# **Introduction**

## ***Overview***

At a seminar within the framework of the project “Updating Climate Change Scenarios for Vietnam,” implemented by the Institute of Meteorology, Hydrology, and Climate Change under the Ministry of Natural Resources and Environment, Mr. Fergus McBean—First Secretary for Climate and Nature at the British Embassy in Vietnam—stated that climate change has posed significant challenges to Vietnam's sustainable development. Rising sea levels, increased storm intensity, and extreme weather phenomena have made Vietnam a vulnerable nation. The recent Yagi storm serves as evidence of the future risks that not only Vietnam but also other countries face due to the impacts of climate change. Therefore, the need for a warning system for extreme climate events in general and floods in particular is especially urgent, particularly for remote areas that lack technical resources. To address this need, I have researched and developed a flood warning system using compact embedded boards such as Arduino Nano integrated with a simple AI model to be able to make informed predictions from collection of environmental data and transmission through the LoRa network, ensuring the ability to monitor environmental changes and suitability for areas with weak telecommunication signals.

## ***Related Work***

Building a system to monitor environmental changes is not a new endeavor; meteorological stations in Vietnam have indeed implemented such systems to track and forecast weather phenomena, thereby supporting forecasting efforts and disaster response. However, despite certain advancements, current systems still have several significant limitations that affect the effectiveness and reliability of environmental monitoring. Firstly, in terms of data transmission protocols, most existing systems still use GSM technology. Although GSM is a popular and easily deployable communication technology, it comes with high operational costs due to dependence on telecommunication service providers. Secondly, the current systems lack complete automation in the data collection and processing processes. Some environmental indicators, such as water depth in rivers, still require manual intervention by staff at measurement stations

## ***Project Tasks***

Task 1: Understand the LoRa communication structure and the various LoRa transmission methods. Gain a comprehensive understanding of how to configure states and customize settings for different communication models.

Task 2: Research and select sensors for the project. Considering the project objectives and the regulations regarding forecasting, warning, disaster communication, and disaster risk levels, there are two crucial parameters: rainfall intensity and water level height. The water level measurement module will include a VL53L0x distance sensor, while the rainfall intensity measurement will utilize a drop-based rain gauge model, for which the KY-025 inductive proximity sensor will be chosen. Additionally, supplementary sensors will be incorporated to provide additional information for the system, including a Rain Drop Sensor and a Thermal Sensor LM35.

Task 3: Connect the selected sensors to the Arduino Nano microcontrollers while exploring the UART, SPI, and I2C communication protocols. The predicted measurement results will be calculated and displayedon specific server

# **Preliminary**

## ***Communication***

### *LoRa network*

#### Introduction

LoRa (Long Range) is a long-range, low-power wireless communication technology developed by Semtech. LoRa is designed to meet the requirements of IoT (Internet of Things) networks, such as wide coverage range, the ability to operate in high-density environments, and energy optimization for battery-powered devices. This technology operates based on the Chirp Spread Spectrum (CSS) modulation technique. While Chirp Spread Spectrum (CSS) is a signal modulation technique in which data is transmitted by adjusting the frequency of chirp pulses. A chirp pulse is a signal with a frequency that changes linearly over time, either increasing or decreasing within a specific frequency range. This enables high efficiency in transmitting signals over long distances and in environments with high interference. In this research, the SX1278 module manufactured by Ebyte will be used. The frequency of LoRa waves is regulated differently in each region and is a fixed parameter. For Vietnam, the regulated frequency is 433 MHz.

#### Structure of the E32-170T30D Module

The E32-170T30D module, developed by Chengdu Ebyte, integrates the SEMTECH RF chip SX1278 with LoRa (Long Range) communication technology. This spread-spectrum modulation technology is designed to achieve long transmission distances, reduce energy consumption, and enhance interference resistance. Its main components include:

A close-up of a chip

AI-generated content may be incorrect.

Figure 2.1 E32-170T30D SX1278

#### Main Parameter

|  |  |  |  |
| --- | --- | --- | --- |
| Main Parameter | Unit | Performance | Remark |
| Working Frequency | MHz | 160-173 |  |
| Transmit Power | dBm | 29.3-30.5 |  |
| Receive Sensivity | dBm | -147±1.0 | Air data rate 2.4kbps |
| Air data rate | bps | 0.3-9.6k |  |
| Measured distance | M | 8000 | Air data rate 2.4kbps |

#### Operating Modes of the Module

**Fixed** Transmission:

Signals are sent precisely to a single server address within a specific channel, while other addresses and channels cannot receive the data.

A diagram of a computer network

AI-generated content may be incorrect.

Figure 2.2 Example for Fixed Transmission

**Broadcast:**

The signal is sent to all addresses lower than the gate within the same channel, while other channels do not receive the data.

A diagram of a computer network

AI-generated content may be incorrect.

Figure 2.3 Example for Broadcas

#### Operation mode

The E32-170T30D module supports 4 operating modes, configured through two control pins, M0 and M1.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Mode | M0 | M1 | Mode Introduction | Remark |
| Normal | 0 | 0 | UART and wireless channel are open, transparent transmission is on | The receiver must work in mode 0 or mode 1. |
| Wake up | 0 | 1 | UART and wireless channel are open, the only difference with mode 0 is that before transmitting data, increasing the wake up code automatically, so that it can awake the receiver under mode 3. | The receiver could be 0, 1 or 2 |
| Power Saving | 1 | 0 | UART close, wireless is under air-awaken mode, after receiving data, UART open and send data | transmitter must be mode 1, unable to transmit in this mode |
| Sleep | 1 | 1 | sleep mode, receiving parameter setting command is available | Receive data and change configuration according to the entered parameters |

Command Parament in Sleep mode:

|  |  |  |  |
| --- | --- | --- | --- |
| No. | Item | Description | Remark |
| 0 | HEAD | Fix 0xC0 or 0xC2, indicating this frame data is a control command. | Must be 0xC0 or 0xC2. - 0xC0: Save the parameters when power-down. - 0xC2: Do not save the parameters when power-down. |
| 1 | ADDH | High address byte of module (default: 00H). | Range: 00H-FFH. |
| 2 | ADDL | Low address byte of module (default: 00H). | Range: 00H-FFH. |
| 3 | SPED | **UART parity bit (bit 7, 6):**   - 00: 8N1 (default).  - 01: 8O1.  - 10: 8E1.  - 11: 8N1 (same as 00). **TTL UART baud rate (bit 5, 4, 3):**  - 000: 1200 bps.  - 001: 2400 bps.  - 010: 4800 bps.  - 011: 9600 bps (default).  - 100: 19200 bps.  - 101: 38400 bps.  - 110: 57600 bps.  - 111: 115200 bps. **Air data rate (bit 2, 1, 0):**  - 000: 0.3 kbps.  - 001: 1.2 kbps.  - 010: 2.4 kbps (default).  - 011: 4.8 kbps.  - 100: 9.6 kbps.  - 101, 110, 111: 19.2 kbps. |  |
| 4 | CHAN | Communication channel. | Range: 00H-36H, corresponding to 160 MHz + CHAN \* 250 KHz. |
| 5 | OPTION | **Fixed transmission enabling bit (bit 7):**  - 0: Transparent transmission mode.  - 1: Fixed transmission mode. **IO drive mode (bit 6):**  - 1: TXD and AUX push-pull outputs, RXD pull-up inputs (default).  - 0: TXD and AUX open-collector outputs, RXD open-collector inputs. **Wireless wake-up time (bit 5, 4, 3):**  - 000: 250 ms (default).  - 001: 500 ms.  - 010: 750 ms.  - 011: 1000 ms.  - 100: 1250 ms.  - 101: 1500 ms.  - 110: 1750 ms.  - 111: 2000 ms. **FEC switch (bit 2):**  - 0: Turn off FEC.  - 1: Turn on FEC (default). **Transmission power (bit 1, 0):**  - 00: 20 dBm (default).  - 01: 27 dBm.  - 10: 24 dBm.  - 11: 21 dBm. |  |

Example of a Command Parameter Packet:

0xc0, 0xff, 0xff, 0x1a, 0x17, 0x44

* HEAD <0xc0>: Save Parameter when power down.
* ADD <0xff-0xff>: Address 65535.
* SPED <0x1a>:
  + Parity: 8N1.
  + Baud rate: 9600 bps.
  + Air rate: 2.4 Kbps.
* CHAN <0x17>: Channel 23.
* OPTION <0x44>:
  + Transmission mode.
  + IO mode: Push-pull.
  + WOR timing: 250ms.
  + FEC: Enabled.
  + Transmission Power: 20 dBm.

### *Zigbee network*

#### Introduction

Zigbee is a wireless communication protocol based on the IEEE 802.15.4 standard, designed for applications requiring low power consumption and low data rates. Operating primarily in the 2.4GHz frequency band, Zigbee is suitable for wireless personal area networks (WPANs) and is widely used in IoT applications such as home automation, industrial control, and sensor networks. A standout feature of Zigbee is its support for mesh networking, which allows devices to communicate through multiple intermediate nodes, thereby expanding the network's coverage area and increasing its reliability. E18-RS1-PCB module is considered for implementing Zigbee communication. This is a system-on-chip (SoC) solution that integrates a microcontroller and a Zigbee-compliant radio transceiver, providing a complete platform for Zigbee applications. The module supports the full Zigbee protocol stack and can be easily interfaced with external sensors or microcontrollers through suitable interfaces.

#### Structure of the E18-RS1-PCB Module

The E18 series module adopts CC2530 RF chip imported from Texas Instruments. The chip integrates 8051single-chip microcomputer and wireless transceiver. Some module models have built-in PA power amplifier toincrease the communication distance. The factory-built firmware implements serial data transparent transmissionbased on the ZigBee3.0 protocol, and supports various commands under the ZigBee3.0 protocol.

A blue and white chip with a white label

AI-generated content may be incorrect.

Figure 2.4 E32-170T30D SX1278

#### Main Parameter:

|  |  |  |  |
| --- | --- | --- | --- |
| Main Parameter | Unit | Performance | Remark |
| Working Frequency | GHz | 2400~2480 |  |
| Transmit Power | dBm | 4.0±0.5 |  |
| Receive Sensivity | dBm | -96.5±1.0 | Air data rate 250kps |
| Air data rate | bps | 250k |  |
| Measured distance | M | 200 | Air data rate 250kbps |

#### Main Parameter

|  |  |  |  |
| --- | --- | --- | --- |
| Main Parameter | Unit | Performance | Remark |
| Working Frequency | MHz | 855-925 |  |
| Transmit Power | dBm | 20±1 |  |
| Blocking Power | dBm | 0~10.0 |  |
| Receive Sensivity | dBm | -115±1.0 | Air data rate 1kbps |
| Air rate data | bps | 1k |  |
| Measured distance | Km | 2.5 | Air data rate 1kbps |

#### Communication Modes

**Same as LoRa, Zigbee has two communication modes, Fix transmission and broadcast**

#### Operation mode

The E18-RS1-PCB module supports 3 operating modes, configured through two control pins, P1.0 and P1.1.

A diagram of a computer chip

AI-generated content may be incorrect.

A diagram of a computer

AI-generated content may be incorrect.

Figure 2.5 E18-MS1-PCB Schematic

|  |  |  |  |
| --- | --- | --- | --- |
| Working Mode | M0 | M1 | Remark |
| Receiver mode | 1 | 0 | Transmitter must be mode 1 |
| Transmitter mode | 0 | 1 | Receiver must be mode 0, unable to transmit in this mode |
| Sleep | 0 | 0 | Receive data and change configuration according to the entered parameters |

### *GSM network*

#### Introduction

GSM, or Global System for Mobile Communications, is a standard for digital cellular networks that has been pivotal in mobile communications since its first implementation in 1991. It supports voice calls, SMS, and data transmission, operating on frequency bands such as 900 MHz and 1800 MHz in Europe, and 850 MHz and 1900 MHz in the US. This technology is particularly valued in IoT applications for its wide coverage and affordability, though it is gradually being replaced by newer standards like 4G and 5G. E30-900M20S module is considered for implementing Zigbee communication

A white card with black text and a qr code

AI-generated content may be incorrect.

Figure 2.6 E30-900M20S

#### Communication Modes

GSM does not have two communication mode like LoRa or Zigbee but transmits directly to the ID which is the phone number identifying the SIM in use.

#### Operation mode

Does not described on datasheet.

A diagram of a computer chip

AI-generated content may be incorrect.

Figure 2.7 E30-900M20S Schematic

## ***WHY IS LORA?***

|  |  |  |  |
| --- | --- | --- | --- |
|  | LoRa | Zigbee | GSM |
| Working Frequency | 160-173 MHz | 2400-2480 GHz | 855-925 MHz |
| Measured distance | 8000 m | 200 m | 2500 m |
| Air rate data | 0.3-9.6k bps | 250k bps | 1k bps |
| Cost | 434.000 VND | 468.000 VND | 570.000 VND + monthly cost |

Zigbee demonstrates superiority with a high transmission speed at an air data rate of 250 kbps; however, according to the Free Space Path Loss (FSPL) formula:

Where:

* *d* is the propagation distance
* *f* is the signal frequency
* *c* is the speed of light

Given Zigbee's very high operating frequency of 2400 GHz, the resulting path loss is also significantly high, making it unsuitable for the measurement model in this project. In contrast, LoRa, with its lower frequency and an acceptable data transmission rate, offers the lowest cost and thus becomes the most suitable choice for the model.

## ***AI Model***

### *Linear Regression*

#### Definition

Linear regression is a statistical method that models the relationship between a dependent variable (what you want to predict) and an independent variable (the input) using a straight line.

#### Algorithm

* y: Dependent variable (the output or prediction).
* x: Independent variable (the input).
* ​: Intercept (the value of y x=0).
* ​: Slope (how much y changes when x increases by 1).
* : Random error (unexplained variation).

The goal is to minimize the sum of squared errors (SSE):

where = ​ is the predicted value.

#### How to Find the Constants

The coefficients ​ and ​ are calculated using the Ordinary Least Squares (OLS) method:

* : Mean of the independent variable x.
* : Mean of the dependent variable y.

This gives the best-fitting line for the data.

### *Multiple Regression*

#### Definition

Multiple regression extends linear regression by using multiple independent variables to predict the dependent variable.

#### Algorithm

* Y: n x 1 vectors of of Dependent variable (the output or prediction).
* X: n x i vectors of Independent variable (the input).
* ​: n x 1 vectors of Coefficients showing the effect of each x on y
* : n x 1 vectors of Random error (unexplained variation).

#### How to Find the Constants

Using OLS, the coefficients are found with matrix algebra:

### *Polynomial Regression*

#### Definition

Polynomial regression models the relationship between the independent variable and dependent variable as a polynomial (e.g., quadratic or cubic), allowing for curved relationships.

#### Algorithm

* y: Dependent variable (the output or prediction).
* : Powers of the independent variable x (where i is the degree).
* ​: Coefficients.
* : Random error (unexplained variation).

Though non-linear in x, it’s treated as linear in the coefficients.

#### How to Find the Constants

Treat as separate independent variables.

Use the OLS method from multiple regression:

### *Conlusion*

|  |  |  |  |
| --- | --- | --- | --- |
| Criteria | Linear Regression | Multiple Regression | Poly Regression |
| Number of Independent Variables | 1 (One independent variable) | 2 or more (Multiple independent variables) | 1 or more, but using higher-order terms of the independent variable. |
| Relationship | Linear | Linear | Can be nonlinear |
| Application | When the data shows a clear linear relationship. | When the data shows a linear relationship between the dependent variable and multiple independent variables. | When the relationship between the dependent and independent variables is nonlinear, requiring curvature. |
| Complexity | Simple, easy to understand and implement. | More complex than Linear Regression, but still a linear model. | A more complex model, can overfit if the degree is too high. |
| Scalability | Limited to linear relationships. | Can scale to handle multiple factors but still a linear relationship. | Can model nonlinear relationships. |
| Data Characteristics | Data should have a clear linear distribution. | Data can have multiple factors influencing the dependent variable. | Data should have a nonlinear relationship or large variations in degree. |
| Risk of Overfitting | Less risk of overfitting if the data fits a linear model. | Can overfit if the number of variables is not controlled. | Higher risk of overfitting if the degree of the polynomial is too high. |
| Optimization Method | |  | | --- | | Solved using the Ordinary Least Squares (OLS) method. | | |  | | --- | | Solved using the Ordinary Least Squares (OLS) method. | | |  | | --- | | Solved using the Ordinary Least Squares (OLS) method. | |

For the project, since it is necessary to provide predicted values based on measurement data, regression models were considered and evaluated. Given that it is almost impossible to have a linear relationship over time for environmental data, Poly Regression is the most suitable model for the project.

Additionally, two other models were considered: Neural Link and Support Vector Regression (regression version of Support Vector Machines). However, due to their complex algorithms, they are highly prone to overfitting when dealing with large environmental datasets. Moreover, these two models are optimized for more complex tasks, such as image processing, and therefore were not selected for use.

## ***Sensor***

### *VL53L0X Laser Distance ToF Sensor*

#### Introduction

The VL53L0X is a Time-of-Flight (ToF) distance measurement module developed by STMicroelectronics, integrating FlightSense™ technology. Using the Time of Flight distance measurement method and the I2C communication protocol

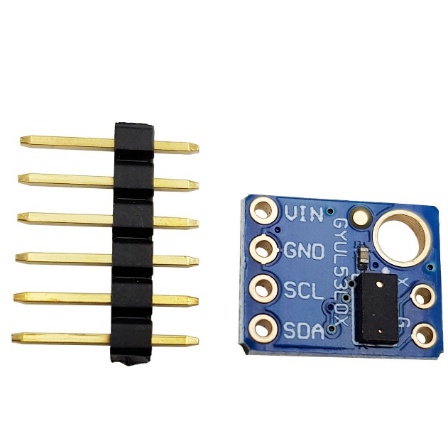
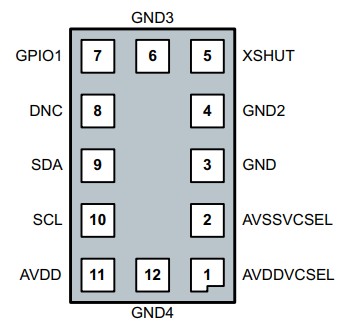


Figure 2.8 VL53L0X

#### Technical specification

|  |  |
| --- | --- |
| Feature | Detail |
| Package | Optical LGA12 |
| Size | 4.40 x 2.40 x 1.00 mm |
| Operating voltage | 2.6 to 3.5 V |
| Operating temperature | -20 to 70°C |
| Infrared emitter | 940 nm |
| I²C | Up to 400 kHz (FAST mode) serial bus Address: 0x52 |

#### Schematic



A diagram of a computer circuit

AI-generated content may be incorrect.

Figure 2.9 VL53L0X schematic

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Pin | Signal Name | |  | | --- | | Signal Type | |  | | | |  | | --- |   Signal Type   |  | | --- | |  | |
| 1 | AVDDVCSEL | Power | |  | | --- | | VCSEL Supply, to be connected to main supply |  |  | | --- | |  | |
| 2 | AVSSVCSEL | Ground | |  | | --- | | VCSEL Ground, to be connected to main ground |  |  | | --- | |  | |
| 3, 4, 6, 12 | GND | Ground | |  | | --- | | To be connected to main ground | |
| 5 | XSHUT | Digital Input | |  | | --- | | Xshutdown pin, Active LOW |  |  | | --- | |  | |
| 7 | GPIO1 | Digital output | |  | | --- | | Interrupt output. Open drain output. |  |  | | --- | |  | |
| 8 | DNC | Digital Input | |  | | --- | | Do Not Connect, must be left floating. |  |  | | --- | |  | |
| 9 | SDA | Digital Input/Output | |  | | --- | | I²C serial data |  |  | | --- | |  | |
| 10 | SCL | Digital Input | |  | | --- | | I²C serial clock input |  |  | | --- | |  | |
| 11 | AVDD | Power | |  | | --- | | Supply, to be connected to main supply |  |  | | --- | |  | |

#### Operating Principle of the I2C Level Shifter.

When power is supplied to VIN, the voltage at will be in the range of 2.6–3.5V; for now, let's assume the value is 3V:

=0.7V

In the case no power, :

= 0.7V

Similarly, the same applies to NMOS Q1B.

#### Distance Calculation Algorithm

Time-of-Flight Principle:

The distance is calculated based on the time it takes for the laser beam to travel to the target and reflect back:

* d: Distance (m).
* c: Speed of light (m/s).
* Δt: Time of flight.

Error Compensation Calibration:

* Offset Calibration: Eliminates errors caused by temperature drift.
* Cross-talk Compensation: Reduces optical noise from the protective glass.