.13	0.02	0.21	
PERCENT SUN	Neversink R	V	13.9	-0.03	0.00	0.72	
PERCENT SUN	Neversink R	V	5.5	0.15	0.02	0.07	
PERCENT SUN	Neversink R	V	0.0	0.28	0.08	0.00	
PERCENT SUN	SUMMARY: VALIDATION TYPE NODES			0.15	0.02	0.03	
OBSERVED SOLAR	Neversink R	V	17.0	-----NOT PRESENT-----			
OBSERVED SOLAR	Neversink R	V	13.9	-----NOT PRESENT-----			
OBSERVED SOLAR	Neversink R	V	5.5	-----NOT PRESENT-----			
OBSERVED SOLAR	Neversink R	V	0.0	-----NOT PRESENT-----			
OBSERVED SOLAR	SUMMARY: VALIDATION TYPE NODES			-----NOT PRESENT-----			
DUST COEFF	Neversink R	V	17.0	0.00	0.00	1.00	
DUST COEFF	Neversink R	V	13.9	0.00	0.00	1.00	
DUST COEFF	Neversink R	V	5.5	0.00	0.00	1.00	
DUST COEFF	Neversink R	V	0.0	0.00	0.00	1.00	
DUST COEFF	SUMMARY: VALIDATION TYPE NODES			0.00	0.00	1.00	
REFLECTIVITY	Neversink R	V	17.0	0.00	0.00	1.00	
REFLECTIVITY	Neversink R	V	13.9	0.00	0.00	1.00	
REFLECTIVITY	Neversink R	V	5.5	0.00	0.00	1.00	
REFLECTIVITY	Neversink R	V	0.0	0.00	0.00	1.00	
REFLECTIVITY	SUMMARY: VALIDATION TYPE NODES			0.00	0.00	1.00	
SHADE	Neversink R	V	17.0	-0.54	0.29	0.00	*
SHADE	Neversink R	V	13.9	-0.48	0.23	0.00	*
SHADE	Neversink R	V	5.5	-0.33	0.11	0.01	
SHADE	Neversink R	V	0.0	0.00	0.00	0.98	
SHADE	SUMMARY: VALIDATION TYPE NODES			-0.13	0.02	0.17	
SOLAR RADIATION	Neversink R	V	17.0	0.20	0.04	0.01	

SOLAR RADIATION	Neversink R	V	13.9	0.24	0.06	0.00	
SOLAR RADIATION	Neversink R	V	5.5	0.30	0.09	0.00	
SOLAR RADIATION	Neversink R	V	0.0	0.21	0.04	0.00	
SOLAR RADIATION	SUMMARY: VALIDATION TYPE NODES			0.14	0.02	0.05	
TIME	Neversink R	V	17.0	-0.50	0.25	0.00	*
TIME	Neversink R	V	13.9	-0.43	0.18	0.00	
TIME	Neversink R	V	5.5	-0.27	0.08	0.03	
TIME	Neversink R	V	0.0	0.24	0.06	0.00	
TIME	SUMMARY: VALIDATION TYPE NODES			0.29	0.09	0.00	
FLOW	Neversink R	V	17.0	0.13	0.02	0.11	
FLOW	Neversink R	V	13.9	0.06	0.00	0.47	
FLOW	Neversink R	V	5.5	-0.06	0.00	0.56	
FLOW	Neversink R	V	0.0	-0.33	0.11	0.00	
FLOW	SUMMARY: VALIDATION TYPE NODES			-0.43	0.18	0.00	

NUMBER OF OBSERVATIONS IS --> 459

DISTANCE	FOR ALL V NODES	0.59	0.35	0.24	*

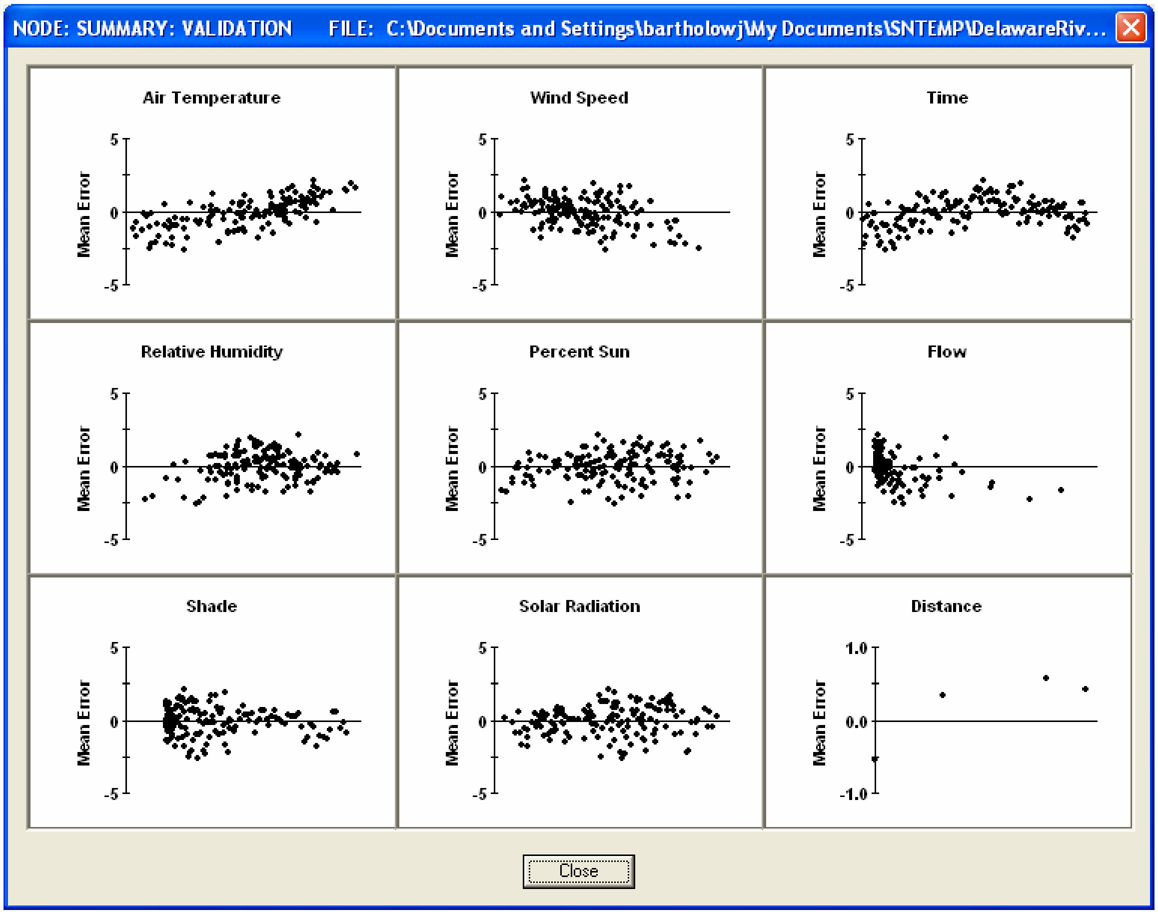

Figure A2. Residual plots for each of the significant model inputs, in this case for all verification nodes on the Neversink River. X-axis represents the breadth of the input data for each variable.

Table A 5. Summary of verification location annual maximum daily goodness-of-fit statistics for the East/West/Mainstem Delaware River for initial model run. Stream distance is measured from the downstream terminus of the study area. USGS gage numbers or DEC UTM coordinates are given in parentheses.

STREAM NAME	NODE TYPE	(KM) STREAM DISTANCE	(R) CORR. COEF.	(C) MEAN ERROR	(+-C) PROB. ERROR	(C) MAX. ERROR	NO. TERMS	REMARKS
MS Delaware R	V	66.5	0.9043	0.50	0.99	-4.67	292	WB DEC Record at Men's Club (465204E/4654777N)
MS Delaware R	V	60.1	0.8749	0.87	1.15	-5.45	427	WB Hale Eddy Gage (01426500)
MS Delaware R	V	57.8	0.8633	1.05	1.09	-5.39	292	WB DEC Record at Roods Creek (470129E/4649820N)
MS Delaware R	V	52.4	0.8125	1.49	1.29	6.00	292	WB DEC Record at Balls Eddy (472525E/4646402N)
MS Delaware R	V	47.0	0.8357	1.95	1.34	6.16	456	WB Hancock Gage (01427000)
MS Delaware R	V	45.6	0.8353	0.51	1.15	4.99	347	WB DEC Record at Shehawken Creek (476332E/46433
EB Delaware R	V	81.3	0.8985	2.37	1.15	6.93	339	DEC Record at Terry's Campground (494290E/46509
EB Delaware R	V	78.6	0.8453	3.47	1.57	8.78	339	DEC Record at Deutch's Flats (491976E/4652531N)
EB Delaware R	V	75.7	0.8915	2.34	1.56	7.96	456	Harvard Gage (01417500)
EB Delaware R	V	75.6	0.8673	2.78	1.50	8.70	319	DEC Record at Harvard Gage (490073E/4652531N)
EB Delaware R	V	46.5	0.9018	0.28	1.33	-6.93	337	DEC Record at Hancock Fireman's Park (476601E/4
MS Delaware R	V	40.7	0.8093	0.04	1.26	4.79	292	DEC Record at Leonard's (477667E/4638781N)
MS Delaware R	V	28.3	0.7911	-0.16	1.29	-6.55	292	DEC Record at Abe Lord Creek (483879E/4635067N)
MS Delaware R	V	19.5	0.7428	1.51	1.53	-7.31	316	DEC Record at Long Eddy (489001/4632656)
MS Delaware R	V	15.4	0.7921	0.55	1.30	-6.94	225	DEC Record at Kellams (490593/4630211)
MS Delaware R	V	15.4	0.8578	-0.03	1.33	-6.84	452	Hankins Gage (01427301)
MS Delaware R	V	12.4	0.7124	0.71	1.47	6.32	200	DEC Record at Hankins (492890/4628883)
MS Delaware R	V	0.1	0.8845	0.55	1.40	-7.12	441	Callicoon Gage (01427510)
SUMMARY: VALIDATION TYPE NODES			0.8605	1.19	1.50	8.78	6114	ALL VALIDATION & CALIBRATION NODES

Table A 6. Summary of verification location annual maximum daily goodness-of-fit statistics for the Neversink River for initial model run. Stream distance is measured from the downstream terminus of the study area. USGS gage numbers or DEC UTM coordinates are given in parentheses.

STREAM NAME	NODE TYPE	(KM) STREAM DISTANCE	(R) CORR. COEF.	(C) MEAN ERROR	(+-C) PROB. ERROR	(C) MAX. ERROR	NO. TERMS	REMARKS
Neversink R	V	17.0	0.6918	1.64	1.07	5.33	313	DEC Record at Woodbourne (533455/4622831)
Neversink R	V	13.9	0.7325	2.82	1.39	7.44	313	DEC Record at Fallsburg (532927/4620177)
Neversink R	V	5.5	0.7894	1.15	1.50	-6.56	313	DEC Record near Thompsonville/Ranch Road (53334
Neversink R	V	0.0	0.8870	-0.29	1.43	-6.56	456	Bridgeville (01436690)
SUMMARY: VALIDATION TYPE NODES			0.7746	1.16	1.58	7.44	1395	ALL VALIDATION & CALIBRATION NODE