UPSTREAM TECH
Powering smart conservation on a changing planet

Modeling the Relationship Between Forest Cover, Temperature, and Water Quality in Paulins Kill River Watershed

The Nature Conservancy New Jersey
June 15, 2020

Alden Keefe Sampson
Technical Director
alden@upstream.tech

Kevin Altamirano
Machine Learning Engineer
kevin@upstream.tech

About Upstream Tech

Upstream Tech is a technology company that builds decision-making tools for environmental conservation to improve natural resource management. By harnessing technological advancements in remote sensing, computer science, and machine learning, we enable efficient analysis and monitoring for conservation organizations focused on water management, wetlands, agriculture, and more.

Executive Summary

The purpose of this project is to understand the relationships between forest cover, rising air temperatures, and implications for water quality in the Paulins Kill watershed of New Jersey. Forest cover change impacts a range of ecological processes and species habitat needs, and this project aims to better understand how a range of possible future conditions would influence water quality. In other words, what level of forest restoration could mitigate the risks to fish and other species associated with a warmer climate? These insights would enable The Nature Conservancy New Jersey (TNC NJ) to plan restoration efforts to ensure that water quality remains within safe ranges for fish species. Outcomes of this work are also intended to enhance existing information in TNC NJ's Floodplain Investment Tool to inform restoration investment decisions.

The water quality metrics evaluated in this study include stream temperature and dissolved oxygen (DO), which must be kept within specific ranges to support river ecosystems. Our analysis included three sites across the Paulins Kill watershed in New Jersey. At each site, we evaluated how eight different forest cover and air temperature scenarios would influence water quality. Specifically, we quantified the frequency of water quality standards being out of compliance and the change in extreme values which pose risks to fish and other members of the riparian ecosystem.

Key Findings

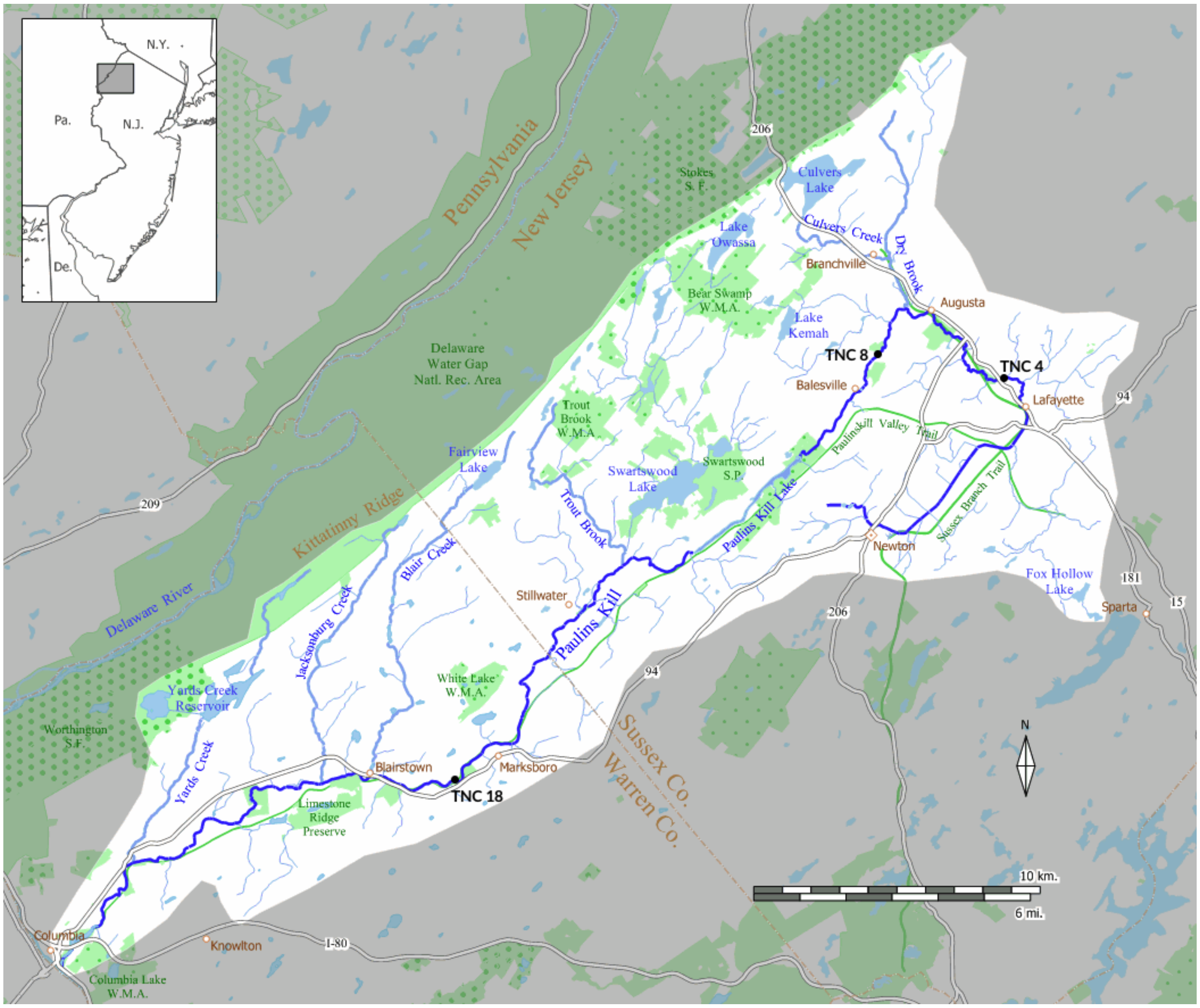
- This study confirms that **forest restoration has the potential to offset the water quality impacts of moderate warming air temperatures**.
- Air temperature is a strong driver of stream temperature and **when air temperatures rise significantly, forest restoration efforts must be substantial to mitigate water quality risks**.
- For all of the eight scenarios evaluated, **dissolved oxygen levels are expected to remain safe** for trout and non-trout based on current water quality standards.
- **Stream temperature metrics are expected to experience greater volatility** and likely be out-of-compliance more often compared to the status quo under several of the future scenarios studied.

- **Scenarios with significant air temperature increase are correlated to higher frequencies of out-of-compliance stream temperatures** based on standards for ideal fish habitat conditions.
- **Machine learning models can learn the complex relationships between environmental conditions and water quality** and aid our understanding of future water quality scenarios.

Approach

Site Selection

We selected three sites in the Paulins Kill watershed for analysis. Gauge sites TNC 4, TNC 8, and TNC 18, were selected to represent how impacts may differ in various locations throughout the watershed.



Map showing the location of the Paulins Kill watershed in the northwest part of New Jersey. Original work by Jim Irwin, November 2006, edited to show study site gauge locations.

Metrics Evaluated

Upstream Tech and TNC NJ developed two metrics to quantify water quality impacts. These metrics provide insight into the risks associated with possible future conditions and the mitigation effects of restoration actions to protect water quality.

- The **frequency** water quality standards (defined below) were not met between April and October, which coincides with the fish spawning and migration period in the Paulins Kill watershed.
- The **magnitude** of water quality extremes during the most stressed time periods in Paulins Kill. These were defined as the The 95th percentile stream temperature and the 5th percentile DO from July to September. They are calculated by taking all of the daily average values during July through September, sorting them from lowest to highest, and selecting the values at the 95th percentile and 5th percentile to illustrate conditions at the extremes.

Water Quality Standards

Standards for water quality classification were taken from the New Jersey Department of Environmental Protection (NJDEP) N.J.A.C. 7:9B-Surface Water Quality Standards released in 2016. For the purposes of this study, the following freshwater standards were evaluated:

- **Stream temperature for trout production** (FW2-TP): Not exceed a daily maximum of 22 degrees Celsius or rolling seven-day average of the daily maximum of 19 degrees Celsius, unless due to natural conditions
- **Dissolved oxygen for trout production** (FW2-TP): Not less than 7.0 mg/L at any time
- **Stream temperature for trout maintenance** (FW2-TM): Not exceed a daily maximum of 25 degrees Celsius or rolling seven-day average of the daily maximum of 23 degrees Celsius, unless due to natural conditions
- **Dissolved oxygen for trout maintenance** (FW2-TM): 24 hour average not less than 6.0, but not less than 5.0 mg/L at any time
- **Stream temperature for non-trout** (FW2-NT): Not exceed a daily maximum of 31 degrees Celsius or rolling seven-day average of the daily maximum of 28 degrees Celsius, unless due to natural conditions
- **Dissolved oxygen for non-trout** (FW2-NT): 24 hour average not less than 5.0, but not less than 4.0 mg/L at any time

Scenarios

To forecast how changes in forest cover and air temperatures would impact water quality variables, including water temperature and dissolved oxygen, we analyzed each of the three project sites using a range of possible future scenarios.

Eight scenarios were evaluated for each of the three drainage areas to understand how these potential future conditions would impact water quality and habitat suitability. Our baseline came from satellite derived land cover maps from NASA's MODIS satellite at each of the selected study sites. We then altered the percentage of forest land cover to simulate a range of future scenarios with differing forest cover percentage levels. Broadly, these scenarios were variations on the themes of forest restoration and increased air temperature. Temperature changes were informed by projections from the Intergovernmental Panel on Climate Change (IPCC)[1], and forest cover changes were based on input from TNC NJ and research conducted by their collaborators on land use change modeling in the Paulins Kill. Air temperature increases were added to the status quo inputs uniformly at each day. Forest cover percent increases were applied to the forest land cover classes already present in the basin, and all other land cover classes were reduced proportionately.

The table below provides more detail on the variables adjusted in all eight of the scenarios. These scenarios were evaluated for each of the three study sites by adjusting input data for the modeling to represent these hypothetical future conditions.

Scenario	Scenario Description	Details
1	Status Quo	No change from current conditions
2	Additional Forest Loss	5% decrease in forest cover
3	Moderate Forest Restoration	2% increase in forest cover
4	Moderate Forest Restoration Moderate Temperature Increase	2% increase in forest cover 0.4° C increase in air temperature
5	Significant Forest Restoration	5% increase in forest cover
6	Significant Forest Restoration Small Temperature Increase	5% increase in forest cover 0.1° C increase in air temperature
7	Significant Forest Restoration Moderate Temperature Increase	5% increase in forest cover 0.4° C increase in air temperature
8	Significant Forest Restoration Significant Temperature Increase	5% increase in forest cover 0.8° C increase in air temperature

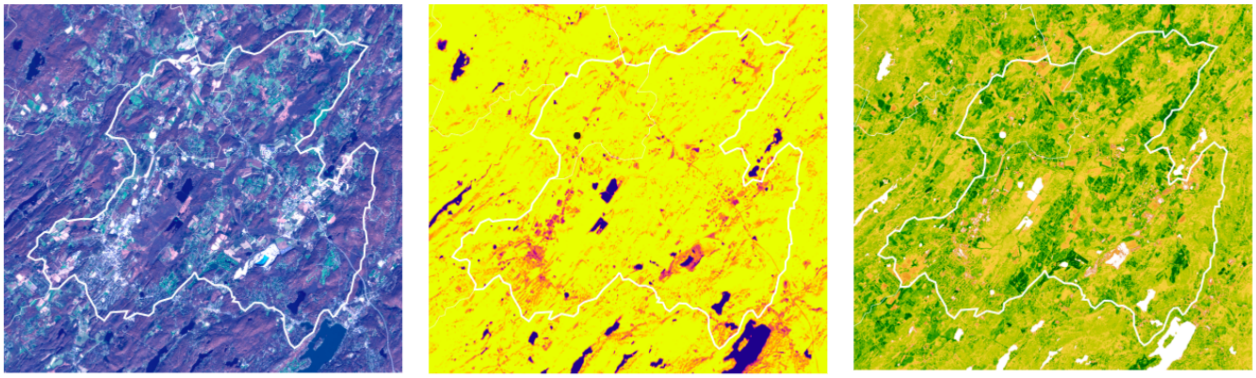
---

[1] Rogelj, Joeri, Malte Meinshausen, and Reto Knutti. "Global warming under old and new scenarios using IPCC climate sensitivity range estimates." *Nature climate change* 2.4 (2012): 248-253.

Modeling Methodology

Input Data

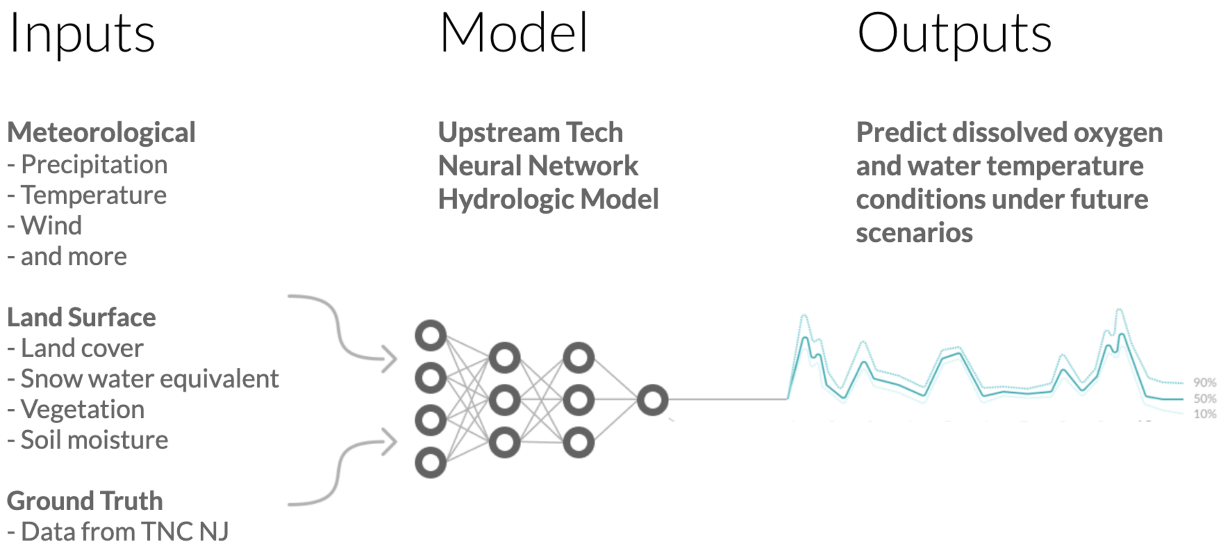
Upstream Tech's hydrologic models use satellite data inputs, combined with weather data, which provide the model with an up-to-date description of hydrologically relevant conditions in the basin, such as vegetation vigor or winter snow cover. The images of the upper region of the Paulins Kill below illustrate the different types of data captured by satellites. Each of the images is produced from the same satellite capture, but are created from observations at different frequencies in the electromagnetic spectrum to highlight different qualities of the land surface. These examples came from European Space Agency's Sentinel-2 constellation, which captures data weekly at a high spatial resolution to provide information about ground conditions with a high level of detail. Our hydrology models use data from a similar satellite, NASA's MODIS satellite, which captures at daily frequency and a lower spatial resolution.



Satellite imagery showing (from left to right): A color image to show ground conditions in the visual spectrum, a visualization of Normalized Difference Water Index to highlight surface water in dark blue, and a visualization of Normalized Difference Vegetation Index showing high vegetation photosynthetic activity in green.

Specifically, our model uses satellite based observations of land cover classification, snow cover, vegetation vigor and day and night land surface temperature. These observations are primarily derived from satellites operated by NASA. In addition to the satellite data inputs, our model also incorporates weather information from National Oceanic and Atmospheric Administration (NOAA)'s North American Land Data Assimilation System (NLDAS) weather model. The vegetation vigor values are calculated from raw satellite reflectance data using the Normalized Difference Vegetation Index, a unitless metric where higher values indicate greater leaf cover and vegetation photosynthetic activity. The snow cover values are calculated using the Normalized Difference Snow Index which reports the fraction of each pixel which is covered in snow. Land cover classifications come from NASA's MODIS satellite based annual land cover data product.

Observed water quality data were accessed from private TNC NJ and United States Geological Survey (USGS) gauge data, which was then aligned with relevant inputs at the same point in time to create a dataset of historic input data paired with ground observations. This enabled the project team to create a training dataset which our machine learning model could then use to detect patterns and relationships. There were some gauge observations that included unusually low measurements likely due to sensor error. We excluded these erroneous values for model training.



Graphic showing the modeling process and components for this study

Machine Learning Model

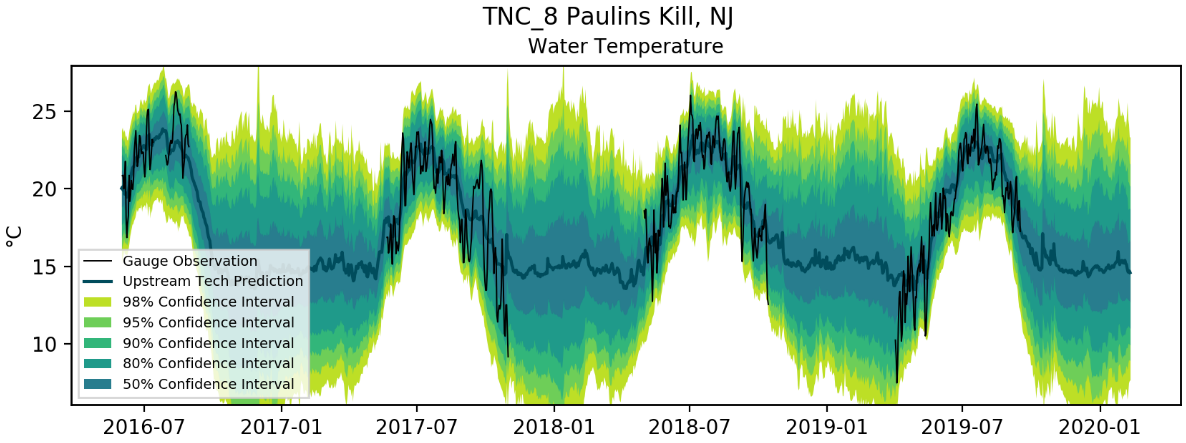
The model used in this analysis is built with a type of neural network building block called Long Short-Term Memory (LSTM). The model works with time-based sequences of data. At each time step in the sequence, the LSTM takes in new inputs, updates a set of internal states it maintains which represent the hydrologic basin conditions, and then finally outputs a prediction.

The model is trained by providing historic training samples (each sample is a time series of inputs paired with gauge measurements) and iteratively updating the strength of the internal connections in the neural network to more accurately predict the desired outputs. Once the model has been trained, it can be applied to new scenarios by providing inputs modified to reflect that scenario. We evaluate the model's predictions on data which was held out of the model training process to understand its accuracy in new situations.

The model training process was conducted in two phases. First, the model was trained on basins from locations throughout the continental United States to learn the general relationships between our inputs and water quality outputs. Second, the model was tuned with gauge data from

TNC NJ private temperature and DO gauges in the Paulins Kill watershed in order to learn the specific nuances of the watershed.

After training was complete, we tested our model using hindcasts, or backtests, to confirm that it was able to recognize patterns and predict temperature and DO effectively. The figure below shows an example hindcast from TNC site 8. TNC NJ's gauge data is drawn as a black line. The dark blue line shows the average value of the model's predicted distribution which is the model's best guess. The bands of color around the dark blue line show the predicted confidence intervals - narrower bands mean the model is more confident at that time.



Example hindcast plot showing a water temperature predictions for TNC site 8

Once the models were trained and tested, we used them to forecast future water quality impacts for the eight scenarios described above at each of the three sites. The input data on percentage forest cover in the area of interest and air temperature were modified and then provided to the model to generate water quality predictions for each scenario. Results were then assessed based on the frequency and magnitude metrics defined above.

Geographic Distribution of Reforestation Actions

It is important to note that the reforestation scenarios in this study were based on forest cover percentages across the full region which drains to each gauge site. This is due to the structure of our current model which models each drainage area as a single "lumped" unit. However, there is scientific research which highlights that targeted reforestation in specific parts of a watershed can yield beneficial water quality outcomes.[2] Deforestation of reforestation of riparian zones in

---

[2] Bond, Rosealea M., Andrew P. Stubblefield, and Robert W. Van Kirk. "Sensitivity of summer stream temperatures to climate variability and riparian reforestation strategies." *Journal of Hydrology: Regional Studies* 4 (2015): 267-279.

particular can influence in-stream conditions significantly.[3][4] Furthermore, studies have confirmed that targeted riparian restoration can indeed offset climate change impacts on fish species.[5] Therefore, it is likely that reforestation efforts targeted in riparian zones would have more significant positive water quality impacts than the basin-wide forest increase scenarios presented in this report.

Results

Stream Temperature Under Evaluated Scenarios

Based on the scenarios evaluated, stream temperature water quality standards for trout maintenance and production showed the most variation in the frequency with which standards were not met under different scenarios. The non-trout stream temperature, on the other hand, was in compliance for all future conditions analyzed at all sites.

The table below shows how these trout maintenance and production standards were impacted by future scenarios, based on the average number of times these metrics were out of compliance between April and October compared to the status quo (where current basin conditions are unchanged). Red indicates increased frequency of the metric being out of compliance, and green indicates improvement in the frequency the river was out of compliance based on water quality standards.

In summary, the results in the table below demonstrate the following.

- Forest restoration has positive effects on stream temperature under all scenarios evaluated.
- Air temperature changes are a driving factor in water temperature.
- If applied evenly across the basin, significant forest restoration is required to offset a small temperature increase. This suggests that targeting forest restoration locations is critical.
- Forest restoration moderates the negative effect of air temperature increases even if it cannot fully counteract them.
- The impacts of different scenarios on the frequency of water quality standards being met is largest at site 18 because the status quo at that site crosses water quality standard thresholds more often than at other sites. This illustrates the importance of both analyzing the frequency (this section) and the magnitude (following section) of water quality impacts.

---

[3] Jones III, EB Dale, et al. "Effects of riparian forest removal on fish assemblages in southern Appalachian streams." *Conservation biology* 13.6 (1999): 1454-1465.

[4] Newton, Michael, and Elizabeth C. Cole. *Linkage between riparian buffer features and regeneration, benthic communities, and water temperature in headwater streams, western Oregon*. General Tech. Rep. PNW-GTR-642. PNW Research Station, USDA Forest Service, Portland, OR, 2005.

[5] Justice, Casey, et al. "Can stream and riparian restoration offset climate change impacts to salmon populations?." *Journal of environmental management* 188 (2017): 212-227.

Site	Scenario Description	Stream temperature for trout maintenance	Stream temperature for trout production
TNC 18	Status Quo	0 (23.0)	0 (108.3)
	Additional Forest Loss	+1.6 (24.6)	+0.3 (108.6)
	Moderate Forest Restoration	-0.6 (22.3)	-1.0 (107.3)
	Moderate Forest Restoration and Moderate Temperature Increase	+3.6 (26.6)	+2.3 (110.6)
	Significant Forest Restoration	-1.6 (21.3)	-2 (106.3)
	Significant Forest Restoration and Small Temperature Increase	-0.6 (22.3)	-0.6 (107.6)
	Significant Forest Restoration and Moderate Temperature Increase	+2 (25.0)	+1.6 (110.0)
	Significant Forest Restoration and Significant Temperature Increase	+8.3 (31.3)	+4.6 (113.0)
TNC 4	Status Quo	0 (1.6)	0 (86.3)
	Additional Forest Loss	0 (1.6)	+0.3 (86.6)
	Moderate Forest Restoration	0 (1.6)	-0.6 (86.6)
	Moderate Forest Restoration and Moderate Temperature Increase	+0.6 (2.3)	+3 (89.3)
	Significant Forest Restoration	0 (1.6)	-0.6 (85.6)
	Significant Forest Restoration and Small Temperature Increase	0 (1.6)	+0.3 (86.6)
	Significant Forest Restoration and Moderate Temperature Increase	+0.6 (2.3)	+3 (89.3)
	Significant Forest Restoration and Significant Temperature Increase	+3.0 (4.6)	+5.6 (92.0)
TNC 8	Status Quo	0 (15.3)	0 (99.3)
	Additional Forest Loss	+1 (16.3)	+0.3 (99.6)
	Moderate Forest Restoration	0 (15.3)	0 (99.3)
	Moderate Forest Restoration and Moderate Temperature Increase	+2.6 (18.0)	+3.0 (102.3)

	Significant Forest Restoration	-0.3 (15.0)	-0.6 (98.6)
	Significant Forest Restoration and Small Temperature Increase	0 (15.3)	0 (99.3)
	Significant Forest Restoration and Moderate Temperature Increase	+2.3 (17.6)	+2 (101.3)
	Significant Forest Restoration and Significant Temperature Increase	+5.6 (21.0)	+6 (105.3)

Change in frequency and average annual frequency (in parenthesis) that water temperature standards were not met. This table shows how often the stream temperature fell within the acceptable range for trout maintenance and production, with red boxes indicating an increase in the number of times water quality standards were not met, green boxes showing improvement in the number of times stream temperature standards were within safe habitat levels, and yellow indicating no change.

Dissolved Oxygen Under Evaluated Scenarios

Unlike stream temperature standards, dissolved oxygen standards were in compliance under the status quo scenarios and remained within safe levels regardless of the scenario evaluated for trout maintenance, trout production, and non-trout standards. There is variation in DO impacts (both positive and negative, detailed below) under different scenarios, but those changes were not large enough to move the river out of compliance for any of the DO standards evaluated.

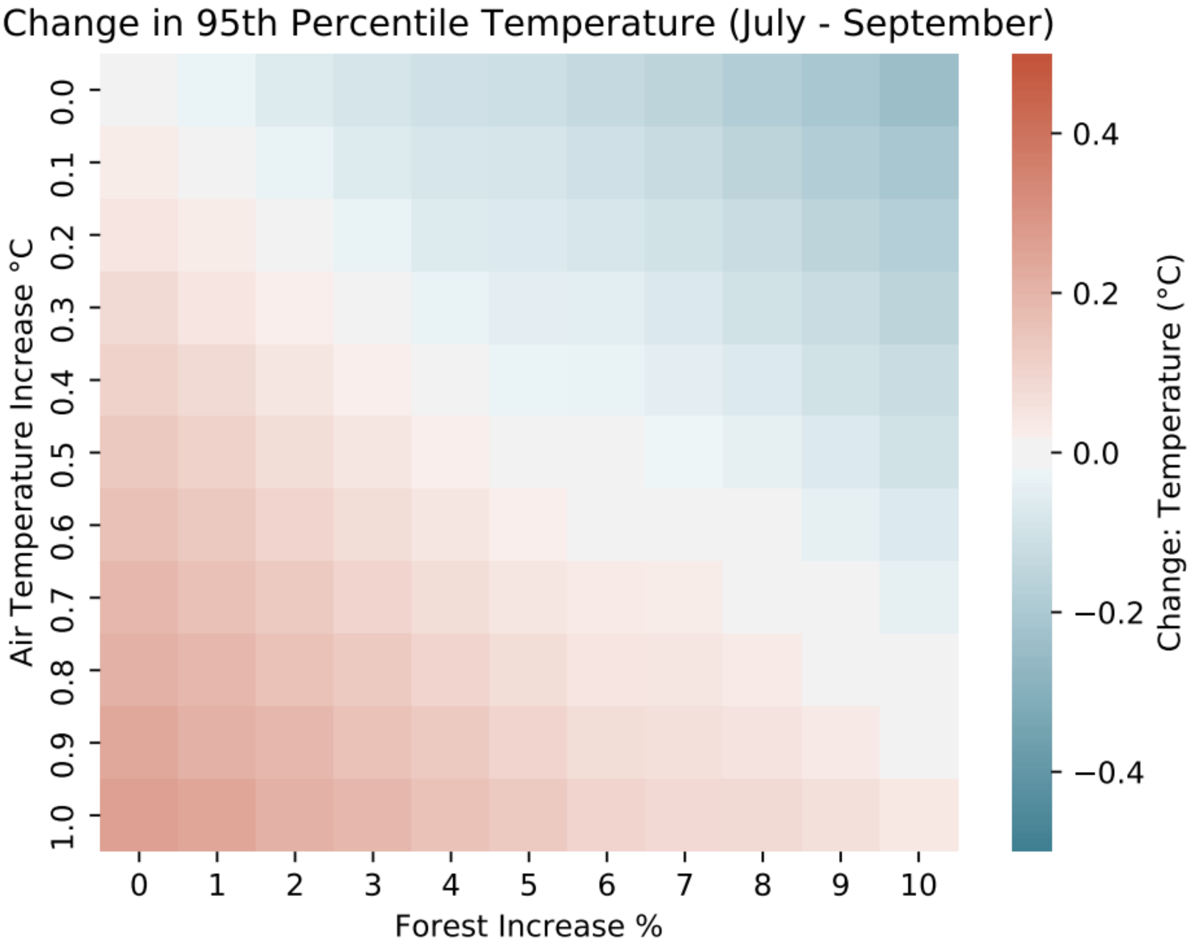
Magnitude of Water Quality Change

The second metric analyzed the magnitude of the most extreme water quality values encountered under each scenario. In the Paulins Kill, July to September is the time period with largest water quality stresses. For this reason, we wanted to further explore the nuances of conditions within this time frame across a myriad of possible air temperature and forest cover scenarios. Specifically, we calculated the 95th percentile (nearly the hottest) water temperature and the 5th percentile (nearly the lowest) DO encountered under each scenario.

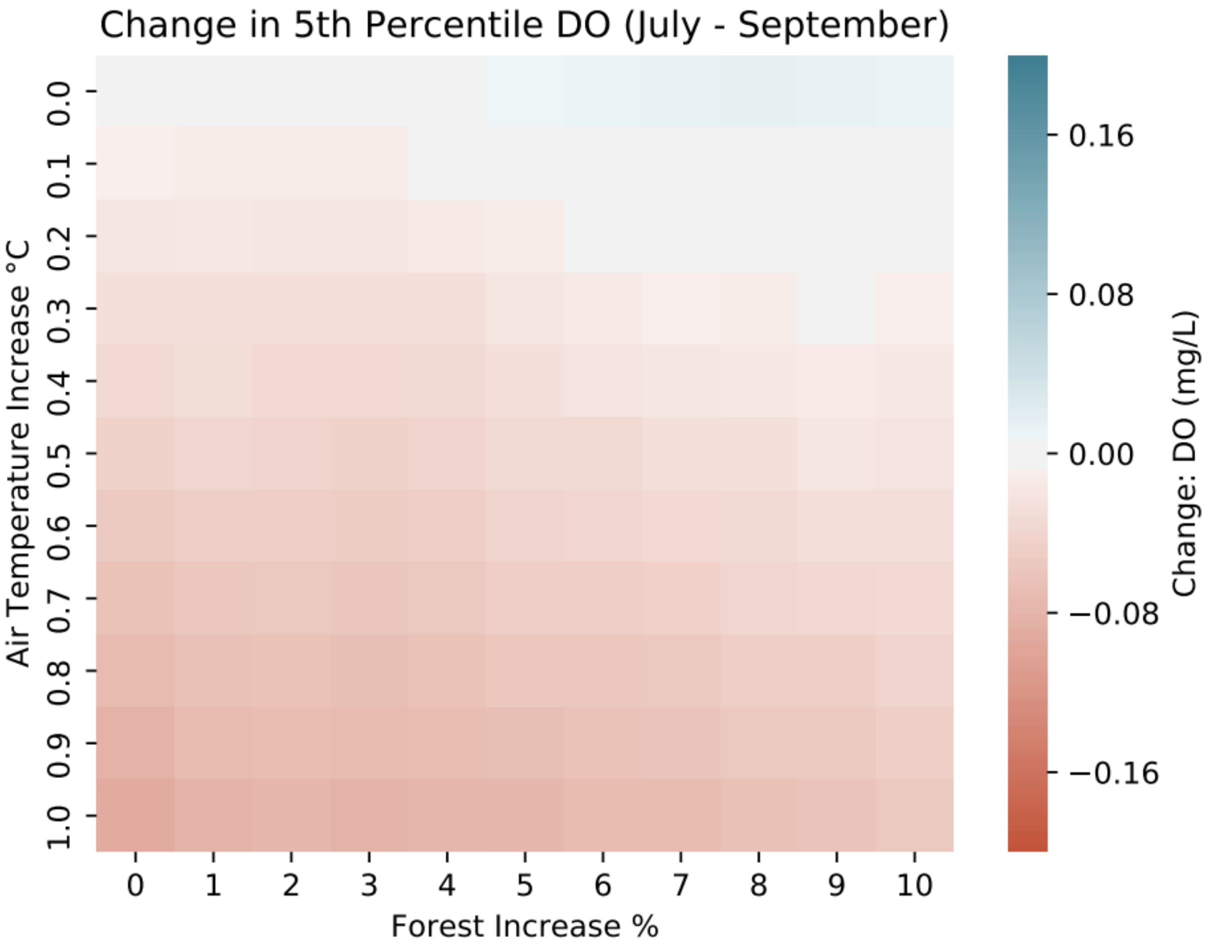
The results are similar, while not identical, across sites so in the interest of space we discuss the results across sites and visualize the plots for TNC site 4 in this section. The plots for sites TNC 8 and TNC 18 can be found in the appendix.

To perform this analysis we developed a gridded visualization of how air temperature and forest temperature related to changes in water temperature and dissolved oxygen levels. This grid includes a total of 100 scenarios. Each of the eight scenarios we analyzed in the table above is represented by one square on the grid, in addition to many additional scenarios to better visualize how air temperature and forest cover interact to impact water quality.

The position of each square defines the amount of air temperature or forest cover increase in that scenario. The color of the square represents the change in water quality under that scenario as compared with the status quo. The status quo is represented by the square in the upper left corner.



Visualization for TNC site 4 showing how air temperature and forest increases influence stream temperature on some of the hottest days of the year. The color of each box in the grid represents the change in stream temperature on hot days against the status quo scenario. At this site, reforestation is effective at offsetting air temperature increases.



Visualization for TNC site 4 showing how air temperature and forest increases influence dissolved oxygen levels in the water. The color of each box in the grid represents the change in dissolved oxygen compared with the status quo. Though all scenarios keep DO within safe levels for trout and non-trout species, this graphic illustrates how reforestation can help regulate DO levels.

In both of the grids above we see how an increase in forest cover across the basin improves both DO and water temperature, while an increase in air temperature has negative effects on water quality. The impact of forest restoration at the TNC 4 site is not as strong of an improvement on DO as it is on water temperature, however DO levels continue to meet water quality standards under all 100 scenarios considered in the grid above. Water temperatures are both the standard that is most stressed and, fortunately, the most impacted by forest cover increases.

Compared to the other two sites studied, TNC 4 stands out as the most responsive to restoration in terms of water temperature. This is likely because the drainage basin for this site has, in percentage terms, the largest amount of development and correspondingly relatively lower forest cover, causing an increase in forest percent to have a larger impact. At other sites the amount of forest or other natural vegetation cover is already relatively higher and we see less improvement from increased forest cover. We again emphasize that these increases in forest cover in our scenarios are spread evenly across the basin and targeted forest restoration in riparian areas is likely to have more pronounced improvements. Still it is encouraging to see the potential for forest restoration to mitigate the worst of potential climate change impacts.

Conclusion

This study elucidates the relationship between basin conditions, climate, and water quality in the Paulins Kill watershed. Overall, we found that dissolved oxygen variations were minimal for all the air temperature and forest cover scenarios evaluated and that dissolved oxygen will remain within compliant levels under these future conditions. Stream temperature, on the other hand, was comparatively more responsive to changing basin and climate scenarios. Increases in forest cover can offset water quality impacts of temperature increases, with regions with larger amounts of development and lower forest cover benefiting more from forest restoration. Overall, increasing the forest cover across a basin improved water water quality outcomes. These results highlight the importance of forest restoration as an adaptation strategy under future warming conditions.

An exciting result of this work exploring scenario impacts on water quality is the demonstration that machine learning models are able to learn the relationships between many environmental conditions and water quality. This study showed how those learned relationships can be investigated to better understand what may happen to our watersheds. This approach can be extended to encompass additional environmental conditions and outcomes.

Though this work highlights insights which can inform decision-making at a high level, these models - like all models - are simplified versions of reality. Particularly with complex natural ecosystems, there are myriad possible interactions or feedback loops which may influence future water quality in unexpected ways. We are continuing to improve our models to capture more detail of the water quality processes within basins to allow us to ask more detailed questions about the drivers of water quality change.

Climate change is posing a host of challenges for our watersheds. As these challenges grow, a scenario analysis approach can help understand the scope of the problem and target effective restoration.

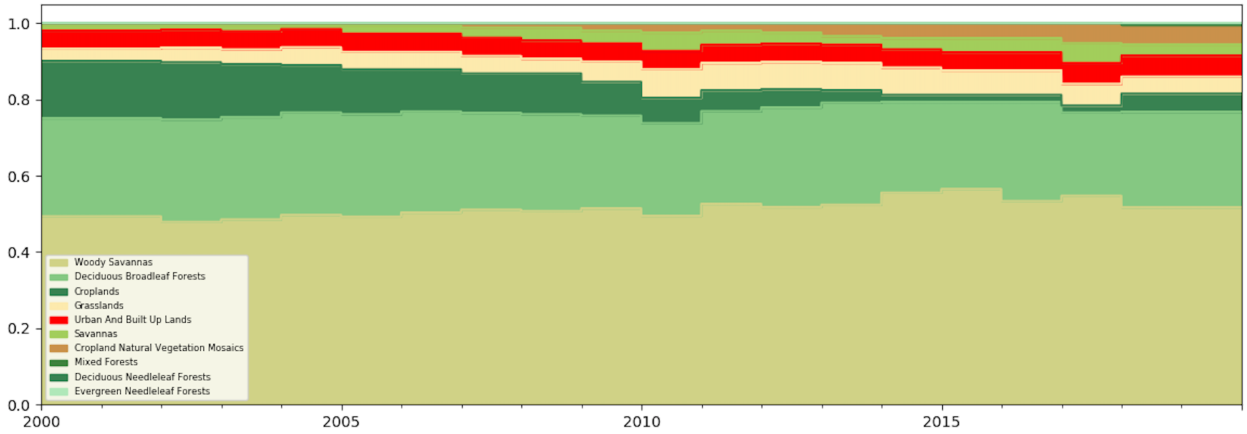
Appendix: Land Cover

This table details the percentages of each drainage basin covered by each land cover category in 2019. TNC 4 has the highest fraction of urban and built up lands, while TNC 18 has the highest fraction of forest cover.

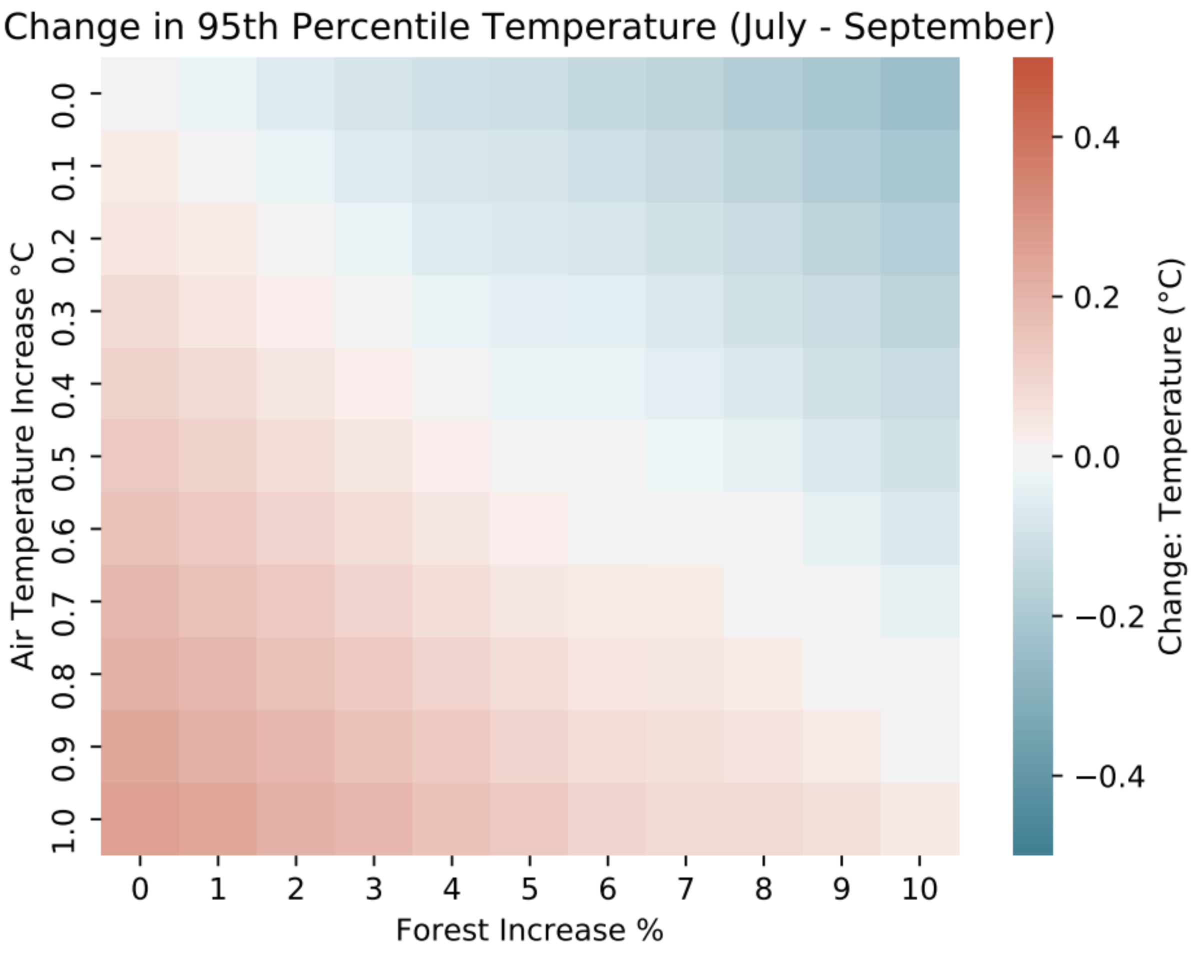
Site	Evergreen Needleleaf Forests	Deciduous Needleleaf Forests	Deciduous Broadleaf Forests	Mixed Forests	Woody Savannas	Savannas	Grasslands	Permanent Wetlands	Croplands	Urban and Built Up Lands	Natural Vegetation Mosaics
TNC 18	0.48%	0.11%	51.10%	0.09%	39.76%	1.95%	0.62%	0.11%	1.55%	1.98%	2.24%
TNC 4	0.13%	0.36%	25.03%	0.01%	51.71%	2.93%	4.45%	0.00%	4.76%	5.51%	5.11%
TNC 8	0.31%	0.03%	32.37%	0.16%	52.15%	3.10%	1.31%	0.16%	2.88%	2.77%	4.78%

Appendix: Individual Site Data

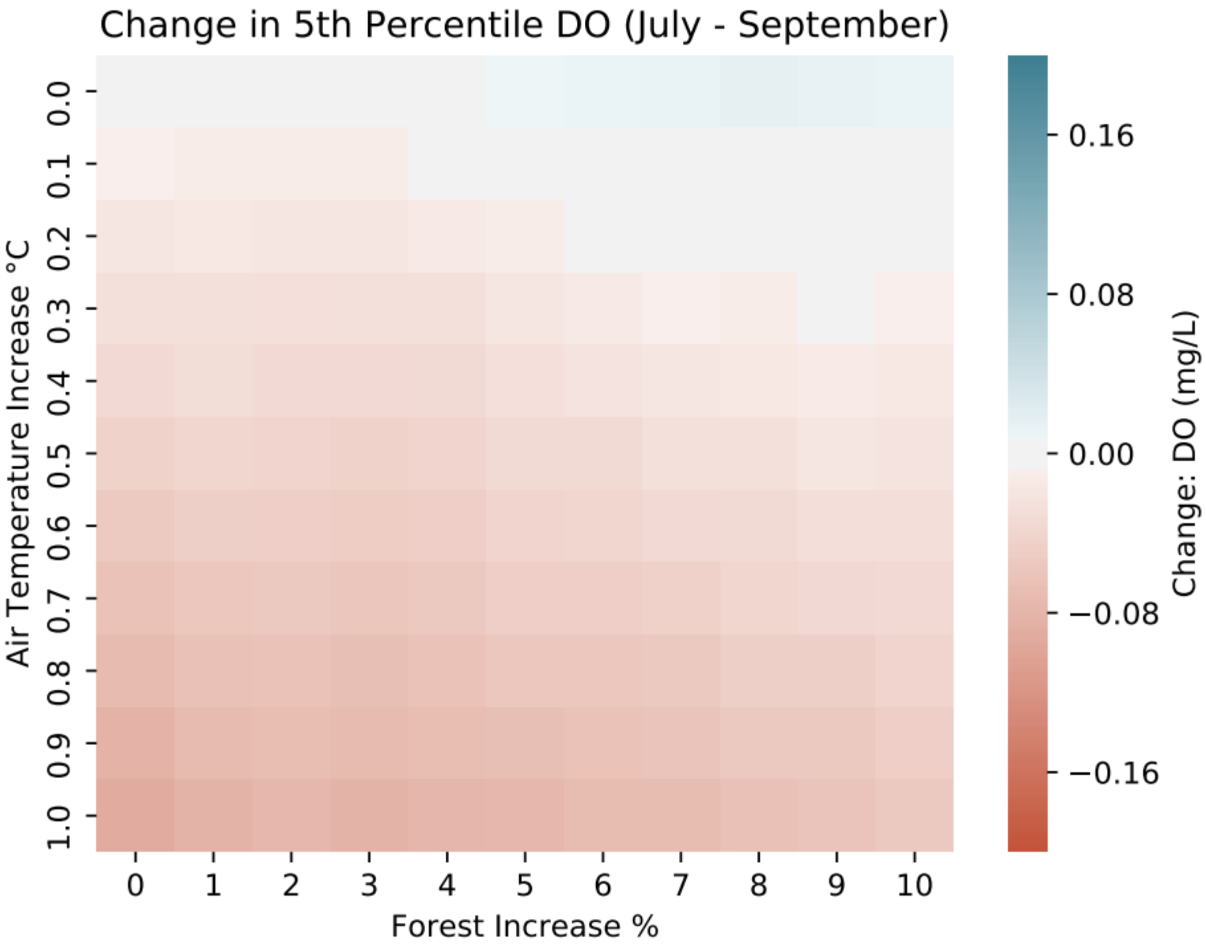
TNC Site 4



Change in land cover types for TNC site 4 over the last two decades, based on MODIS satellite data, with woody savannas as the largest proportion followed by deciduous broadleaf forests.

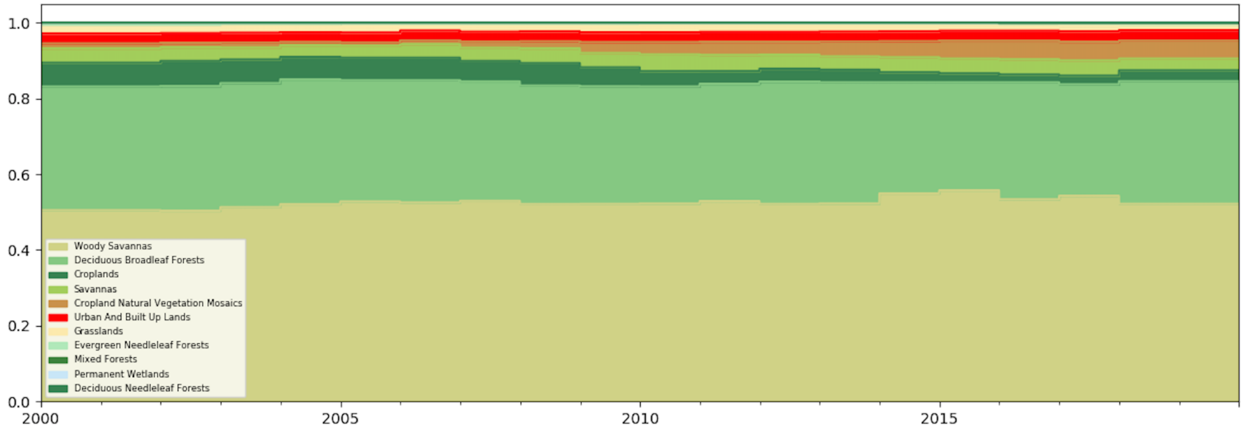


Visualization for TNC site 4 showing how air temperature and forest increases influence stream temperature on some of the hottest days of the year. The color of each box in the grid represents the change in stream temperature on hot days against the status quo scenario. At this site, reforestation is effective at offsetting air temperature increases.

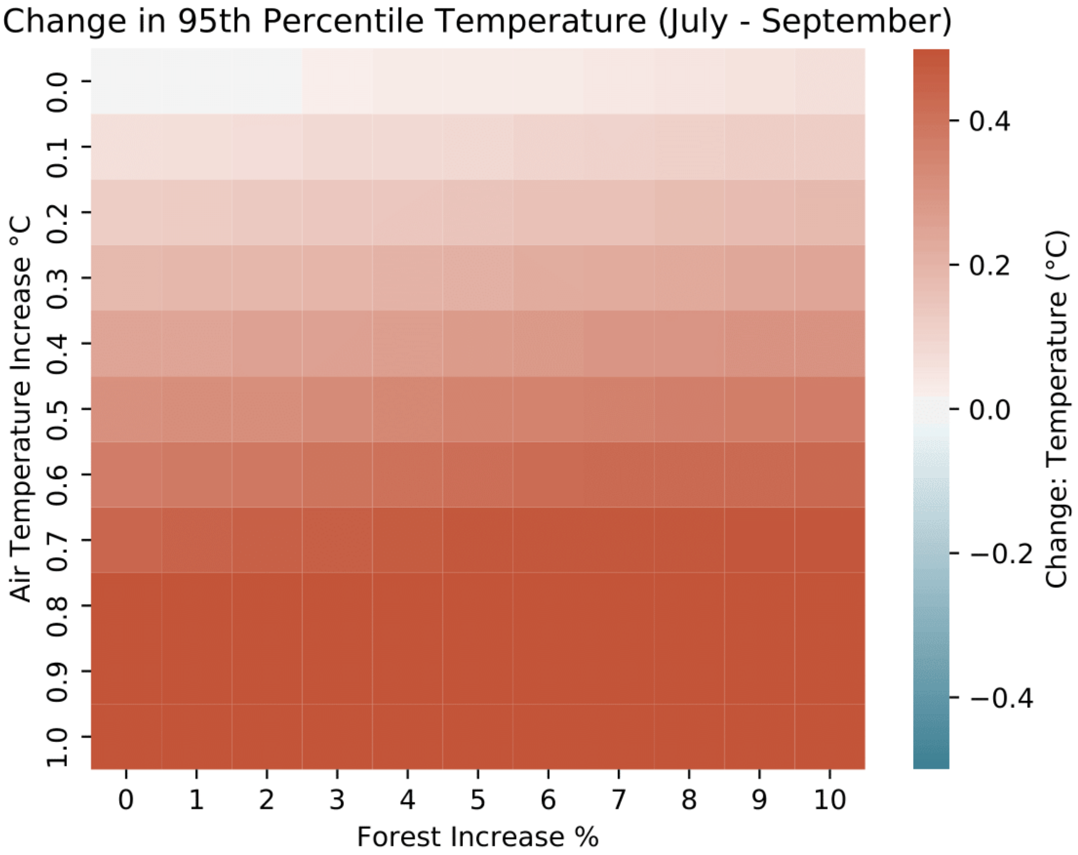


Visualization for TNC site 4 showing how air temperature and forest increases influence dissolved oxygen levels in the water. The color of each box in the grid represents the change in dissolved oxygen compared with the status quo. Though all scenarios keep DO within safe levels for trout and non-trout species, this graphic illustrates how reforestation can help regulate DO levels.

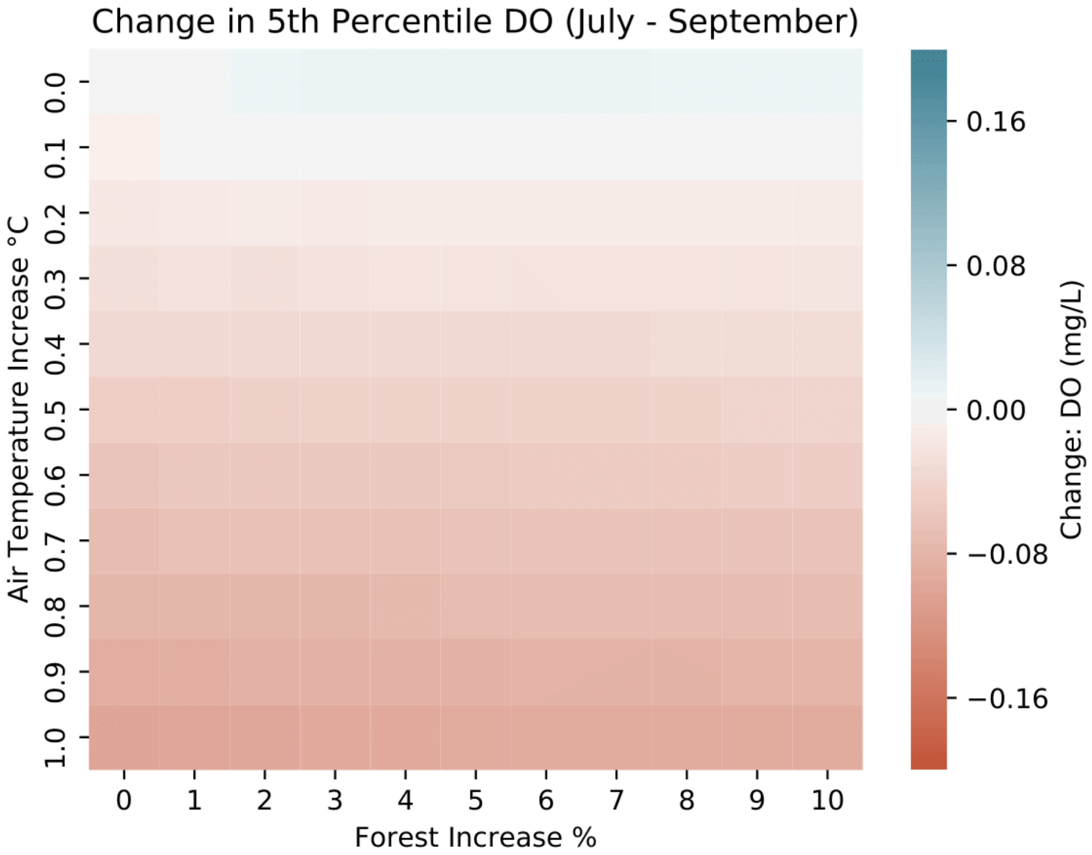
TNC Site 8



Change in land cover types for TNC site 8 over the last two decades, based on MODIS satellite data, with woody savannas (or abandoned farm fields) as the largest proportion followed by deciduous broadleaf forests.

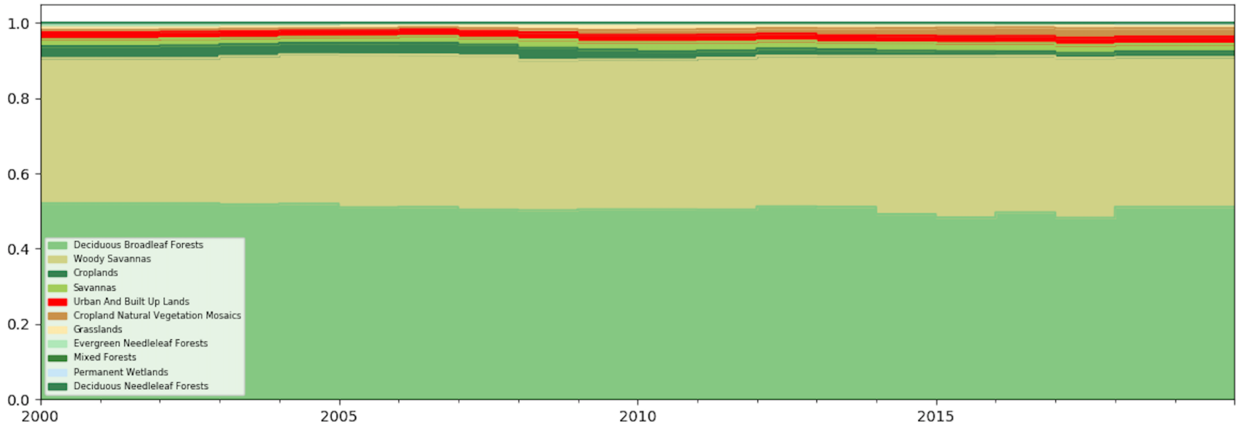


Visualization for TNC site 8 showing how air temperature and forest increases influence stream temperature on some of the hottest days of the year. The color of each box in the grid represents the change in stream temperature on hot days against the status quo scenario. At this site, forest and other natural land cover are already relatively high fractions of the basin and air temperature is the primary determinant of stream temperature.

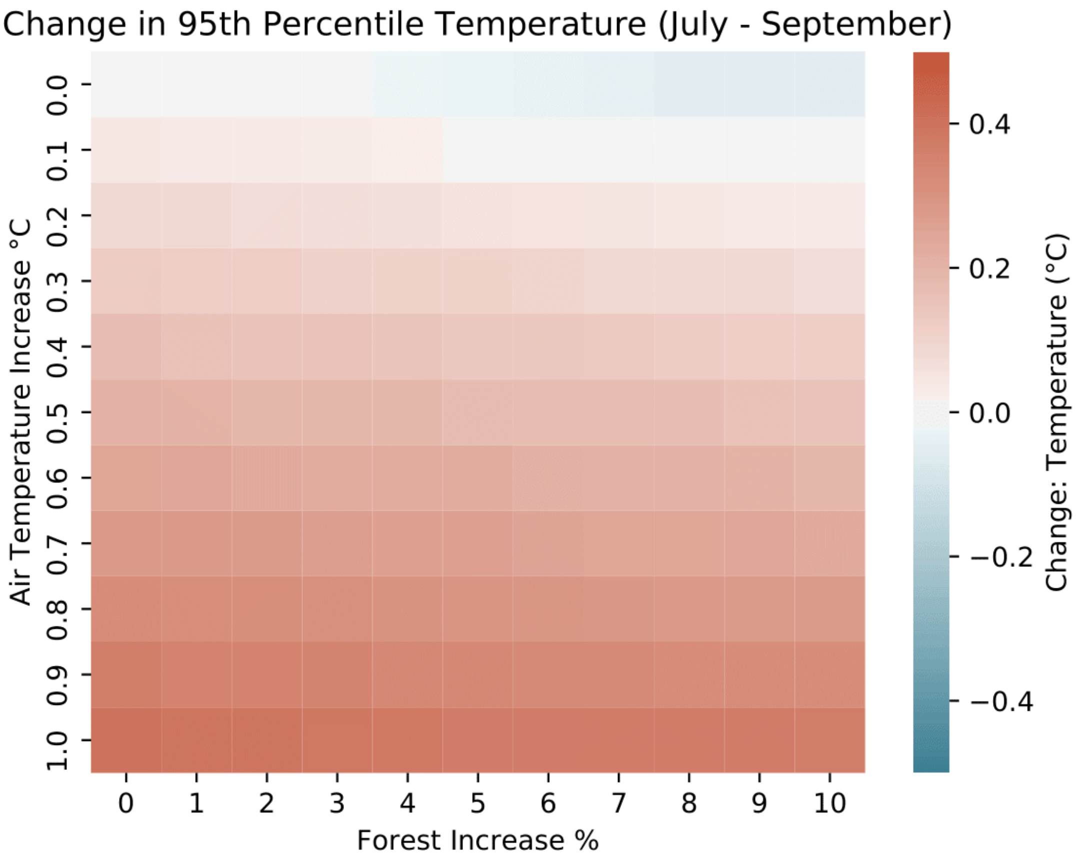


Visualization for TNC site 8 showing how air temperature and forest increases influence dissolved oxygen levels in the water. The color of each box in the grid represents the change in dissolved oxygen compared with the status quo. At this site, forest cover increases have a small but measurable positive effect on DO.

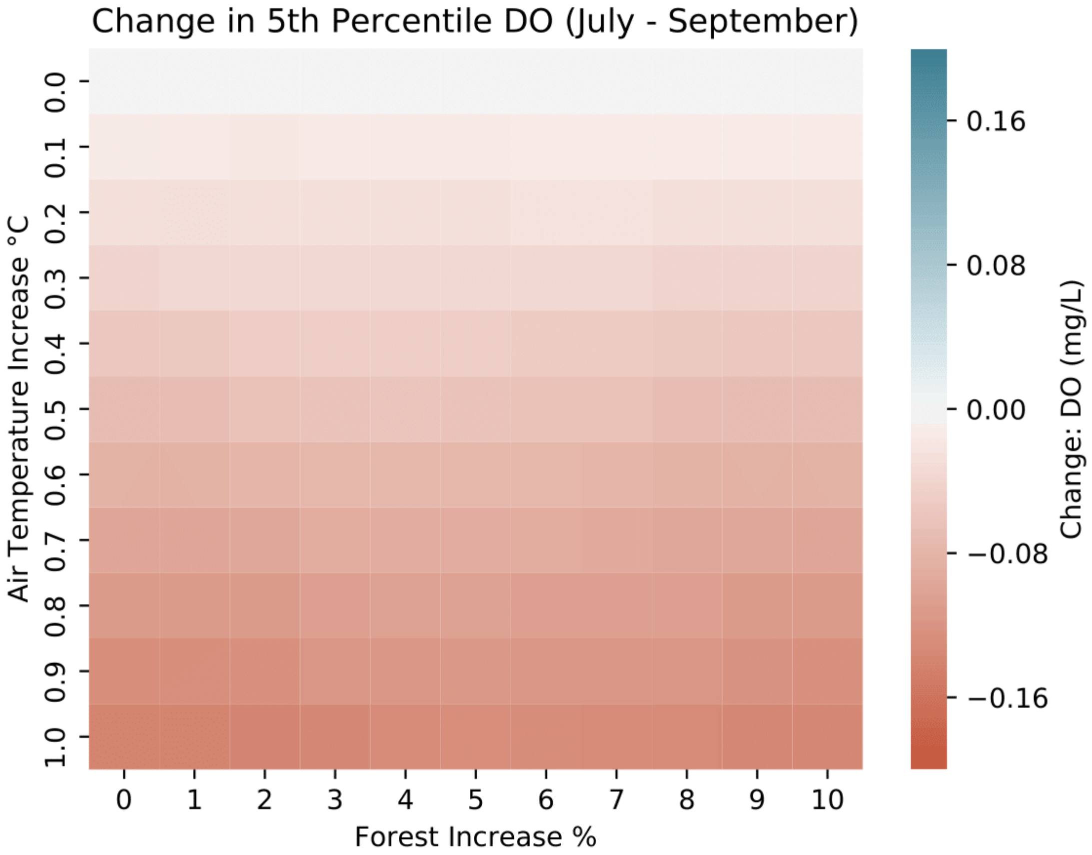
TNC Site 18



Change in land cover types for TNC site 18 over the last two decades, based on MODIS satellite data, with a similar proportion of deciduous broadleaf forests and woody savannas (often former farm fields).



Visualization for TNC site 18 showing how air temperature and forest increases influence stream temperature on some of the hottest days of the year. The color of each box in the grid represents the change in stream temperature on hot days against the status quo scenario. At this site, reforestation has some effect, though this is limited at large temperature increases nearing 1°C.



Visualization for TNC site 18 showing how air temperature and forest increases influence dissolved oxygen levels in the water. The color of each box in the grid represents the change in dissolved oxygen compared with the status quo. At this site, forest cover increases have a limited positive effect on DO.