

FLUX TO FLOW: A CLEARER VIEW OF EARTH'S WATER CYCLE

VIA NEURAL NETWORKS AND SATELLITE DATA

BY

ALBERT ERIC LARSON

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

CIVIL AND ENVIRONMENTAL ENGINEERING

UNIVERSITY OF RHODE ISLAND

2023

DOCTOR OF PHILOSOPHY DISSERTATION
OF
ALBERT ERIC LARSON

APPROVED:

Thesis Committee:

Major Professor Ali Shafqat Akanda
 Soni Pradhanang
 Thomas Boving
 Brenton DeBoef
DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND

2023

ABSTRACT

Earth's weather and climate changes dynamically, and humans must monitor the evolution of our environment or fail to plan and adapt. Water is our most valuable resource, and as such it is of the utmost importance to have a continuous view of its movements as water cycles over, in, and under land, in the atmosphere, and in the sea. One way that we observe the water cycle is with satellite measurements. Unfortunately, the resolution of these instruments is typically too coarse to visualize many transient phenomena of interest. If their vision is fine enough, the instrument captures only a small space of Earth at any given time. Furthermore, it is valuable to connect satellite-derived water cycle measurements of large watershed basins with ground truth observations of the rivers flowing through these basins. These connection operations, though, are time and cost prohibitive or limited in performance, with efficient functions veiled behind black box closed-source solutions. Herein, all of the above facts are investigated under the veil of a single name, Flux to Flow (F2F). The name Flux to Flow encapsulates how the work takes several fluxes of geospatial data and transforms them into more coherent flows of knowledge. F2F performs metrologically well in the operation of streamflow forecasting when focused solely on a single or a few hydrological monitoring nodes at a time. Based on the results, future work might entail scaling up of the system to many computing nodes running in tandem, observing more adjacent outputs, or in its application as a standard device in connected municipal systems.

ACKNOWLEDGMENTS

Thank you to my wife Caitlyn for her steadfast love and for being my best friend.

Thank you to my family for their lifelong encouragement of creative self-expression, and for keeping learning, aesthetics & the pursuit of truth at the center of the conversation.

Thank you to Ali for sharing a bounty of time and energy in an effort to help me. Thank you for holding me accountable, and for persistently moving the bias back to water as the focal point.

Thank you to Soni, Tom, and Jason for your willingness to serve on my committee and for being fountains of wisdom.

Thank you to Joe Goodwill for the lessons in grading and teaching; to Peter Cornillon, Gavino Puggioni, and Marco Alvarez for their help honing the technical tools; to Kevin Bryan and Gaurav Khanna for their support with the research computing resources; and to Farah, Hanan, Cameron, Mohammed, Lukas, Darya, and Wulan for being great lab mates.

Thank you to all the fantastic teachers I've had over the years, and to the greater University of Rhode Island community for being a vibrant and colorful ecosystem.

PREFACE

This thesis follows the manuscript format. Each of the three within carry a similar structure as is typical of the structure one might produce for an academic journal. The three cords form one strand that benefit from being studied together.

Manuscript 1, “Discerning Hydroclimatic Behavior with a Deep Convolutional Residual Regressive Neural Network”, looks at four United States basins and a single streamflow measurement per basin. It also considers just one neural network architecture called Flux to Flow. Upon further consideration, we have realized that the structure itself is more aptly called dcrrnn and pronounced discern. It is called this for two reasons: 1, because of the nature of a trained mind’s ability to learn how to discern the truth from a flurry of information; and 2, because it is an acronym representing the phrase “deep convolutional residual regressive neural network”. The name dcrrnn was accidental and an obfuscation device just to conceal as manuscript 1 was submitted to a double-blind submission. It was undesirable to share the name Flux to Flow because of our ties to the name on the web already. As the work has progressed, dcrrnn is now understood more appropriately as a very specific neural network construction, whereas F2F encompasses the macroscopy of the work. Though there is only a single architecture used in this manuscript, it is both dcrrnn and Flux to Flow. This study considers about seven years of data, considering a daily time scale.

Manuscript 2, “Deep Convolutional Residual Regressive Neural Networks and Sea Surface Temperatures from Aqua and Argo in the 2000s”, focuses in greater detail on a single water-focused essential climate variable (sea surface temperature). The deployed experiments, similar to manuscript 1, features solely the dcrrnn architecture under the name Flux to Flow. The time series studied is relatively short in time, only considering a single year of monthly measurements; however, the size of the geography studied is quite large, considering big pieces of the Atlantic, Pacific, and

Indian oceans.

Manuscript 3, “Holistic Water Cycle Analysis via the Confluence of Climate Model, Satellite, Ground Truth, and Machine Learning Signal Processing Technologies: Two North American Transboundary River Watersheds”, is best understood as the confluence of manuscripts 1 and 2. We fuse measurements of sea surface temperature with measurements of land flow and create images that do not contain nan values, a sometimes frustrating numerical data structure component. We compare the performance of using these fused images against the original technique from manuscript 1 of simply clipping and z-scoring land surface flows and neglecting the ocean. We look at more output targets per moment than manuscript 1, but fewer than manuscript 2. We also use several different neural network constructions to compare the dcrrnn structure against other simpler neural network structures.

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGMENTS	iii
PREFACE	iv
TABLE OF CONTENTS	vi
LIST OF FIGURES	ix
LIST OF ABBREVIATIONS	xi
MANUSCRIPT 1	1
1.1 Abstract	2
1.2 Keywords	2
1.3 Introduction	2
1.4 Materials & Methods	3
1.4.1 Study Areas	3
1.4.2 Satellite Derived Observations	4
1.4.3 Ground Truth Measurements	4
1.4.4 Data Collection and Preprocessing	5
1.4.5 Treatment	5
1.5 Results	6
1.6 Discussion	8
1.7 Conclusion	11
1.8 Code Availability	12
1.9 References	12
1.10 Appendix	17

MANUSCRIPT 2	21
2.1 Abstract	22
2.2 Introduction	22
2.3 Materials & Methods	25
2.3.1 Sea Surface Temperature (SST)	25
2.3.2 Aqua	25
2.3.3 AMSR-E	25
2.3.4 MODIS	27
2.3.5 Ground Truth Measurements	28
2.3.6 Treatment	29
2.4 Results	32
2.5 Discussion	33
2.6 Conclusion	35
2.7 Code Availability	37
2.8 References	37
2.9 Appendix	40
MANUSCRIPT 3	46
3.1 Abstract	47
3.2 Keywords	47
3.3 Introduction	48
3.4 Materials & Methods	49
3.4.1 Yukon River Watershed	49
3.4.2 Columbia River Watershed	51
3.4.3 Data Sources	53
3.4.4 Preprocessing	54
3.4.5 Treatments	56
3.5 Results & Discussion	57

3.6	Conclusion	65
3.7	Code Availability	66
3.8	References	66
3.9	Appendix	72

LIST OF FIGURES

1.1	Drainage basins under investigation	17
1.2	NLDAS daily surface and subsurface flows	17
1.3	Strip chart and histogram plots of z-scored gauged streamflow observations	18
1.4	Neural network sample output	19
1.5	Kernel Density Estimates of the 2,268 experiments. Left shows grand NSE and KGE.	20
1.6	3-D merit plots, basin delineated by color. Intensity of color indicates higher frequency within bin range.	20
2.1	L2 AMSR-E SST field, March 3, 2004, no coordinate system	40
2.2	L2 AMSR-E SST, plotted with available coordinates and world map .	40
2.3	L3 AMSR-E file plotted without supplied coordinate system	41
2.4	L3 AMSR-E file plotted with available coordinates and world map .	41
2.5	L3 daily MODIS file containing only high quality flagged pixels	42
2.6	Monthly L3 MODIS image containing only high quality observations .	42
2.7	Sample training monthly observation; January 2010 MODIS day observation of the Hawaiian Islands; segmented into 100 x 100 pixel regions.	43
2.8	December 2010. The top panel shows the distances from Argo and MODIS to predicted, optimized prediction, AMSR-E and each other for all nine experiments. Two bottom panels are plots of training and validation losses during neural network training.	44
2.9	Relatively “good” perceptual change, Pacific Night case	45
2.10	Relatively “poor” perceptual change, Indian Day case	45
3.1	Extents of the Yukon and Columbia River watersheds and gauge locations.	72

3.2	Hydrographs of all gauged streamflow data. Second and fourth plots are zoomed in versions of the same colors in the first and third plots. Both have low discharges whose details are lost amongst the larger portions of the river.	73
3.3	Sample observation of surface flow (Q_s), subsurface or groundwater flow (Q_{sb}), $Q_s + Q_{sb}$, flows merged with SST but represented with a single colormap, and flows merged with SST represented with two colormaps highlighting each physical parameter's unique dynamic range.	74
3.4	Histograms of model predictability across all experiments delineated by whether SST is included as part of the input or not.	75
3.5	Histograms of test results for Columbia experiments deconstructed by neural network architecture at the time of training.	76
3.6	Histograms of test results for Yukon experiments deconstructed by neural network architecture at the time of training.	76
3.7	Columbia & Yukon experiments using dcrrnn and the 1,000 neuron single hidden layer neural networks, disaggregated by lag.	77
3.8	Columbia & Yukon experiments using dcrrnn and the 1,000 neuron single hidden layer neural networks, disaggregated by z-scored vs. non-z-scored.	78
3.9	National Drought Mitigation Center Weekly Output, March 2, 2023, Drought Monitor Output.	79

LIST OF ABBREVIATIONS

F2F, Flux to Flow

dcrnn, deep convolutional residual regressive neural network

USGS, United States Geological Survey

NASA, National Aeronautics and Space Administration

AMSR-E Advanced Microwave Scanning Radiometer for EOS

NLDAS, National Land Data Assimilation System

GLDAS, Global Land Data Assimilation System

MODIS, Moderate Resolution Imaging Spectroradiometer

EOS, earth observation system

kg/m², kilograms per square meter

TTS, train test split

NSE, Nash Sutcliffe efficiency

KGE, Kling Gupta efficiency

f³s⁻¹, cubic feet per second

A, actual

M, model

AM, acceptable measurements

σ , standard deviation

KDE, kernel density estimate

.nc, netCDF file

GRACE, Gravity Recovery and Climate Experiment

UTC, Coordinated Universal Time

REMSS, Remote Sensing Systems

PODAAC, Physical Oceanography Distributed Active Archive Center

CMIP, Coupled Model Intercomparison Project

MNIST, Modified National Institute of Standards and Technology

CIFAR, Canadian Institute for Advanced Research

MAE, Mean Average Error

MSE, Mean Squared Error

RMSE, Root Mean Squared Error

GPU, graphics processing unit

K, degrees Kelvin

MW, megawatts

GCD, Grand Coulee Dam

NWIS, National Water Information System

DS, dataset

DOI, digital object identifier

ECV, essential climate variable

SDG, sustainable development goals

L0-L4, levels of NASA data

nan, non-numerical

SST, sea surface temperature

RS, remote sensing

CO₂, carbon dioxide

Q_s, surface flow

Q_{sb}, subsurface flow

m, meters

km, kilometers

nn, neural network

i/o, inputs and outputs

EPA, Environmental Protection Agency

ESI, evaporative stress index

US, United States of America

ETo, reference evapotranspiration

VIIRS, Visible Infrared Imaging Radiometer Suite

GOES, Geostationary Operational Environmental Satellite

nb, *nota bene*

ReLU, rectified linear unit

Δ , difference between two numbers

MANUSCRIPT 1

Discerning Hydroclimatic Behavior
with a Deep Convolutional Residual
Regressive Neural Network

Albert Eric Larson

Abdeltawab Hendawi

Soni Pradhanang

Thomas Boving

Ali Shafqat Akanda

1.1 Abstract

The impact of human driven climate change continues to manifest itself daily in the form of extreme events and conditions like droughts, floods, heatwaves and storms. These events sustain uncertainty, risk, and loss to the global ecosystem. Better forecasting tools are mandatory to calibrate our response to these events. These tools help adapt to natural hazards in watersheds and adapt to the planet's dynamic environment. Here, we present a platform based on neural networks called Flux to Flow (F2F) for obtaining, visualizing, and analyzing the basin response of watersheds to water cycle fluxes and their extremes. We examine four United States drainage basins of acreage greater than one million each: Bear, Colorado, Connecticut, and Mississippi. We simulate watershed basin response to the varying climates and magnitudes of hydroclimatic fluxes they face. We combine model and ground observations of water cycle fluxes of precipitation, soil moisture, surface runoff, and sub-surface baseflow. Experiments modulating the lever of time lag between remotely sensed and ground truth measurements are performed to assess the metrological limits of this forecasting device. The resultant grand mean Nash Sutcliffe and Kling Gupta efficiency values are both of greater value than 90%. Our results show that Flux to Flow can become a powerful resource to simulate and forecast hydroclimatic extremes and resulting watershed responses and natural hazards in a globally changing climate.

1.2 Keywords

water, climate, sustainability, supervised representation learning, societal considerations

1.3 Introduction

Water is connected to and connects all living things on Earth. It is wielded to power electronic devices, enables plants, food, and animals to grow, serves as the living

and recreational space for all creatures, and is nourishment to the human body. It has been both the subject of, platform for, and weapon of choice in numerous conflicts. Global hydraulic infrastructure is highly variable. Dirty water can be a source of disease and death. It is branded, modified, and sold at differing levels of purity and concentration. The cost of equipment to control the flow of water is high, maintenance is frequent, and changes in demand and supply for water as a resource are constant sources of concern.

Human activities have changed and continue to change Earth's environment. The changes are visible in both short (meteorological) and long (climatological) time scale responses ([34]). As the temperature of our home planet increases, the amount of snow and sea ice loses volume over time ([30]; [25]), sea levels rise and swallow up once inhabited land ([35]; [32]), storms intensify ([20]), droughts last longer ([37]), floods become more severe ([24]; [14]), animal populations go extinct ([28]), and the availability of freshwater becomes more unreliable ([9]). Concurrently, manmade Earth observation and control systems continue to improve ([6]; [26]).

Here, we approach the topic of watershed modeling with a deep neural network. We observe the connections between model output of four United States drainage basins to actual gauged in the river measurements. All basins are larger than a million acres and thus provide ample data to observe how changes in runoff and subsurface flow impacts the quantity of water discharging from the major river within the basin. Given our results, we envision future work applying the same tools to study and consider all of Earth's watersheds at fine fidelity.

1.4 Materials & Methods

1.4.1 Study Areas

Four United States drainage basins with areas of greater than one million acres each were selected as study areas and are shown in Figure 1.1. The Bear River and

Connecticut River watersheds are significantly smaller than either the Mississippi River or the Colorado River basins. The satellite imagery used observes approximately 100 square kilometers of area (on the order of 25,000 acres) in each pixel.

1.4.2 Satellite Derived Observations

For each basin there are two input images extracted from raw data obtained through the NASA Goddard Earth Sciences Data and Information Services Center. The raw data is National Land Data Assimilation System (NLDAS) model output. NLDAS is a project run by several United States based institutions and universities. NLDAS takes continental scale meteorological data parameters (e.g., air temperature, wind speed, surface pressure, precipitation, incoming radiation, specific humidity) as input and deterministically creates water and energy flux layers as outputs. The NLDAS project in its second phase applies several different water and energy balance algorithms to create flux outputs from one common set of meteorological inputs. Here, the Noah water and energy budget algorithm is used. The channels of interest are components of water flux, specifically surface and sub-surface runoff, as they collectively represent the lateral movement of liquid water along and under the surface towards the terminal drainage point at a given point in time ([39]; [23]).

1.4.3 Ground Truth Measurements

Concurrent with the two NLDAS channels is a single gauged in the river streamflow measurement. Daily streamflow measurements from sites near the terminus of each basin are obtained from the United States Geological Survey's National Water Information System. The USGS operates nearly 30,000 daily streamflow data collectors ([7]). Sites were selected based on the availability, proximity to the terminal point of the basin, and relative continuity of data. Gaps in data collection are solved with linear interpolation.

1.4.4 Data Collection and Preprocessing

For this study, we looked at the time range starting on January 1, 2015, until the most recent output available, March 1, 2022. The NLDAS model output is available in a monthly and hourly product. We combine the hourly data available for surface and subsurface streamflow into a daily product. The raw hourly NLDAS product with all variables is a directory sized 351 gigabytes comprised of 62,805 hourly files. The summing and extraction of lateral flows shrunk the total file size by a factor of more than 150. Each raw data file consumes 5.8 megabytes of disk space, while each daily surface and subsurface flow extraction 822.7 kilobytes. Filtered data consumes only 2.1 gigabytes and can easily be held on a graphical processing unit when trained with the neural network. File size decreases further when clipped to a particular basin. Images are z-scored relative to themselves, while gauged streamflow data is z-scored relative to the entire time series of seven years. Whitening has been shown to improve the performance of training a neural network ([19]; [5]).

1.4.5 Treatment

For this experiment, we constructed a deep, convolutional, residual, regressive neural network. The images of Earth's surface and subsurface water flow are passed through this network. Eventually, the transformed images reach a destination where the image shapes have been constrained in size to match that of the target of the input pair; here, the target is one pixel as the daily value for gauged streamflow is a single physical measurement. The problem is one of regression because the prediction of streamflow is continuous and can theoretically be any value greater than zero. We use convolutional neural networks because our input to the network is a sequence of two channel images ([31]). We also use residual learning, which allows us to make the network very deep but control the opacity of the initial structure of the image. This makes training faster ([12]). Rectified linear unit activation functions are applied in

all but the last layer of nodes, and batch normalization is used in the residual layers ([1]; [16]). Batch normalization is like the z-score treatment in our preprocessing step. Finally, we selected a variant of stochastic gradient descent for optimization of the neural network nodes ([2]; [21]).

1.5 Results

Hourly NLDAS model outputs of surface and subsurface flow are summed to daily accumulations over the time span of January 1st, 2015, to March 1st, 2022. This time series is 2,617 long comprised of two channel images. Channels are surface and subsurface flow measured in units in kilograms per square meter. Units are analogous to the weight of water in a location. Sample observation output from each basin capturing flow behavior on June 6th, 2021, is displayed in Figure 1.2. The effects of spatial resolution are apparent, as the Bear River and Connecticut River basins have pronounced rectangular edges due to their relatively small size. This pixelation effect is not present in the Mississippi River and Colorado River observations of lateral flow from the basin view at this constrained figure size.

Gauged streamflow measurements of the four target rivers are significantly different in magnitude from one another; therefore, we process each with a z-score treatment to center their mean values around the number zero and standardize each increasing and decreasing integer around intervals of standard deviation. Figure 1.3 shows plots of the gauged streamflow measurements of each basin are plotted in two ways. The four individual strip charts show the change in streamflow over time, and the single overlapping histograms show how often actual measurements in the respective basin occur relative to the average discharge. This is a single dimensional z-scoring system. We also perform a two dimensional treatment to each of the input channels, surface, and subsurface streamflow. Whereas the 1-D treatment uses the entire time series of gauged streamflow measurements for computation, 2-D z-

scores are computed based on a single image at a time. Modifiable hyperparameters controls of the network are basin under observation, lag in time between input and output datasets, number of training epochs (forward and backward passes of the neural network) and the ratio of training data to testing data. There is also an override for stopping the model training early when the training data has a Nash Sutcliffe efficiency (NSE) value of a variable efficiency percentage.

Figure 1.4 shows a sample output from one configuration of the neural network. The topmost graph illustrates the time series of discharge measurements in cubic feet per second of the Bear River. This graph is rotated ninety degrees relative to its sibling hydrograph in Figure 1.3. There is a notable seasonality to this streamflow measurement of Bear. Spring brings melting snowpack in the nearby mountainous terrain and subsequent increases in neighboring river flows. Spring melting snow in 2021 appears more subdued than all other years observed. The Bear River drainage basin is located in between the Great Salt Lake and Yellowstone National Park in the Rocky Mountain region of the United States. The eponymously named river flows in a counterclockwise loop.

The second row plots each modeled observation in the time series against its respective actual measurement. On the left is a study of the model output ordered on the x-axis from low to high flows and corresponding actual measurement on the y-axis. The right plot retains the same axis labels, but instead observes spatial proximity of values. Darker points are more commonly occurring ranges of flow. The left plot also contains two lines of best fit, the ideal or desired line found from the data, and the actual line of fit as exists between the actual gauged streamflow and the neural network model output of streamflow from surface and subsurface flow.

The third and final row shows epochal values during the neural network training process. On the left, the average error between the actual measurements and network output declines as the model goes through its iterations of propagate and backpropa-

gate. Concomitant with error vs. epoch is efficiency vs. epoch. As the error declines towards zero, the NSE measurement increases towards 100%. Here, neural network set to stop at an NSE value of 95%, which occurs in the sixth epoch.

We perform nine iterations of the configuration of 252 experiments. For each of the four basins, there are sixty three experiments per iteration based on nine possible values of lag and seven possible values of training and test data split, equating to 2,268 individual runs of the same neural network. Each experiment either stops when the measurement of average NSE of the training dataset within an epoch equals 95% (bottom right, Figure 1.4) or the total number of epochs of back and forward propagation of the entire basin dataset reaches 100. Computations are constrained to a single node with two central processing units, a single NVIDIA GeForce RTX 2080 Ti graphical processing unit, and no more than 130 gigabytes of random access memory. Our platform is written in the python programming language and managed with the miniconda package manager. The total run time to compute the experiments within was 83.0 hours.

1.6 Discussion

The results presented indicate relatively favorable performance of the neural network architecture when transforming of surface and subsurface flow into a prediction of basin gauged streamflow; the kernel density estimates (KDE) in Figure 1.5 and Figure 1.6 illustrate this point. We executed a total of more than 2,200 experiments in total using the common architecture. We use two hydrological metrics: Kling Gupta (KGE) and Nash Sutcliffe (NSE) ([27]; [11]; [10]; [22]). For each of these metrics, the peak resultant merit value of the 2,268 experiments is greater than ninety percent with a standard deviation of less than 0.06. The results are tolerant to lagging the data beyond the residence time of water in the atmosphere ([38]; [8]).

Others have observed the changing water quantity of the Mississippi. One study

used NLDAS data focused on a subsection of the Mississippi with a higher quantity of streamflow target sites ([29]). Another group considers a different data system altogether for watershed modeling on the upper Mississippi basin ([4]). Some groups suggest that NLDAS is too simplistic and decided to create their own blend. They take a much broader approach than the scope of the problem observed here ([36]). The same is true for another study, where the study considers several different models and about 1,000 small river basins ([3]). Some use meteorological data as a predictor for electric outages, as seen in a study looking at Connecticut. They, too, use the Nash Sutcliffe efficiency as a figure of merit ([40]) but are approaching the problem with a different lens. Their target is a smaller population and the risk of being without electric power.

This process can be expanded in different ways. Our study relies on the internal programming of NLDAS to compute surface and subsurface flow. There is much uncertainty in these observations based on the natural heterogeneity of the land surface. We do not look at the independent influence of any single forcing variable. Take snow, for example. In large mountain proximal basins such as those near the Rocky Mountains or Himalayan ranges, accumulation of subzero degrees Celsius water in solid form provides a continuous upland buffer tank for the communities with which the river down land serves. As the relative presence of carbon dioxide increases and the land temperature responds in agreement, the duration and scale of snow melt and sea ice responds. It is challenging to equate with exact certainty how much solid water exists. To a degree, interpolating satellite data with gauged data is sufficient, but these apparatuses are challenging to maintain in cold temperatures or in places of very high altitude. One could elect to observe more individual locations as targets, therefore making the relationship no longer image to single value at a given time, but instead image to image. There are studies that consider the impact of slow moving oceanic and atmospheric abnormalities upon the hydrology of the land. Inde-

pendent variables include the Madden-Julian oscillation ([18]), the El Niño–Southern Oscillation ([15]), and the Atlantic meridional overturning circulation ([17]).

While the NLDAS product used here is of a particular spatial fidelity, the Global Land Data Assimilation System is coarser in its resolution. It is beneficial to the scientific community to have a clearer picture of the meteorological forcing and environmental responses in the ocean, land, air, and mixed interfaces. One could use this framework to fuse the high resolution NLDAS product with the global GLDAS product and evaluate the result according to one common set of metrics. The software could be packaged and ported to use with an already existent embedded in situ mesh system to provide forecasting information. Instead, one might consider looking at a different time signature, such as seasonally decomposed but over several years. Instead, one might introduce higher resolution localized water quality data into the model. By tracking environmental statistical anomalies relative to other points in time and relative to the global community, municipal decision makers can clue into the trajectory of their land, their structures, and their constituents within. The choice to retreat is not to be approached lightly, but in some instances is becoming the necessary measure ([33]; [13]). This intelligence can also be placed in the hands of consulting engineers to distribute in new and existing infrastructure. Logic is necessary to manage assets of complex hydraulic systems (pumps, motors, chemical feed, aeration, dewatering, gates, valves), and digital twin systems are becoming fashionable.

Lessons must be learned from events on both sides of the water quantity spectrum such as the 2022 Pakistan and Mississippi floods on one end and the 2017 Cape Town South Africa water crisis on the other. The opportunities to improve our monitoring systems are many; however, more people are needed in the conversation to consider how we might better cooperate with the environment.

1.7 Conclusion

Using modern techniques and data systems, we introduce a fresh perspective to studying and understanding the water cycle with a learned representation. Our results show that a deep convolutional residual regressive neural network combined with water flux and gauged streamflow data comes to an optimized state, exhibiting strong forecasting performance according to standard hydrological statistical figures of merit. Through the careful use of visuals and data management, this solution is poised to approach with success other locations, degrees of fidelity, time scales, and parameters of interest in the greater climate observatory community.

1.8 Code Availability

Scripts are available at <https://github.com/albertlarson/f2f>

1.9 References

- [1] Abien Fred Agarap. “Deep learning using rectified linear units (relu)”. In: *arXiv preprint arXiv:1803.08375* (2018).
- [2] Shun-ichi Amari. “Backpropagation and stochastic gradient descent method”. In: *Neurocomputing* 5.4–5 (1993), pp. 185–196.
- [3] Xitian Cai et al. “Assessment of simulated water balance from Noah, Noah-MP, CLM, and VIC over CONUS using the NLDAS test bed”. In: *Journal of Geophysical Research: Atmospheres* 119.24 (2014), pp. 13–751.
- [4] Manyu Chen et al. “Effect of watershed delineation and climate datasets density on runoff predictions for the Upper Mississippi River Basin using SWAT within HAWQS”. In: *Water* 13.4 (2021), p. 422.
- [5] Zhi Chen, Yijie Bei, and Cynthia Rudin. “Concept whitening for interpretable image recognition”. In: *Nature Machine Intelligence* 2.12 (2020), pp. 772–782.
- [6] Nicholas H Crisp et al. “The benefits of very low earth orbit for earth observation missions”. In: *Progress in Aerospace Sciences* 117 (2020), p. 100619.
- [7] Melvin D Edwards, Arthur L Putnam, and Norman E Hutchison. *Conceptual design for the National Water Information System*. Tech. rep. US Geological Survey, 1986.
- [8] Luis Gimeno et al. “The residence time of water vapour in the atmosphere”. In: *Nature Reviews Earth & Environment* 2.8 (2021), pp. 558–569.
- [9] Peter H Gleick and Heather Cooley. “Freshwater Scarcity”. In: *Annual Review of Environment and Resources* 46 (2021), pp. 319–348.

- [10] Hoshin Gupta and Harald Kling. “On typical range, sensitivity, and normalization of Mean Squared Error and Nash-Sutcliffe Efficiency type metrics”. In: *Water Resources Research* 47.10 (2011).
- [11] Hoshin Gupta et al. “Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling”. In: *Journal of hydrology* 377.1-2 (2009), pp. 80–91.
- [12] Kaiming He et al. “Deep residual learning for image recognition”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.
- [13] Miyuki Hino, Christopher B Field, and Katharine J Mach. “Managed retreat as a response to natural hazard risk”. In: *Nature climate change* 7.5 (2017), pp. 364–370.
- [14] Yukiko Hirabayashi et al. “Global flood risk under climate change”. In: *Nature climate change* 3.9 (2013), pp. 816–821.
- [15] Dunxin Hu et al. “Pacific western boundary currents and their roles in climate”. In: *Nature* 522.7556 (2015), pp. 299–308.
- [16] Sergey Ioffe and Christian Szegedy. “Batch normalization: Accelerating deep network training by reducing internal covariate shift”. In: *International conference on machine learning*. PMLR. 2015, pp. 448–456.
- [17] M Ionita et al. “Long-term drought intensification over Europe driven by the weakening trend of the Atlantic Meridional Overturning Circulation”. In: *Journal of Hydrology: Regional Studies* 42 (2022), p. 101176.
- [18] Xianan Jiang et al. “Fifty years of research on the Madden-Julian Oscillation: Recent progress, challenges, and perspectives”. In: *Journal of Geophysical Research: Atmospheres* 125.17 (2020), e2019JD030911.

- [19] Juha Karhunen et al. “A class of neural networks for independent component analysis”. In: *IEEE Transactions on neural networks* 8.3 (1997), pp. 486–504.
- [20] Thomas R Karl, Neville Nicholls, and Jonathan Gregory. “The coming climate”. In: *Scientific American* 276.5 (1997), pp. 78–83.
- [21] Diederik P Kingma and Jimmy Ba. “Adam: A method for stochastic optimization”. In: *arXiv preprint arXiv:1412.6980* (2014).
- [22] Wouter JM Knoben, Jim E Freer, and Ross A Woods. “Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores”. In: *Hydrology and Earth System Sciences* 23.10 (2019), pp. 4323–4331.
- [23] Xu Liang et al. “A simple hydrologically based model of land surface water and energy fluxes for general circulation models”. In: *Journal of Geophysical Research: Atmospheres* 99.D7 (1994), pp. 14415–14428.
- [24] P Christopher D Milly et al. “Increasing risk of great floods in a changing climate”. In: *Nature* 415.6871 (2002), pp. 514–517.
- [25] Chao Min et al. “The emerging Arctic shipping corridors”. In: *Geophysical Research Letters* (2022), e2022GL099157.
- [26] Viorel Minzu, Said Riahi, and Eugen Rusu. “Optimal control of an ultraviolet water disinfection system”. In: *Applied Sciences* 11.6 (2021), p. 2638.
- [27] J Eamonn Nash and Jonh V Sutcliffe. “River flow forecasting through conceptual models part I—A discussion of principles”. In: *Journal of hydrology* 10.3 (1970), pp. 282–290.
- [28] Camille Parmesan, Terry L Root, and Michael R Willig. “Impacts of extreme weather and climate on terrestrial biota”. In: *Bulletin of the American Meteorological Society* 81.3 (2000), pp. 443–450.

- [29] Junyu Qi, Qianfeng Wang, and Xuesong Zhang. “On the use of NLDAS2 weather data for hydrologic modeling in the Upper Mississippi River Basin”. In: *Water* 11.5 (2019), p. 960.
- [30] Yue Qin et al. “Agricultural risks from changing snowmelt”. In: *Nature Climate Change* 10.5 (2020), pp. 459–465.
- [31] Waseem Rawat and Zenghui Wang. “Deep convolutional neural networks for image classification: A comprehensive review”. In: *Neural computation* 29.9 (2017), pp. 2352–2449.
- [32] Florian Sévellec, Alexey V Fedorov, and Wei Liu. “Arctic sea-ice decline weakens the Atlantic meridional overturning circulation”. In: *Nature Climate Change* 7.8 (2017), pp. 604–610.
- [33] Anne R Siders. “Social justice implications of US managed retreat buyout programs”. In: *Climatic change* 152.2 (2019), pp. 239–257.
- [34] Peter Stott. “How climate change affects extreme weather events”. In: *Science* 352.6293 (2016), pp. 1517–1518.
- [35] Claudia Tebaldi et al. “Extreme sea levels at different global warming levels”. In: *Nature Climate Change* 11.9 (2021), pp. 746–751.
- [36] Hoang Tran et al. “A hydrological simulation dataset of the Upper Colorado River Basin from 1983 to 2019”. In: *Scientific Data* 9.1 (2022), pp. 1–17.
- [37] Emily Underwood. *Models predict longer, deeper US droughts*. 2015.
- [38] Ruud J Van Der Ent and Obbe A Tuinenburg. “The residence time of water in the atmosphere revisited”. In: *Hydrology and Earth System Sciences* 21.2 (2017), pp. 779–790.

- [39] Youlong Xia et al. “Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products”. In: *Journal of Geophysical Research: Atmospheres* 117.D3 (2012).
- [40] Feifei Yang, Diego Cerrai, and Emmanouil N Anagnostou. “The effect of lead-time weather forecast uncertainty on outage prediction modeling”. In: *Forecasting* 3.3 (2021), pp. 501–516.

1.10 Appendix

Figure 1.1: Drainage basins under investigation

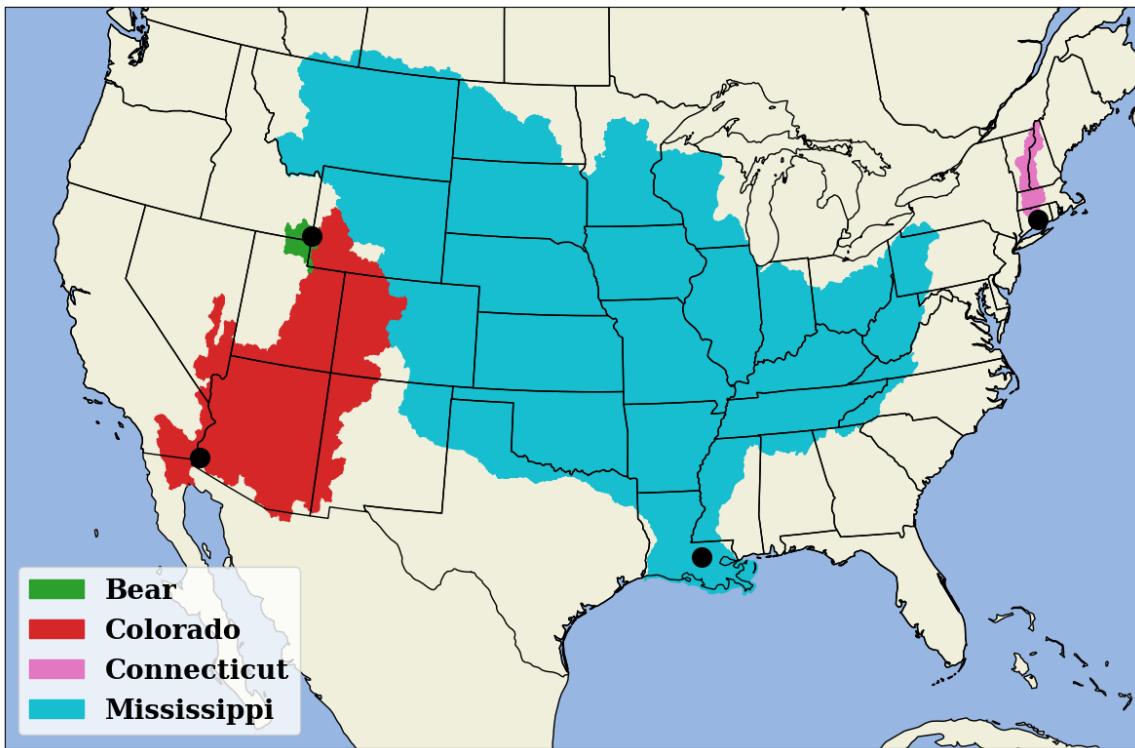


Figure 1.2: NLDAS daily surface and subsurface flows

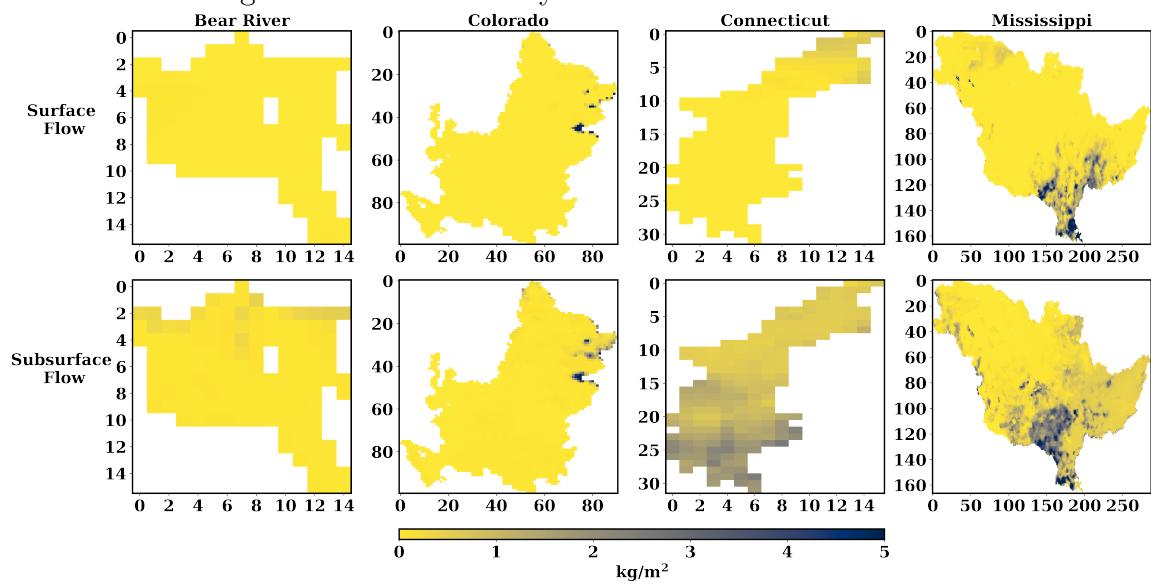


Figure 1.3: Strip chart and histogram plots of z-scored gauged streamflow observations

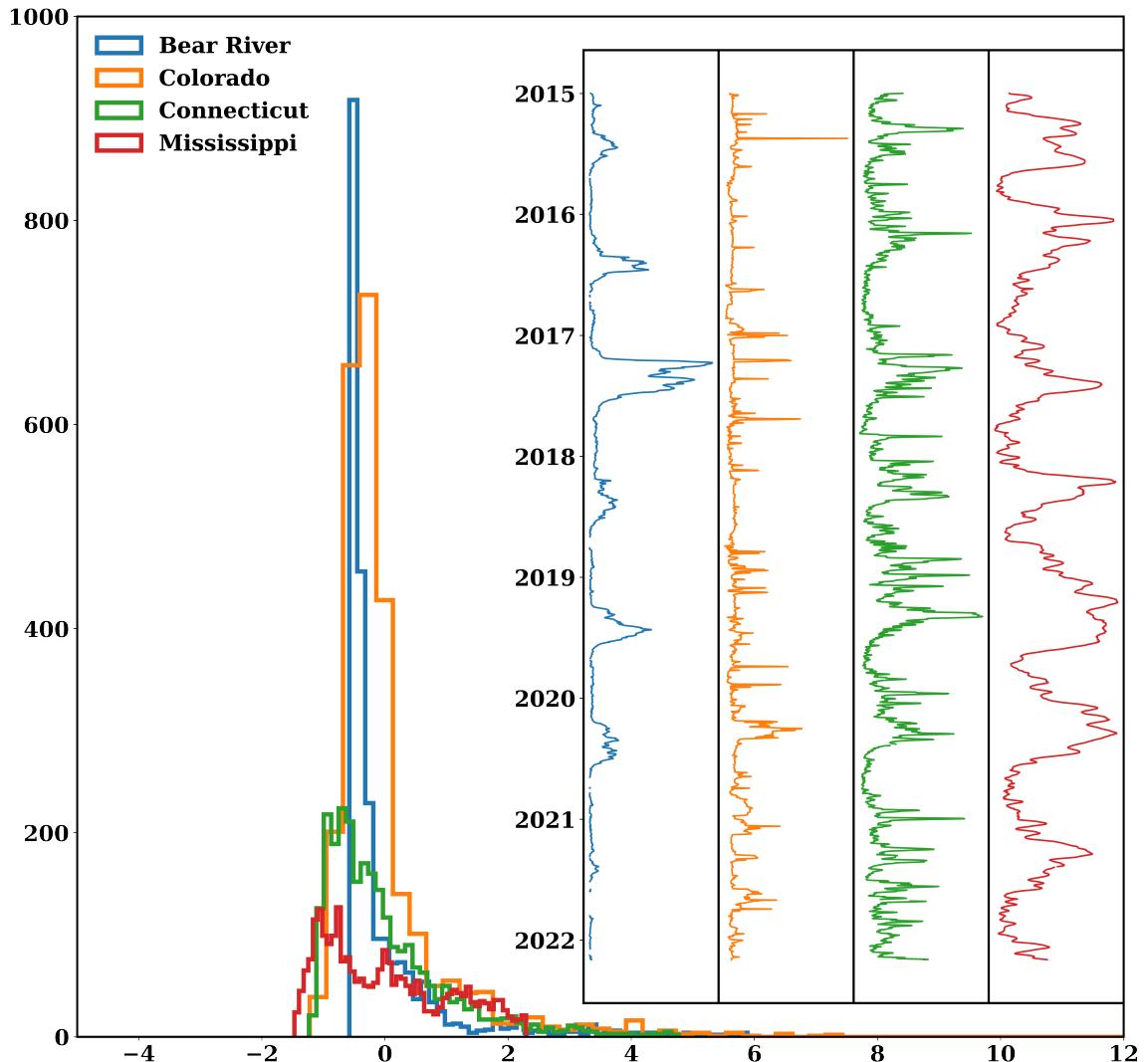


Figure 1.4: Neural network sample output

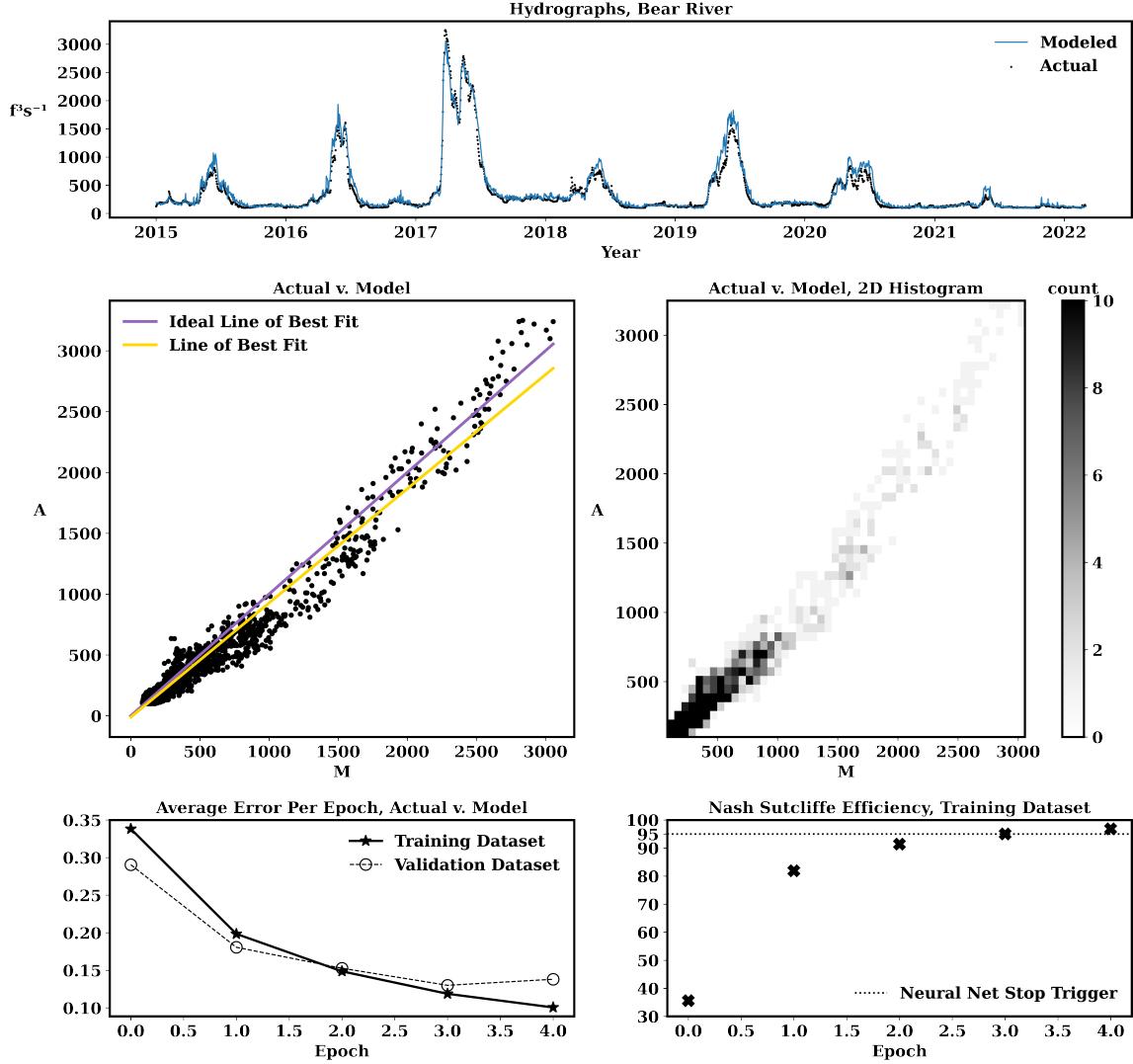


Figure 1.5: Kernel Density Estimates of the 2,268 experiments. Left shows grand NSE and KGE.

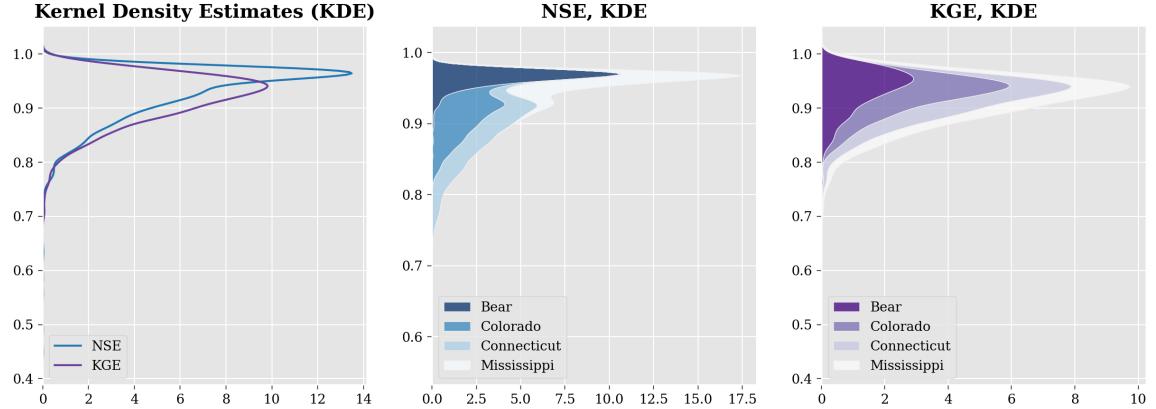
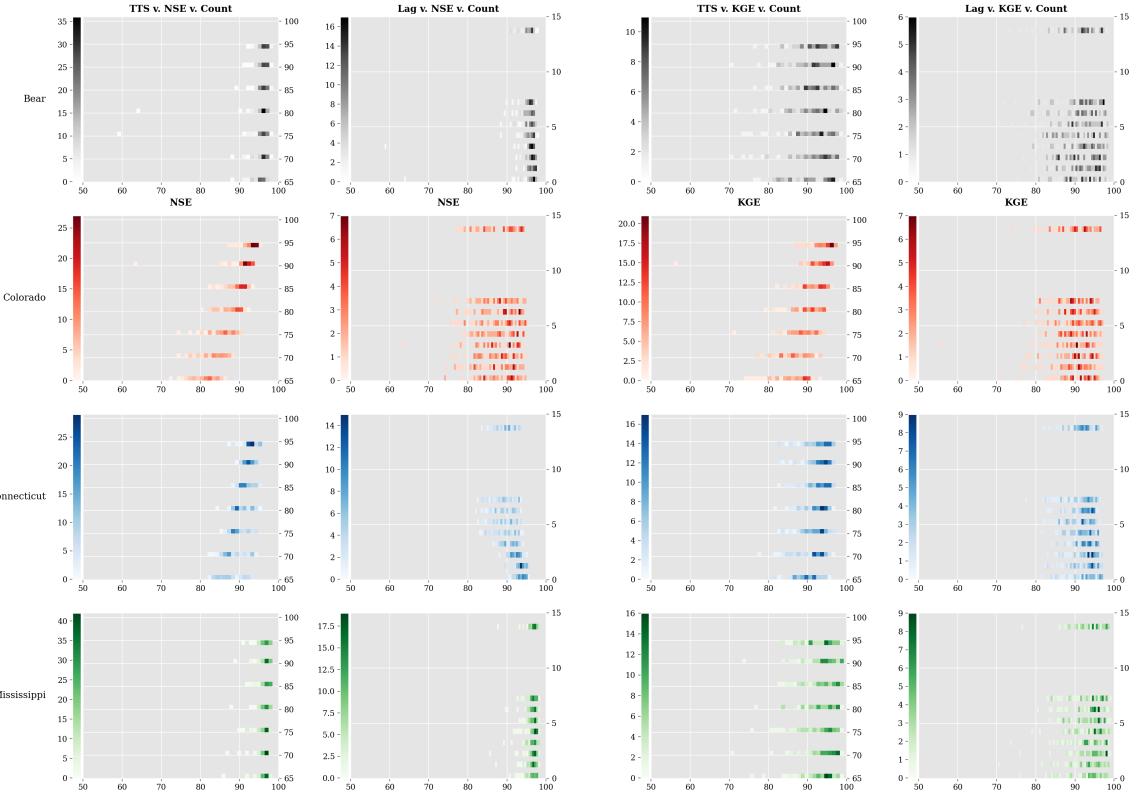


Figure 1.6: 3-D merit plots, basin delineated by color. Intensity of color indicates higher frequency within bin range.



MANUSCRIPT 2

Deep Convolutional Residual Regressive Neural Networks
and Sea Surface Temperatures from Aqua & Argo in the 2000s

Albert Eric Larson and Ali Shafqat Akanda

2.1 Abstract

Sea surface temperature (SST) is an essential climate variable that can be measured via ground truth, remote sensing, or hybrid “model” methodologies. Here, we celebrate the progress of high resolution sea surface temperature via the acknowledgement of a few technological advances from the late 20th and early 21st century. Specifically, we observe three snapshots of twelve monthly SST measurements in 2010 as measured by the passive microwave radiometer AMSR-E, the visible and infrared monitoring MODIS instrument, and the in situ Argo dataset ISAS. We author simple scripts to perform the functions of an extract, transform, and load system. We experiment with a machine learning technology known as deep convolutional residual regressive neural networks and attempt to fuse AMSR-E and MODIS into a superior product. Looking forward, we hope to integrate F2F with future satellite missions such as Surface Water Ocean Topography (SWOT) or Interferometric Synthetic Aperture Radar (InSAR) to enhance the precision of coastal regions observations of water.

2.2 Introduction

Under the auspices of its 193 member states, the World Meteorological Organization holds of particular importance certain physical measurements of Earth and the surrounding environment. Each parameter is singled out by being promoted under the acronym ECV, denoting the physical measurement to be an Essential Climate Variable. Furthermore, the same organization has established goals and deadlines for the participating countries to sustain all forms of life on Earth, known as sustainable development goals or SDGs.

Evidence continues to mount that human beings through industrialization have modified and are continuing to significantly modify the climate. However, the modern cause for concern is the rate at which our climate has changed rather than the Boolean of has it or has it not. Measurements of carbon dioxide (CO₂) tell the story:

detected values of atmospheric CO₂ have increased by 50% of the starting value at the advent of industrialization ([3]). Invariant to latitude and longitude, the impacts are felt everywhere. Earth's response to our stimuli manifests in the form of heat waves, stronger storms, longer periods of drought, greater impulses of meteorological water accumulation over land, and a general increase in environmental variability. While in wealthy communities, modern civil infrastructure serves as a boundary layer to environment-related catastrophes, the poor and powerless are unequally yoked. One must consider also the importance of the ecology itself. As humanity conquers the environment, in what state are the creatures of the atmosphere, land and oceans? What does the next five, ten, five hundred years look like at the current rate? If deemed unacceptable, what changes can be made to mitigate or adapt to implications of past and present poor actions? What are the global environmental quality standards? How can standards be enforced in unequal nation states?

There are many global environmental observation systems that study the entire Earth from a distance. The sheer volume of effort and observation output can be gleaned by world wide web crawling one example: the details of the Coupled Model Intercomparison Project website. Through this project, expert parties from all over the globe share the effort of simulating Earth by focusing on their separate silos whilst having common tunnels to assemble, communicate, benchmark, and improve. Remote sensing (RS) instruments have recently (1950 – present) grown in frequency of occurrence, capability, availability, and affordability. Attached to planes, balloons, spacecraft, or other autonomous means, these devices capture images in a controlled fashion over medium to large portions (swaths) of the land, atmosphere, and ocean. The growth of global RS data is pivotal to our ability to continuously view the macroscopic climate system. The output of these RS devices can be studied with a variety of optical techniques, in most cases altered via some logic to provide a more value-added product depending on the data consumer's needs.

For example, consider the delineation of processing levels provided in the peer-reviewed document coinciding with the deployment of the ECOSTRESS instrument ([5]). A lower value processing level (L0, L1) indicates raw measurements of electromagnetic radiation from the Sun through Earth’s atmosphere off the land / water surface, back up to through the atmosphere to the satellite instrument. In turn, a higher value (L2 – L4) indicates the generation of physical parameters, gridding of swath data, and data assimilation with other parameters. In the instance of ECOSTRESS, the higher levels refer to land surface temperature, emissivity, measures of evapotranspiration, and at Level 4 evaporative stress and water use indices.

Here we investigate the temperature of the ocean’s surface (SST). This target is chosen because of its focus on the behavior of water in the environment, its importance in numerical weather and climate forecasting, and detectability via satellite-mounted RS instrumentation. Also, it has matched continuous ground truth temporospatial measurements that can be investigated for intercomparison of dataset bias, variances, and uncertainties. While it is not practical to grid the entire surface of Earth with sensors, in many situations and places it is extremely valuable to use arrays of dense, spatially linked precise in situ measurements. Satellite observations have a coarse resolution and can miss many interesting small-scale anomalies within the hydrosphere.

We compare the raw satellite observations to the lower resolution but more precise measurements of sea surface temperature. We apply a treatment to the lower resolution but generally more available satellite instrument (AMSR-E), setting its target output to be the higher resolution MODIS product. Our hypothesis is that fusing the AMSR-E data to MODIS data will create a product that is closer in performance to MODIS than its AMSR-E input.

2.3 Materials & Methods

2.3.1 Sea Surface Temperature (SST)

The origin of SST as a continuously monitored variable began when Benjamin Franklin captured measurements of the ocean as he traversed the Atlantic, acquiring data and synthesizing these observations into information about the Gulf Stream. There is a rich history from that point forward to the present day, and a thorough review available in the literature ([15]).

SST is largely academically segregated into the field of physical oceanography. Nevertheless, the connection between ocean, atmosphere, and land are intimately intertwined. There is a trickling down from global SST anomalies like the El-Niño Southern Oscillation, the Madden Julian Oscillation, the Atlantic meridional overturning circulation, western boundary currents, gyres, and eddies to the availability of food, energy, and clean drinking water. Consequently, better understanding and application of oceanic parameters in consideration of land-based hydrology means improved forecasting and preparation for the future.

2.3.2 Aqua

The Aqua satellite was launched on May 4, 2002 ([17]). Upon it, two instruments sit: AMSR-E and MODIS. Both, among other things, are designed to study the temperature of the ocean. The measurements obtained as the satellite is moving from South Pole towards North always crosses the equator at approximately 1:30 PM local time nadir (directly below the satellite). In the downward portion of the orbit, the satellite crosses the equator at 1:30 AM local time nadir.

2.3.3 AMSR-E

AMSR-E is a passive microwave radiometer ([12]). The acronym stands for Advanced Microwave Scanning Radiometer for Earth Observing System. There are

several products produced on top of the raw radiance data collected by this instrument, and the AMSR-E data is processed by different ground stations depending on the parameter of interest. The produced datasets contain latitude, longitude, several physical parameters (e.g., SST, Albedo, soil moisture) as well as other pertinent metadata. As it pertains to sea surface temperature, AMSR-E is available in Level 2 or Level 3 products, and as part of Level 4 assimilation system output.

To detail a sample, one single Level 2 netCDF (.nc) file containing AMSR-E data was downloaded. The record selected is that of March 3rd, 2004, with a UTC time of 01:07:01. The file contains three coordinates (latitude, longitude, and time) and thirteen data variables.

Each variable is a single matrix comprised of columns and rows of measurements. The important distinction here is that the data structure is stored to reflect the path of the orbit. See Figure 2.1. When the sea surface temperature is plotted as it sits in the matrix, it is difficult to discern what is transpiring. There appears to be some curvature of the measurements, but other than that little is known to an untrained eye beyond the title and colormap.

Inclusion of the latitude and longitude coordinates, as well as a global basemap generates a clearer picture as seen in Figure 2.2. A single Level 2 AMSR-E SST file contains matrices representing one full orbit around the globe. Each file holds partially daytime and partially nighttime observations. Because of diurnal warming, it is desirable to separate the nighttime and daytime passes. Furthermore, many analyses are comprised of an ensemble of satellite observations from different platforms such as this one. A grid makes for more orderly computations at large spatial scales. Certainly, one could elect to grid every observation to the AMSR-E or MODIS native product coordinate system. With our experiments, we choose the path of rectangular gridding. We consider the Level 3 product because of the interest in spatial relationship across large geographic scales and variable time (daily, weekly, seasonally, yearly,

generationally).

The Level 3 equidistant rectangular gridded product is accessed via the NASA Jet Propulsion Laboratory's Physical Oceanography Distributed Active Archive Center (PODAAC) and was produced by Remote Sensing Systems of Santa Rosa, California. This product comes in 25 km resolution and is delineated by daytime and nighttime passes of the satellite. The time series runs from June of 2002 until October of 2011 when the AMSR-E instrument ceased functioning. Figure 2.3 illustrates the point that even without explicitly defining the coordinate system in the visualization, the matrix of SST values is already placed in proper spatial order. Figure 2.4 reinforces the fact that little change occurs with the inclusion of latitude and longitude coordinates when plotted on a rectangular grid.

We accumulate the daily daytime and nighttime readings from AMSR-E into monthly products. While a monthly product has already been produced for use with the Climate Model Intercomparison Project, AMSR-E_CMIP5 as available from the does not delineate between daytime and nighttime. Here, we simply compute the monthly average for daytime and nighttime passes on a pixel-wise basis for each month. We finally re-grid the AMSR-E data to the MODIS L3 grid.

2.3.4 MODIS

MODIS, or Moderate Resolution Imaging Spectroradiometer, measures thirty-six different radiance bands in the infrared and visible ranges of the electromagnetic spectrum ([4]). Level 3 sea surface skin temperature as obtained from MODIS comes in 4 kilometer and 9 km products, and is derived from a subset of the thirty-six radiance bands. The products are available in daily, average of eight days, and monthly products. They are also delineated by daytime and nighttime passes of the Aqua's polar-orbiting nature. SST products deriving from MODIS are further specified by the length of the waves within the thermal infrared range used to derive the measure-

ment: longer waves (11–12 microns) and middling waves (3–4 microns). The MODIS documentation state that the 3–4 micron wave SST product is less uncertain, but only usable at night because of the daytime sun glint impact on 3–4 micron waves. We use the long wave 11–12 micron infrared measurements to keep constant the source of both daytime and nighttime passes.

The MODIS Aqua Level 3 SST Thermal IR Monthly 4km V2019.0 product comes with latitude and longitude coordinates, SST values and per pixel quality measurements denoting when contamination is likely. The grid is equidistant rectangular, a match with the AMSR-E grid but at a finer original resolution. Of the over thirty million pixels for an entire day of 4 km MODIS pixels, 90% of them in the random sample selected here are deemed contaminated and filtered out (Figure 2.5). This contrasts with the 50% loss of AMSR-E pixels. This great loss in pixels due to quality is attributed to cloud contamination. To compensate, we use the monthly product (Figure 2.6) where a greater amount of time has transpired, allowing for a higher probability of clean global coverage. A randomly sampled MODIS monthly image yields 50% loss, in line with the AMSR-E daily product and much improved upon relative to the daily MODIS observation files.

2.3.5 Ground Truth Measurements

For a source of ground truth data, we selected the “In Situ Analysis System” (ISAS) dataset obtained from the University of California’s Argo repository and produced by a consortium of French institutions ([7]). An important constraint for this work was to obtain only the surface level measurement of temperature at the highest frequency available during the years of both AMSR-E and MODIS. These products are provided in a gridded format are used to observe temperature measurements at many depth levels. In the publication attached to the ISAS dataset ([7]), the target physical quantity is steric height and ocean heat content; with these as their target

output, gridded depth-dependent temperature is stored as a byproduct. The 0.5 degree monthly dataset is presented in a Mercator projection, slightly different than the AMSR-E and MODIS grids. Mercator lines of longitude have a uniform distance in between them; the distance between latitudes from the equator changes. Identical to AMSR-E, we re-grid this data to the MODIS grid and coordinate system.

2.3.6 Treatment

The treatments we apply to the data are several configurations of one common concept: neural networks. Neural networks are not new, but the growth of graphical processing units (hardware) has enabled them to flourish in software. Neural networks are a type of learned representation. A structure is fed connected input and target pairs. Based on the predictive quality of the initial network structure, an error between the neural network output and the target occurs. This error is in turn fed to an optimization algorithm that iteratively and slightly alters each “neuron” of the initial network structure until it reaches a designated optimal state. Via many small calculations and the simultaneous application of statistical mechanics, neural networks are known to provide qualities like that of a brain, such as capturing spatial eccentricities and temporal changes in sets of related images. Neural networks are applied to a range of tasks from the more mundane such as learning a quadratic equation, to the more cutting edge, like extreme event forecasting or cancer detection.

Transfer learning has become commonplace in the field of machine learning ([21]). Transfer learning places an emphasis on creating reusable treatment structures for others to build on top of without inadvertently causing the audience to get lost in possibly unimportant details. We employ transfer learning to create a complex configuration with a relatively short learning curve. The neural network is characteristically deep, convolutional, residual, and regressive. Our construct is inspired by the work of residual networks ([9]). However, our problem is one of a regressive nature. Sea

surface temperature has a continuous temperature range that it exists within. This is a notable difference to some of the more common introductory neural network examples, such as those associated with the MNIST and CIFAR datasets where the number of possible outputs is very small ([22]; [14]). Loss functions associated with regressive problems are constrained to just a couple: mean absolute error (MAE) and mean squared error (MSE). The calculation of the loss function must be differentiable. This is due to the optimization component of neural networks. The literature is rich with publications regarding neural network optimizers, as well as the general mechanics of neural nets ([1]; [13]).

Once neural network architecture and hyperparameters are chosen, training and validation data is loaded into the network. While training the neural network, close observation is made of the reduction in error between training input and output as the neural network begins to optimize or learn. We also monitor the validation dataset at each training iteration. The learning process stops once the training and validation data has been passed through the network a certain number of times, or epochs. When prototyping or pilot-testing the experiment set to be carried out, one should test with a very short number of epochs and a larger sum of epochs to see where good performance meets fast time of computation.

After training, the optimized neural network structure is intentionally frozen. Before the point of freezing, the neurons of the network can be adjusted for optimization, like a student asking a teacher for advice when studying. The frozen state and inference imposed upon it is like a student being prompted with a pop quiz and no teacher assistance. This test or input data are similar enough to the training that the teacher believes the student will have success in passing the test according to the selected merit (mean squared error, the loss function). After the test, the performance of the model is evaluated and a decision is made regarding next logical steps in the research.

A neural network can become biased to its training inputs. It starts to memo-

rize the training dataset, which does not make for a generally applicable algorithm. Avoidance of biasing comes at the cost of variance ([8]). Applying dropout is one technique to systematically prevent system bias by simply “turning off” a certain percentage of random neurons at each iteration of the algorithm ([10]; [19]). Another approach is the application of early stopping. The loss function of a neural network typically looks like a very steep curve down to a flat bottom. Rather than allow the network to persist in the flat bottom for long and become overfit, simple logic can be employed to stop training early when the network shows evidence that it has reached an optimal state. Percentage of data split between training and testing proportions is another relevant training hyperparameter. A larger proportion of the dataset being part of the training portion could lead to overfitting of the model and lack of generalized predictability. On the other hand, insufficient training data might lead to an inability to adequately characterize the reality of the data pairs.

The image sets subjected to treatment are on the large side computationally. Holding many one million or nine million pixel images within the memory of a single graphical processing unit becomes intolerable to the device. One could elect to use multiple GPUs or a compute node with a great provision of memory. Here, we constrain the experiment to a single GPU and cut the images up into smaller pieces of square data. Our patch size is fixed at 100 x 100, though this is a tunable hyperparameter. Figure 2.7 shows a Pacific Ocean study region, highlighting Hawaii and regions east. While this image is too large to process directly in the neural network, we can solve this problem by creating the eighteen patches of 100 x 100 pixels, representing the 300 x 600 pixel region under observation.

Neural networks do not function when nan values are present in any of the images. We enacted a broad treatment to the AMSR-E and MODIS images, computing the mean of the entire image, excluding the nan values. Then, where the nan values are present, we replace them with the mean value. This has the convenient byproduct of

introducing into the neural network many training pairs where the input and output are simply comprised of the average global SST value as obtained via the AMSR-E and MODIS instrument.

2.4 Results

We prepared the SST data for nine different cases all studying the year 2010: the **A**tlantic, **P**acific, and **I**ndian Oceans segmented by monthly **D**ay, **N**ight, and **H**ybrid (nb, hybrid means day and night images averaged together) observations of SST (nb, first letters of the previous are bolded to call attention for their use in Figure 2.8). We train the neural network on the first ten months of the year, validate with the eleventh month, and test with the twelfth month. A training session runs for 100 epochs. Each image in the geographically constrained time series is 300 pixels x 600 pixels in size, divided up into eighteen 100 x 100 pixel segments to incrementally feed the neural net.

In all instances, both training and validation loss functions drop by several orders, indicating successful training without overparameterization as would be indicated by low training error but high validation error. Performance of network as it relates to the test data, December 2010, is seen in the upper plot of Figure 2.8. The goal is for measurements labeled Pred or Optim to bring the RMSE value between AMSR-E and Argo or MODIS down. Pred is the prediction directly from the neural network. The Optim case takes the Pred and performs a band pass on the signal. If there are any measurements outside three standard deviations of the mean, they are replaced with nan values instead. This is a device meant to combat some of the challenges with coastal artifact.

In every case, the RMSE between the optim and MODIS is higher than the AMSR-E input. In some instances, the test case of December does make an optim output that is closer to Argo than the input AMSR-E. In some cases, though, it makes a

worse performing product with regards to Argo than either AMSR-E or MODIS. A bright spot is that the optim output is closer to MODIS than the Argo product. See Figures 2.9 and 2.10 for samples of how the RMSE translates to actual transformation of the images.

2.5 Discussion

While the continual development of a relatively open extract, transform, and load ([2]) system along with creation of the actual destination for the loaded data (namely the neural network treatment and posttest analytics) is certainly a plus, the results of this study request future engagement. In every case of this study, the neural network appears to struggle with coastal regions. This is due to the nature of the land sea boundary layer in all these datasets. At the presence of land, the raw data (as they are downloaded as .nc files) are given a non-number (nan) designation. Neural networks weren't designed optimally for the currently produced segregated RS datasets. The datasets appear to be manufactured with the understanding that one group of people are still more interested in ocean behaviors, another land behavior. For the purpose of training a neural network using the convolutional flavor, images with no nans are needed. As referenced earlier, steps were taken during the training process to circumvent the presence of land by substituting those pixels temporarily with the local mean value. Another option is the application of the substitution of the nan values with the mean as computed by the entire "scene" or day. There is the potential and a likelihood that the substitution of these values is introducing a source of structured noise. This noise might be leading to the higher than preferable test performance as denoted in Figure 2.8. Furthermore, it is probable that this structured noise is hindering training of the neural network process itself.

As it relates to computer vision tasks such as this set of experiments, the use of mean squared error as a loss function has been called into question as an appropriate

target ([20]). Their results certainly warrant some concern, and our experiments have some corroboration with their findings. Our images are single channel inputs and can be considered grayscale pictures. When displaying SST images, we use a colormap based on what we know to be the physical limits of the parameter itself. This is a different approach than typical of image based machine learning techniques. Alternate loss functions to the standards baked into PyTorch are available ([11]). These functions require the inputs to be either between 0 and 1 (grayscale) or 0 and 255 (color images). Another avenue is the pursuit of physics-based loss functions ([18]). While neural networks are a useful tool, they alone are not a silver bullet, especially as it relates to geophysics. However, the neural network community has a keen interest in computational efficiency.

We applied a land mask generated from the MODIS instrument. Aqua has far surpassed its projected useful life span and was designed before the new millennium. The new Surface Water Ocean Topography (SWOT) mission launched in 2023 ([6]; [16]). It will bring many new insights to the hydrology community. Among those insights are a more precise global picture of Earth’s coastal regions.

Only a fraction of the available data was observed in this study. The ISAS Argo dataset was a single file attached to a DOI address of over fifty gigabytes. We extracted simple the surface layer of this dataset. There is great value in consideration of SST depth layers. Furthermore, we studied monthly time series images of all three raw datasets. AMSR-E and MODIS each have near complete global pictures within two to three days. These datasets are then transformed in different ways and can lose fidelity by various types of decimation such as regridding from swaths to squares, uncertainty in formulas used for conversion from base input to high level (L2 - L4) physical parameter, or by forms of compression.

Cloud cover is a persistent factor at play within the community. The question of “is the measurement (pixel) currently observed impacted in an undetectable way?”

can't fully go away, because even at the hyperlocal "nowcasting" time scale there is missed detection of events. Because of the pervasive challenges, people need to come together. Great global solutions require more cooperation, engagement, and the act of building bridges with one another. Missions like the International Space Station, Artemis, Landsat, and GRACE, and SWOT are only examples of what global cooperation can result in. Here at the microscopic level of single parameter consideration, we need more of the same type of teamwork. Dynamic collaboration amongst many stakeholders raises economic efficiencies. Improvement of just a single ECV requires takes the participation of a deep supply chain. Water is life. It connects the global ecosystem in nearly every facet, from food supplies, health of exotic wild animals, to the manufacturing of semiconductor chips and the treatment of industrial water. A clearer perspective is always welcomed to help sustain life. The missions of Aqua and Argo certainly achieved their planned missions in that way.

2.6 Conclusion

Sea surface temperature is an essential climate variable and crucial to understand the movement of water throughout the hydrosphere. The beginning of the 21st century marked a new frontier in the measurement of SST via the Aqua mission and Argo program. We observed three overlapping datasets focused on measurements of sea surface temperature: AMSR-E microwave measurements, MODIS infrared measurements, and ISAS Argo float in situ measurements. We focus the study on three large oceanic regions: Indian, Pacific, and Atlantic. We used Flux to Flow, an extract, transform, load, treat, and evaluation framework based around a deep convolutional residual regressive neural network. We attempt to transform the coarser resolution satellite product towards the finer one and intercompare all datasets. While the neural network performs well according to its typical loss functions, we find that the presence of frequent nan values, the limitations of mean squared error as a loss func-

tion used in computer vision tasks, and the sheer size of output target quantities compounded with high desired precision results in limited success when applying the neural network transformation. We see the future of this framework including other treatment algorithms, experiments where fewer output target values are considered, or computing resources run in tandem to build a bigger network for a potential better grasp on the transformation process.

2.7 Code Availability

Scripts are available at https://github.com/albertlarson/f2f_sst

2.8 References

- [1] Marcin Andrychowicz et al. “Learning to learn by gradient descent by gradient descent”. In: *Advances in neural information processing systems* 29 (2016).
- [2] Srividya K Bansal and Sebastian Kagemann. “Integrating big data: A semantic extract-transform-load framework”. In: *Computer* 48.3 (2015), pp. 42–50.
- [3] A Eldering et al. “The Orbiting Carbon Observatory-2 early science investigations of regional carbon dioxide fluxes”. In: *Science* 358.6360 (2017), eaam5745.
- [4] Wayne E Esaias et al. “An overview of MODIS capabilities for ocean science observations”. In: *IEEE Transactions on Geoscience and Remote Sensing* 36.4 (1998), pp. 1250–1265.
- [5] Joshua B Fisher et al. “ECOSTRESS: NASA’s next generation mission to measure evapotranspiration from the international space station”. In: *Water Resources Research* 56.4 (2020), e2019WR026058.
- [6] Lee-Lueng Fu et al. “The SWOT (Surface Water and Ocean Topography) mission: Spaceborne radar interferometry for oceanographic and hydrological applications”. In: *Proceedings of OCEANOBS* 9 (2009), pp. 21–25.
- [7] Fabienne Gaillard et al. “In situ-based reanalysis of the global ocean temperature and salinity with ISAS: Variability of the heat content and steric height”. In: *Journal of Climate* 29.4 (2016), pp. 1305–1323.
- [8] Stuart Geman, Elie Bienenstock, and René Doursat. “Neural networks and the bias/variance dilemma”. In: *Neural computation* 4.1 (1992), pp. 1–58.

- [9] Kaiming He et al. “Deep residual learning for image recognition”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.
- [10] Geoffrey E Hinton et al. “Improving neural networks by preventing co-adaptation of feature detectors”. In: *arXiv preprint arXiv:1207.0580* (2012).
- [11] Segrey Kastryulin et al. “Image quality assessment for magnetic resonance imaging”. In: *arXiv preprint arXiv:2203.07809* (2022).
- [12] Toneo Kawanishi et al. “The Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E), NASDA’s contribution to the EOS for global energy and water cycle studies”. In: *IEEE Transactions on Geoscience and Remote Sensing* 41.2 (2003), pp. 184–194.
- [13] Diederik P Kingma and Jimmy Ba. “Adam: A method for stochastic optimization”. In: *arXiv preprint arXiv:1412.6980* (2014).
- [14] Alex Krizhevsky and Geoff Hinton. “Convolutional deep belief networks on cifar-10”. In: *Unpublished manuscript* 40.7 (2010), pp. 1–9.
- [15] PJ Minnett et al. “Half a century of satellite remote sensing of sea-surface temperature”. In: *Remote Sensing of Environment* 233 (2019), p. 111366.
- [16] Rosemary Morrow et al. “Global observations of fine-scale ocean surface topography with the Surface Water and Ocean Topography (SWOT) mission”. In: *Frontiers in Marine Science* 6 (2019), p. 232.
- [17] Claire L Parkinson. “Aqua: An Earth-observing satellite mission to examine water and other climate variables”. In: *IEEE Transactions on Geoscience and Remote Sensing* 41.2 (2003), pp. 173–183.
- [18] Apostolos F Psaros, Kenji Kawaguchi, and George Em Karniadakis. “Meta-learning PINN loss functions”. In: *Journal of Computational Physics* 458 (2022), p. 111121.

- [19] Nitish Srivastava et al. “Dropout: a simple way to prevent neural networks from overfitting”. In: *The journal of machine learning research* 15.1 (2014), pp. 1929–1958.
- [20] Zhou Wang and Alan C Bovik. “Mean squared error: Love it or leave it? A new look at signal fidelity measures”. In: *IEEE signal processing magazine* 26.1 (2009), pp. 98–117.
- [21] Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang. “A survey of transfer learning”. In: *Journal of Big data* 3.1 (2016), pp. 1–40.
- [22] Han Xiao, Kashif Rasul, and Roland Vollgraf. “Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms”. In: *arXiv preprint arXiv:1708.07747* (2017).

2.9 Appendix

Figure 2.1: L2 AMSR-E SST field, March 3, 2004, no coordinate system

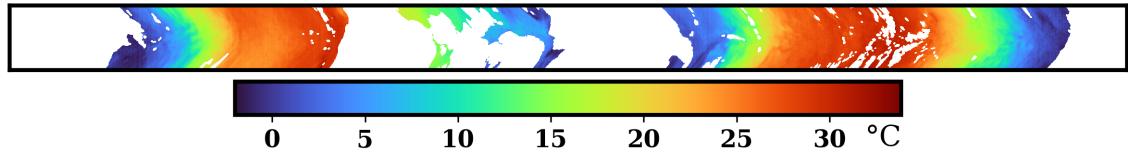


Figure 2.2: L2 AMSR-E SST, plotted with available coordinates and world map

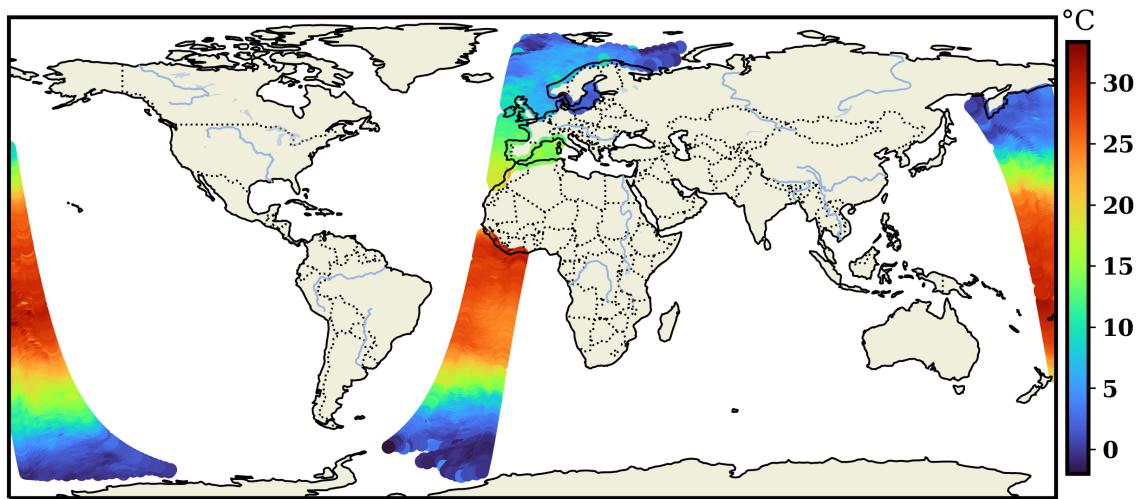


Figure 2.3: L3 AMSR-E file plotted without supplied coordinate system



Figure 2.4: L3 AMSR-E file plotted with available coordinates and world map

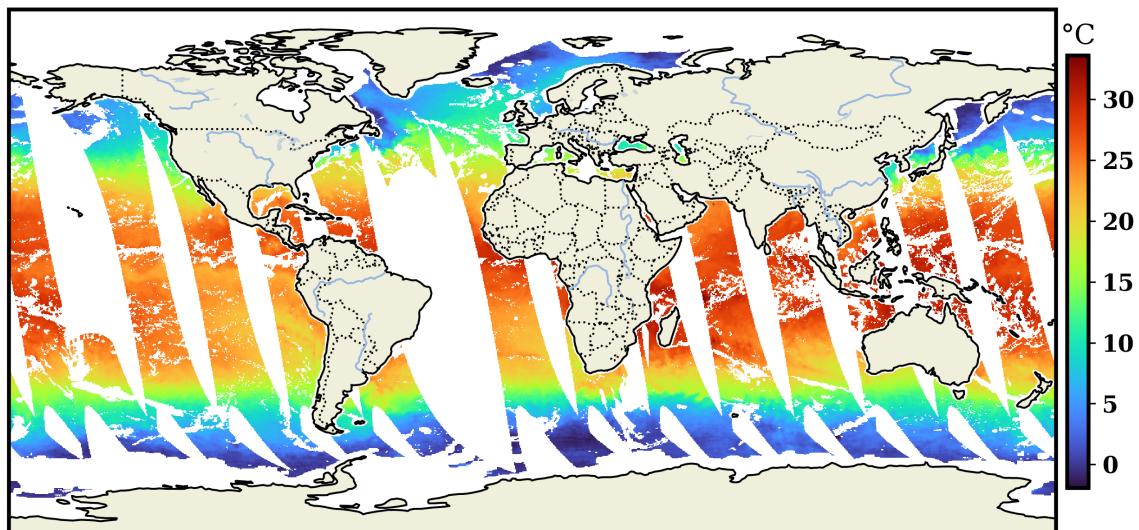


Figure 2.5: L3 daily MODIS file containing only high quality flagged pixels

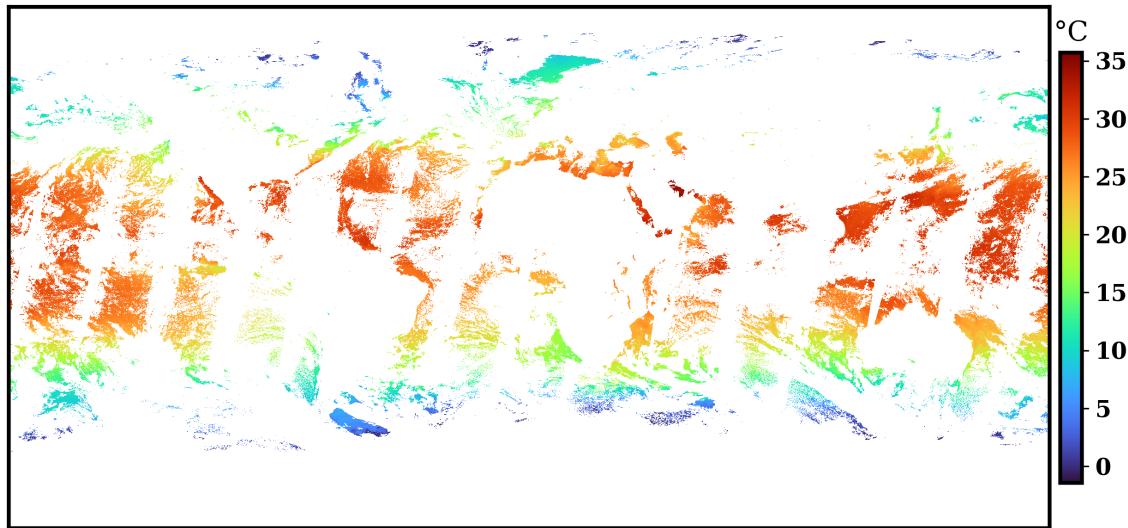


Figure 2.6: Monthly L3 MODIS image containing only high quality observations

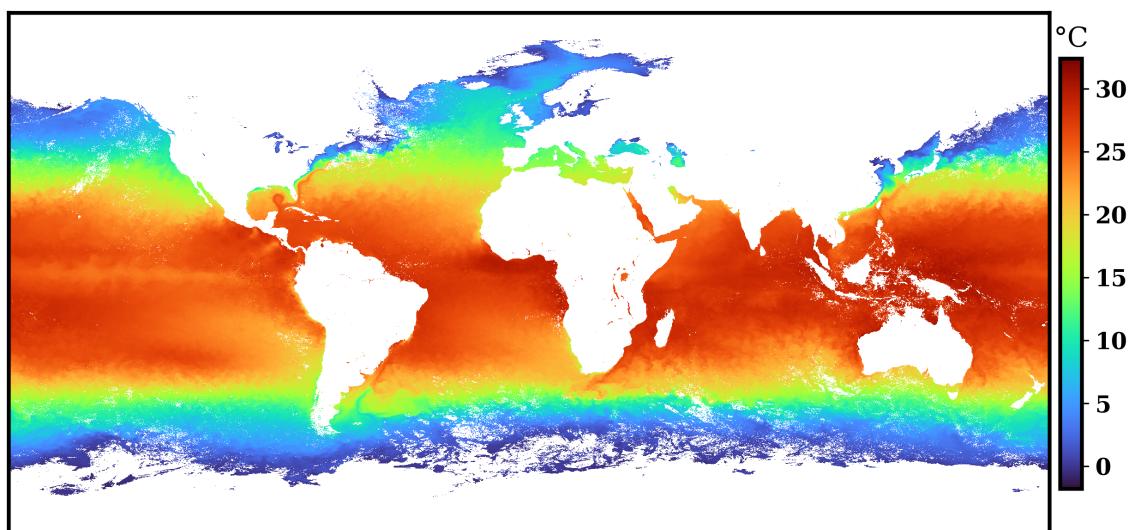


Figure 2.7: Sample training monthly observation; January 2010 MODIS day observation of the Hawaiian Islands; segmented into 100×100 pixel regions.

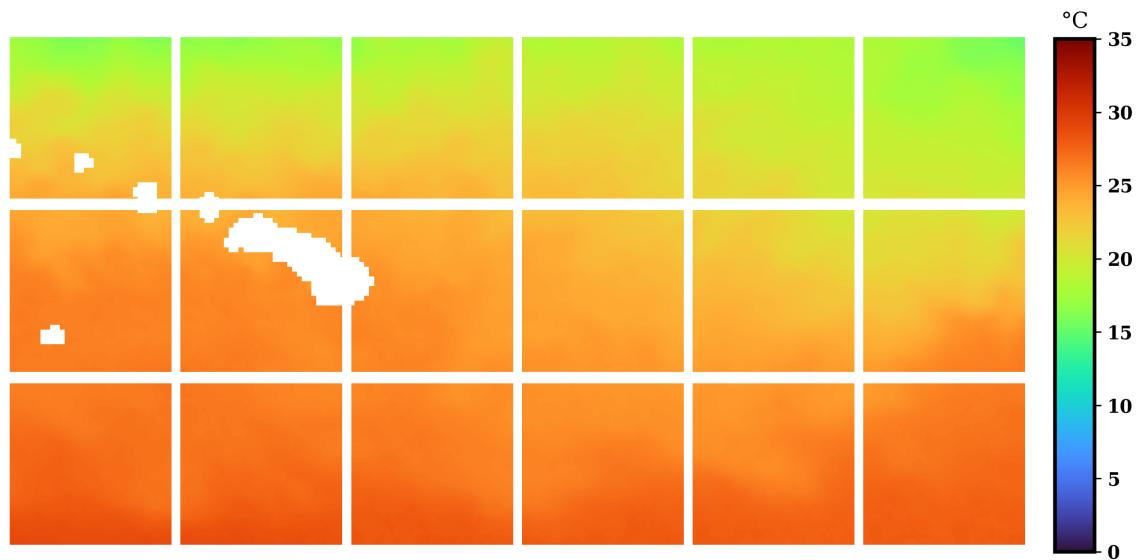


Figure 2.8: December 2010. The top panel shows the distances from Argo and MODIS to predicted, optimized prediction, AMSR-E and each other for all nine experiments. Two bottom panels are plots of training and validation losses during neural network training.

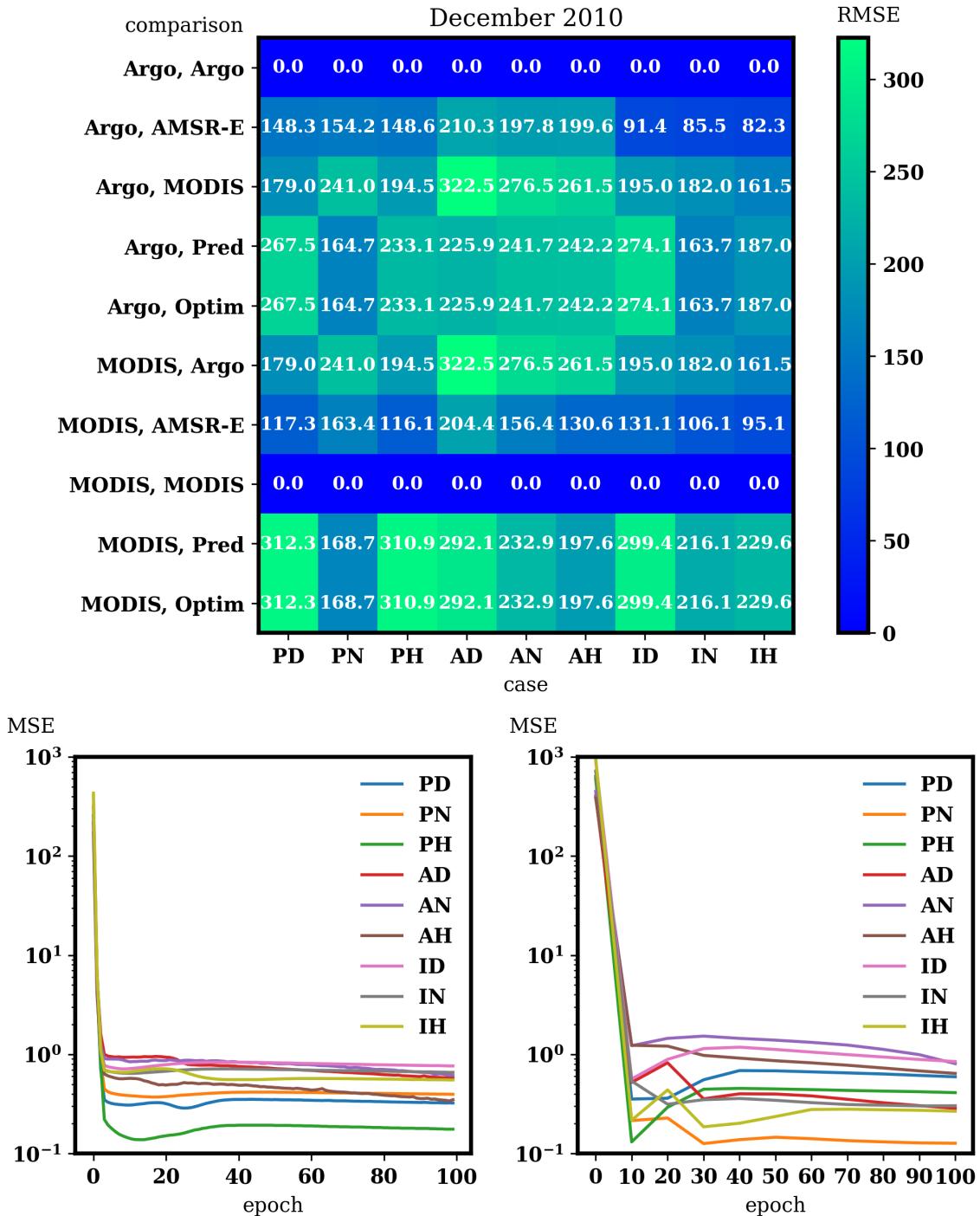


Figure 2.9: Relatively “good” perceptual change, Pacific Night case

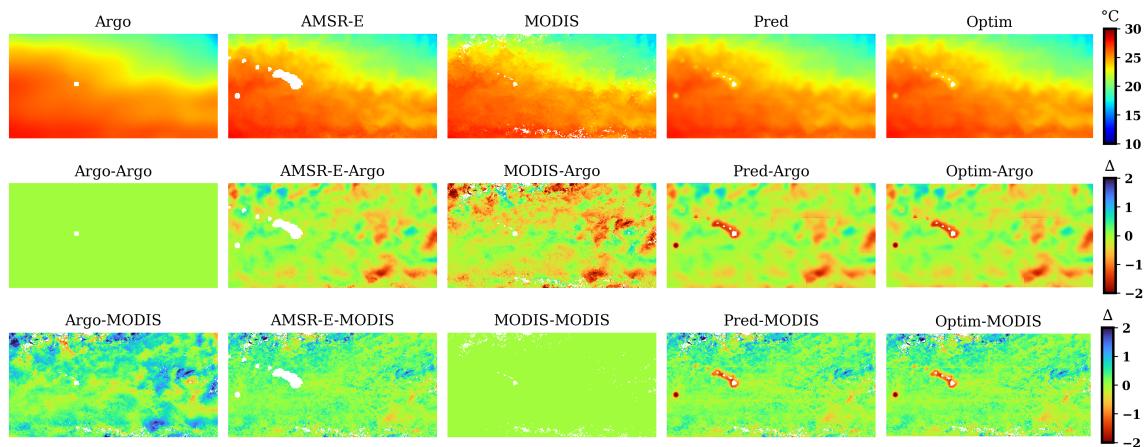
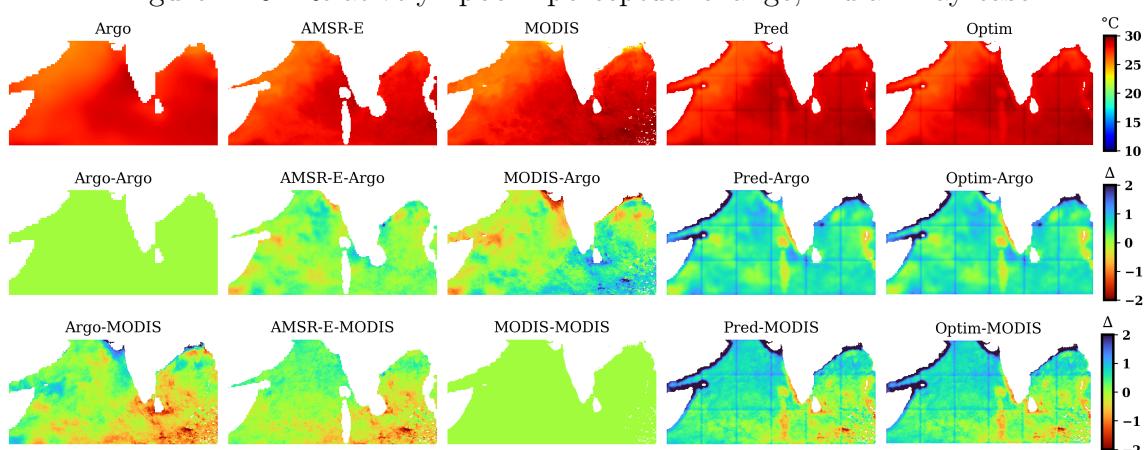


Figure 2.10: Relatively “poor” perceptual change, Indian Day case



MANUSCRIPT 3

Holistic Water Cycle Analysis via the Confluence of Climate Model, Satellite,
Ground Truth, and Machine Learning Signal Processing Technologies: Two
North American Transboundary River Watersheds

Albert Eric Larson

Soni Pradhanang

Thomas Boving

Ali Shafqat Akanda

3.1 Abstract

Water continuously cycles throughout the land, ocean, and atmosphere. Accordingly, it is important for hydrological analyses to consider water as it moves throughout the entire hydrosphere, and not just a single facet of the process. We use neural networks as a device to transform geospatial observations of water quantity and quality into forecasts of ground truth streamflow measurements. Two very large transboundary basins, the Columbia River and Yukon River, are subjects of this investigation. We first describe the basins. Then we create two datasets for each basin: one with coupled surface flow, subsurface flow, and sea surface temperature of the basin adjacent oceans; and another with simply the surface and subsurface flow land measurements constrained to the definitive boundaries of the delineated watershed. Finally, we load these datasets into Flux to Flow (F2F), our neural network test platform. Our results indicate that, even with the smallest neural network we try (four neurons only), use of sea surface temperature greatly improves forecasting of monthly streamflow from up to two years lag between the input images and the output gauged streamflow measurements. We see the future use of the F2F pattern having more output targets and likely requiring multiple compute nodes to scale the work. We discuss and identify drought monitoring as a suitable next step. We believe this work has only scratched the surface regarding the integration of land and ocean parameter datasets to fields devoid of non-numerical observations.

3.2 Keywords

water, water quantity, streamflow, sea surface temperature, GLDAS, MODIS, Columbia River, Yukon River, z-score, Nash Sutcliffe efficiency, neural networks

3.3 Introduction

Depending on the location, there are different water related signals that humans can witness and measure. These time variant water signals include but are not limited to parameters such as rainfall, snowfall, cloud cover, fog, ice, soil moisture, streamflow, river height, body of water temperature (lakes, seas, ponds, oceans), tidal timing and intensity, waterbody color, turbidity, and trace chemical concentrations in the water. These variables are commonly cloistered into variables of the land, variables of the atmosphere, and finally but certainly not least important variables of the ocean. Together they form the hydrosphere. We focus on variables of the land and variables of the ocean; however, we affirm the impact of the atmosphere upon land and ocean. Even at long time scales, there are certain parts of the globe that are constantly obscured from viewing with traditional bands of the electromagnetic spectrum due to noise such as clouds, wind, or other interference. This fact places an emphasis on the need for continuation of innovation in satellite measurements, ground sensors, and the spectrum of devices in between. The state of the science is still in many instances ‘flying blind’; Earth’s environment is dynamic and complex and modern solutions need to grow to combat modern problems.

We consider two extra-large basins. As an added complexity, we target trans-boundary watersheds. The motivation behind the use of transboundary watersheds is to simulate differences or nuance when leaving the gates of our national data systems in the acquisition of gauged river discharge measurements. In this instance, our journey is a short one to the northern riparian of the United States: Canada. Our view considers seventeen years of data, spanning years at monthly snapshots as opposed to the shorter end of the time spectrum (seconds, minutes, hours, days, or weeks). The two basins selected herein (Yukon River, Columbia River) are unique because of, among other features, their: 1, far from equatorial latitude; 2, proximity to the integration of land and ocean; 3, great range in quantity of water infrastructure

between the two habitats; 4, large range of elevation change.

We study surface and subsurface flow as computed for one of the GLDAS datasets. Additionally, we couple the two GLDAS datasets to MODIS sea surface temperature, creating a global dataset mostly devoid of non-number pixels. Of interest is the predictability of several gauges at a time when using the clipped version of GLDAS versus our custom ocean-enabled version of GLDAS. Furthermore, we modulate the lever of z-scored datasets versus non-z-scored datasets. We find that combining surface flow with subsurface flow (derived measurements of the changing weight of water in a region) and sea surface temperature has a markedly improved performance in predicting gauged streamflow than the datasets containing only surface and subsurface flow.

The wide lens guiding motivation behind this study is three-fold: 1, to improve the reader's understanding and capability to 'quickly' visualize the pertinent available macroscopic modern water cycle monitoring data sets; 2, to highlight state of the science level data typically used by experts for the generation and interpretation of global policy targeted outputs; and 3, to present a narrative of water (hydrology, limnology, oceanography) as the focal point of study because of its status as a primary building block of life.

3.4 Materials & Methods

3.4.1 Yukon River Watershed

The Yukon River and greater watershed are situated within part of the state of Alaska, and the eponymously named Canadian territory known simply as Yukon (Figure 3.1). A thorough baseline hydrological artefact of the Yukon River exists ([6]). The headwaters of the Yukon River emerge from the constellation of finger style lakes (e.g., Atlin Lake, Marsh Lake, Teslin Lake, Gladys Lake) known customarily as the Southern Lakes region along the northwestern border of the Canadian province of British Columbia. Sampled along the route of the river, one is likely to find bedrock,

confined and unconfined aquifers, a host of unique soil profiles, vegetation, permafrost, and a not to be understated flurry of flora and fauna friendly to subarctic climatology. At a microscopic, chemical level, the frequent boreal forest attributes concurrent with the frigid temperatures are well understood ([40]). Zooming out to a wider view than just the Yukon, evidence continuously mounts regarding the impacts of the changing climate. As goes the rest of the world, the Yukon feels an impulse. Nitrogen and phosphorous pollution are two of the more notorious organic chemical components that when unmanaged can wreak havoc on a watershed, causing eutrophication, loss of biodiversity ([7]). It's a short walk down the primrose path to a river lacking in a once endemic species of fish ([27]).

The Yukon River is famous for its salmon. In the last several years, the Yukon has come under public scrutiny for declining populations of chinook and chum in the Yukon and its neighbor river the Kuskokwim. A combination of bycatch in pollack fisheries, overfishing, marine heatwaves and algae blooms are causing the environment to become potentially less suitable for fish. There are prevalent modern terrors such as “forever chemicals” and the ubiquity of microplastics in the hydrosphere to the degree that many careers are being dedicated to the study as it relates to contamination of our waterways ([45]). It is a wonder at all sometimes that our river ways can sustain any life given the way some humans steward the gift. Humanity has much work to do to repair and improve the current status quo.

Environmental flows, or “e-flows” is “the quantity, timing, and quality of water flows required to sustain freshwater and estuarine ecosystems and the human livelihoods and wellbeing that depend on these ecosystems” ([3]). Considering a perspective filtered through the comprehension of e-flows, the Yukon has natural advantages. It is in a higher latitude, close to the North Pole (there's a town called North Pole in Alaska along the Yukon). Because of the high latitudes, constantly cold (equatorially relative) temperatures, and sometimes very long dark periods of time, there's a

tendency for life forms to shirk away from this environment. As such, the river is less prone to farming, heavy nutrient loading, or industry strain. That is not to say that its observation is in vain or without merit, far from it. Large swaths of previously unoccupied or lightly occupied land are important ones to observe with greater clarity the forcings and results of the natural global Earth phenomena as it is subjected to human intervention.

Physically, the Yukon River and surrounding land has all the features one desires in a naturally occurring surface water: beautiful lakes, high mountain scenery, sinuous meanders, novel woodland creatures. For example, erosion drives the creation of oxbow lakes, small slices of river cut off after anomalous water patterns drive unusual erosion in a suitable soil substrate. The Yukon is home to many cyclically migrating birds, as well as a habitat for many bears ([35]). Bears are attracted to the Yukon's most notorious creatures: the chinook and chum populations traversing the watercourse.

3.4.2 Columbia River Watershed

South of the Yukon Basin, at the southern edge of British Columbia and the boundary between the contiguous US and Canada lies the beginning of the Columbia River (Figure 3.1). The watershed is split into two major parts by the Cascade Mountain range running northerly from Oregon up into British Columbia. Land wise, the basin is primarily east of the Cascades. The western half of the larger Columbia drainage basin as divided by the Cascades finds a large share of water in the flows of the Willamette River. The waters of the Willamette stem from Waldo Lake, and it moves in a northerly direction before merging with the formal Columbia River. The Willamette unfortunately came under great scrutiny for the exhaustion of its goodwill to the point of receiving the designation of one of the United States most polluted waterways ([34]). There are efforts to restore the waterway that have

logged twenty-five active years and are beginning to restore these damaged arteries and veins of water.

The Willamette meets the Columbia in Portland, Oregon, and there is a short portion that flows north and west into the Pacific Ocean, catching a view of Mount Saint Helens to the east as the Columbia traces the boundary between the states of Oregon and Washington. Prior to Portland, the Columbia has a long meandering path through Washington and the Canadian province of British Columbia. Moving against the current, one follows the river in an easterly direction until coming upon the Tri-Cities in Washington. Kennewick, Pasco, and Richland are the three towns at the confluence of the Columbia and the other two of its other major tributaries (Snake River and Yakima River). All proverbial roads of the Columbia move upwards in elevation, and the coherence of a singular river eventually dissipates. Where Yakima for example weaves back into the Cascade Range, there are many finger lakes serving as the upper riparian.

Snake River is the largest tributary of the Columbia, its starting line in Wyoming trekking a westward course just north of Bear River. The Snake River's headwaters are outflows of Yellowstone. The Columbia River and Snake River share a potentially overlooked but crucial feature from their convergence backwards: a vector of hydraulic projects along both of their paths, none more notable than the Grand Coulee Dam (GCD). GCD is a hydroelectric power generating dam, with a power generating capacity of 6,809 megawatts (MW)). GCD is the largest power station (not just hydro, but of any type) in the United States. Though, it is dwarfed in size to the Chinese Three Gorges Dam, which boasts an installed capacity of 22,500 MW.

The eponymous Columbia ultimately has its headwaters in British Columbia where it forms from a series of three finger lakes: the smaller Slocan Lake and the protracted-in-length Upper Arrow Lake and Kootenay Lake. The three direct suppliers of water are amongst the Columbia Range of mountains. Furthering the point,

there is also a Mount Columbia in the range.

3.4.3 Data Sources

This study compiles together several large portions of data from five different institutions. NASA provides the two land variables, surface and subsurface flow. These variables are derived by integrating several meteorological forcings into the Noah water balance algorithm. The Noah land surface model has a rich history dating back to the late 1980s with formal named implementations released in the middle 1990s ([11]). Surface and subsurface flow are produced measurements of unit kilograms per square meter and represent the spatial weight of water at a given time in each location ([52]). Noah factors known soil properties, solar radiation, precipitation, wind, pressure, humidity, and changes in the purses of snowpack (if applicable to the region).

The sea surface temperature product used is one of the monthly outputs from the MODIS instrument on the Aqua satellite. In this instance, the thermal longer wave infrared band observed by MODIS is selected ([37]). Furthermore, we use the data product that only looks at nighttime passes of the MODIS sensor. Daytime passes captured are ignored because of the potential diurnal warming effect that manifest in daytime observations. The skin of the ocean can become surficially inflated. Our preference is to a measurement of the ocean closer to the signature of the depths of the waters rather than the strength of the sun on a given day. Solar radiation is already encompassed in the calculation of land flows, so we attempt to limit its impact.

Gauged streamflow for both basins are obtained through the United States Geological Survey and Canada's bureau of climate and the environment ([47]; [10]). In all instances, we constrained ourselves to near continuous basin measurements over the entirety of the time series. The requirement of continuity (no non-number pixels

and preferably no interpolation between months) plays a part in the quantity of measurements per basin. The study length is 210 single month measurements, beginning with July 2002 (coinciding with the first available monthly measurement of MODIS) and ending with December 2019. Measurements of actual streamflow in cubic feet per second for each basin are provided in the Figure 3.2.

3.4.4 Preprocessing

We run two distinct workflows across the time series for the Columbia and the Yukon basin. Motivated by a desire to understand whether the coupling of monthly SST to the GLDAS measurements of water flux across land impacts the speed and ability of the neural networks to learn algorithms connecting the inputs and outputs (i/o), we create datasets that are comprised just of surface and subsurface flow, and those that combine time aligned SST with the surface and subsurface flow measurements.

The process of combining land and surface measurements is illustrated in Figure 3.3. In the top row, we have one measurement of surface (Q_s) and subsurface (Q_{sb}) flow. In the second row first column, we have added those two measurements together pixelwise. In the second column, second row, the combined Q_s and Q_{sb} are integrated with the corresponding monthly measurement of SST. In row two column two, having the single colormap representing all numbers provided causes a “washed out” effect where the Q_s and Q_{sb} measurements disappear because their physical range here are roughly between 0.0 and 1.0 kilograms per meter squared, as opposed to the wider 0.0 to 20.0°C range. Concurrently, row three captures the dynamic ranges of water flux over land, and the temperature of the ocean’s surface temperature.

In their respective raw formats, the land and ocean variables have different resolutions and different grids for each pixel of data. This fact makes assimilation of the data in its raw state impossible. The fix for this problem is bilinear regridding. We

retain the MODIS grid and resolution of approximately 9 km x 9 km per pixel. To facilitate regridding, we utilize xESMF, a high level python package hooked into the Fortran, C, and C++ based Earth System Modeling Framework ([55]). We input the source image (Qs and Qsb), the destination grid (SST), and the type of regridding (bilinear). The algorithm provides as an output our source image in the destination grid.

We also use a windowing technique to eliminate nan values from the MODIS SST images. The PyTorch software suite has the function to create tiles of uniform smaller images from a large input ([42]). We unfold each MODIS raster into smaller non-overlapping squares. Within each tile, we set the software to find any non-number values and replace them with the average of the available numerical values in the tile. One could simply take the mean of the entire global image and replace non-number values that way; however, this is prone to the creation of unusual bias due to the wide variance of sea surface temperature based on location of observation. By using this tile method, we limit potential bias through the consideration of pixels only that are in relative proximity to each other. After gap filling, we fold the tiles back up into the original image size.

Our control preprocessing in the study follows the same chain found in the first iteration of Flux to Flow. We clip the measurements of Qs and Qsb to the geographical constraints captured in the publicly available shapefiles for each basin. To do this, we involve a few different publicly available software packages. We open the shapefile (a digital vector storage format) with GeoPandas ([26]), convert the vector file to a GeoJSON file with shapely, and clip each measurement of surface and subsurface flow with rasterio ([18]) and xarray ([23]). As opposed to the coupled land and ocean images, we do not add the measurements of surface and subsurface flow together. We fill any non-number values with the mean of the available measurements.

3.4.5 Treatments

To investigate the capabilities and limitations of neural networks as they pertain to the calibration and prediction of streamflow from land and surface measurements of water, we run 1,920 experiments total, 960 for each basin. The 960 experiments are spread amongst five neural network configurations of increasing complexity. Of the 192 configurations per neural net, we run experiments splitting the datasets into training data and test data splits of 70% / 30%, 80% / 20%, 90% / 10%, and 95% / 5%. The more data that is in the training set, the higher likelihood of better performance in the testing regime. We have six different lag profiles (zero lag, one month lag, three month or “one season” lag, six month or “two season” lag, one year lag, and two years lag). We test between z-scoring and not z-scoring the input channels. Lastly, we have an option whether to shuffle or not shuffle the training data. Shuffling data is known empirically to improve statistical learning ([33]).

Four of five neural networks have a single hidden layer. Single hidden layer neural networks, with enough neurons in the hidden layer, are theorized to be capable of universal function approximation ([22]). Concurrently, we evaluate the feasibility of using a single hidden layer of differing sizes (4 neurons, 30 neurons, 200 neurons, and 1,000 neurons) to approximate the transformation algorithm between location specific water flux to flow measurements. To facilitate these networks, we reshape each input raster of shape height times width into a vector of one by (height times width). The vector passes through the hidden layer, receives signal conditioning via rectified linear units (ReLU), and outputs six or seven (based on the basin under observation) predictions for each of the streamflow gauges. Based on the difference between the predictions of streamflow output by the model and the actual measurements of streamflow according to the mean squared error loss function, backpropagation and adjustments of the neurons of the network are made using the Adam optimizer ([30]). Regardless of the neural network architecture, the learning process (forward propaga-

tion, the calculation of loss, backpropagation and updating of weights) is performed 100 times (otherwise known as 100 epochs) across the training dataset. Model testing simply makes a single pass through the unseen testing data (the most recent in time data, quantity dictated by the lag of the test case and the percentage of data devoted to testing).

The fifth and most complex network is a convolutional structure loosely based on a modern image classification architecture ([20]). Each of our input channels are forced into a convolutional space (sometimes referred to as the latent space) where the input images are considered in many dimensions by the machine. This function allows for automatic feature learning by the machine of potentially important structural details or patterns unique to the SST fields. To hedge against the potential for the machine to miss the mark, this latent space learning is containerized, added to the input that was used to feed latent space learning, and then fed into a conventional neural network structure that bottlenecks down the input images to 100 neurons, fifty neurons, twenty neurons, ten neurons, and finally the output size of either six or seven neurons. In contrast to the image classification network, our structure has a regressive target output. A regressive target output simply imputes that the value of streamflow can be any continuous value (within reason) greater than zero.

3.5 Results & Discussion

The 1,920 experiments were split into their respective basins, the Columbia and Yukon. Each set of 960 experiments ran on their own node comprised of at least one NVIDIA GeForce RTX 2080 Ti GPU and 64 GB of CPU RAM. Due to its slightly larger per image pixel size (200 x 800 for the SST enabled, 38 x 143 for Qs and Qsb only), the Yukon test set ran for approximately twenty eight hours while the Columbia basin (200 x 600 for SST enabled, 46 x 57 for Qs and Qsb only) tests lasted only 21.5 hours.

In developing the test set of experiments, we again encountered the implications of the “washed out” effect as seen in the preprocessing portion of this document. Upon feeding the single channel hybrid ocean temperature and land flux images into a standard z-score treatment of our network, we found that regardless of the complexity of the network and with no lag of the data, performance was very poor. A z-score is computed by mean centering the data under observation ($0.0 =$ mean of the data) and dividing by the standard deviation (z-scores between -1.0 and 1.0 are all of those within one standard deviation away from the mean). The values of Q_s and Q_{sb} have a relatively small range to that of SST. Therefore, the dynamics of the Q_s and Q_{sb} were overshadowed by the larger range of measured SST values. One can’t simply mix two physical parameters and then compute a z-score. They have two different dynamic ranges. Dynamic range is the general placement of a collection of values on the number scale, as well as and possibly more importantly the distance between their largest and smallest observed value. Our fix consisted of going back to the preprocessing step, computing the z-scores for SST and $Q_s + Q_{sb}$ individually, and then and only then merging them into a single time series image for each month.

In Figure 3.4 we present aggregate results of the testing data ran through their respective trained models. Ignoring all the other parameters (lag, quantities of training and test data, use or abstention from the use of z-scoring, shuffled or non-shuffled data), there is an evident distinction. The use of SST coupled with Q_s and Q_{sb} changes the performance of neural network streamflow prediction. Nash Sutcliffe efficiency, or NSE, is a metric frequently used to evaluate hydrological modeling efforts ([38]). A negative or near zero value implies a poor model prediction, whereas a value close to 1.0 indicates a strong penchant to accurately predict the actual streamflow. Calculations of NSE herein only consider testing data, not training data. To be clear, let’s consider an imaginary example of a different time scale. A time series of 100 days in a row split into an 80% / 20% train test split means that the model (here the

neural network) learns an algorithm from the first eighty days and then makes inferences about the latter twenty days. Factoring in the training data when computing NSE would give NSE a positive boost, but it wouldn't necessarily be considered fair or unbiased. One can make a very large neural network that simply memorizes the input data, which would give the impression of perfect performance. This is fine if that was the desired task. As a baseline or “mic check” of your system, it is a reasonable device; however, it would likely have little to no practical application beyond getting the metrology for your experiment dialed in. Therefore, we only consider test data.

In Figures 3.5 and 3.6, we disaggregate the results from Figure 3.4 into their respective individual neural network configuration. In the instance of the Columbia, for both non-SST and SST experiments, as the complexity of the neural network increases, the performance of the network with regards to the unseen test data measured by NSE increases. With SST present, performance climbs faster. In column three of Figure 3.5, the single hidden layer network comprised of 200 neurons, the non-SST experiments have a centered NSE value around -1.0, whereas the SST experiments caused a jump to a range between 0.35 and 0.90. The value of SST is even more opaque in the Yukon experiments. In Figure 3.6, the data shows that the simple change from four to thirty neurons combined with the SST-enabled data gives results of model predictability well over 0.0. The non-SST datasets don't see this performance until dcrrnn (pronounced discern and the shorthand name for our fine-tuned deep neural network) enters the proverbial training arena.

In Figures 3.7 and 3.8, we shed the non-SST experiments and smaller single hidden layer configurations; instead, we just focus on experiments with SST and one of two neural net architectures (dcrrnn or the 1,000 neuron single hidden layer neural network). Specifically, we filter and plot two small datasets: 1, a disaggregation of the remaining experiments per basin based on lag between input and output data; and 2,

a disaggregation based on the Boolean variable of z-scored inputs versus non-z-scored inputs.

Both figures indicate the same reality as in the original F2F document. Civil infrastructure creates its own set of nuance and challenges to the modeling realm. The Yukon River is a more continuously wild place. Handling the two different sets of signal chains is like handling the sound profile of a strictly raw acoustic guitar signal versus handling an electric guitar buried underneath a dozen different pedal effects. Though the acoustic player has bends, twists, turns, and hopefully nuance to their playing with time, there is a more finite (yet relatively natural) sonic realm emoting from the acoustic player's hands. The same might be said for the Yukon in its current state, and there is a certain beauty about the natural movements of water and sediment in an unconstrained basin such as the Yukon. It may experience times of drought or flood, but there is not yet any fashioned mass of concrete capable of holding back very high (relative to the baseflow of the river) quantities of the Yukon waters for long periods of time, and consequentially the hydrograph has little artificial complexity.

The same cannot be said for the Columbia River. Since greater control of Columbia's flows sit with single-point sources, and these are manmade devices, optimized performance of F2F in a basin such as the Columbia River basin would benefit from much more data at the inlet and outlet of each point source. A next step might be just considering a single basin (or single State) and accumulating all available hydraulic data. The EPA has its Enforcement and Compliance History Online (ECHO) data available as a representational state transfer (REST) software architecture. In another run of the land data assimilation system (i.e., a hypothetical NLDAS 3 or GLDAS 3), it would be of value to see the impacts of programmatically including a few hundred or thousand continuous point-source or ground truth measurement sources. Certainly, calibration of this many outputs would require either: 1, much

compute time on a single graphics processing unit; or 2, and probably more likely, the use of many GPUs in tandem.

Reference is made in the Materials & Methods to a preference of input images of the ocean associated with the deeper depths of the waters rather than the strength of the Sun on a given day. Our study certainly is self-limited in the depth field. The ocean at places is very, very deep. Soil has geospatial delicacies; the vadose zone is where science is witnessing the fallout from humanity’s continued mishandling of chemicals. These facets beget their own studies, which is outside the scope of this investigation. We acknowledge that although this study shows promising results regarding the use of advanced neural network technology in the forecasting of monthly streamflow out to a lag of two years, true temper of the devices will only come through testing other datasets, time scales, and locations.

Another water-focused avenue we see F2F moving towards is drought propagation. Historically, humanity’s records of drought monitoring are very good, due in part to various backwards looking methodologies. For example, a gridded, spatially interpolated dataset observing the standardized precipitation index (SPI) dating back to 1895 is available ([50]). The U.S. drought monitor time series dates to 2000 ([48]). There is even one digital product measuring the Palmer modified drought index at a 0.5 / 5x5 (degree / square kilometers) resolution dating back to the beginning of the common era, 0 CE, over 2000 years ago ([17]).

Drought is intimately linked to water scarcity, and this topic has never been more relevant. Water scarcity represents one of three things: 1 – the condition where no drinking water infrastructure exists; 2 – the condition where the drinking water infrastructure is inhumane; and 3 – the condition where available water resources are used unsustainably over a long period of time. In the modern scientific community, drought is science based, whereas water scarcity is rooted in policy, management, and justice as it relates to the global state of water infrastructure ([49]). These

differences notwithstanding, drought and water scarcity are symbiotic in their nature. The destruction of natural ecosystems catalyzes further drought conditions ([16]). As the dynamics of the global climate change Earth’s structure at the air, land, and oceanic interfaces, the occurrence and severity of drought conditions are becoming in some cases greatly exacerbated ([29]; [39]). Therefore, it is necessary to understand the propagation of drought in time as it starts in its most benign state, and how it can take a turn for the worse and become an extreme event.

There are more than fifty different indices related to drought. The National Drought Mitigation Center (NDMC) highlights five digitally, and six in the original software system companion paper ([48]). In all of them, keeping everything else equal, a decline in value equilibrates to a worsening drought condition. A simple model given these eight channels of data could be constructed by averaging the outputs at a given location and time, followed by a whitening technique such as z-scoring and segmentation of the result into bins yielding a single image representation. The timescale of the ‘percent of normal’ statistic changes in response to certain drought condition steps (e.g., from extreme to exceptional drought, the consideration of the condition switches from degree of specialness over a six month period to specialness over a twelve month period). Consideration of prior observations with thoughtful control of influence decay of said observation over time falls under the umbrella of drought propagation concerns. There are other open source software repositories that have already replicated many of the formal algorithms. One such algorithm of note is the Penman-Monteith equation of reference evapotranspiration (ETo). This algorithm was standardized by the American Society of Civil Engineers after benchmarking against a panel of twenty indices for reference evapotranspiration ([1]; [44]).

Drought propagation studies look at the teleconnections between meteorological, agricultural, hydrological, and socioeconomic drought. Many studies are classified under the drought propagation umbrella, such as those of historical nature that mon-

itor oceanic phenomena like the El Niño Southern Oscillation ([43]; [4]), or the use of global greening strategies as a mitigation effort ([41]; [14]). Studies of drought propagation cover the entire globe, from China ([19]; [24]; [54]) to South America ([4]; [5]; [9]), Central Asia ([21]), Africa ([12]; [51]; [15]) and the United States ([2]).

Figure 3.9 presents a sample output from the National Drought Mitigation Center. Their output has six seven different color based classification codes available to diagnose a region. Attempting the application of Flux to Flow to generate these images using the hybrid fields would only be a slight departure from the current body of experiments. Having such a finite quantity of outputs means that the loss function needn't be of the regressive variety but could take advantage of one of the many classification-based loss functions ([25]), or potentially of the burgeoning transformer-based modeling domain ([28]).

This experiment has several components that distinguish itself: 1, use of optimized computing performance and efficiency; 2, knowledge and selectivity amongst the richness of the geospatial data landscape; 3, consideration at several orders of magnitude of neural network complexity; 4, the openness of the technology; 5, consideration of many experiment samples to root out bias; and 7, use primarily of a collaborative green computing facility. The driving research questions behind this set of experiments were: 1, what impact does the application of sea surface temperature have upon the prediction of streamflow of large watershed basins; 2, how many overlapping, continuous, monthly measurements of gauged streamflow in two transboundary basins can we easily obtain, and what are the characteristics of these measurements; 3, how does Flux to Flow perform when using monthly datasets and greater than one but less than ten output targets? We find that sea surface temperature does have a positive impact upon neural network model predictability of streamflow. Each of these large basins was limited to less than ten. The measurements of discharge vary, indicating a wide dynamic range in the study of these basins. This is expected, as

a backyard with a brook is never far away, and wisely planned cities are commonly built around plentiful coiffeurs of surface water resources.

In the future, we may consider different ground truth data sources, particularly the EPA ECHO repository. We also see modern and future satellite missions on the horizon. Of note is the ECOSTRESS satellite and its signature evaporative stress index (ESI) product ([8]; [13]). Following on the success of NASA’s ECOSTRESS is a planned joint initiative between the Indian Space Research Organization and the French Space Agency, known as Thermal infraRed Imaging Satellite for High-resolution Natural resource Assessment or ‘TRISHNA’, and the Copernicus Land Surface Temperature Monitoring (LSTM) mission ([32]; [31]). Both are targeting the latter half of the 2020s for mission launch and will improve the temporospatial coverage of a similar product to ECOSTRESS, which has a resolution of seventy meters by 70 m x 70 m per pixel. This is more than 10,000 times finer resolution than the monthly MODIS product used within that has a pixel resolution of 9 km x 9 km per pixel. However, MODIS has a much more frequent revisit time, seeing the same location multiple times before ECOSTRESS does. This fact is part of the motivation behind such follow on missions. In order to study changes of the planet during the day at very fine resolutions from satellites that are orbiting around Earth, we will need multiples. This practice of redundancy has been already put to work in several satellite missions, such as the two MODIS instruments on the Aqua and Terra satellite platforms, the multiple GOES, Sentinel, Landsat, GRACE, and VIIRS missions.

We lastly see the potential for the integration of equations such as those used in the companion papers to NLDAS’s streamflow, the Saint-Venant equations ([53]; [46]). Physics-based machine learning is a field of its own. While we are pleased with the use of the convolutional, image focused networks, we see the benefits of more closely knitting historical hydrological advances into the bones of the neural network

and think it could help move all experiments to the highest echelon of performance according to NSE ([36]).

3.6 Conclusion

Herein, we investigated the use of neural network architectures and how they can be applied to river flow forecasting of two transboundary watersheds. Our inputs to the networks were derived fields of meteorologically forced surface and subsurface flow, and gauged streamflow data obtained from United States and Canadian bureaus associated with hydrological monitoring. Flow fields only show observations for the land, which can create issues and cause limitations to the quality of neural net model generation and performance outcomes. In response, we also fused time-concomitant sea surface temperature fields to the GLDAS observations of flow, finding a marked difference in the prediction of streamflow. To bolster this study, we considered two different basins of vastly different human imprints and found a clear watermark when the basin is sufficiently artificially modified. We see a continuation of this work by locals to other nations using their continuous gauged streamflow data, the fusion of higher resolution datasets, translation for use as a drought monitoring prediction system, or potential scaling of the system to consider many more streamflow sensors.

3.7 Code Availability

Scripts are available at https://github.com/albertlarson/f2f_holistic

3.8 References

- [1] Christine M Albano et al. “A Multidataset Assessment of Climatic Drivers and Uncertainties of Recent Trends in Evaporative Demand across the Continental United States”. In: *Journal of Hydrometeorology* 23.4 (2022), pp. 505–519.
- [2] Tushar Apurv and Ximing Cai. “Drought propagation in contiguous US watersheds: A process-based understanding of the role of climate and watershed properties”. In: *Water Resources Research* 56.9 (2020), e2020WR027755.
- [3] Angela H Arthington et al. “The Brisbane declaration and global action agenda on environmental flows (2018)”. In: *Frontiers in Environmental Science* 6 (2018), p. 45.
- [4] Erika Berenguer et al. “Tracking the impacts of El Niño drought and fire in human-modified Amazonian forests”. In: *Proceedings of the National Academy of Sciences* 118.30 (2021), e2019377118.
- [5] Alena G Bevacqua et al. “Spatial and temporal patterns of propagation from meteorological to hydrological droughts in Brazil”. In: *Journal of Hydrology* 603 (2021), p. 126902.
- [6] Timothy P Brabets, Bronwen Wang, and Robert H Meade. “Environmental and hydrologic overview of the Yukon River Basin, Alaska and Canada”. In: *Water-Resources Investigations Report* 99 (2000), p. 4204.
- [7] Stephen R Carpenter et al. “Nonpoint pollution of surface waters with phosphorus and nitrogen”. In: *Ecological applications* 8.3 (1998), pp. 559–568.

- [8] Kerry Cawse-Nicholson and Martha Anderson. “ECOSTRESS Level-4 DisALEXI-JPL Evaporative Stress Index (ECO4ESIALEXI) User Guide”. In: (2021).
- [9] Joel E Correia. “Between flood and drought: Environmental racism, settler waterscapes, and Indigenous water justice in South America’s Chaco”. In: *Annals of the American Association of Geographers* 112.7 (2022), pp. 1890–1910.
- [10] ECCC. *Real-Time Hydrometric Data - Water Level and Flow - Environment Canada* — [wateroffice.ec.gc.ca](https://wateroffice.ec.gc.ca/mainmenu/real_time_data_index_e.html). https://wateroffice.ec.gc.ca/mainmenu/real_time_data_index_e.html. [Accessed 10-Mar-2023]. 2023.
- [11] M.B. Ek et al. “Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model”. In: *Journal of Geophysical Research: Atmospheres* 108.D22 (2003).
- [12] Johan P Enqvist and Gina Ziervogel. “Water governance and justice in Cape Town: An overview”. In: *Wiley Interdisciplinary Reviews: Water* 6.4 (2019), e1354.
- [13] Joshua B Fisher et al. “ECOSTRESS: NASA’s next generation mission to measure evapotranspiration from the international space station”. In: *Water Resources Research* 56.4 (2020), e2019WR026058.
- [14] Njidda Mamadu Gadzama. “Attenuation of the effects of desertification through sustainable development of Great Green Wall in the Sahel of Africa”. In: *World Journal of Science, Technology and Sustainable Development* (2017).
- [15] Solomon H Gebrechorkos et al. “Variability and changes in hydrological drought in the Volta Basin, West Africa”. In: *Journal of Hydrology: Regional Studies* 42 (2022), p. 101143.
- [16] Augusto Getirana, Renata Libonati, and Marcio Cataldi. “Brazil is in water crisis—it needs a drought plan”. In: *Nature* 600.7888 (2021), pp. 218–220.

- [17] E.P. Gille et al. *NOAA/WDS Paleoclimatology - Living Blended Drought Atlas (LBDA) Version 2 - recalibrated reconstruction of United States Summer PMDI over the last 2000 years*. 2017. DOI: 10.25921/7XM8-JN36. URL: <https://www.ncei.noaa.gov/metadata/geoportal/rest/metadata/item/noaa-recon-22454/html>.
- [18] Sean Gillies. “Rasterio documentation”. In: *MapBox*, July 23 (2019).
- [19] Yi Guo et al. “Elucidating the effects of mega reservoir on watershed drought tolerance based on a drought propagation analytical method”. In: *Journal of Hydrology* 598 (2021), p. 125738.
- [20] Kaiming He et al. “Deep residual learning for image recognition”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.
- [21] Sarah Ho et al. “A new approach to quantify propagation time from meteorological to hydrological drought”. In: *Journal of Hydrology* 603 (2021), p. 127056.
- [22] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. “Multilayer feedforward networks are universal approximators”. In: *Neural networks* 2.5 (1989), pp. 359–366.
- [23] Stephan Hoyer and Joe Hamman. “xarray: ND labeled arrays and datasets in Python”. In: *Journal of Open Research Software* 5.1 (2017).
- [24] Shuzhe Huang et al. “Drought propagation modification after the construction of the Three Gorges Dam in the Yangtze River Basin”. In: *Journal of Hydrology* 603 (2021), p. 127138.
- [25] Katarzyna Janocha and Wojciech Marian Czarnecki. “On loss functions for deep neural networks in classification”. In: *arXiv preprint arXiv:1702.05659* (2017).
- [26] Kelsey Jordahl et al. “geopandas/geopandas: v0. 8.1”. In: *Zenodo* (2020).

- [27] David R Kanter et al. “Nitrogen pollution policy beyond the farm”. In: *Nature Food* 1.1 (2020), pp. 27–32.
- [28] Salman Khan et al. “Transformers in vision: A survey”. In: *ACM computing surveys (CSUR)* 54.10s (2022), pp. 1–41.
- [29] Andrew D King et al. “The role of climate variability in Australian drought”. In: *Nature Climate Change* 10.3 (2020), pp. 177–179.
- [30] Diederik P Kingma and Jimmy Ba. “Adam: A method for stochastic optimization”. In: *arXiv preprint arXiv:1412.6980* (2014).
- [31] Benjamin Koetz et al. “High spatio-temporal resolution land surface temperature mission-a copernicus candidate mission in support of agricultural monitoring”. In: *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*. IEEE. 2018, pp. 8160–8162.
- [32] J-P Lagouarde et al. “The Indian-French Trishna mission: Earth observation in the thermal infrared with high spatio-temporal resolution”. In: *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*. IEEE. 2018, pp. 4078–4081.
- [33] Kangwook Lee et al. “Speeding up distributed machine learning using codes”. In: *IEEE Transactions on Information Theory* 64.3 (2017), pp. 1514–1529.
- [34] Jessica I Lundin et al. “Legacy habitat contamination as a limiting factor for Chinook salmon recovery in the Willamette Basin, Oregon, USA”. In: *Plos one* 14.3 (2019), e0214399.
- [35] Emma Marris. *Meet the "Ice Bears" of the Yukon* — *sierraclub.org*. <https://www.sierraclub.org/sierra/2021-6-winter/feature/meet-ice-bears-yukon>. [Accessed 08-Mar-2023]. 2021.

- [36] Ben Moseley, Andrew Markham, and Tarje Nissen-Meyer. “Finite Basis Physics-Informed Neural Networks (FBPINNs): a scalable domain decomposition approach for solving differential equations”. In: *arXiv preprint arXiv:2107.07871* (2021).
- [37] NASA. *MODIS Aqua Level 3 SST Thermal IR Monthly 9km Nighttime V2019.0*. <https://cmr.earthdata.nasa.gov/virtual-directory/collections/C2036877952-POCLOUD/temporal>. [Accessed 10-Mar-2023]. 2023.
- [38] J Eamonn Nash and Jonh V Sutcliffe. “River flow forecasting through conceptual models part I—A discussion of principles”. In: *Journal of hydrology* 10.3 (1970), pp. 282–290.
- [39] Gustavo Naumann et al. “Increased economic drought impacts in Europe with anthropogenic warming”. In: *Nature Climate Change* 11.6 (2021), pp. 485–491.
- [40] Jonathan A O'Donnell et al. “Source water controls on the character and origin of dissolved organic matter in streams of the Yukon River basin, Alaska”. In: *Journal of Geophysical Research: Biogeosciences* 115.G3 (2010).
- [41] Farn Parungo et al. “Gobi dust storms and the Great Green Wall”. In: *Geophysical Research Letters* 21.11 (1994), pp. 999–1002.
- [42] Adam Paszke et al. “Pytorch: An imperative style, high-performance deep learning library”. In: *Advances in neural information processing systems* 32 (2019).
- [43] Thomas C Piechota and John A Dracup. “Drought and regional hydrologic variation in the United States: Associations with the El Niño-Southern Oscillation”. In: *Water Resources Research* 32.5 (1996), pp. 1359–1373.
- [44] M Richards. *PyETO*. <https://pyeto.readthedocs.io/en/latest/>. 2019.
- [45] Erika D Schreder and Mark J La Guardia. “Flame retardant transfers from US households (dust and laundry wastewater) to the aquatic environment”. In: *Environmental science & technology* 48.19 (2014), pp. 11575–11583.

- [46] Theodor Strelkoff. “Numerical solution of Saint-Venant equations”. In: *Journal of the Hydraulics division* 96.1 (1970), pp. 223–252.
- [47] USGS (US Geological Survey). “National water information system”. In: (2016).
- [48] Mark Svoboda et al. “The drought monitor”. In: *Bulletin of the American Meteorological Society* 83.8 (2002), pp. 1181–1190.
- [49] Anne F Van Loon and Henny AJ Van Lanen. “Making the distinction between water scarcity and drought using an observation-modeling framework”. In: *Water Resources Research* 49.3 (2013), pp. 1483–1502.
- [50] Russell S. Vose et al. *NOAA’s Climate Divisional Database (nCLIMDIV)*. 2014.
DOI: 10.7289/V5M32STR. URL: <https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.ncdc:C00005>.
- [51] Shirley Walters. “Learning about climate justice in times of drought and Covid-19”. In: *International Journal of Lifelong Education* 41.1 (2022), pp. 45–59.
- [52] Youlong Xia et al. “Comparative analysis of relationships between NLDAS-2 forcings and model outputs”. In: *Hydrological Processes* 26.3 (2012), pp. 467–474.
- [53] Youlong Xia et al. “Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products”. In: *Journal of Geophysical Research: Atmospheres* 117.D3 (2012).
- [54] Yu Zhang et al. “Agricultural drought prediction in China based on drought propagation and large-scale drivers”. In: *Agricultural Water Management* 255 (2021), p. 107028.
- [55] Jiawei Zhuang. “xESMF Documentation”. In: (2019).

3.9 Appendix

Figure 3.1: Extents of the Yukon and Columbia River watersheds and gauge locations.

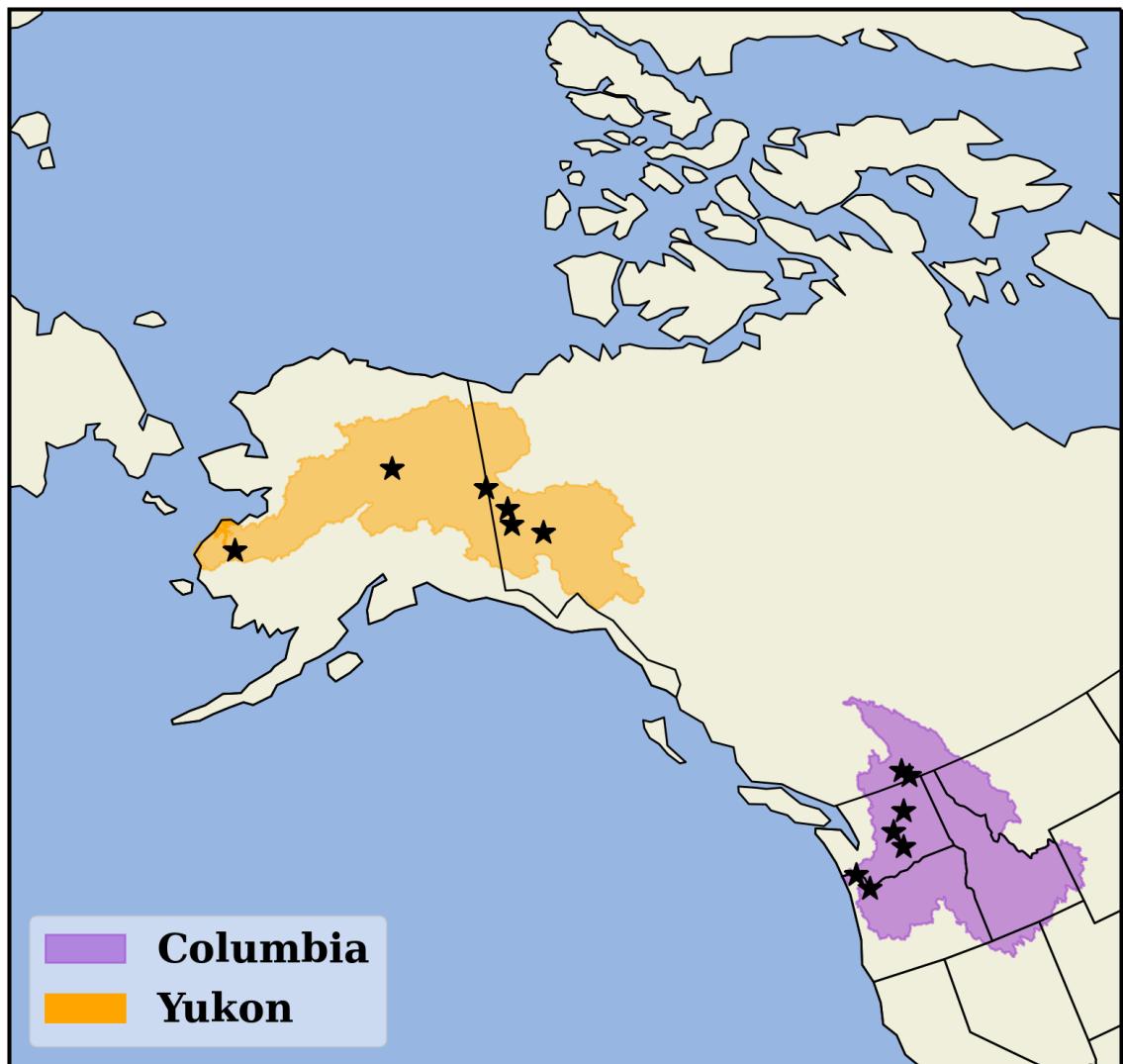


Figure 3.2: Hydrographs of all gauged streamflow data. Second and fourth plots are zoomed in versions of the same colors in the first and third plots. Both have low discharges whose details are lost amongst the larger portions of the river.

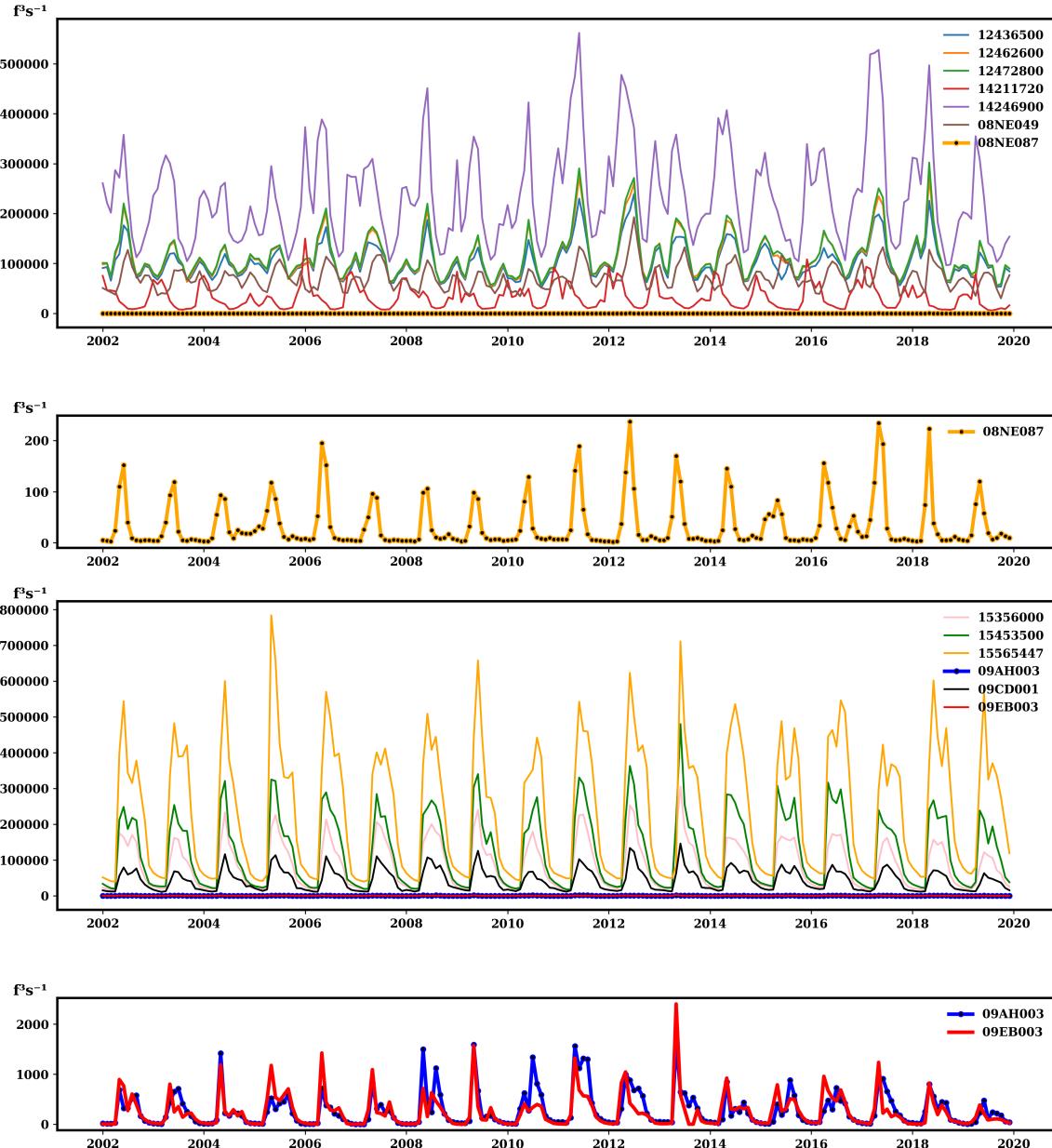


Figure 3.3: Sample observation of surface flow (Q_s), subsurface or groundwater flow (Q_{sb}), $Q_s + Q_{sb}$, flows merged with SST but represented with a single colormap, and flows merged with SST represented with two colormaps highlighting each physical parameter's unique dynamic range.

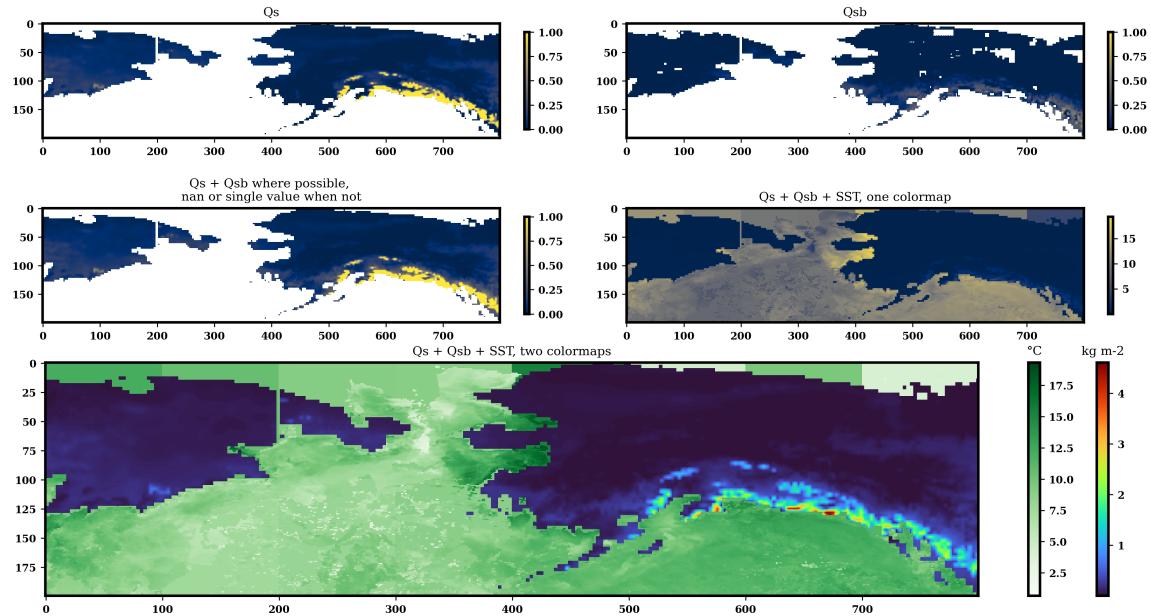


Figure 3.4: Histograms of model predictability across all experiments delineated by whether SST is included as part of the input or not.

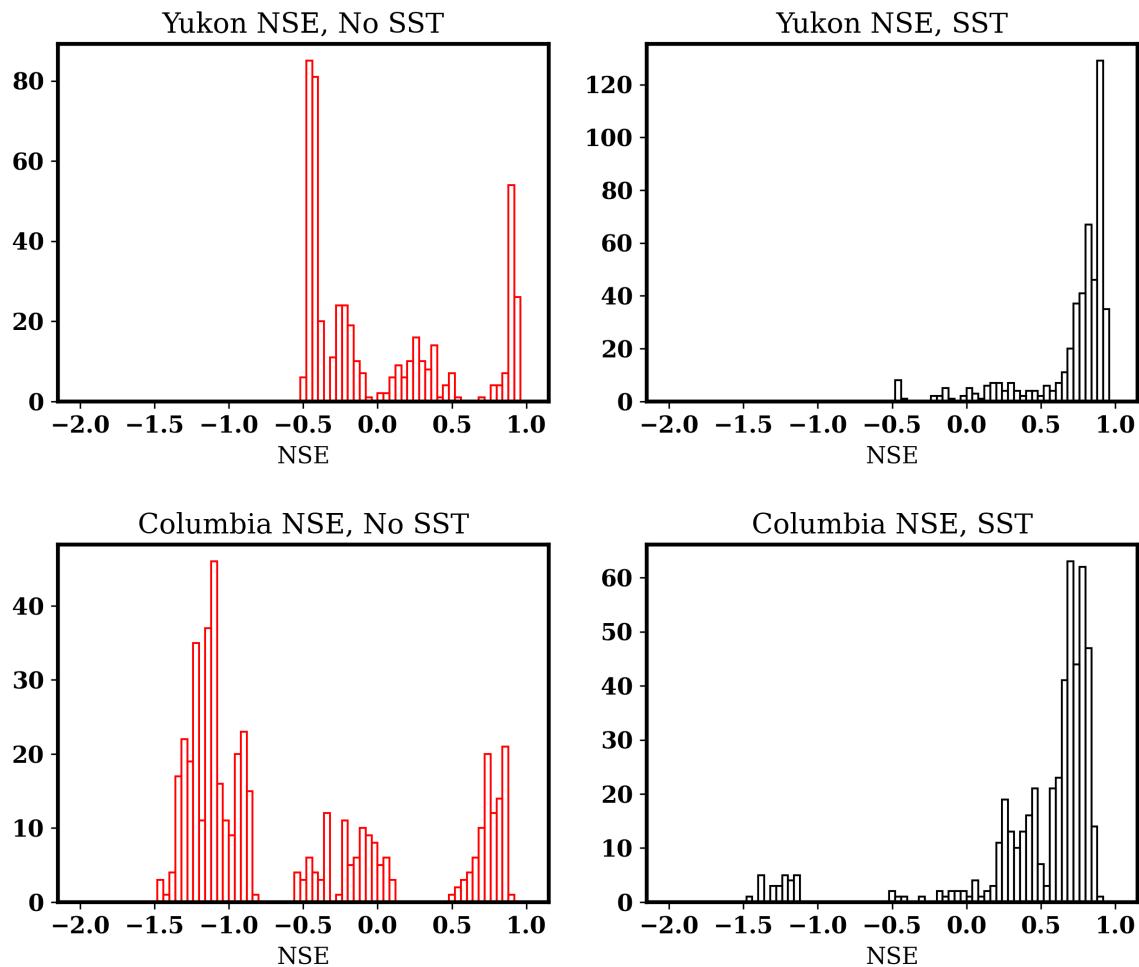


Figure 3.5: Histograms of test results for Columbia experiments deconstructed by neural network architecture at the time of training.

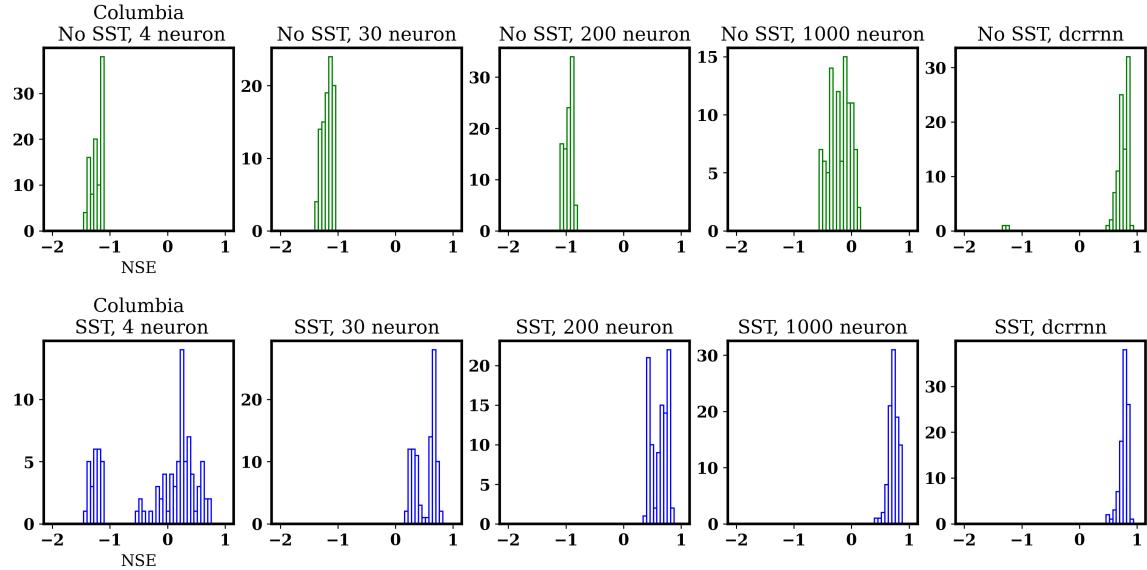


Figure 3.6: Histograms of test results for Yukon experiments deconstructed by neural network architecture at the time of training.

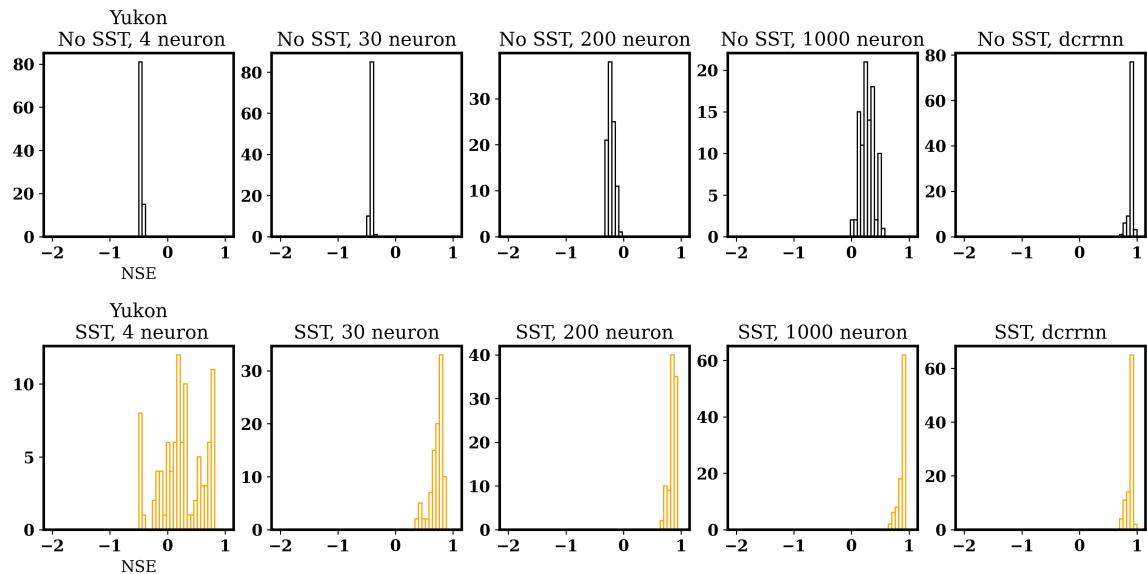


Figure 3.7: Columbia & Yukon experiments using dcrrnn and the 1,000 neuron single hidden layer neural networks, disaggregated by lag.

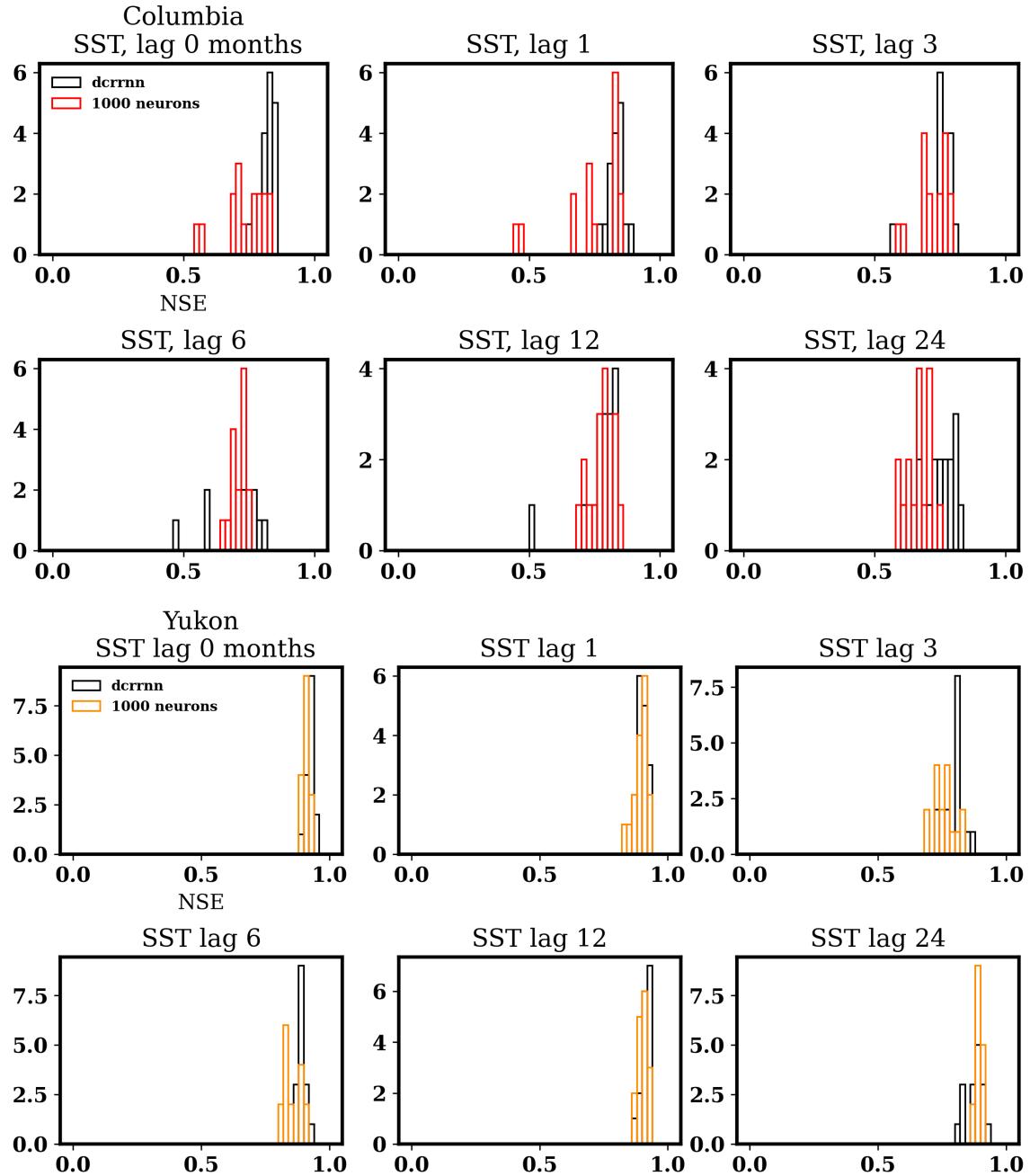


Figure 3.8: Columbia & Yukon experiments using dcrrnn and the 1,000 neuron single hidden layer neural networks, disaggregated by z-scored vs. non-z-scored.

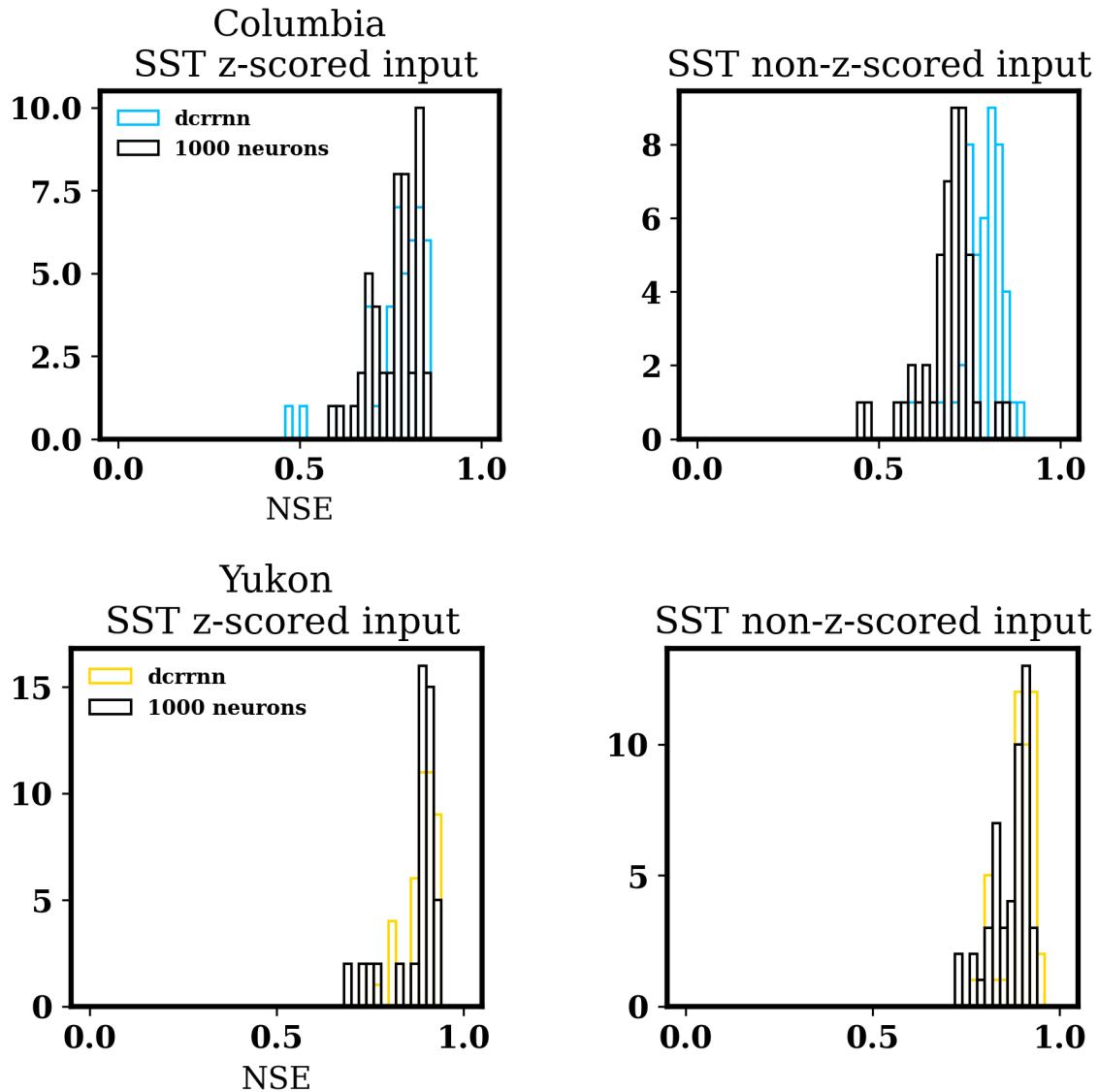


Figure 3.9: National Drought Mitigation Center Weekly Output, March 2, 2023, Drought Monitor Output.

