FINAL REPORT ~ FHWA-OK-22-06

FLOOD MONITORING AND PREDICTION SYSTEM DEVELOPMENT

Hazem Refai, Ph.D.
Abdullah Mousselli
Amer Zayegh, Ph.D. Candidate

School of Civil Engineering and Environmental Science (CEES)
The University of Oklahoma
Norman, Oklahoma

December 2022

OKLAHOMA
Transportation

The Oklahoma Department of Transportation (ODOT) ensures that no person or groups of persons shall, on the grounds of race, color, sex, religion, national origin, age, disability, retaliation or genetic information, be excluded from participation in, be denied the benefits of, or be otherwise subjected to discrimination under any and all programs, services, or activities administered by ODOT, its recipients, sub-recipients, and contractors. To request an accommodation please contact the ADA Coordinator at 405-521-4140 or the Oklahoma Relay Service at 1-800-722-0353. If you have any ADA or Title VI questions email ODOT-ada-titlevi@odot.org.

The University of Oklahoma herein certifies that, as of the date of submission to the Oklahoma Department of Transportation, the electronic version of this report meets all federal and state online accessibility requirements pertaining to:

United States Code Title 29, Section 794d vocational rehabilitation rights to access technology, and Oklahoma State Statutes Title 62, Section 34.20 duties relating to communications systems.

and is ready for posting online.

Flood Monitoring and Prediction System Development

FINAL REPORT ~ FHWA-OK-22-06

ODOT SPR2 ITEM NUMBER 2314

Submitted to:

Office of Research and Implementation

Oklahoma Transportation

Submitted by:

Hazem Refai, Ph.D.
Abdullah Mousselli
Amer Zayegh, Ph.D. Candidate
School of Electrical and Computer Engineering (ECE)
The University of Oklahoma

December 2022

TECHNICAL REPORT DOCUMENTATION PAGE

1. REPORT NO.	2. GOVERNMENT ACCESSION NO.	3. RECIPIENT'S CATALOG NO.
FHWA-OK-22-06		

4. TITLE AND SUBTITLE	5. REPORT DATE
Flooding Monitoring and Prediction System Development	Dec 2022
	6. PERFORMING ORGANIZATION CODE

7. AUTHOR(S): Hazem Refai, Ph.D., Abdullah Mousselli, and Amer Zayegh,	8. PERFORMING ORGANIZATION REPORT Click here to enter text.

9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. WORK UNIT NO.
The University of Oklahoma 660 Parrington Oval, Norman, OK. 73019	11. CONTRACT OR GRANT NO. ODOT SPR Item Number 2314

12. SPONSORING AGENCY NAME AND ADDRESS	13. TYPE OF REPORT AND PERIOD COVERED
Oklahoma Transportation Office of Research and Implementation 200 N.E. 21st Street, Rm. 3A7 Oklahoma City, OK 73105	Final Report Oct 2021 - Sep 2022
	14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

Click here to enter text.

16. ABSTRACT

The first half of this report details the design of a flooding emulation system to evaluate LiDAR-, radar-, and ultrasonic-based devices for monitoring water level rise and measure water flow speed. The second half of the report describes the incorporation of NOAA data with data collected at ODOT weather station to predict the potential of flooding.

17. KEY WORDS	18. DISTRIBUTION STATEMENT
Flooding, LiDAR, Radar, ultrasonic, weather station, rain.	No restrictions. This publication is available from the Office of Research and Implementation, Oklahoma DOT.

19. SECURITY CLASSIF. (OF THIS REPORT)	20. SECURITY CLASSIF. (OF THIS PAGE)	21. NO. OF PAGES	22. PRICE
Unclassified	Unclassified	53	N/A

Form DOT F 1700.7 (08/72)

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

SYMBOL	WHEN YOU KNOW	MULTIPLY BY	TO FIND	SYMBOL
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl	fluid ounces	29.57	milliliters	m
oz	gallons	3.785	liters	L L
gal	cubic feet	0.028	cubic meters	m^3
ft^3	cubic yards	0.765	cubic meters	m^3
vd	NOTE: volumes greater than 1000 L shall be shown in m^3			
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
°F	Fahrenheit	5 (F-32)/9 (F-32)/1.8	Celsius or	°C
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	$candela/m^2$	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

SYMBOL	WHEN YOU KNOW	MULTIPLY BY	TO FIND	SYMBOL
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
°C	Celsius	1.8C+32	Fahrenheit	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	$candela/m^2$	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380.
(Revised March 2003)

Contents

List of Figures

List of Tables

Chapter 1: Introduction

The need for a flood alerting and forecasting system is critical, especially given the effects of increasingly severe weather over the last ten years. Flash floods, particularly, can be deadly due to how quickly they develop and how difficult they are to predict. Consequently, floods are now recognized as the most fatal natural disaster in the United States. Floods occurring between 1980 and 2022 (as for October 11, 2022 are ranked second in cost for weather-related disasters in Oklahoma (see Figure 1), with around 2 to 5 billion dollars (NCEI.Monitoring.Info@noaa.gov, 2022).

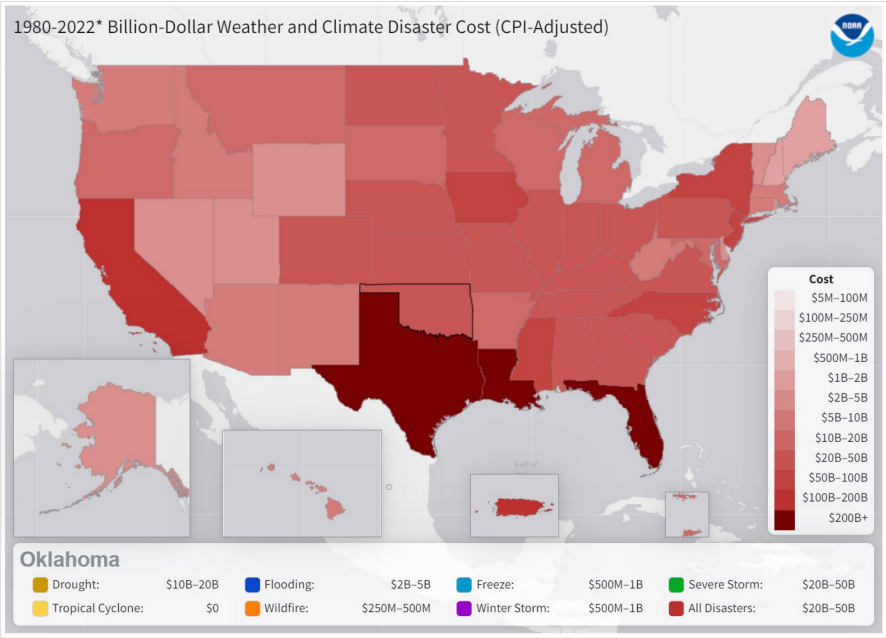

Figure 1. Cost of weather-related disasters in Oklahoma from 1980 through 2022.

Spring and early summer months of March through July have the highest occurrence of annual floods in Oklahoma. On April 17, 2013, [2] Oklahoma cities and towns of Medicine Park, Meers, Chickasha, and Newcastle experienced a single supercell storm that caused between four and seven inches of rain in only two hours and resulted in flash flooding. The month of December is least to have an annual flood. (Sauer, 1974, WRIR 73-52).

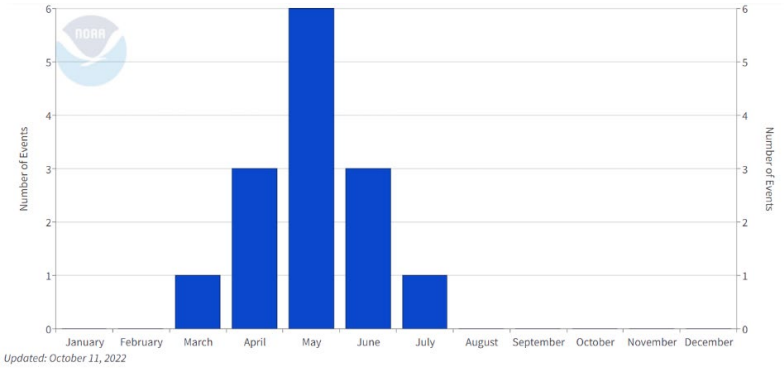

Figure 2. Annual flood event in Oklahoma.

Systematic monitoring of flood data (i.e., measurement of stage and discharge rate) in Oklahoma began between 1930 and 1940. Stations were installed to capture flow characteristics of streams (*Oklahoma Flood Watch active* 2022). It is important to know that a flash flood can develop in dry areas not associated with a body of water as a result of intense rainfalls. Regular monitoring stations have difficulty monitoring these conditions. Climate records show that Oklahoma's rainfall varies from year to year, characterized by seasonal differences. Wintertime precipitation tends to be widespread and stratiform in nature [1]; summertime precipitation is almost convective in nature, resulting from thunderstorms. Autumn and spring offer both stratiform and convective precipitation. In addition to seasonal rains, Oklahoma can experience storms that vary in severity, depending on rain duration and rate. Such storms could be classified into four major types:

1. Ordinary cell: forms from a single, small cell with high rain rate for short duration resulting in low chance of flood.
2. Multi-cell cluster: consists of multi storm clusters that can produce a significant amount of rainfall in small geographic areas
3. Multicell line: consists of a squall line and chance of extending for hundreds of miles, producing massive amounts of rainfall, and forming new cells that will likely cause flash floods. This type is the most common in Oklahoma.
4. Supercell thunderstorm: characterized by a single cell that can persist for hours despite its small size and produce extremely heavy precipitation with flash floods

This report evaluates market-available, industrial-based flood monitoring and alerting systems developed and deployed by several private companies and federal agencies. These technologies include weather-based flood prediction techniques and currently utilized sensors to measure the level and flow of water bodies. Data-based machine learning (ML) algorithms used weather parameters from 15 Road Weather Information Systems (RWIS) deployed on road-bridge cross intersections along Oklahoma highway I-35. Data was processed and analyzed during, before, and after historical flash flood events

reported by the Oklahoma Department of Transportation (ODOT) and the National Weather Service (NWS). Water level- and velocity data was collected to enhance the short-term flood prediction model and extend the data-driven forecast. An open channel testbed has been designed and fabricated with help of the operation team at OU-Tulsa campus to facilitate experimental scenarios of water flow and level changes using multiple contactless candidates. The objective was to improve the accuracy and suitability of flood prediction and monitoring systems. Integrating water stage and flow data resulted in a more efficient and generalized model for detecting all types of flash flood systems.

Chapter 2: Related Work

This section features ongoing research and development of flood monitoring and alerting technologies, especially contactless water level and water-surface velocity measuring techniques and ML-based flood prediction models. Researchers in [3] built a risk assessment model using a Random Forest (RF) ML approach for determining each factor's contribution to a flood occurrence. Results showed that three-day precipitation, distance to the river, and stream power index had the highest risk scores based on the model. The proposed model [3] generated a prediction of flood event with ~71% accuracy. Researchers from United Kingdom utilized cumulative and rainfall intensity data in [4] to forecast the severity of flood occurrence. Data were classified into three categories, namely normal, abnormal, and severe. An RF classifier model outperformed other ML models with an AUC score of approximately 85%. A literature review of ML applications designed for short-term and long-term flood predictions was conducted in [5]. Results in the journal article highlight methods like Artificial Neural Networks (ANN), RF, Support Vector Machine (SVM), and many other single and hybrid ML methods used to build real-time, short-term (i.e., hourly), and long-term (i.e., monthly, seasonally) flood prediction models. Models presented in [6] utilize factors like precipitation rate, drop size, river water level (i.e., stage), and streamflow (i.e., surface velocity or stream flow index). Precipitation rate and drop size are crucial for predictions, as they affect rising water levels. In [6], an acoustic Dendrometer was used to measure drop size distribution. Sensor light rays determine rain drop size and speed. A series of signal conditioning and piezoelectric sensors were placed under the surface of a Zinc slab to prevent direct contact with water and to increase surface area for data collection. Surface diameter was 13.7 cm with a surface area of 294,6733 cm^2. Notably, the sensor can read from 0.25 m to 6.5 mm. Rainfall was divided into two groups: 1) stratiform and 2) convective rain. The stratiform travels less than 25 mm/h and lasts more than an hour, while convective rain experiences speeds above 25 mm/h with a duration of just a few minutes. Rain that falls to the ground is defined with intensity R, which is the number of points collected with diameter D within a one-hour interval. Total precipitation is dependent upon the speed and direction of raindrops on the surface of earth. Rise in water level during rain varies depending on conditions. In one instance, there was a rise in river level of 0.01 meters per hour in stratiform rain and 0.4 meters per hour in convective rain.

Monitoring water level (i.e., stage) at creeks and rivers is crucial for early detection of floods; therefore, a continuous monitoring of surface level water is required. This can be accomplished using a variety of methods. Pressure sensors, ultrasonic sensors, radar sensors, optical sensors, light detection, and ranging (LIDAR) sensors are used most often. These methods provide accurate and reliable data for water level measurements. Using contact sensors (e.g., pressure sensors and pressure transducers) for water level measurements [7], [8] has major drawbacks, as these sensors are unable to independently measure static

and dynamic pressures. In fact, measurements would be affected by sensor orientation with respect to water flow, as well as debris, rocks, and extreme weather conditions. It is important to know that sensors also require periodic maintenance and testing. Consequently, non-contact measuring technologies prove to be more efficient and reliable. Among these technologies, ultrasonic rangefinders, optical sensors, and LIDARs are most common. Researchers in [9] demonstrated the development of an early flood detection system using ultrasonic and infrared sensors. The system consisted of six passive infrared (PIR) sensors and one ultrasonic sensor connected to a microcontroller platform. It was installed on a pole at the edge of a bridge, looking down at both creek and road directions. All six infrared sensors were utilized to measure both ground and actual temperature. The ultrasonic sensor was used to measure distance to objects below. Both ground temperature and water level measurements aided in building an ANN model for predicting flood. Multiple neural network architectures were evaluated. Results demonstrated that air temperature variance affects the speed of sound, which, in turn, affects ultrasonic sensor measurements. Estimating measurement corrections using collected ground temperature led to accurate ANN models. Researchers concluded that ANN models proved to be more accurate than nonlinear regressions, fuzzy-based models, and decision-trees. Optical water level measurement is a relatively new method for flood detection and monitoring—one which has been investigated more widely in recent years. The study in [10] utilized enhanced water level detection based on the edge detection principle for developing a flood detection system. Day and night cameras, an infrared projector, white board, processor unit, data logger, and data transmission unit were utilized. The developed optical gauge was assessed at several sites under different conditions (e.g., rain and heavy tropical rain events, snow, fog, daytime, and nighttime). Results were compared with other conventional sensors, such as radar and pressure sensors. Accuracy was approximately 95% with 1 cm margin of error for water depth of 75 cm. Researchers in [11] utilized a scene-text recognition approach wherein the system consisted of four main components, namely water level riser, graduated flood gauge, camera, and microcontroller processing unit. Camera images were first preprocessed by reducing image size to 100 KB. Next, scaled images were processed using computer vision (CV) and image processing techniques, including background extraction, template matching, and region of interest trimming. After determining the region of interest, dominant color detection and a digit recognition neural network were leveraged to measure water level change. The approach performed extremely well in daytime and showed only a few limitations during nighttime, primarily due to lack of riser visibility at night. A similar approach was proposed for real-time flood detection in [12] using video surveillance. After the frame was acquired, background subtraction was applied. Due to the nature of flash floods, flooded water might contain debris and small objects (e.g., trees and rocks), which can be subtracted from the foreground using morphological closing techniques. During the next stage, an eight-directional, connected-component labeling algorithm was used to determine volume size of a body of

water to estimate and categorize the flood. Results showed that flash floods rarely have a laminar flow. Given the development of time-of-flight distance sensors, non-water parts make distinguishing flash floods much easier. The use of LIDAR sensors for distance measurement has become more common due to its low-cost, high-energy efficiency, and small measurement footprint. LIDAR-based distance sensing is common for digital terrain mapping [13]. The principle of LIDAR ranging relies on the rugosity of the reflective surface for generating an ideal reflection of the incipient laser beam. As with radar, LIDAR ranging measures time-of-flight, albeit using higher frequency waves for greater pulse intensities. Near-infrared (NIR) light is most often used for this purpose. Wavelengths of 900–1,100 nm (270–330 THz) are common, due to the low cost of lasers operating in this wavelength range and the lower energy density than typical for visible spectrum [14]. LIDAR distance sensing accuracy is determined by leveraging two characteristics: 1) pulse width and 2) method accuracy for measuring flight time. LIDAR has rarely been used to measure water levels. A primary difference between terrestrial LIDAR applications and water-level LIDAR applications is the lower reflectivity of water surfaces for infrared radiation compared to solid objects. A calm water surface may behave as a purely specular surface, which produces reflections instead of dispersing incipient energy. In practice, a mirror-like river surface will reflect ∼2–3% of an infrared radiation incident beam [15], while the remainder is transmitted through the water. It is important to note that the reflectivity of water changes with an increase in water rugosity.

Although researchers have rigorously pursued non-contact water level measuring techniques, attention on non-contact surface velocity measuring techniques is scarce and lacking. A concept suggested by [16] and tested in July 2020 used Doppler LIDAR sensors to measure water velocity. Experiments were performed at two river sites. Obtained results were highly accurate and showed a residual velocity standard deviation of 0.07 m/s. The proposed system was used to monitor surface velocities during flood events, and the measurements provided by the doppler LIDAR were later used to estimate water discharge (i.e., flow rate).

Chapter 3: Industrial Flood Monitoring and Alerting Systems

In this section, the main industrial flood monitoring and alerting systems currently deployed in the USA are compared to identify the major differences among them. Valarm [18], the flood alert system widely deployed in the East Coast, offers industrial Internet of Things (IoT) monitoring techniques, including water resources management, pollution and air quality monitoring, and fluid and water usage. Valarm's IoT water level system is composed of three sensors to monitor water level (i.e., ultrasonic level sensors, pressure transducers, and RADAR level sensors). Ultrasonic sensors are critical, depending on monitoring scenarios deployed. Pressure sensors measure the weight pressure of water above them. Radar sensors use electromagnetic waves for monitoring. This type of sensors is more expensive, as its radiation can penetrate objects that might interfere with level measurements. Figure 3 illustrates the Valarm system and software interface.

The U.S. Geological Survey (USGS) and the National Weather Service (NWS) jointly maintain a flood warning system [19] that issues four categories of alerts:

1. Flood Watch: a flood is anticipated within 12-48 hours.
2. Flood Warning: flooding is expected across a large region or is actively taking place.
3. Flash Flood Watch and Warning: potential for flooding due to heavy rain or dam failure; and
4. Flood Statement: flooding is expected along major streams where there is no threat to people or property.

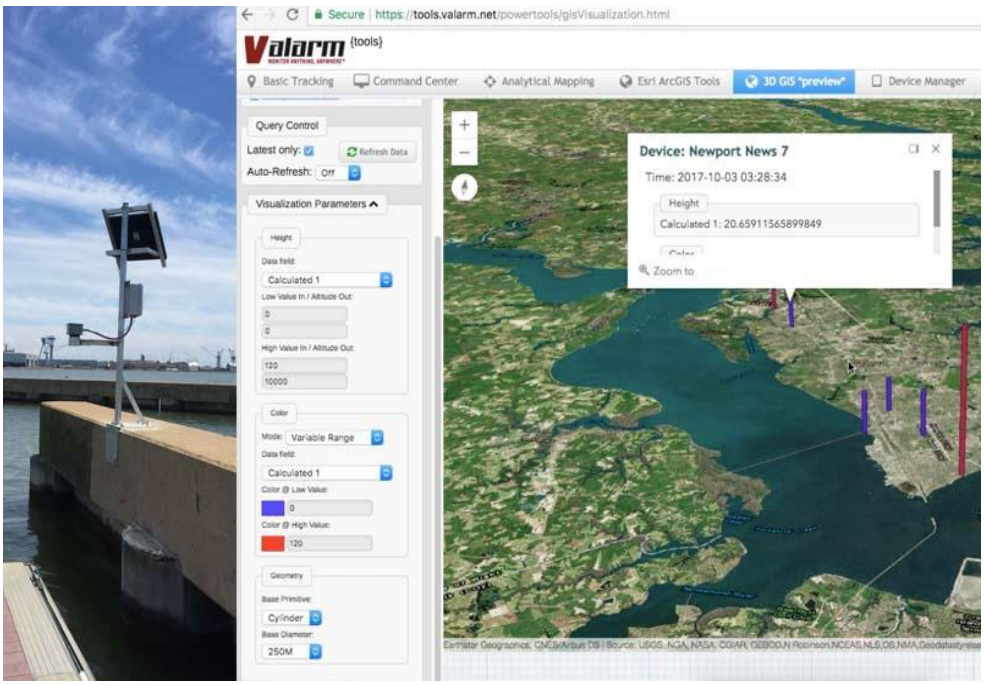

Figure 3. Valarm flood monitoring and alerting system.

This flood warning system is based on data from local rainfall, stream level, and streamflow; it requires three steps, including data collection via gaging, data processing, and dissemination of flood warning information. Alert gages perform both sensing and communicating tasks. Sensors are employed to detect changes in precipitation and water levels that affect parameters. Notably, system activity can vary widely from one location to another. Sensors and instruments used, as well as number of gage sites and their locations, are based on the nature of application and the size of the coverage area. The station is typically installed on riverbanks or a standing structure, such as a pier or bridge. Gages can also be built into standpipes, as shown in Figure 4.

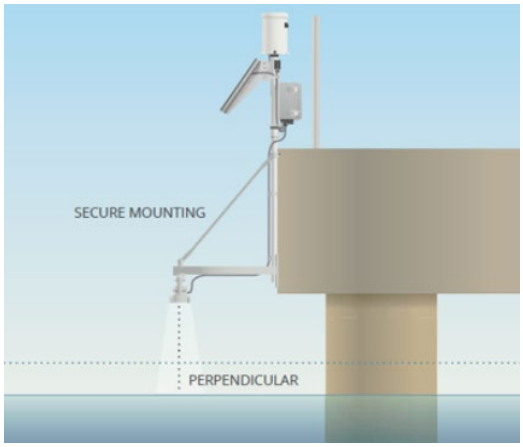

Figure 4. USGS and NWS flood monitoring and alerting system.

Each station has a tipping bucket rain gauge that collects precipitation. The data logging system houses the data logger, telemetry module, and power supply. The telemetry module provides access data in real time, transmitting 'ALERT' via radio frequencies. Alert and measured parameters are transmitted via email when parameters exceed predefined limits. The system provides an advanced and powerful real time monitoring system using advanced products selected for quality, reliability, and value. Each station is equipped with four devices:

1. RLS Radar Water Level Sensor from OTT, which uses radar pulse technology to measure area depth.
2. HSA TB3 tipping bucket rain gauge, which offers precipitation data with less than 3% margin of error.
3. NexSens 3100-MAST wireless telemetry system, which contains the data logging system, cellular modem, and solar charging.
4. WQData LIVE web datacenter, which supports remote access to collected data from any device.

The "Water Detection Warning System" by ELTEC [20], shown in Figure 5, detects wet, slippery, and/or flooded roads. The system utilizes a bronze-housed sensor positioned at the desired highway detection point. When water touches the steel electrodes, an alert signal is sent and a warning light flashes

continuously. The system sends an alert to report a potential flood to the central office. For wet road detection, an embedded sensor warns drivers on a slippery road to slow down prior to reaching the affected area.

Figure 5. ELTEC flood road monitoring system.

YSI, a Xylem brand technology, developed a flood alerting system which is currently deployed by major U.S. government environment monitoring agencies, including USGS and the United States Army Corps of Engineers (USACE). The YSI flood alert system includes a radar level and velocity sensor rain gauge, as well as a data acquisition and transmission cabin.

Figure 6 depicts YSI system's four main components, including:

1. Submersible vented pressure sensor (e.g., Dwyer SBLT2-10-40) for monitoring water level;
2. Tipping bucket rain gauge for measuring rainfall [H-3401] (SDI-12 output data parameters represent accumulated rainfall since last measurement, raw bucket count since last measurement, total accumulation, and total daily accumulation);
3. Side looking radar water level and velocity sensor (e.g., SonTek-SL3000) to determine water discharge (e.g., measure water velocity change, while water level remains fixed); and
4. System cabin for data acquisition and transmission.

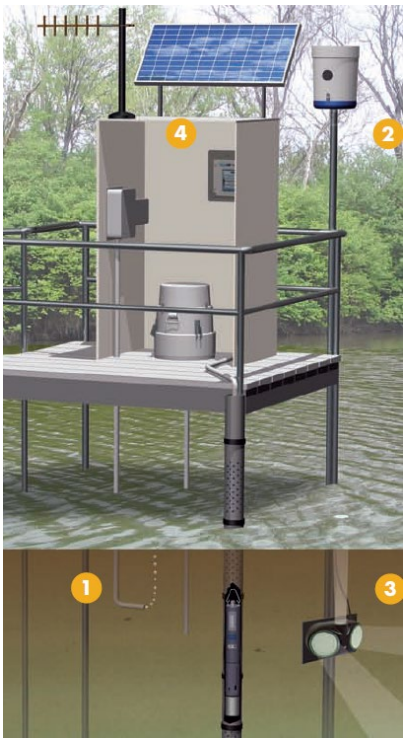

Figure 6. YSI flood monitoring and alerting system.

YSI's flood alert system provides multiple data acquisition methods, as well as accurate early detection for flood events. Despite these advantages, the system lacks the ability to monitor road surfaces, as it is deployed within bodies of water. It also requires continuous maintenance. Deploying the system in water makes it vulnerable to extreme weather and environmental conditions, as well as confounding elements like rocks and debris, which are likely during flash flood events.

Chapter 4: Design and Development of an Open-Channel Water Flow Emulator

Weather conditions are not the only factor in determining the occurrence of flood events. Terrain of the water body and the hydrological-topographical characteristics of the area play a vital role in determining a flood probability for a given area, as well [21]. Thus, obtaining field measurements is essential. Experiments conducted under such conditions are, however, often difficult to replicate due to several uncontrollable factors (e.g., background flows and surface topography). Having the capability to perform and repeat hydrodynamic experiments in a controlled setup is essential for accurately evaluating water level and water-surface velocity measuring technologies under study. Before implementing an experimental testbed, it is particularly important to understand the characteristics of the water surface profiles in open channel flows to emulate a natural body of water flow. A study in [22] classifies water surface profiles based on two criteria: Hydraulic slop and Hydraulic Curve. Water surface profile can then be classified into various categories. A USGS study analyzed the hydrologic and geospatial characteristics of the lower part of the Arkansas river. Findings indicate that in high-energy depositional environments (e.g., stream and river channels; creeks) the riverbed tends to have a mild slope, which means that the surface water flow is not affected by the underwater flow. This study indicated that we can build a water tank with equal depth (i.e., horizontal slope) and still emulate a high-energy depositional environments water surface flow. Figure 7 shows a 3D illustration of the water tank utilized in the emulation environment at OU-Tulsa. Dimensions of 1.7 m (70 inches) in length, 1.6 m (62 inches) in depth, and 1m (40 inches) in width were selected to meet sensor specifications under evaluation (e.g., field of view and minimum water depth).

The emulator has one inlet and four outlets at different levels. A height adjustable arm is attached to the tank where sensors are placed. The emulator tank has a capacity of 2700 L (700 gallons), has been fabricated with the help of the OU-Tulsa operation team. Three-quarter-inch plywood was used in favor of other available building materials (e.g., plexiglass) due to its durability and cost. The interior surface of the tank was covered with a water-proof epoxy layer to prevent leaking, and the exterior surface was painted to prevent degradation and erosion.

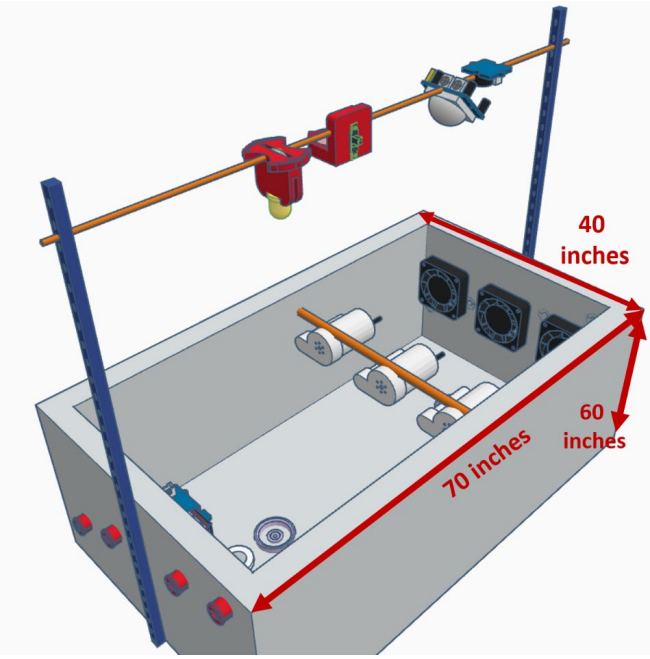

Figure 7. Emulator setup and dimensions.

To ensure sensors will be deployed in a body of water for field application, mounting was extended using an adjustable pole for evaluating the candidate over higher levels from the water surface that are similar to actual heights utilized by ODOT (see Figure 8).

Figure 8. Water-side setup using adjustable pole.

To produce an accurate evaluation, a realistic environmental setup is necessary. Propeller water pumps were utilized to emulate currents in high-energy reefs. Various modes were generated by changing the frequency of the pump to emulate crashing waves and tidal surges. Figure 9 explains various types of waves and their corresponding frequencies.

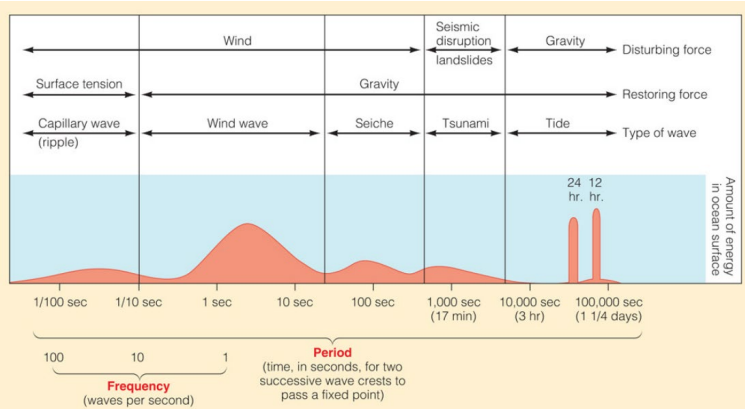

Figure 9. Water wave types.

Three Aquality LBS-50 [24] propeller pumps were installed to generate ample energy and turbulence for creating water conditions typical of a river channel or a creek. The pumps use an adjustable sine wave for fine-tuning. Sensors were combined using a wired primary/secondary device communication protocol. An accurate ground truth sensing mechanism was established for comparing accuracy of the various non-contact water levels and surface velocity technologies under evaluation. Contact sensors, such as pressure sensors and pressure transducers, proved most accurate for measuring water levels under lab experimental setup [7], [8]. Given significant drawbacks of this system under real-world conditions (e.g., over sensitivity to their orientation with respect to water flow; disturbance by debris, rocks, and extreme weather conditions; and necessary periodic maintenance and calibration), the experimental setup provided the most accurate results when compared to other traditional measuring techniques and technologies. The Blue Robotics Bar02 ultra high-resolution depth pressure sensor [25] shown in Figure 10 was calibrated using a ruler. It continuously measures and logs water level data. Sensor communicates over I2C with 3.3 V logic level.

Figure 10. Bar02 depth pressure sensor.

The Digiten G ¾ Hall effect sensor [26] (see Figure 11) continuously logged water flow rate data. Three flow meters were used to measure the average water velocity at the point of interest, while water flows in and out through the inlet and outlet. Water current drives the wheel with attached magnet to turn. Subsequently, the magnetic field rotation triggers Hall effect which generates high- and low-level square waves.

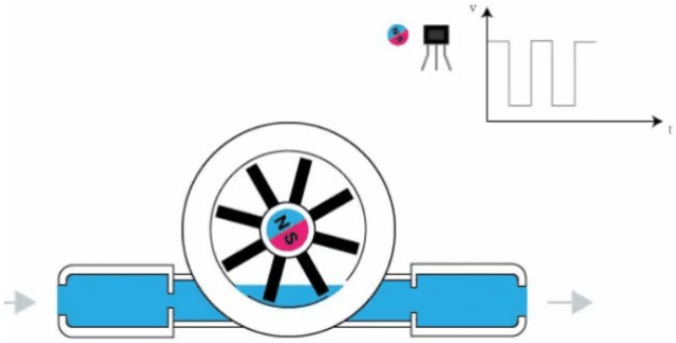

Figure 11. Hall sensor working principle.

Figure 12 illustrates the ground truth sensors setup in the emulator tank.

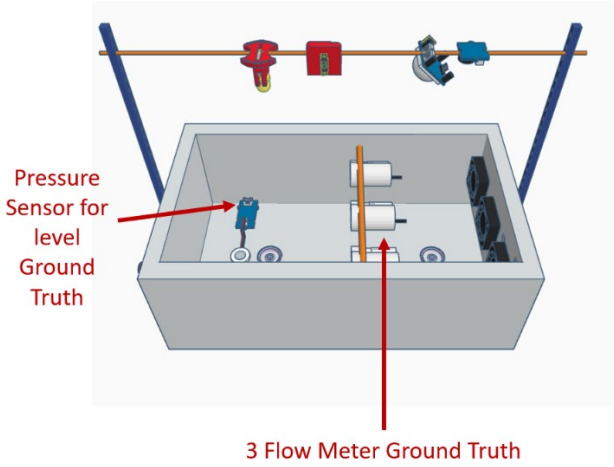

Figure 12. Ground truth setup.

Flow meters were calibrated via a bucket and stopwatch method. Logging flow data from more than one point at the same time (e.g., inlet, outlet, and area of interest flow rate) is difficult, as there is no internal clock that can be forcibly synchronized through an external datalogger. However, synchronization is necessary, as flow rate constantly changes in response to variable waveforms that generate high-energy flow form, as discussed above. General Purpose Input/Output (GPIO) controller interface was used as interrupt. In this way, code forces an external interrupt on the GPIO pins connected to the sensors whenever a detected pulse triggers the pulse counter following an interrupt (i.e., measuring period). The number of pulses is used to calculate flow rate.

It is important to differentiate between the water velocity V, measured in m/s, and the flow rate Q, measured in L/min. Velocity V is equal to flow rate Q divided by cross-sectional area A. For ease of calculation, we typically assume that the area of interest is circle shaped, as shown in Figure 13. Hence, the sensor footprint for most non-contact water velocity sensing technologies is circular or square shaped.

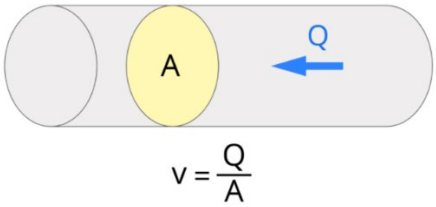

Figure 13. Water velocity calculation.

4.1. Water Level Measuring Technologies

Three kinds of water-level sensors, namely ultrasonic, RADAR and LiDAR, were evaluated. This section details specifications for each evaluated technology, along with experimental setups and results. Various experimental scenarios were established by changing parameters, such as distance between sensors and water surface (i.e., sensor height); rate of water level change; and characteristics of changing the water surface and flow.

NovaLynx Ultrasonic sensor [27] has been utilized for level measurements. This sensor communicates over RS232 8N1 ASCII protocol with a sensing range between 0.5 m up to 10 m. Because the measurement cycle is designed to be triggered by setting the enable status to high, the enable pin has been connected directly to the power source to activate continuous measuring. Each measuring cycle lasts between three and five seconds, depending on the number of attempts for successful measurement. Figure 14 illustrates the bottom view of the sensor.

The sensor generates four values at its output, as follows. First is air temperature indicated in Kelvin. Ultrasonic sensor accuracy is affected by air temperature. Accordingly, the sensor uses temperature to accommodate distance error measurements based on an offset factor calculated internally. Second is measurement cycle duration in milliseconds, and third is distance to the target (e.g., water surface).

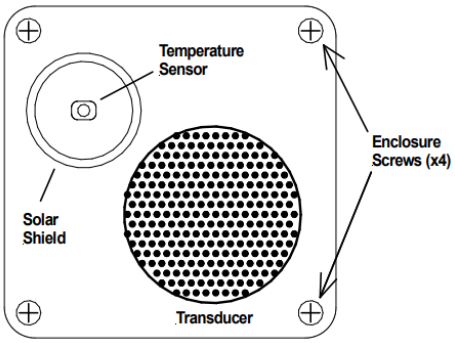

Figure 14. Bottom view of NovaLynx ultrasonic water level sensor.

This value is calculated using the following equation:

$$Q = S + (F * (K - 273))$$

Q: Speed of sound after factoring out the temperature effect.

S: Speed of sound in dry air at 20 c.

F: Constant (0.607).

K: Temperature in Kelvin.

Distance to target is then calculated based on Time of Flight (ToF), using following equation:

$$D = T * Q$$

where T is the time of travel. The fourth value the sensor outputs is number of measurements to produce a successful measurement. To assure measurement accuracy, the sensor was mounted vertically to position the beam perpendicular to the water surface.

The second water level measuring technology evaluated was the OTT RADAR Level Sensor (RLS) [28]. This sensor communicates over SDI-12 ASCII protocol and can also interface through RS485 8N ASCII protocol. Measurement range is between 0.4 m and 35 m. The RLS sensor can generate various parameters at its output, based on application and user requirements. The sensor has two modes of measurements. The first returns the target level measured to a reference level (e.g., water level related to a level zero). The second mode, which was used in our experiments, measures the distance between the sensor and the water surface.

The RLS sensor returns the following parameters:

- Time, in seconds, the sensor required to obtain measurements
- Level / distance value (i.e., outputs +99999999, given an invalid measurement).
- Status of the measurement:
    - 0: acceptable measurement
    - 2: no target detected
    - 4: internal error
    - 8: high variance of individual measurements
    - 16: SDI-12 interface interruption
- SNR value in dB (noting SNR > 15 indicates good signal quality with well-chosen mounting location and alignment to the water surface). Figure 15 depicts the RLS.

Figure 15. OTT RADAR level sensor (RLS).

The third sensor utilized for level measurements was TruSense S330 LiDAR, which is a highly sensitive precision laser sensor used to measure liquid surfaces within 0.46 cm to 50 m. This sensor utilizes infrared laser light of 905nm wavelength emitted from the transmitter lens. The reflected beam from the liquid surface returns to the receiver lens, and then exact distance is calculated by comparing return time to the speed-of-light constant (i.e., Time of Flight). The ability of a laser sensor to measure a liquid surface depends on the liquid's reflectance and angle of incident. Hence, the LiDAR sensor was mounted to ensure the beam remained perpendicular to the water surface. Optical diffusing lenses were installed to improve results under emulator experimental conditions (i.e., range to target less than 10 meters with turbulent water surface during measurement). Figure 16 shows the sensor setup and measurement principle.

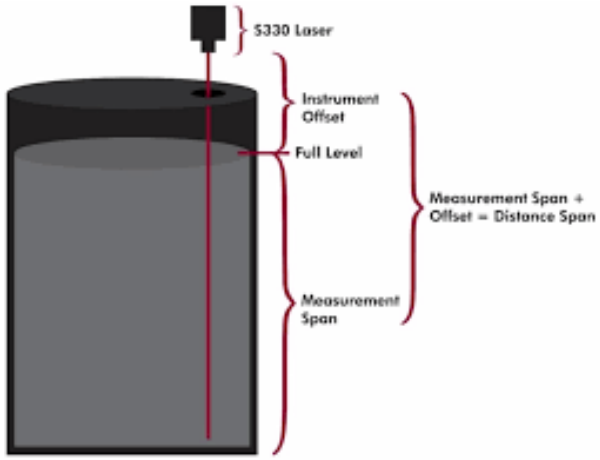

Figure 16. TruSense S330 LiDAR sensor and measurement principle.

The S330 sensor communicates to the data logger over RS-232 ASCII protocol. The sensor can return different values at its output, based on the configuration set by the user:

- Distance to the water surface
- Time stamp since powered on
- Power return, which could be an indicator of good positioning and sensor alignment over the water surface.

Despite the sensor's simplicity and accuracy, S330 LiDAR is extremely sensitive to optical interference and temperature. Thus, the manufacturer recommends installing the sensor in an enclosed environment, like a pipe, when utilized in outdoor applications. Doing so will reduce interference of various light resources (e.g., sun, streetlights, and the like) and avoid thermal shutdown of the sensor due to temperature rise.

4.2. Level Measurements Experimental Scenarios and Evaluation Results

To evaluate sensor accuracy and efficiency, various experiments were conducted by changing setup parameters, such as:

- distance between sensors and water surface,
- water level change, and
- changing water surface and flow characteristics.

Accuracy of several non-contact candidates were compared using a Mean Absolute Error (MAE) score against ground truth measurements. Each experiment was repeated three times to ensure repeatability of the established scenario.

In the first set of experiments, distance between the water surface and the mounted sensors was altered. Water surface was smooth with no inlet, outlet, or propellers turned on. This set aimed to evaluate candidate performance over different heights against specifications stated by manufacturers. Table 1 presents MAE for each water level measuring technology at different distances between mounted sensors and water surface. Results of these experiments indicated that ultrasonic technology was more accurate when the sensor is mounted near to water surface (e.g., less than 1.5 m).

Table 1
MAE of Level Sensors

Distance between sensor and water surface (cm)	*MAE Ultrasonic*	*MAE Radar*	*MAE LiDAR*
512 cm	2.54 cm	1.27 cm	1.72 cm
470 cm	2.37 cm	1.22 cm	1.51 cm
415 cm	2.29 cm	1.17 cm	1.43 cm
352 cm	2.21 cm	1.1 cm	1.31 cm
210 cm	2.15 cm	1.07 cm	1.21 cm
190 cm	1.92 cm	1.05 cm	1.04 cm
171 cm	1.45 cm	0.97 cm	0.93 cm
149 cm	0.27 cm	0.81 cm	0.82 cm
123 cm	0.27 cm	0.81 cm	0.81 cm

By contrast, RADAR and LiDAR had higher accuracy and more stability when mounted at a higher distance from water surface. Results confirmed manufacturer specifications listed by each sensor.

For the second set of experiments, water level was changed through inlet (i.e., ascending) and outlet (i.e., descending). Water level and water velocity measurements were recorded.

Table 2
MAE of Level Sensors (Ascending Water Level)

Candidates Height	*MAE Radar Level*	*MAE LiDAR Level*	*MAE Ultrasonic Level*
210 cm	*1.07 cm*	*1.21 cm*	*2.3 cm*
352 cm	*1.13 cm*	*1.41 cm*	*2.9 cm*
415 cm	*1.54 cm*	*1.73 cm*	*3.21 cm*
470 cm	*1.78 cm*	*1.95 cm*	*3.44 cm*
512 cm	*1.96 cm*	*2.11 cm*	*3.62 cm*

Table 2, Figure 17, and Figure 18 confirm that both RADAR and LiDAR achieved better MAE score than the ultrasonic sensor, which had high differences of measurements when compared to ground truth.

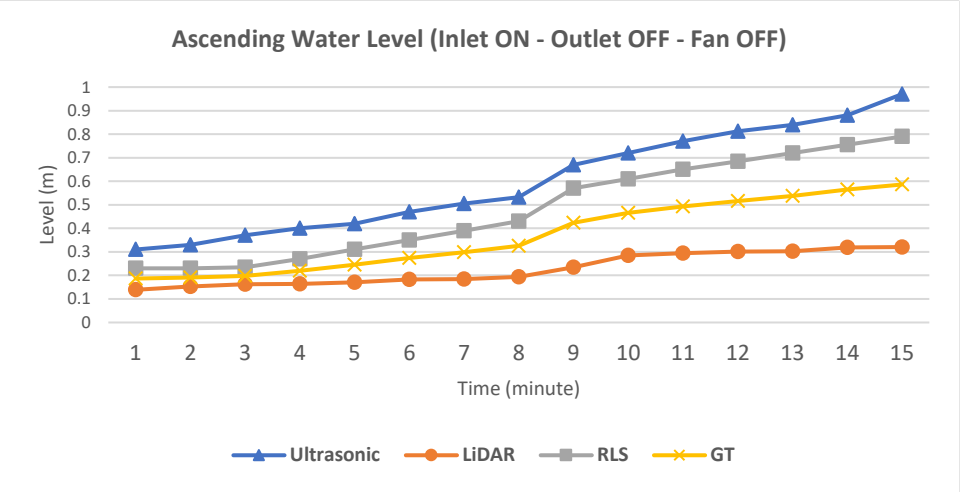

Figure 17. Water level measured by candidates vs. ground truth.

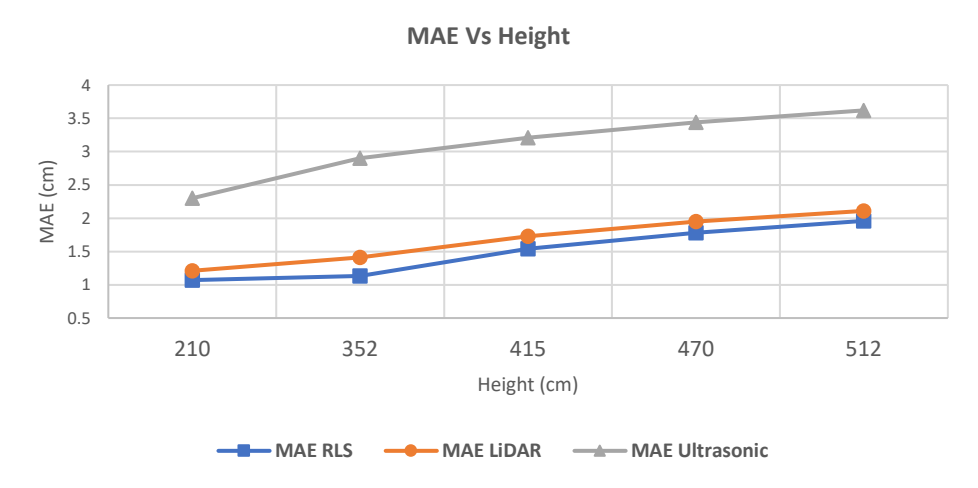

Figure 18. MAE score vs. height of candidates.

Similarly, experiments were conducted to evaluate the non-contact technologies when water level decreased (i.e. descending). The experiment used two different configurations—the first with both inlet and outlet turned on, wherein inlet flow rate was less than outlet flow rate as water level was decrementing.

Table 3
MAE of Level Sensors (Descending Water Level – Inlet ON, Outlet ON)

Candidates Height	*MAE Radar Level*	*MAE LiDAR Level*	*MAE Ultrasonic Level*
10 cm	1.98 cm	2.27 cm	2.95 cm
352 cm	2.32 cm	2.97 cm	3.64 cm
415 cm	2.76 cm	3.24 cm	4.21 cm
470 cm	2.98 cm	3.66 cm	4.87 cm
512 cm	3.06 cm	3.89 cm	5.32 cm

Table 3, Figure 19 and Figure 20 demonstrate results for the first configuration. Like the previous experiment, RADAR and LiDAR outperformed the ultrasonic sensor when compared to ground truth.

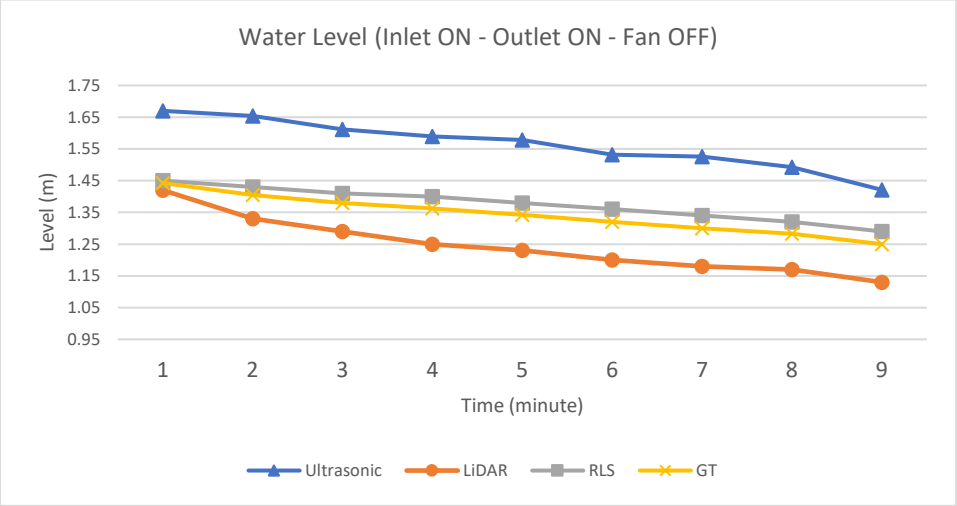

Figure 19. Water level measured by candidate s vs. ground truth.

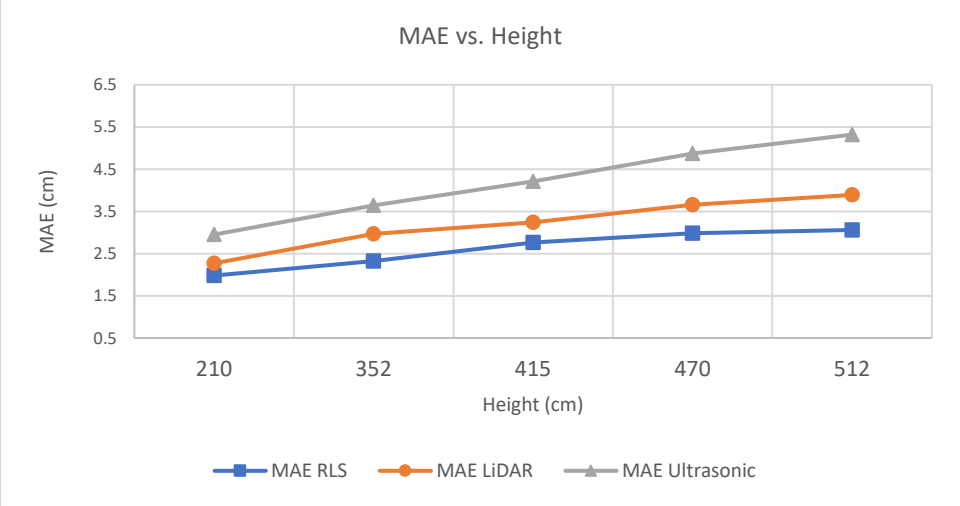

Figure 20. MAE score vs. candidates' height.

In the second configuration the water level was decreased as the inlet was turned off and outlet was on, which permitted the emulator to discharge water. Table 4, Figure 21, and Figure 22 confirm results obtained in previous experiments, wherein the ultrasonic sensor had lower performance than RADAR and LiDAR.

Table 4

MAE of Level Sensors (Descending Water Level – Inlet OFF, Outlet ON)

Candidates Height	*MAE Radar Level*	*MAE LiDAR Level*	*MAE Ultrasonic*
210 cm	1.02 cm	1.24 cm	1.53 cm
352 cm	1.21 cm	1.4 cm	1.78 cm
415 cm	1.44 cm	1.65 cm	1.98 cm
470 cm	1.71 cm	1.89 cm	2.34 cm

Candidates Height	*MAE Radar Level*	*MAE LiDAR Level*	*MAE Ultrasonic*
512 cm	1.99 cm	2.12 cm	2.67 cm

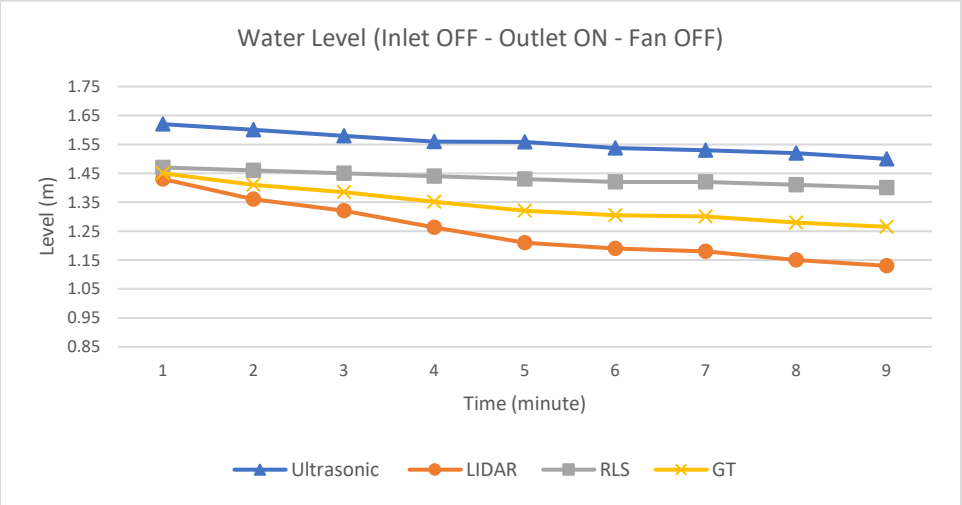

Figure 21. Water level measured by candidates vs. ground truth.

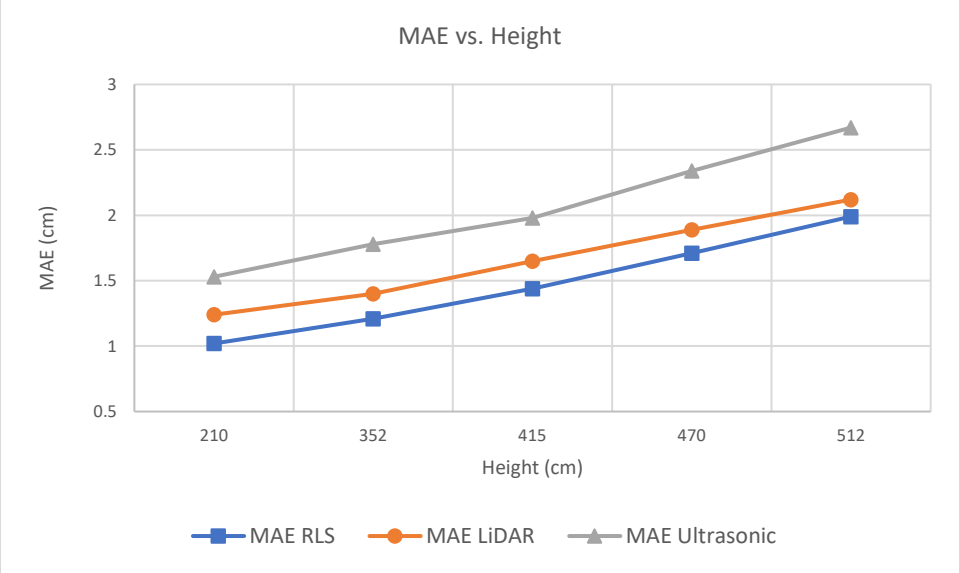

Figure 22. MAE score vs. candidates' height.

The next set of experiments evaluated non-contact candidate performance under different water surface profiles. Three water propellers generated surface turbulence in three modes:

- High energy surface profile: creek surface flow under rough weather conditions.
- Low energy surface profile: creek surface flow under normal weather conditions.
- Tidal surface profile: water surface flow of a river.

Figures 23, 24, and 25 show the water surface of high energy and tidal profiles, respectively. Tables 5, 6, and 7 demonstrate MAE scores for low energy water, high energy, and tidal surface profiles, respectively. For this set of experiments, both inlet and outlet were turned off.

Figure 23. Low energy surface profile.

Table 5

MAE of Level Sensors (Low Energy Surface Profile – Inlet OFF, Outlet OFF)

Candidates Height	*MAE Radar Level*	*MAE LiDAR Level*	*MAE Ultrasonic*
210 cm	0.48 cm	0.57 cm	0.89 cm
352 cm	0.66 cm	0.71 cm	1.02 cm
415 cm	0.93 cm	0.99 cm	1.87 cm
470 cm	1.03 cm	1.11 cm	2.11 cm
512 cm	1.15 cm	1.48 cm	2.43 cm

Figure 24. High energy surface profile.

Table 6

MAE of Level Sensors (High Energy Surface Profile – Inlet OFF, Outlet OFF)

Candidates Height	*MAE Radar Level*	*MAE LiDAR Level*	*MAE Ultrasonic*
210 cm	0.81 cm	0.87 cm	1.12 cm
352 cm	0.98 cm	1.2 cm	1.87 cm
415 cm	1.12 cm	1.7 cm	2.23 cm
470 cm	1.58 cm	2.43 cm	2.84 cm
512 cm	2.11 cm	3.21 cm	3.07 cm

Figure 25. Tidal (pulse) surface profile.

Table 7

MAE of Level Sensors (Tidal (Pulse) Surface Profile – Inlet OFF, Outlet OFF)

Candidates Height	*MAE Radar Level*	*MAE LiDAR Level*	*MAE Ultrasonic*
210 cm	0.63 cm	0.7 cm	0.96 cm
352 cm	0.75 cm	1.1 cm	1.44 cm
415 cm	0.81 cm	1.32 cm	1.73 cm
470 cm	0.97 cm	1.87 cm	2.32 cm
512 cm	1.22 cm	2.13 cm	2.56 cm

MAE scores of the candidates under evaluation highlight that disturbing the water surface reduced the accuracy of the ultrasonic sensor, however, the turbulence did not have a significant effect on RADAR and LiDAR accuracy.

4.3. Water Surface Velocity Measuring Technologies

OTT Surface Velocity RADAR (SVR) [29] was utilized for water surface velocity measurements. This sensor interfaced with the controller (data logger) over SDI-12 ASCII protocol. Sensor footprint ranged between 0.5 m and 35 m. The sensor can be mounted at a nominal vertical angle of 30 or 45 degrees. A 30-degree angle was used in our experiments to ensure that sensor beam is directed toward the point of interest. Alignment must be parallel and, ideally, against the main flow direction. Figure 26 illustrates SVR mounting setup.

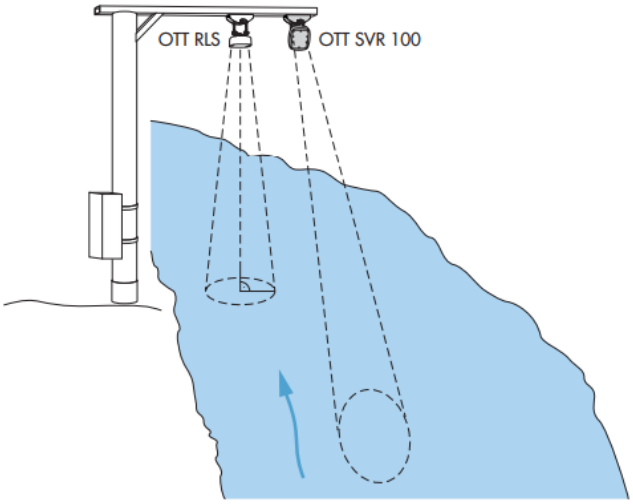

Figure 26. SVR and RLS mounting setup.

The functional principle of the sensor was based on Doppler effect. A transmitter emitted electromagnetic (EM) pulses with a frequency of 24.4 GHz. Given that the surface water is moving and has a minimum roughness, EM pulses will be reflected with a slight frequency shift, and then collected by the receiver antenna. The sensor uses an algorithm to calculate average flow velocity based on the difference between radiated and received frequencies.

The following equation is used to calculate the filtered speed.

$$v(f) = v(t) * Q + v(t-1) * (1-Q)$$

where V(f) is filtered velocity, v(t) is current velocity, and Q is constant (0.3).

Sensor sensitivity could be configured within the range between 10 and 46. The higher the sensitivity, the more accurate the measurements when the water surface is smooth. Flow direction filter can be set towards the sensor, away from the sensor, or deactivated.

SVR sensor produces the following parameters at its output: average flow velocity; current flow velocity; sensor tilt angle in degrees; signal quality index—a value between 0 and 3, where 0 represents an excellent signal and 3 represents a poor signal; vibration index—a value between 0 and 3, where 0 indicates no

device vibration and 3 indicates significant device vibration; and SNR value defined as the logarithmic ratio of the average power of the signal to the average noise power of the interfering signal.

4.4. Water Surface Velocity Measurements Experimental Scenarios and Evaluation Results

Water surface velocity measurements have been conducted along with level measurements; thus the experimental scenarios for surface velocity sensor are identical to level sensors. Figure 27 shows the surface water velocity for the first set of experiments (i.e., ascending water level [Inlet ON – Outlet OFF – Fan OFF]) against flow ground truth.

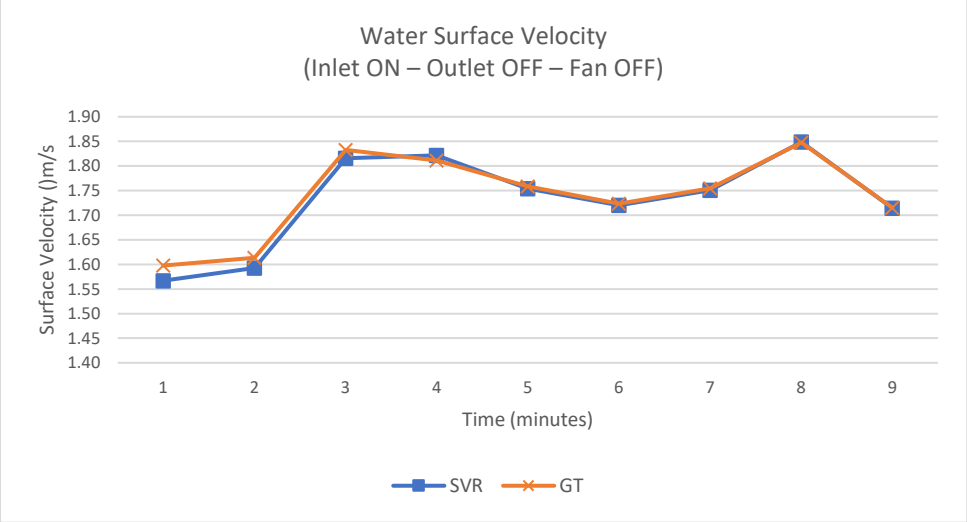

Figure 27. Water surface velocity vs. flow ground truth (ascending water level).

Table 8 displays MAE score of SVR sensor against the flow ground truth.

Table 8

SVR MAE Score vs Ground Truth (Ascending Water Level – Inlet ON, Outlet OFF, Fan OFF)

Candidates Height	MAE Radar (SVR)
210 cm	*0.016 m/s*
352 cm	*0.018 m/s*
415 cm	*0.022 m/s*
470 cm	*0.029 m/s*
512 cm	*0.038 m/s*

Table 9, Table 10, Figure 28 and Figure 29 show surface water velocity for the second set of experiments (e.g., descending water level) against flow ground truth. As aforementioned, this set of experiments includes two types of configurations:

- Inlet ON – Outlet ON – Fan OFF (Table 9 and Figure 28) and

- Inlet OFF – Outlet ON – Fan OFF (Table 10 and Figure 29).

Table 9

MAE of SVR vs. Ground Truth (Descending Water Level – Inlet ON, Outlet ON, Fan OFF)

Candidates Height	*MAE Radar (SVR)*
210 cm	0.009 m/s
352 cm	0.011 m/s
415 cm	0.017 m/s
470 cm	0.024 m/s
512 cm	0.032 m/s

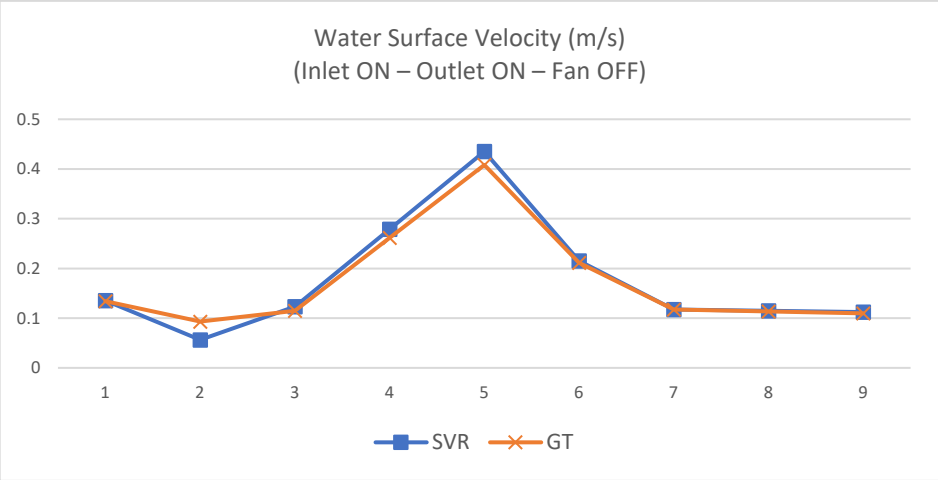

Figure 28. Water surface velocity vs. flow ground truth (descending water level – Inlet ON, Outlet ON, Fan OFF).

Table 10

MAE of SVR vs. Ground Truth (Descending Water Level – Inlet OFF, Outlet ON, Fan OFF)

Candidates Height	*MAE Radar (SVR)*
210 cm	0.08 m/s
352 cm	0.08 m/s
415 cm	0.087 m/s
470 cm	0.091 m/s
512 cm	0.097 m/s

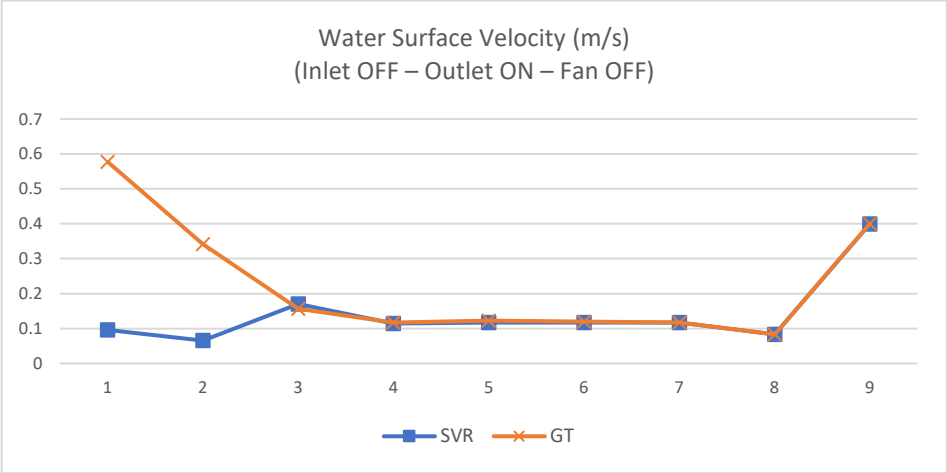

Figure 29. Water surface velocity vs. flow ground truth (descending water level – Inlet OFF, Outlet ON, Fan OFF).

A subsequent set of experiments were conducted under turbulent water surface conditions to evaluate SVR sensor performance under different flow conditions. Turbulence modes were identical to those generated for level measurement candidates. Tables 11, 12, and 13 demonstrate MAE score for low energy water, high energy, and tidal surface profiles, respectively. For this set of experiments, both inlet and outlet were turned off.

Table 11

MAE of SVR Sensor (Low Energy Surface Profile – Inlet OFF, Outlet OFF)

Candidates Height	MAE Radar (SVR)
210 cm	*0.8 m/s*
352 cm	*0.78 m/s*

Candidates Height	MAE Radar (SVR)
415 cm	*0.83 m/s*
470 cm	*0.91 m/s*
512 cm	*0.97 m/s*

Table 12

MAE of SVR Sensor (High Energy Surface Profile – Inlet OFF, Outlet OFF)

Candidates Height	MAE Radar (SVR)
210 cm	*0.14 m/s*
352 cm	*0.15 m/s*
415 cm	*0.18 m/s*
470 cm	*0.23 m/s*
512 cm	*0.28 m/s*

Table 13

MAE of SVR Sensor (Tidal (pulse) Surface Profile – Inlet OFF, Outlet OFF)

Candidates Height	MAE Radar (SVR)
210 cm	*0.11 m/s*
352 cm	*0.13 m/s*
415 cm	*0.16 m/s*
470 cm	*0.21 m/s*
512 cm	*0.24 m/s*

Results of SVR exhibited consistent and reliable accuracy under all experimental scenarios. As SVR accuracy is affected by tilt angle and the footprint of its measuring surface, having an embedded tilt and vibration sensor alongside a signal quality index and an SNR index is extremely beneficial for determining correct sensor installation.

Chapter 5: Flash Flood Occurrence Forecasting using Weather Data

This section details the system's initial modeling approach. Weather attributes previously recorded by the RWIS system deployed by ODOT on I-35 highway were used to build an ML model using flood mitigating factors that had been proven influential in other studies discussed in the related work section of this report (e.g., precipitation intensity).

Various short-term and long-term forecasting approaches were explored, and results for the best performing model are presented below.

5.1. Analyzing Weather Conditions in Severe Storms

Many studies have analyzed severe storm variables [22], [23], [24]. Oklahoma storms have typically been accompanied by a significant temperature drop [1]. Storms are often accompanied by gusty winds immediately before rainfall commences. This factor is known to cause flash floods.

Daily historical weather data recorded before, during, and after flash flood events reported by ODOT and the NWS in Norman was analyzed, as was time series historical data retrieved from ODOT's RWIS. Data included ambient temperature, air pressure, humidity, wind speed, gust wind speed, brightness, and precipitation intensity. These factors were used to build a flash flood forecast system. RWIS is an intelligent system designed and developed by Dr. Hazem Refai's research team at the OU-Tulsa campus for the purpose of monitoring weather and road conditions. Fifteen stations are currently deployed by ODOT along Oklahoma's I-35 corridor at road-bridge cross sections. The distance between each is approximately 16 miles (see Figure 30).

Each station consists of six sensing components. The CLIMA [26] weather station measures weather parameters (e.g., wind speed and direction), air temperature and pressure, humidity, brightness, and precipitation intensity/type. The second component is composed of two InfraRed surface sentinel sensors [27] that measure surface temperatures for both road and bridge. The third houses two underground temperature probes [28] for recording sub-surface temperature at 2-inch and 6-inch depths. Oklahoma's flash floods are often the result of intense precipitation with extended duration and close repetition of precipitation and/or a combination of these factors.

Over the past several years, such extreme weather events occur more frequently due to the rising global temperature and subsequent climate change [1], [24].

On May 16th, 2021, a flood event reported by ODOT on SH-77 near lake Murray became the focus of intense analysis. A resulting flood occurred six miles from RWIS station 35ST015.

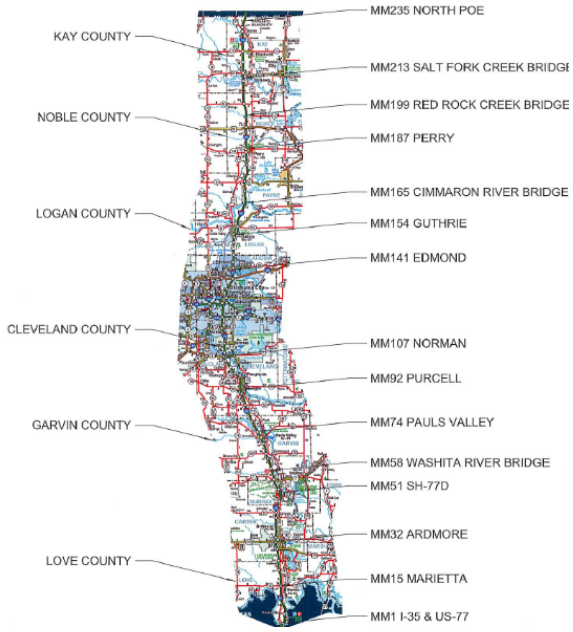

Figure 30. Oklahoma Hwy 1-35 RWIS station locations.

Humidity has been established as a parameter correlated with severe storms. Figures 31 and 32 show graphs of temperature and humidity from RWIS station 35ST015 on May 16, 2021, where a flood event was reported at 18:00.

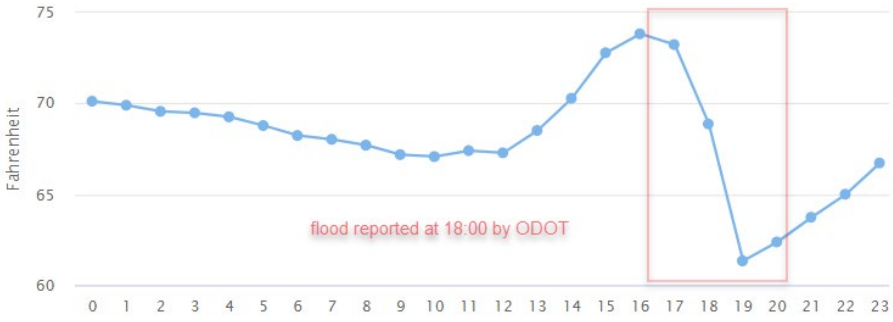

Figure 31. Temperature data reported by station 35ST015 on May 16, 2021.

Figure 32. Humidity data reported by station 35ST015 on May 16, 2021.

Other weather parameters (e.g., atmospheric pressure, wind speed, and gust wind speed) were captured before, during, and after severe storms and flood events. Flash floods are caused by intense precipitation, extended duration, close repetition of precipitations, and/or a combination of these factors. Figure 33

illustrates precipitation data gathered on May 16th, 2021, for station 35ST015 before flood event was reported at 18:00.

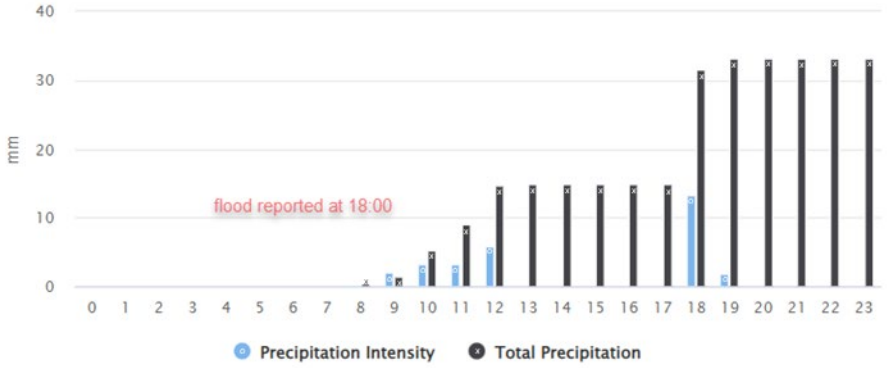

Figure 33. Precipitation data reported by station 35ST015 on May 16, 2021.

These types of flood events—resulting from of an extended duration of precipitation, as well as an intense amount of precipitation within a short period of time—are typically easier to predict than flash floods resulting from very high precipitation intensity occurring in small period of time (e.g., usually within an hour). This type of flood event is referred to as an "extreme flash flood" and is more dangerous and much harder to predict. Similar sudden weather changes were observed in different locations where flash floods were reported. Figure 34 and 35 show the temperature and humidity data recorded by 35ST199, where a flash flood was reported on Oct. 10, 2021.

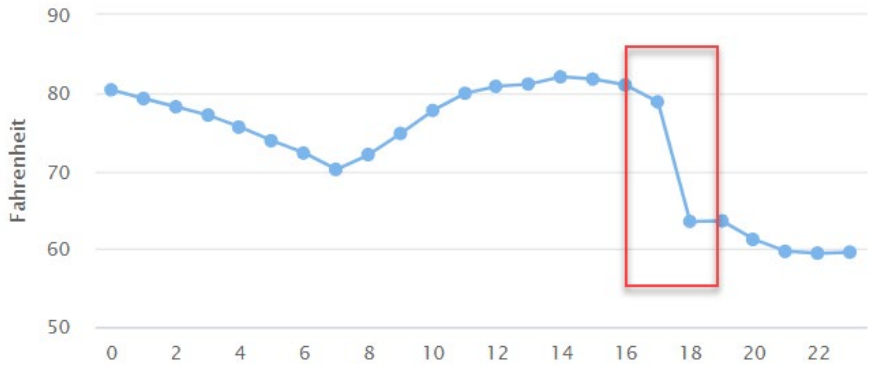

Figure 34. Precipitation data reported by Station 35ST199 on May 10, 2021.

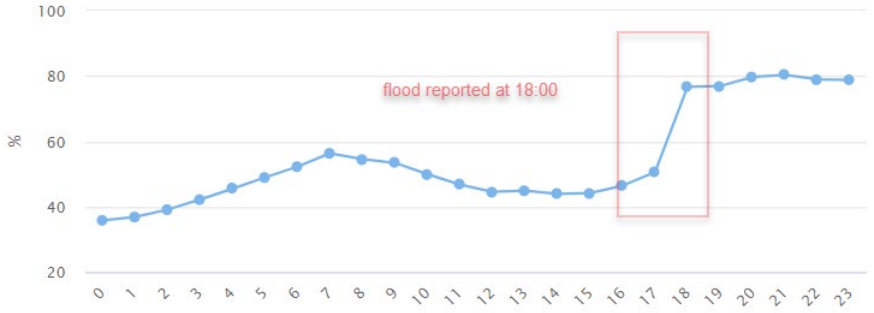

Figure 35. Humidity data reported by station 35ST199 on May 10, 2021.

In this project, we used historical records of flash flood events reported by ODOT. Additional historical flood data for 2021 was retrieved from the National Center for Environmental Information (NCEI). Table 14 displays Oklahoma flood event records for 2020 and 2021, along with associated RWIS stations.

Table 14

Flash Flood Historical Data in Oklahoma for 2020-2021

County	*Date*	*Start*	*End*	*Location*	*Nearby RWIS Station*
Love	01/10/2020	16:00 CST	19:00 CST	[33.93 -97.12]	35ST001
Love	1/16/2020	12:00 CST	23:00 CST	[33.93, -97.12]	35ST001
Love	2/12/2020	02:00 CST	15:00 CST	[34.03,-97.07]	35ST001
Love	09/01/2020	10:00 CST	13:00 CST	[33.93, -97.12]	35ST001
Love	04/29/2021	01:00 CST	04:00 CST	[33.83, -96.41]	35ST015
Love	05/16/2021	18:00 CST	22:00 CST	[33.99, -97.12]	35ST015
Love	05/25/2021	07:00 CST	11:00 CST	[34.29, -96.49]	35ST015
Carter	03/19/2020	07:00 CST	11:00 CST	[34.38, -98.26]	35ST032
Carter	09/01/2020	07:00 CST	11:00 CST	[34.54, -97.92]	35ST032
Murray	05/15/2020	16:00 CST	20:00 CST	[34.53, -97.96]	35ST051
Garvin	03/19/2020	09:00 CST	12:00 CST	[34.70, -97.19]	35ST074
Garvin	05/15/2020	15:00 CST	19:00 CST	[34.73, -97.23]	35ST074
Garvin	04/27-28/2021	23:00 CST	03:00 CST	[34.73, -97.25]	35ST092
Garvin	04/28/2021	06:00 CST	09:00 CST	[34.73, -97.25]	35ST092
Cleveland	03/19/2020	07:00 CST	11:00 CST	[34.38, -98.26]	35ST092
Cleveland	08/31/2020	19:00 CST	20:00 CST	[35.20, -97.49]	35ST092
Cleveland	04/28/2021	07:00 CST	11:00 CST	[34.76, -96.66]	35ST107
Cleveland	05/27/2021	16:00 CST	20:00 CST	[35.84, -97.42]	35ST107
Cleveland	06/28/2021	15:00 CST	23:00 CST	[35.18, -97.51]	35ST092
Cleveland	07/01/2021	15:00 CST	18:00 CST	[35.21, -97.45]	35ST107
Oklahoma	03/19/2020	00:00 CST	03:00 CST	[35.49, -97.47]	35ST136
Oklahoma	09/01/2020	08:00 CST	11:00 CST	[35.58, -97.26]	35ST136
Oklahoma	05/27/2021	16:00 CST	19:00 CST	[35.68, -97.17]	35ST136
Oklahoma	06/27/2021	03:00 CST	06:00 CST	[35.46, -97.52]	35ST124
Oklahoma	07/01/2021	11:00 CST	14:00 CST	[35.67, -97.53]	35ST124
Oklahoma	07/10/2021	21:00 CST	23:00 CST	[35.40, -97.51]	35ST107
Oklahoma	08/13/2021	14:00 CST	17:00 CST	[35.50, -97.55]	35ST124
Oklahoma	10/10/2021	19:00 CST	22:00 CST	[35.56, -97.61]	35ST124

Each RWIS station was mapped to a nearby (i.e., within 20 miles) NWS station, which adds to validation accuracy. Figure 36 shows the location of both RWIS station 35ST015 and NWS weather station KOARDMOR4. Note that the two stations are separated by approximately 12 miles and are situated 10 miles from a reported flooded area.

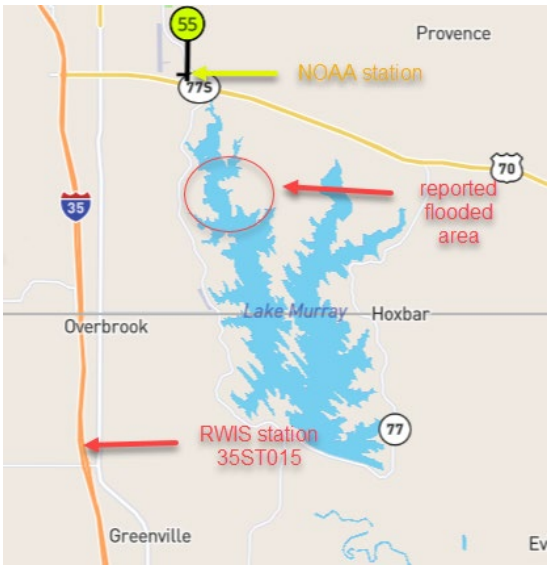

Figure 36. Relative location of NOAA and RWIS stations.

Figure 37 exhibits KOARDMOR4 NWS station and 35ST015 RWIS station temperature variance.

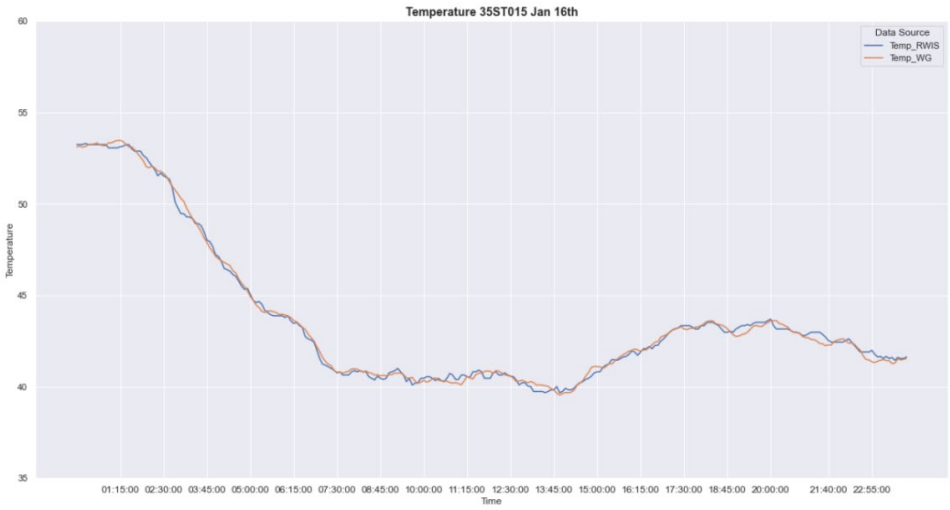

Figure 37. NWS and RWIS temperatures comparison.

Similar validation processes were conducted for additional weather data recorded by RWIS (e.g., humidity, precipitation, wind speed, and the like). Table 15 shows Mean Absolute Percentage Error (MAPE) comparison for RWIS stations and NWS weather stations.

Table 15

MAPE of NWS vs. RWIS Parameters

Attribute	MAPE
Temperature	%1.13
Air pressure	%3.95
Humidity	%4.08
Precipitation	%3.62
Wind Speed	%2.21

After validating the accuracy of weather data measured by ODOT RWIS, data was preprocessed to develop and train ML and recurrent neural network (RNN) models for predicting flash flood events based on weather data.

5.2. Data Preprocessing and Preparation

Various weather features were examined and analyzed to better understand the relevance of each, as well as their distributions. Figure 38 depicts the boxplot of distributions for each feature. Three y-axis scales were utilized to visually emphasize variability.

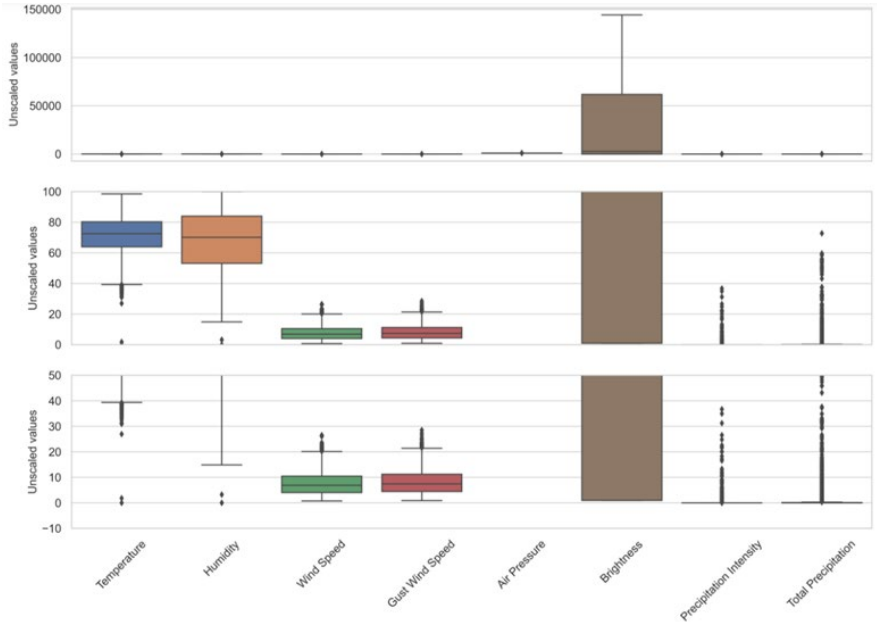

Figure 38. Comparison of weather features at different y-axis scales.

Figure 38 clearly demonstrates that brightness significantly dominates other features. Notably, such dominance in magnitude might overestimate the feature's contribution to the model, especially when using ML algorithms. The result could produce a biased or inefficient model. To solve this, all weather features were fitted through Z-score normalization. Figure 39 shows the scaled distribution for all weather features.

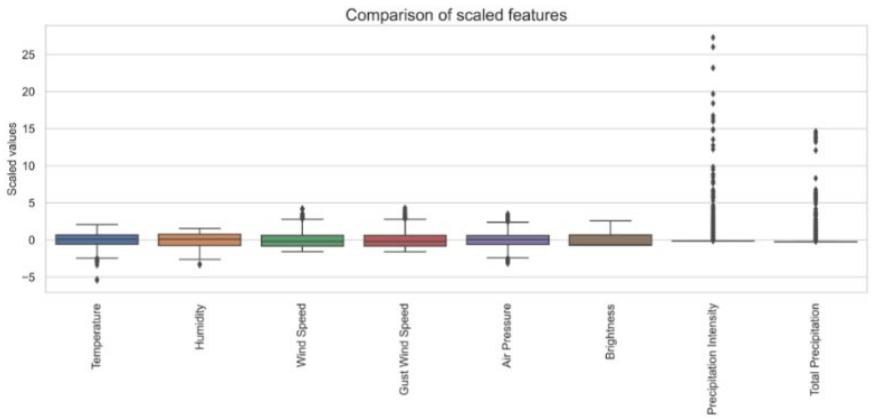

Figure 39. Comparison of scaled features.

Precipitation intensity and total precipitation features are discrete relative to some outlier values. Values can be anticipated, as extreme precipitation events are rare in nature. Class distribution for the 6158 instances was plotted in our seasonal data, which was composed of data measured from station 35ST107 between March 1 and Sept. 30, 2021, and all other historical flood data reported in 2020 and 2021 from other station locations. Data was classified as a "flood," when the event was reported during period, or "safe," when no flood event was reported. Figure 40 shows only 138 flood instances of the 6,158 weather events were identified.

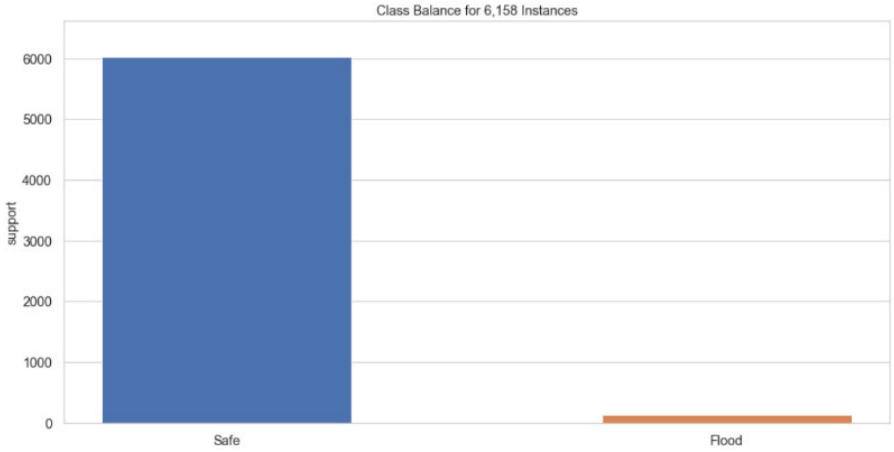

Figure 40. Class balance for 6,158 instances.

An analysis was performed to further comprehension about the relationship between data attributes and storm class. Pearson correlation was initially conducted. Pearson correlation coefficient (PCC) is a measure of linear correlation between two data sets. Simply put, PCC is the ratio between the covariance of two variables and the product of their standard deviations—essentially a normalized measurement of covariance, such that the result has a value between −1 and 1. As with covariance, the measure reflects a linear correlation of variables.

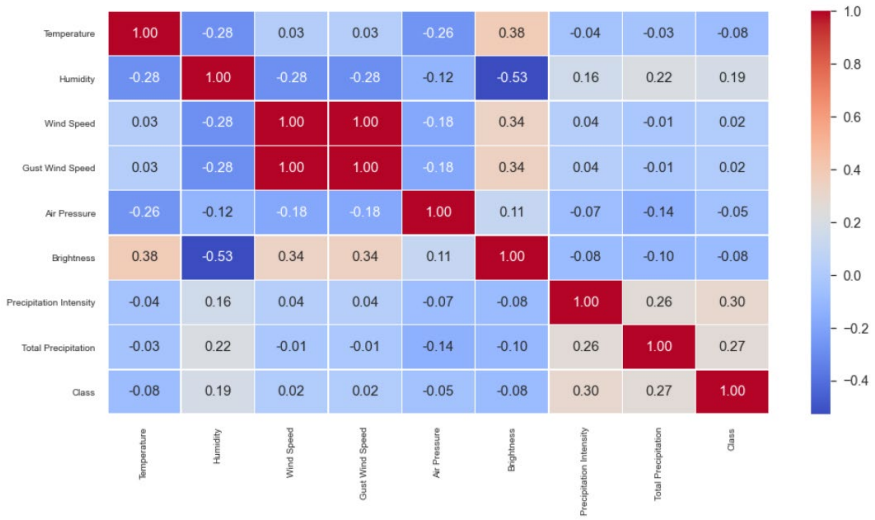

Figure 41. Pearson correlation map of data attributes versus storm class.

Figure 41 shows positive correlation between the class index and precipitation data attributes; it also demonstrates a correlation between class index and the attribute of humidity data. A negative correlation was observed between temperature and air pressure. Wind speed and gust wind speed were highly correlated. One-dimension and two-dimension data analysis were conducted.

A one-dimensional ranking of features utilized a ranking algorithm that considered only one feature at a time (e.g., histogram analysis). The Shapiro-Wilk algorithm was utilized to assess the distribution normality of instances with respect to the feature. Figure 42 shows that most attributes had a relatively normal distribution. The exceptions are precipitation intensity and total precipitation, which can be expected, as these attributes are discrete and include many zero values.

A two-dimensional, feature ranking leverages a ranking algorithm that considers pairs of features at a given time (e.g., joint plot analysis). Feature pairs are then ranked by score and visualized using the lower left triangle of a feature co-occurrence matrix. A covariance ranking algorithm was utilized, which attempted to compute the mean value of deviation products of variates from their respective means. Covariance loosely attempted to detect a colinear relationship between features.

Analysis of covariance (ANCOVA) is a standard statistical methodology that combines the features of both variance (ANOVA) and linear regression analysis for determining whether there is a difference in some response variables between two or more groups.

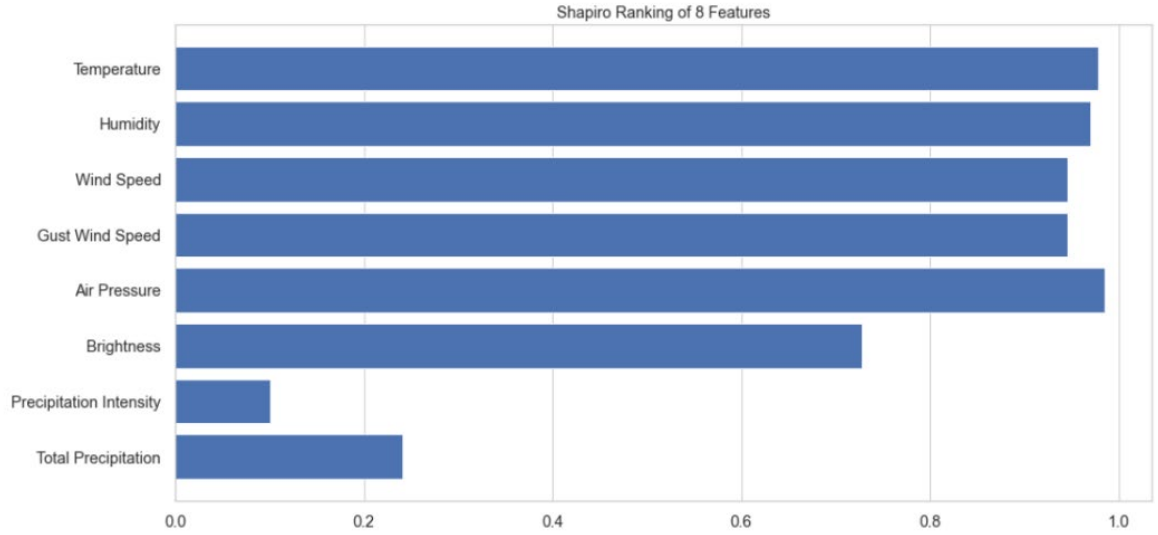

Figure 42. Shapiro-Wilk ranking of weather data attributes.

Figure 43 shows a high inverse correlation between 1) temperature and air pressure and 2) temperature and humidity, which supports initial observations during storms.

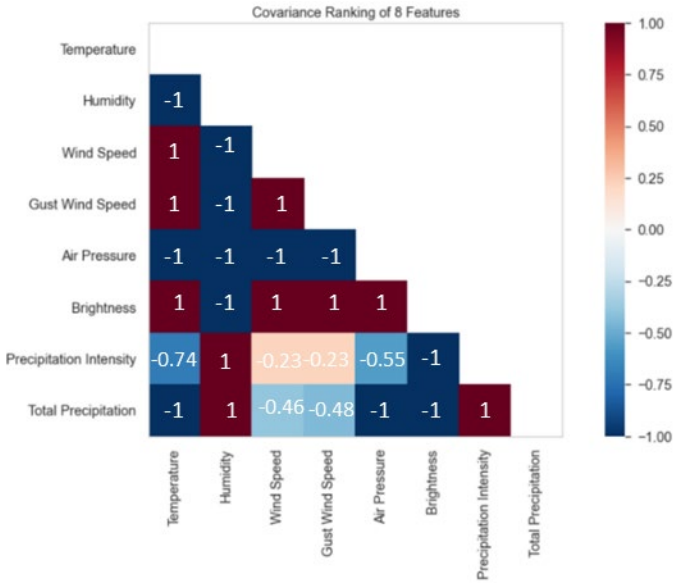

Figure 43. Pairwise covariance correlation between all data attributes.

PCA visualization analysis was conducted. The PCA decomposition visualizer utilizes principal component analysis to decompose high dimensional data into two or three dimensions so that each instance can be plotted in a scatter plot. Using PCA suggests that the projected dataset can be analyzed along axes of principal variation, and 2) can be interpreted to determine if spherical distance metrics can be utilized.

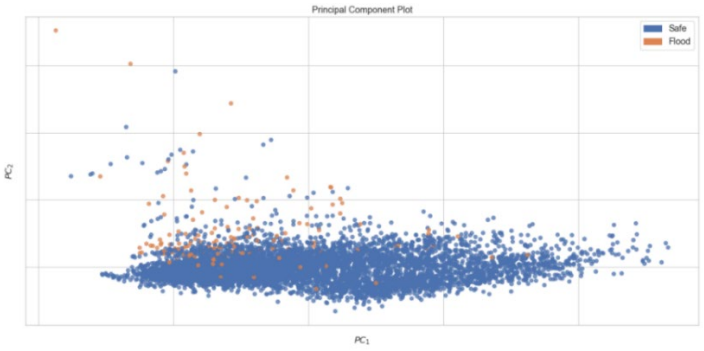

Figure 44. 2D PCA visualization.

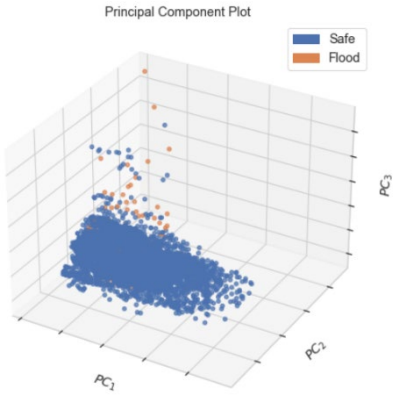

Figure 45. 3D PCA visualization.

2D and 3D PCA visualization is shown in Figures 44 and 45, respectively. Notably, there is little separation for some data points belonging to class flood. To further analyze data point separation, a manifold visualization technique was assessed. Figure 46 shows that most flood class data points can be separated based on the precipitation intensity value of the specific data record.

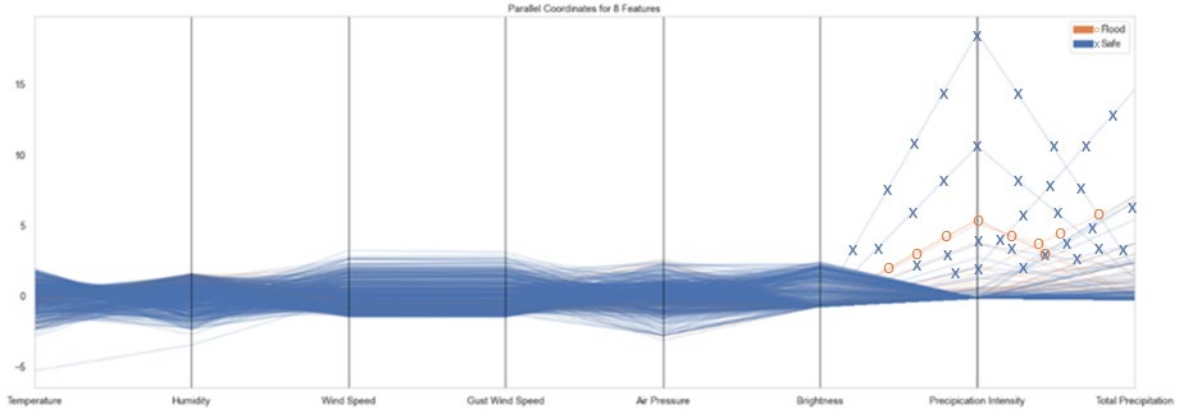

Figure 46. Parallel coordinates visualizer for all data records.

5.3. Applying Machine Learning Algorithms on Time-Series Datasets

Since ML algorithms lack memory capabilities and, instead, treat each row (i.e., record) of data separately. Feature engineering was performed on the dataset to manage time series data. Features used for embedding time concept of the time series data were added to the dataset. Four extra features for each weather attribute were added. Multiple lag values were tested, and our initial selection of a four-hour sequence input produced the most accurate model. Our initial analysis showed that a four-hour window of data is most suitable for capturing weather changes that accompany flash flood events, as shown in Figure 47.

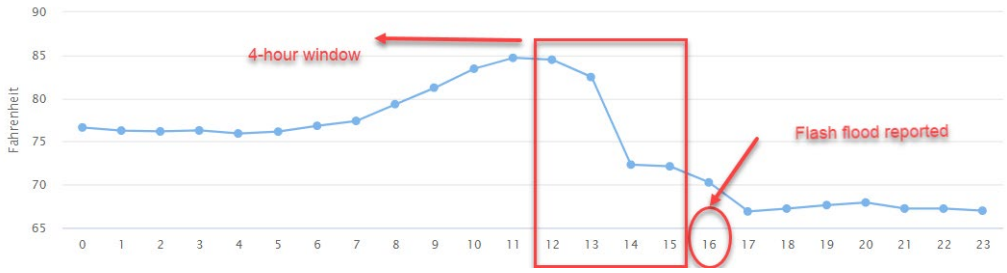

Figure 47. Temperature graph on May 27th, 2021.

T-1, T-2, T-3, and T-4 represent the difference between current measurement at time T and the measurement at time T-n, where n=1,2,3,4 represent a four-hour window. Four arguments were introduced for this function.

5.4. Supervised Machine Learning Classification

Binary classification approach was adopted for this project, wherein two classes were present:

1) 0 representing a safe class where no flood event is predicted; and

2) 1 representing a predicted flood event.

Each row of data represents a sequence of four-hour historical data for each feature (X-1, X-2, X-3, X-4), where X are input features, including temperature, humidity, wind speed, gust wind speed, brightness, air pressure, precipitation intensity, and total precipitation. y represents the class for the next hour and is a binary attribute, where y=0 indicates no flood event (i.e., safe) and y=1 indicates flood event. The dataset was split into 75% for training and 25% for testing. Because the dataset was imbalanced, it was important to split target classes accordingly to ensure the same percentage representation for class distribution in both training and testing datasets. A stratify split approach was used to ensure that both the training and testing sets had a proportion of examples in each class present in the training-testing split percentage. A K-fold cross validation approach was utilized in the training process with k=5. The performance for each method was then summarized using score metrics defined in terms of true-false positive and true-false negative.

5.4.1. Decision Tree Classifier

Decision tree consists of tree branches wherein each branch is considered as a conditional statement. A hierarchy is then formed based on a feature importance score. Final classification occurs at the leaves of the decision tree. Figure 48 and Table 16 show promising results, with an AUC score of 89% and a recall value of 79% for classifying flood class. Decision tree algorithms are immune to multicollinearity by nature and work well with small-size datasets. As a result, performance is superior when compared to other classifiers.

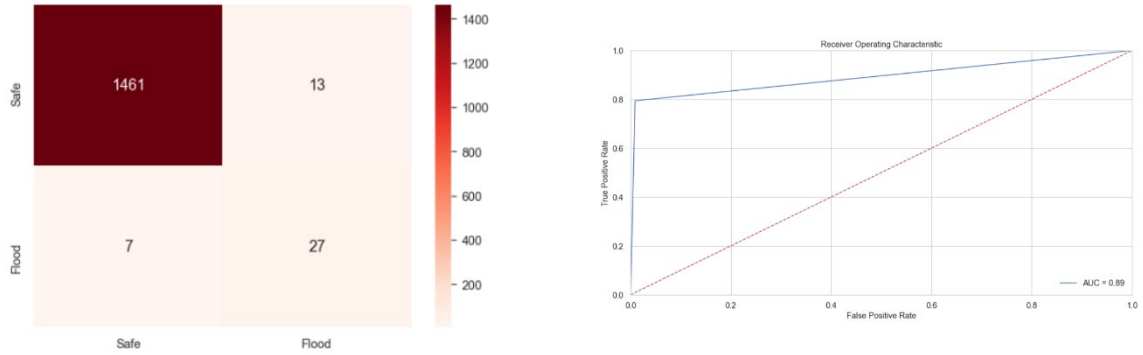

Figure 48. Confusion matrix and ROC plot for decision tree classifier.

Table 16

Classification Scores for Decision Tree Classifier

Class	Precision	Recall	F-1 Score	Support
Safe	1	0.99	0.99	1474
Flood	0.68	0.79	0.73	34

5.4.2. Support Vector Machine (SVM) Classifier

SVM attempts to find the best split for fitting data points belonging to different classes. The decision for the split is based on the relative position in relation to a supporting boundary—known as the hyperplane—which maximizes the distance between data points from different classes.

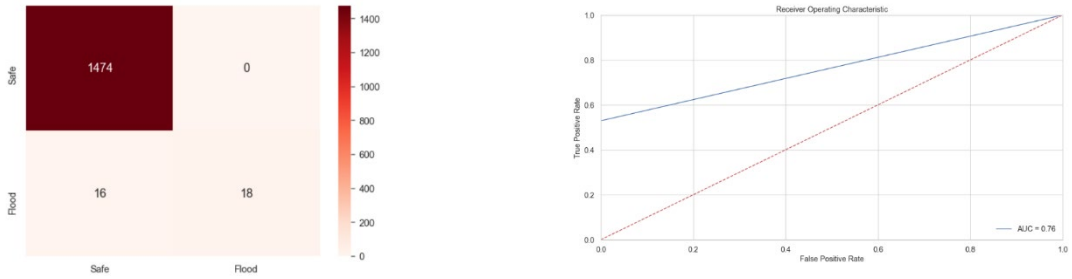

Figure 49. Confusion matrix and ROC plot for SVM classifier.

Table 17

Classification Report for SVM Classifier

Class	Precision	Recall	F-1 Score	Support
Safe	0.99	1	0.99	1474
Flood	1	0.53	0.69	34

The SVM classifier showed poor results when classifying flood class due to imbalanced datasets (e.g., recall value was 53%, and AUC score was 76%.

5.4.3. Random Forest (RF)

RF is a bagging technique that utilizes several decision trees on subsets of the given datasets. The Random Forest algorithm uses the prediction from each tree and—based on majority vote—predicts the final class.

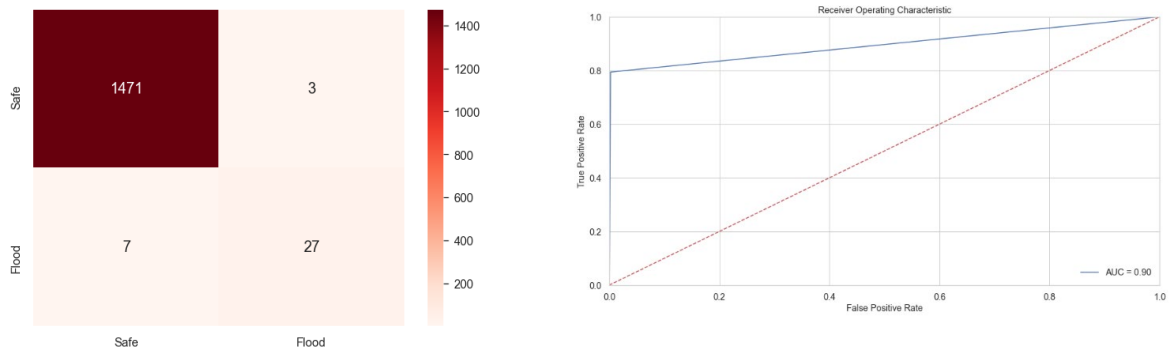

Figure 50. Confusion matrix and ROC plot for RF classifier.

Table 18

Classification Report for RF Classifier

Class	Precision	Recall	F-1 Score	Support
Safe	1	0.99	1	1474
Flood	0.90	0.79	0.84	34

5.4.4. XGboost

XGboost is an ensemble of decision trees like the Random Forest algorithm. However, with XGboost each new tree depends on the evaluation of the previous tree. This sequential way of adding decision tree classifiers is called “boosting.” Unlike traditional boosting, the XGBoost runs tree optimization in parallel, which makes it faster compared to other boosting and ensemble methods. A grid search method was tested to tune the optimal set of hyperparameters for an XGBoost classifier.

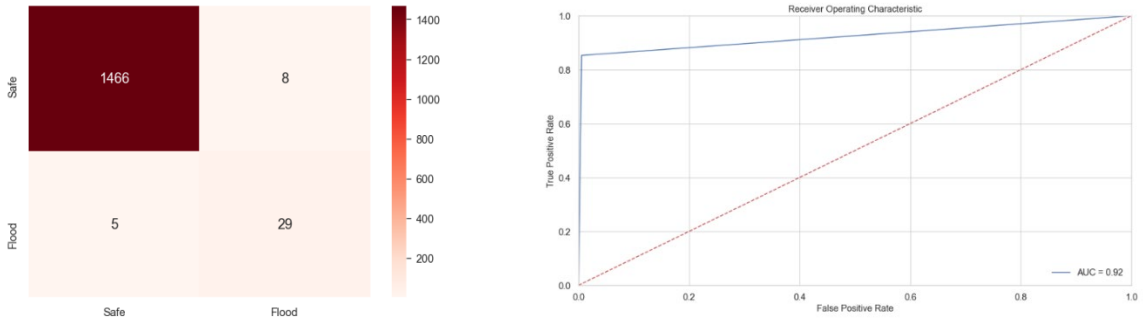

Figure 51. Confusion matrix and ROC plot for XGBoost classifier.

Table 19

Classification Report for XGboost Classifier

Class	Precision	Recall	F-1 Score	Support
Safe	1	0.99	1	1474
Flood	0.78	0.85	0.82	34

Figure 52 shows the feature importance score, which represents the number of split decisions used by each feature. Clearly, a one-hour difference in precipitation had the most significant importance, with 18 split decisions based on its value. The one-hour difference of humidity and gust wind speed also played a key role in building the model by highlighting the importance of features added to represent change over a window of time.

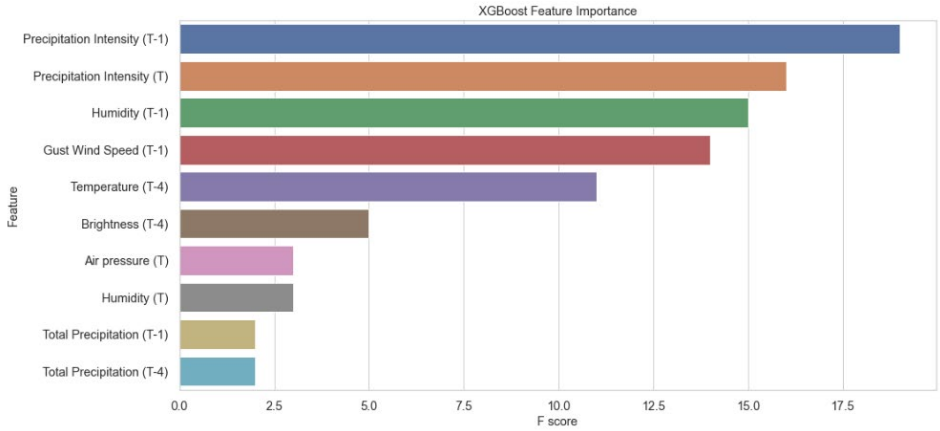

Figure 52. Feature important based on the XGBoost classifier.

When comparing the scores of different ML algorithms used to predict a flash flood event based on weather data, it became obvious that the XGboost ensemble classifier (with 85% accuracy for detecting the flood class and 92% accuracy for AUC) outperformed other algorithms.

This project emphasizes the importance of understanding the characteristics of weather conditions in flood event prediction before choosing suitable classification algorithms and forecast time windows. It is important to note, however, that the modeling that produced acceptable results was only a one-hour forecast model, as weather changes occur during extreme precipitation events.

Chapter 6: Conclusion

This project aimed to investigate flood monitoring and prediction techniques/technologies for the purpose of developing and deploying several stations across the state of Oklahoma at locations where historical flood events occurred. This work introduced a novel approach for evaluating non-contact flood detection technologies. An open-channel, testbed emulator was designed and developed, and several non-contact water level and velocity candidates were evaluated under various conditions. The novel testbed aids in developing new water surface sensing technologies. Results demonstrated that RADAR and LiDAR technologies had the most accurate readings with an MAE score of around 2 cm and an MAE of 0.1 m/s. We believe that collecting both rain rate and water level will provide a highly accurate prediction for floods. Notably, this information must be obtained during a precipitation event. In this way, the system cannot issue a warning before the rain commences. To enhance prediction duration, Doppler SVR can track storms and issue an earlier warning of upcoming storms in the direction of the site location.

An ML-based, one-hour flood forecasting system was developed using historical weather data from RWIS and historical flood events from ODOT and the NWS. The most accurate model was an RF ensemble model with 85% accuracy in forecasting flood events and 92% accuracy in AUC score.

References:

[1] “Climate of Oklahoma.” Oklahoma Climatological Survey, Oklahoma Climatological Survey, climate.ok.gov/index.php/site/page/climate_of_oklahoma.

[2] US Department of Commerce, NOAA. Daily Historical Weather Information for the NWS Norman Forecast Area, NOAA's National Weather Service, 14 Apr. 2020, www.weather.gov/oun/wxhistory.

[3] Wang, Z.; Lai, C.; Chen, X.; Yang, B.; Zhao, S.; Bai, X. Flood hazard risk assessment model based on random forest. J. Hydrol. 2015, 527, 1130–1141.

[4] M. Khalaf et al., "A Data Science Methodology Based on Machine Learning Algorithms for Flood Severity Prediction," 2018 IEEE Congress on Evolutionary Computation (CEC), 2018, pp. 1-8, DOI: 10.1109/CEC.2018.8477904.

[5] Mosavi, Amir, Pinar Ozturk, and Kwok-wing Chau. "Flood Prediction Using Machine Learning Models: Literature Review." Water (Basel) 10.11 (2018): 1536. Web.

[6] M. S. Yadnya and I. W. Sudiartha, "Measurement of drop size distribution rain in mataram utilize dendrometer acoustic for floods prediction," 2016 International Seminar on Intelligent Technology and Its Applications (ISITIA), Lombok, 2016, pp. 107-110, doi: 10.1109/ISITIA.2016.7828642.

[7] C. Lai, Y. Lo, J. Yur and C. Chuang, "Application of Fiber Bragg Grating Level Sensor and Fabry-Pérot Pressure Sensor to Simultaneous Measurement of Liquid Level and Specific Gravity," in IEEE Sensors Journal, vol. 12, no. 4, pp. 827-831, April 2012, DOI: 10.1109/JSEN.2011.2161075.

[8] Basha, Elizabeth & Ravela, Sai & Rus, Daniela. (2008). Model-based monitoring for early warning flood detection. SenSys'08 - Proceedings of the 6th ACM Conference on Embedded Networked Sensor Systems. 295-308. 10.1145/1460412.1460442.

[9] M. Mousa, X. Zhang and C. Claudel, "Flash Flood Detection in Urban Cities Using Ultrasonic and Infrared Sensors," in IEEE Sensors Journal, vol. 16, no. 19, pp. 7204-7216, Oct.1, 2016, doi: 10.1109/JSEN.2016.2592359.

[10] Lin, Yan-Ting & Lin, Yi-Chun & Han, Jen-Yu. (2018). Automatic Water-Level Detection Using Single-Camera Images with Varied Poses. Measurement. 127. 10.1016/j.measurement.2018.05.100.

[11] B. Basnyat, N. Roy and A. Gangopadhyay, "A Flash Flood Categorization System Using Scene-Text Recognition," 2018 IEEE International Conference on Smart Computing (SMARTCOMP), Taormina, 2018, pp. 147-154, doi: 10.1109/SMARTCOMP.2018.00085.

[12] A. Filonenko, Wahyono, D. C. Hernández, D. Seo and K. Jo, "Real-time flood detection for video surveillance," IECON 2015 - 41st Annual Conference of the IEEE Industrial Electronics Society, Yokohama, 2015, pp. 004082-004085, DOI: 10.1109/IECON.2015.7392736.

[13] A. H. Özcan and C. Ünsalan, "LiDAR Data Filtering and DTM Generation Using Empirical Mode Decomposition," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 10, no. 1, pp. 360-371, Jan. 2017, DOI: 10.1109/JSTARS.2016.2543464.

[14] Fernandez-Diaz, Juan & Carter, William & Shrestha, R. & Glennie, Craig. (2014). Now You See It… Now You Don't: Understanding Airborne Mapping LiDAR Collection and Data Product Generation for Archaeological Research in Mesoamerica. Remote Sensing. 6. 9951-10001. 10.3390/rs6109951.

[15] Paul, J. D., Buytaert, W., & Sah, N. (2020). A technical evaluation of lidar-based measurement of river water levels. Water Resources Research, 56, e2019WR026810. https://doi.org/10.1029/2019WR026810.

[16] Tamari, Serge & Guerrero-Meza, Vicente. (2016). Flash Flood Monitoring with an Inclined Lidar Installed at a River Bank: Proof of Concept. Remote Sensing. 8. 834. 10.3390/rs8100834.

[17] F. Alimenti et al., "Noncontact Measurement of River Surface Velocity and Discharge Estimation With a Low-Cost Doppler Radar Sensor," in IEEE Transactions on Geoscience and Remote Sensing, vol. 58, no. 7, pp. 5195-5207, July 2020, DOI: 10.1109/TGRS.2020.2974185.

[18] Flood Warning Systems, Water Level Monitoring, And Alerting. Last accessed on May 28, 2022, from https://www.valarm.net/flood-warning-systems/

[19] Fondriest Environmental, Inc. "Flood Warning Systems." Fundamentals of Environmental Measurements. 17 Sep. 2015. Web. < https://www.fondriest.com/environmental-measurements/parameters/water-quality/water-temperature

[20] Traffic Flood Warning Systems: Flood Detection Systems. (n.d.). Last accessed on May 28, 2022, from https://elteccorp.com/products/warning-systems/flood-warning-systems/

[21] Lee, Jae-Yeong and Kim, Ji-Sung, Detecting Areas Vulnerable to Flooding Using Hydrological-Topographic Factors and Logistic Regression, Applied Sciences, 2021, https://www.mdpi.com/2076-3417/11/12/5652

[22] Namoto, Shinsuke & Ishikawa, Tadaharu & Kojima, Takashi. (2018). Characteristics of water surface profiles of open channel flow with levee overtopping. International Journal of Safety and Security Engineering. 8. 484-492. 10.2495/SAFE-V8-N4-484-492.

[23] Hydrogeologic Characteristics and Geospatial Analysis of the Water-Table Changes in the Lower Part of the Arkansas River. United States Geological Survey.

[24] Aqua Dream 24v Dc Sine Wave Maker Pump Marine Aquarium Coral Reef 4800 GPH [LBS-50]. Retrieved May 18, 2022, from https://aquadreamusa.com/collections/lbs-series/products/aquility-wave-maker-4800-gph-aq-lbs-50

[25] Bar02 Ultra High Resolution 10m Depth/Pressure Sensor. Retrieved May 18, 2022, from https://bluerobotics.com/store/sensors-sonars-cameras/sensors/bar02-sensor-r1-rp/

[26] DIGITEN G1" Water Flow Sensor, Food-Grade Hall Effect Sensor Flow Meter Flowmeter Counter 1-60L/min. Retrieved May 18, 2022, from https://www.digiten.shop/products/digiten-g3-4-water-flow-hall-sensor-switch-flow-meter-1-60l-min

[27] 260-700 Ultrasonic Water Depth Sensor. Retrieved May 18, 2022, from https://novalynx.com/store/pc/260-700-Ultrasonic-Snow-Depth-Sensor-p275.htm

[28] OTT RLS - Radar Level Sensor. Retrieved May 18, 2022, from https://www.ott.com/products/water-level-1/ott-rls-radar-level-sensor-861/

[29] OTT SVR – Surface Velocity Radar. Retrieved May 18, 2022, from https://www.ott.com/products/water-flow-3/ott-svr-100-2406/