# Datasets & Ideas

Broad idea: predict water

Predictions from Google Maps vs. from data (potential final product / goal)

## Data Sources

* <https://www.waterqualitydata.us/>
* https://wbwaterdata.org/dataset
* <https://www.kaggle.com/anbarivan/indian-water-quality-data>
* <https://www.kaggle.com/adityakadiwal/water-potability>
* <https://www.kaggle.com/ozgurdogan646/water-quality-dataset>

## Previous Research

* **INDIAN PAPER: READ FOR GREAT METHODOLOGY BREAKDOWN** [**https://www.hindawi.com/journals/abb/2020/6659314/#data-availability**](https://www.hindawi.com/journals/abb/2020/6659314/#data-availability)
* <https://www.datascience2000.in/2021/10/water-quality-prediction-using-machine.html>
* <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/96WR03529>
* https://www.researchgate.net/profile/Pirro-Icka/publication/287957321\_Evaluation\_of\_Water\_Quality\_Index\_for\_Drinking\_Water/links/5923fd63aca27295a8aad7c1/Evaluation-of-Water-Quality-Index-for-Drinking-Water.pdf

**KEY SOURCE FOR POTABILITY:**

<https://www.epa.gov/wqc/national-recommended-water-quality-criteria-human-health-criteria-table>

* **National Recommended Water Quality Criteria - Human Health Criteria Table**

**Great Resource:** <https://environmentalsystemsresearch.springeropen.com/articles/10.1186/s40068-016-0053-6>

**What are Water Quality Standards? EPA:**

https://www.epa.gov/standards-water-body-health/what-are-water-quality-standards

**CDC:** https://www.cdc.gov/healthywater/drinking/public/regulations.html

# Key Measures - Water Quality:

### **What is a Fresh Water Quality Index?**

The Freshwater Quality Index (WQI) is a tool developed by scientists to help evaluate the quality of water in these streams and rivers. It summarizes large amounts of water quality data into a single "score" from 1 to 100. Higher scores reflect cleaner water.

**RESEARCH PAPER ON WQI:** [**http://pubs.sciepub.com/ajwr/1/3/3/index.html**](http://pubs.sciepub.com/ajwr/1/3/3/index.html)

**WATER QUALITY CLASSIFICATION:**

<https://www.dec.ny.gov/chemical/23853.html>

# WQI and WQC Calculation:

# Context/ Definitions:

**PH**:

* Measure of Water Acidity: pHs of less than 7 indicate acidity, whereas a pH of greater than 7 indicates a base.
* **Potability**: pH level of water sources should be at a pH measurement level between 6.5 to 8.5 on a scale that ranges from 0 to 14. The best pH of drinking water sits right in the middle at a 7.
  + Water that has a pH range outside of this recommendation may be a sign that the water contains heavy, toxic metals or chemicals due to a higher pH level

**Hardness:**

* Water hardness refers to the amount of minerals, particularly, of Calcium and Magnesium, found in water. || Capacity of water to precipitate soap in mg/L. Minerals are naturally found in water; however, when they exceed 120mg/L, the high level of minerals begins to have negative effects. This is also when water is called hard, as opposed to being soft water, with only up to 60mg/L of minerals, or moderately hard water, with 60mg/L and 120mg/L amount of minerals.
* **Potability**: The general rule of thumb is to drink clean water, with hardness being somewhere in the middle of soft and hard, 60 mg/L to 120 mg/L. Some also advise to not go beyond 170 mg/L, which indicates very high levels of calcium and magnesium.
* But according to health authorities, there is no general health advice on restricting hardness in drinking water for health and safety reasons.

**Solids (ppm == mg/L): Total Dissolved Solids [TDS]**

* **Potability:** 500 ppm is the recommended maximum amount of TDS for your drinking water. Any measurement higher than 1000 ppm is an unsafe level of TDS. If the level exceeds 2000 ppm, then a filtration system may be unable to properly filter TDS.
  + The water with high TDS value indicates that water is highly mineralized. Desirable limit for TDS is 500 mg/l and maximum limit is 1000 mg/l which prescribed for drinking purpose
* a TDS meter says 100 ppm, which means that from one million particles, 100 are dissolved ions and 999,900 are water molecules.
* High values of TDS in ground water are generally not harmful to human beings, but high concentration of these may affect persons who are suffering from kidney and heart diseases. Water containing high solid may cause laxative or constipation effects

**Sulfate:**

* Cathartic effects are commonly reported to be experienced by people consuming drinking-water containing sulfate in concentrations exceeding 600 mg/litre although it is also reported that humans can adapt to higher concentrations with time (US EPA, 1985). Dehydration has also been reported as a common side-effect following the ingestion of large amounts of magnesium or sodium sulfate (Fingl, 1980).
* Potability: Sulfate levels above 250 mg/L may make the water taste bitter or like medicine. High sulfate levels may also corrode plumbing, particularly copper piping. In areas with high sulfate levels, plumbing materials more resistant to corrosion, such as plastic pipe, are commonly used.
* Sulfate mainly is derived from the dissolution of salts of sulfuric acid and abundantly found in almost all water bodies. High concentration of sulfate may be due to oxidation of pyrite and mine drainage etc. Sulfate concentration in natural water ranges from a few to a several 100 mg/liter, but no major negative impact of sulfate on human health is reported. The WHO has established 250 mg/l as the highest desirable limit of sulfate in drinking water.

**Conductivity:**

* Conductivity is a measure of water's ability to pass an electrical current.
* Conductivity is affected by the presence of inorganic dissolved solids (ions) and temperature. Pure water does not possess a lot of conductivity but as more ions are added the conductivity increases. The more ions (inorganic dissolved solids such as sodium, chloride, nitrate, phosphate, sulfate, magnesium, calcium, iron, and aluminum) that are present in water the higher its conductivity.
* Similar to ions, the warmer water is the higher its conductivity. As such, conductivity is reported at 25° C. This is known as specific conductance.
* Higher conductivity should be negatively related to potability

**Trihalomethanes [Chlorine byproducts]**:

* Chlorine is used to disinfect drinking water. This prevents regrowth of microorganisms, or contamination from an outside source, such as during a water main break.
* Total trihalomethanes (TTHM) are a group of disinfection byproducts that form when chlorine compounds that are used to disinfect water react with other naturally occurring chemicals in the water. They are colorless, and will evaporate out of the water into the air.
* **Potability**: The USEPA and MassDEP have set an MCL for TTHM of 80 parts per billion (ppb) or micrograms per liter (ug/L) as an annual average.
* THM are considered to be possibly carcinogenic to humans by USEPA because of evidence of carcinogenicity in experimental laboratory animals and limited evidence in people. Some of the individual chemicals that comprise TTHM have also caused other effects in experimental laboratory animals following high levels of exposure, including toxicity to the liver, kidneys, neurological and reproductive systems.

**Turbidity**: The turbidity of water depends on the quantity of solid matter present in the suspended state. It is a measure of light emitting properties of water and the test is used to indicate the quality of waste discharge with respect to colloidal matter. The mean turbidity value obtained for Wondo Genet Campus (0.98 NTU) is lower than the WHO recommended value of 5.00 NTU.

**Organic Carbon**: The water quality criteria for total organic carbon are 2 mg/L for treated water and 4 mg/L for source water. The criteria should not be exceeded at any time in drinking water systems that use chlorination for disinfection.

**Potability:**

**B.O.D:** Represents the amount of oxygen consumed by bacteria and other microorganisms while they decompose organic matter under aerobic (oxygen is present) conditions.

**Fecal coliform:** The presence of fecal coliform bacteria in aquatic environments indicates that the water has been contaminated with the fecal material of humans or other animals.

* The total coliform group has been selected as the primary indicator bacteria for the presence of disease causing organisms in drinking water. It is a primary indicator of suitability of water for consumption. **If large numbers of coliforms are found in water, there is a high probability that other pathogenic bacteria or organisms exis**t. The WHO and Ethiopian drinking water guidelines require the absence of total coliform in public drinking water supplies.

**From:** https://www.datascience2000.in/2021/10/water-quality-prediction-using-machine.html

1. pH value:

PH is an important parameter in evaluating the acid–base balance of water. It is also the indicator of acidic or alkaline condition of water status. WHO has recommended a maximum permissible limit of pH from 6.5 to 8.5. The current investigation ranges were 6.52–6.83 which are in the range of WHO standards.

2. Hardness:

Hardness is mainly caused by calcium and magnesium salts. These salts are dissolved from geologic deposits through which water travels. The length of time water is in contact with hardness producing material helps determine how much hardness there is in raw water. Hardness was originally defined as the capacity of water to precipitate soap caused by Calcium and Magnesium.

3. Solids (Total dissolved solids - TDS):

Water has the ability to dissolve a wide range of inorganic and some organic minerals or salts such as potassium, calcium, sodium, bicarbonates, chlorides, magnesium, sulfates etc. These minerals produced an unwanted taste and diluted color in the appearance of water. This is the important parameter for the use of water. The water with high TDS value indicates that water is highly mineralized. The Desired limit for TDS is 500 mg/l and maximum limit is 1000 mg/l which is prescribed for drinking purposes.

4. Chloramines:

Chlorine and chloramine are the major disinfectants used in public water systems. Chloramines are most commonly formed when ammonia is added to chlorine to treat drinking water. Chlorine levels up to 4 milligrams per liter (mg/L or 4 parts per million (ppm)) are considered safe in drinking water.

5. Sulfate:

Sulfates are naturally occurring substances that are found in minerals, soil, and rocks. They are present in ambient air, groundwater, plants, and food. The principal commercial use of sulfate is in the chemical industry. Sulfate concentration in seawater is about 2,700 milligrams per liter (mg/L). It ranges from 3 to 30 mg/L in most freshwater supplies, although much higher concentrations (1000 mg/L) are found in some geographic locations.

6. Conductivity:

Pure water is not a good conductor of electric current rather it's a good insulator. Increase in ions concentration enhances the electrical conductivity of water. Generally, the amount of dissolved solids in water determines the electrical conductivity. Electrical conductivity (EC) actually measures the ionic process of a solution that enables it to transmit current. According to WHO standards, EC value should not exceed 400 μS/cm.

7. Organic\_carbon:

Total Organic Carbon (TOC) in source waters comes from decaying natural organic matter (NOM) as well as synthetic sources. TOC is a measure of the total amount of carbon in organic compounds in pure water. According to the US EPA < 2 mg/L as TOC in treated / drinking water, and < 4 mg/Lit in source water which is used for treatment.

8. Trihalomethanes:

THMs are chemicals which may be found in water treated with chlorine. The concentration of THMs in drinking water varies according to the level of organic material in the water, the amount of chlorine required to treat the water, and the temperature of the water that is being treated. THM levels up to 80 ppm is considered safe in drinking water.

9. Turbidity:

The turbidity of water depends on the quantity of solid matter present in the suspended state. It is a measure of light emitting properties of water and the test is used to indicate the quality of waste discharge with respect to colloidal matter. The mean turbidity value obtained for Wondo Genet Campus (0.98 NTU) is lower than the WHO recommended value of 5.00 NTU.

10. Potability:

Indicates if water is safe for human consumption where 1 means Potable and 0 means Not potable. (0) Water is not safe to drink and (1) Water is safe to drink.