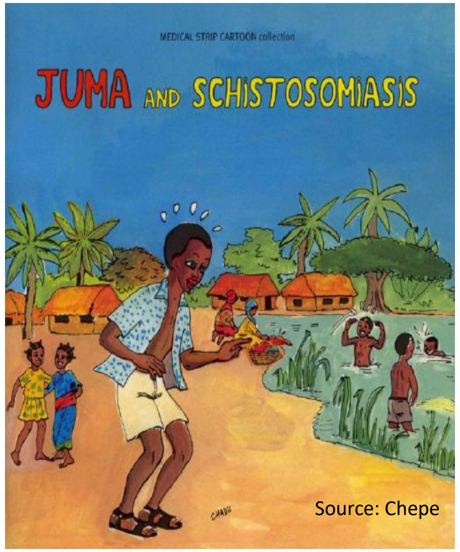


Schistosomiasis prevention undermined by water point forecasting

Prof. Cheikh Ahmadou Bamba GUEYE
Université Cheikh Anta Diop, Dakar, Senegal

Trieste – November 2024 – ICTP

Joint work with my students
Fiorenantsoa Santatra, Bassirou Kassé, Gildas Zougmoré

Roadmap

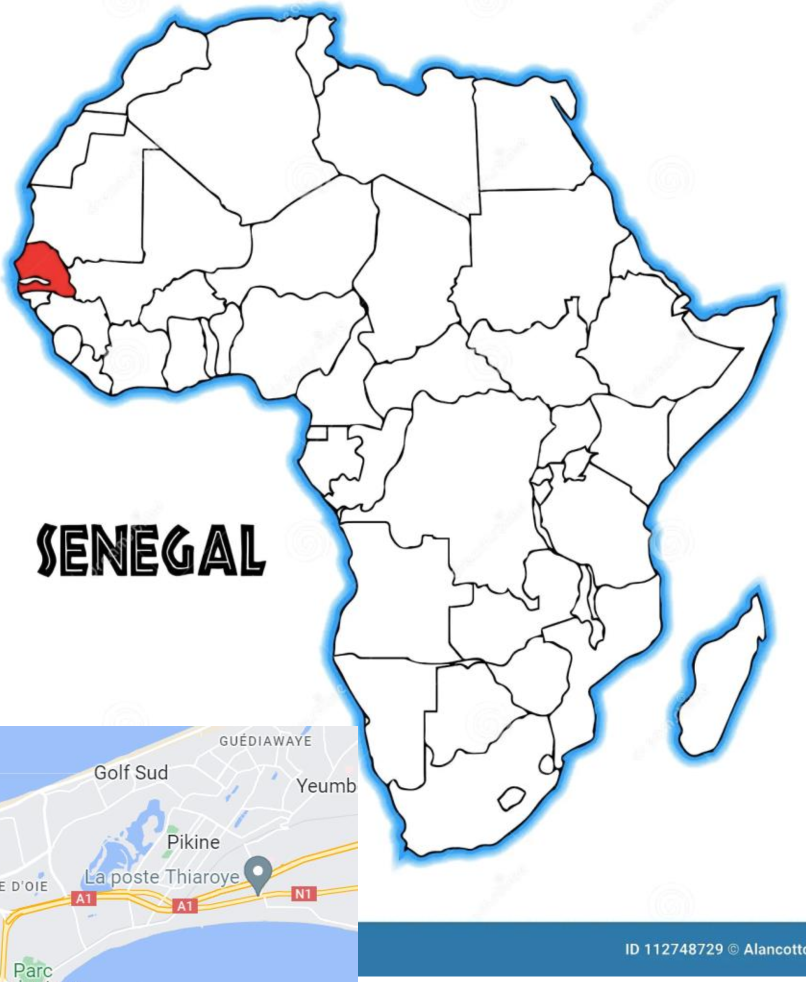
- Introduction
- Schistosomiasis overview
- Infection and transmission
- IoT-based physico-chemical water points monitoring
- Data fusion based on snails control and environmental management

UCAD in a nutshell

- Université Cheikh Anta Diop (UCAD) is the largest university in french-speaking West Africa

- UCAD has more than 100,000 students, 1412 university lecturers and researchers, 1500 technical staff, 6 faculties and 8 institutes

- UCAD stands for West Africa’s excellence in science and medicine

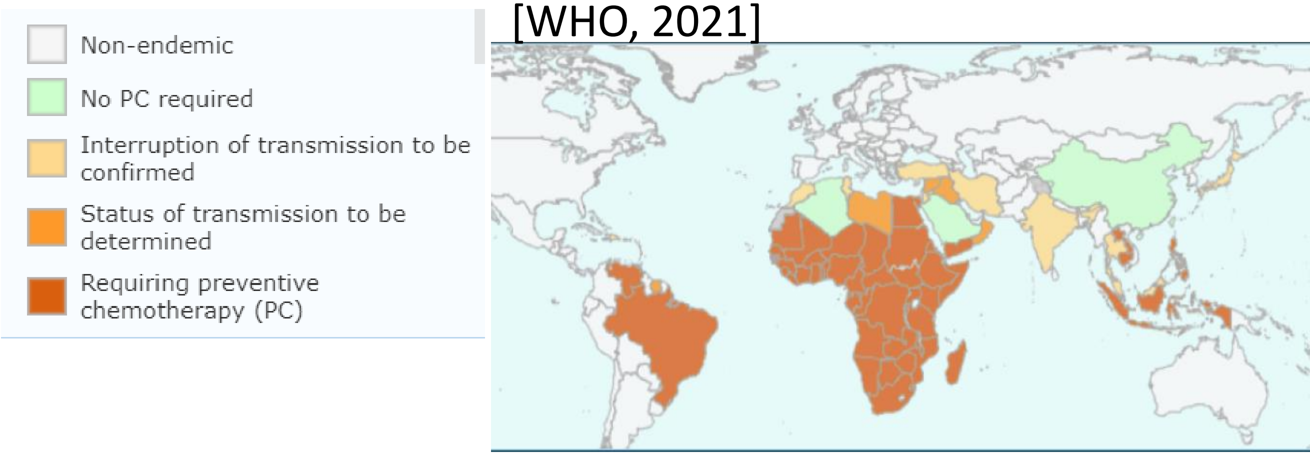


Networking Services & Telecommunications Team

- Related topics
  - Improving sustainability concept in underserved areas
  - Internet of Things
  - Self-organized systems
- Selected projects
  - Smart agriculture, health, and livestock in developing countries in West Africa
  - FerloNet : sense, collect and disseminate information related to pastoral and other activities in Senegal
  - COWShED: COmmunication within White Spots for brEeDers

Schistosomiasis around the world

[WHO, 2021]



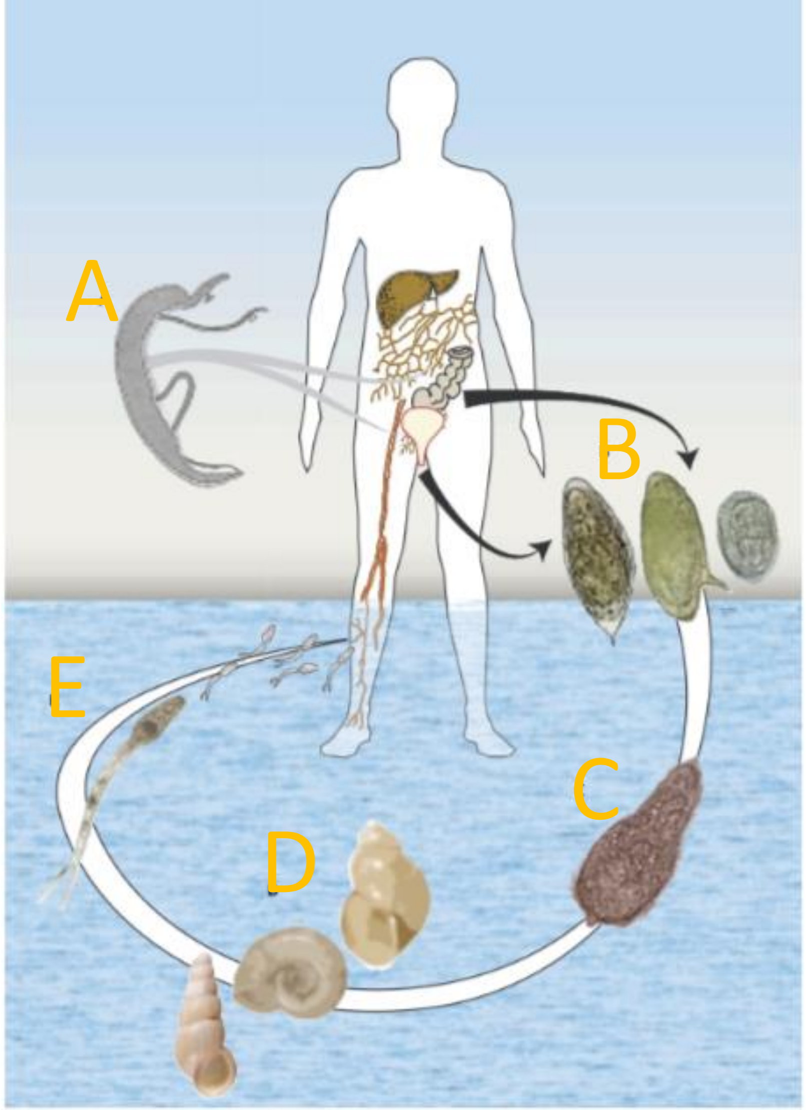
- **Schistosomiasis** or **bilharzia** affects more than 251.2 million people distributed over 78 world countries [WHO, February 2023]
- A parasitic disease caused by parasitic worms transmitted by snails living in warm water
- In sub-Saharan Africa, 207.2 million people are affected which represent roughly 82.5% of the people reported to be infected all over the world
- More than 700 million people live in endemic areas
- Substantial socioeconomic impacts on impoverished communities

Senegal’s context

- The prevalence rate varies between 0.3% and 1%
- According to Senegal, after malaria, schistosomiasis is the second disease that calls for admission to hospital
- Spread of hydraulic developments (river dam, irrigation, …)
- Daily life activities in rural areas are done around water points
- Prevalence peaks of over 80% in the north, whereas, 50% in the center of the country in seasonal transmission hotspot

Senegal river (1800 km)

Schistosomiasis transmission life cycle



Source: Gryseels et al.

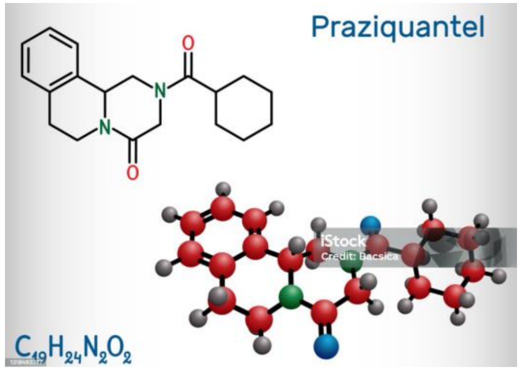
A. Cercariae transform into males or females **schistosomes**, pair up and then migrate to their region of the venous blood system. Females leave the males and moves to smaller venules closer to the lumen of the intestine or bladder to lay her eggs (6 weeks after infection)

B. Eggs (viable up 1 week) are excreted in the urine or faeces of an infected person

C. Eggs that reach fresh water hatch, releasing a **miracidium** which, to develop further must infect a snail of the correct species within 24h

D. Asexual multiplication takes place in **intermediate hosts snail**, and results in the release of cercariae into water about 4–6 weeks later

E. Cercariae actively swim around and can penetrate actively a skin through proteolytic secretion of a definitive host up to 72 h



Drug administration

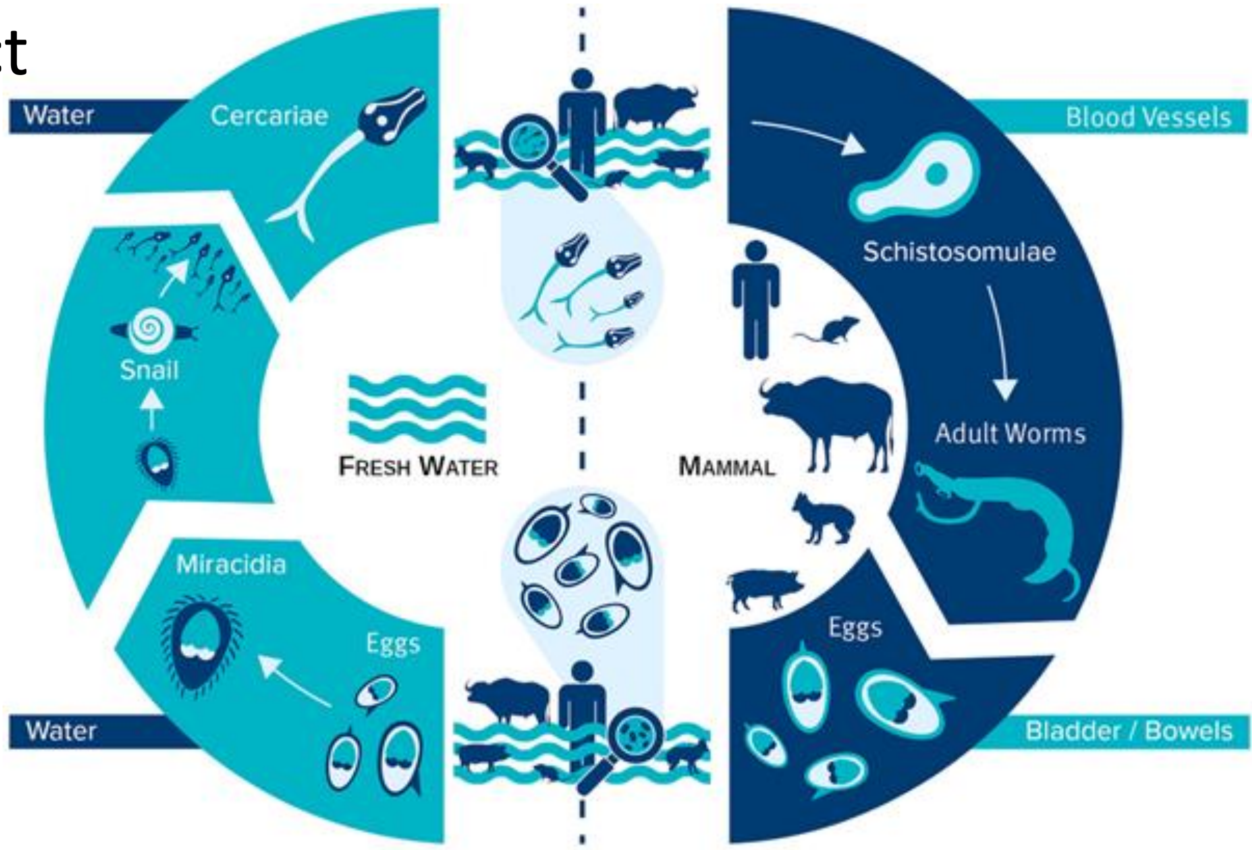
- Praziquantel (PZQ) is the main strategy for controlling Schistosoma haematobium and Schistosoma mansoni, the species responsible for the urogenital and intestinal forms of the disease, respectively
- Hybrid parasite strains between human and animal species potentially resistant to PZQ
- Lack of awareness of the parasite reservoir wrt children under 5 years
- Without symptoms and morbidity, rural people that have Positive serology do not go to hospital
    - Human host the virus and participate actively in schistosomiasis spreading
- Lack of information in real time of infected water points
- The prevalence is still high, jeopardizing the elimination goal

Motivations

- Consider Internet of Things based applications that enable continuous and remote monitoring of water points
- Assess the impact of climate change on the dynamic behavior and population densities of snails vectors
- Leverage machine learning tools to detect anomalies that deviate from common statistical properties of a distribution and then trigger alerts
- Develop decision tool for the early detection and prevention of schistosomiasis in different transmission settings

At the beginning … (2016)

- Sensor-based Bilharzia Detection Project funded by IDRC
- Team partners:
  - Halima Elbiaze (UQAM, Canada)
  - Moussa Diallo (ESP, Senegal)
  - Bassirou Kasse (UCAD, Senegal)
  - Ibrahima Niang (UCAD, Senegal)
  - El Kabir Lo (UCAD, Senegal)



Adapted from graphic development by SCI foundation, originally adapted from illustration by Genome Research Limited.

Favorable factors for bilharzia transmission life cycle

- Bilharzia life cycle is based on physical and chemical factors
  - Physical factors include temperature, solar irradiance, water movement, water-level fluctuation and desiccation, depth of water
  - Chemical factors, such as salinity, ion balance, hydrogen concentration, are measured within water points where intermediate hosts that transmit bilharzia are living
- Water **temperature**, water **pH**, **solar irradiance**, have significant effect on **freshwater snails** population dynamics [McCreesh et al.]
- Diurnal **light intensity** has a great impact on **eggs maturation** [World Health Organization]
- It exists a correlation between **light intensity** and **snails breeding** [Safaa Al-Asadi]
- Solar irradiation in Senegal is roughly equals to 242 $\mu mol/m^2/s$ which represents a solar irradiance of 5.8 $kWh/m^2/day$

Favorable factors for bilharzia transmission life cycle (suite)

- Optimum water pH for Bulinus snails (resp. Biomphalaria) lies between 6.0 to 6.5 (resp. 7.0 to 8.2)
- Optimal temperature that enables favorable snails breeding's (oviposition, larvae, youth) is measured between 25°C to 28°C
- By measuring fixed parameters, we can whenever determine:
  - if the condition are favorable for eggs maturation
  - whether intermediate hosts freshwater snails be alive, growth up optimally
  - even if successful breeding of snails is possible

Anomalies detection test

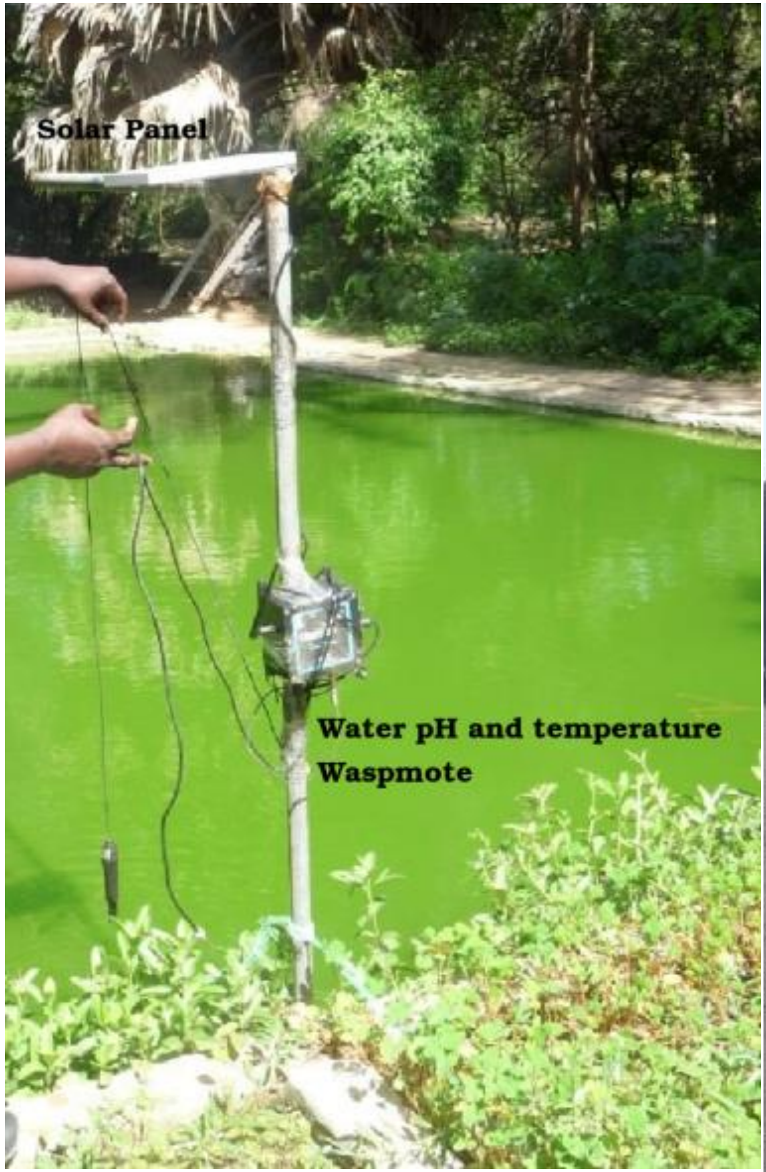
- True positives (resp. true negatives) are positives (resp. negatives) that have been correctly reported
- A **false negative** refers to a wrongly unsuspected sample
- A **false positive** is a sample that has been wrongly suspected
- **False Positive Rate** (FPR) is the proportion of all samples that have been wrongly reported as positive

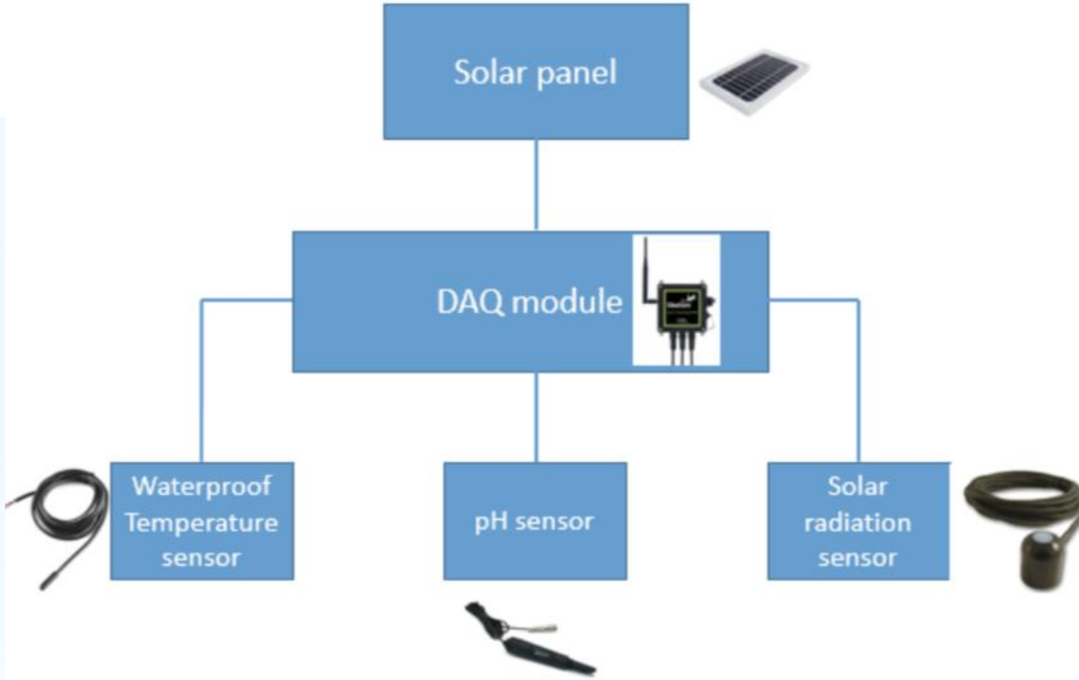
$$\mathcal{FPR} = \mathcal{T}_{\mathcal{FP}} / (\mathcal{T}_{\mathcal{TN}} + \mathcal{T}_{\mathcal{FP}})$$

- **True Positive Rate** (TPR) is the proportion of samples that have been rightly reported as anomalies

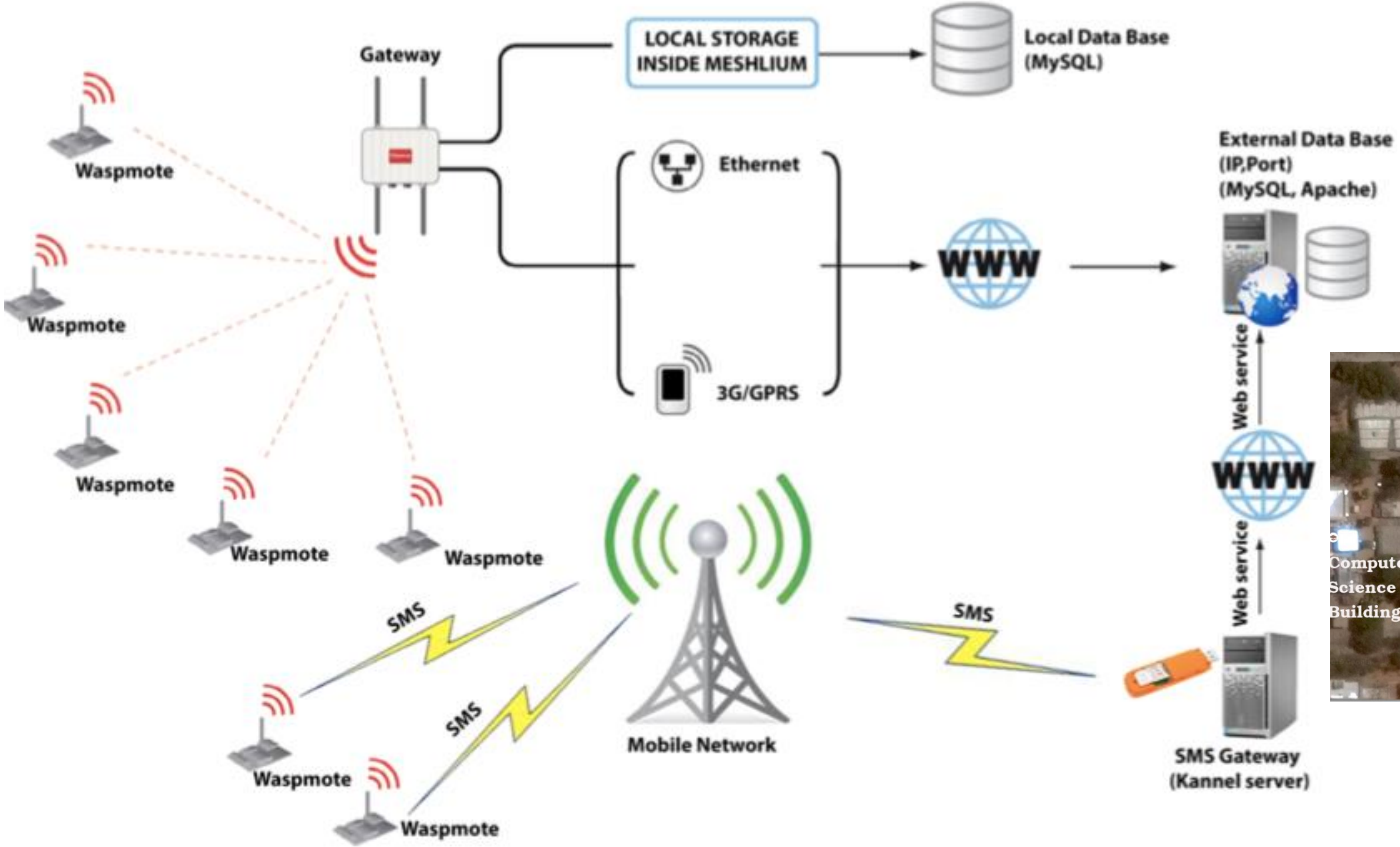
$$\mathcal{TPR} = \mathcal{T}_{\mathcal{TP}} / (\mathcal{T}_{\mathcal{FN}} + \mathcal{T}_{\mathcal{TP}})$$

Data acquisition module





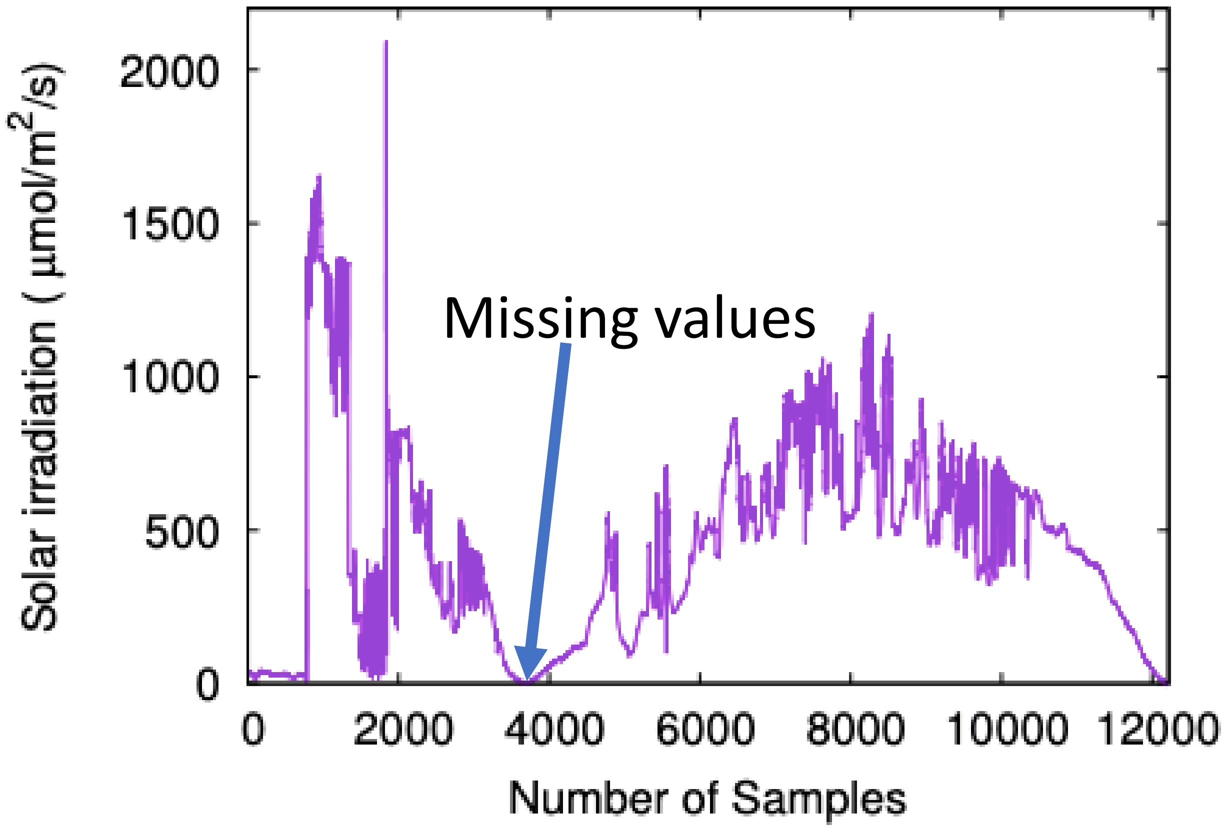
Sensors-Based Bilharzia Detection architecture (2017)

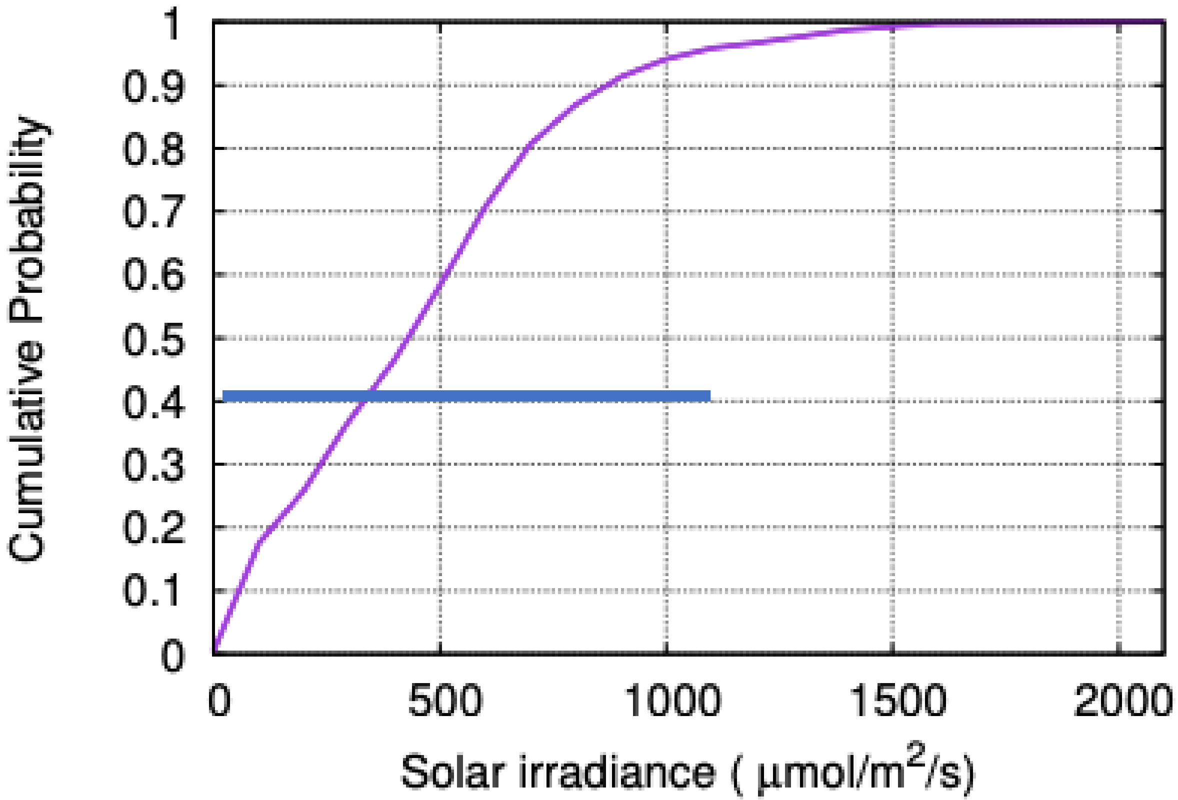


Experimental settings

- ZigBee radio transmission between *Waspmote* and gateway
- *Waspmote* device embeds an *ATmega1281* microcrontroller running at 14 MHz, a Xbee-ZB-pro transmitter, sensors
- Frame size is equal to 128 octets and contains ID's *Waspmote*, frame type, frame number, type of sensor, measured value, battery voltage, timestamp, etc.
- Sampling interval is fixed to 1 mn (resp. 2 mn) for Water pH and temperature (resp. solar irradiance) during 2 weeks
- Means water pH, water temperature, solar irradiance are computed within a fixed interval of 10 minutes
- 2160 samples are obtained and clustered within different sliding windows equal to either 1h, or 2 h, or 3h, or 6 h
- R language and environment is used

Solar irradiance parameters

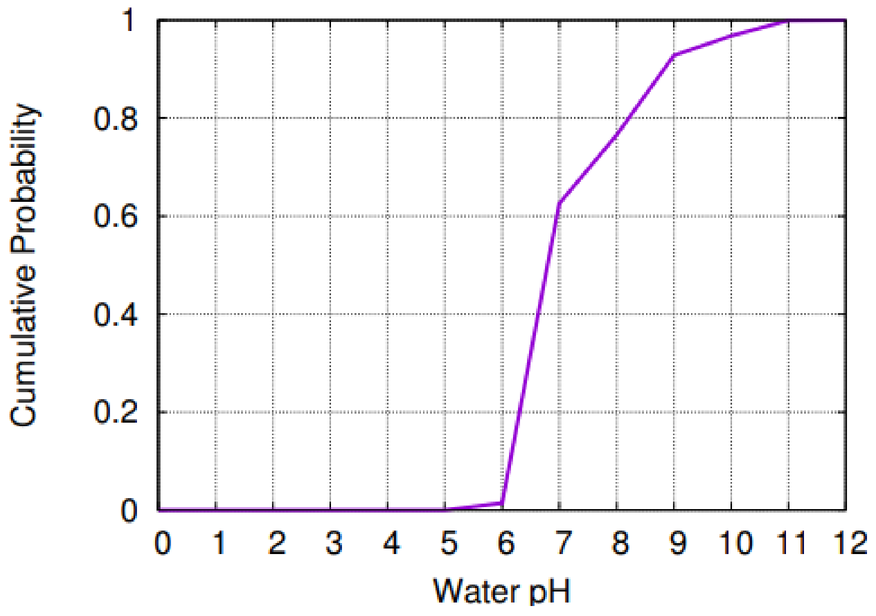


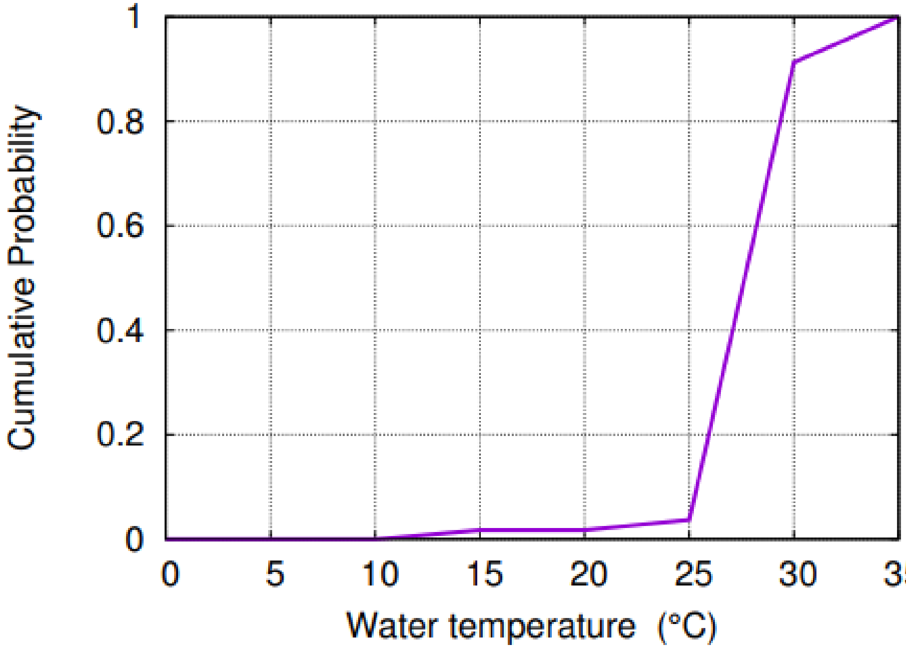


- More than 40% of measurements have a irradiation upper than 500 µmol/m2/s
- Very high light intensity

Temperature and pH results

- 20% samples have a pH value upper than 8 (6.5 to 8.5 are favorable)
- 40% of samples are upper than 28°C which means an unfavourable breeding conditions (25°C to 28°C are optimal)
- A couple of days have favorable conditions for snails breeding's





Performance of anomalies detection technique

(a) Random Forest (RF)

Sliding windows (Data training set collection)	1 hour	2 hours	3 hours	6 hours
$\mathcal{FPR}$ (%)	0	0	0	0
$\mathcal{TPR}$ (%)	88.89	100	100	100

(b) Artificial Neural Network (ANN)

Sliding windows (Data training set collection)	1 hour	2 hours	3 hours	6 hours
$\mathcal{FPR}$ (%)	5.7	11.93	9.32	6.9
$\mathcal{TPR}$ (%)	100	100	100	100

(c) Support Vector Machines (SVM)

Sliding windows (Data training set collection)	1 hour	2 hours	3 hours	6 hours
$\mathcal{FPR}$ (%)	0	0	0.84	0
$\mathcal{TPR}$ (%)	100	100	100	100

In case of high FPR, water point will be unjustifiably in quarantine, and thus not used by rural population according to their day-to-day life activities

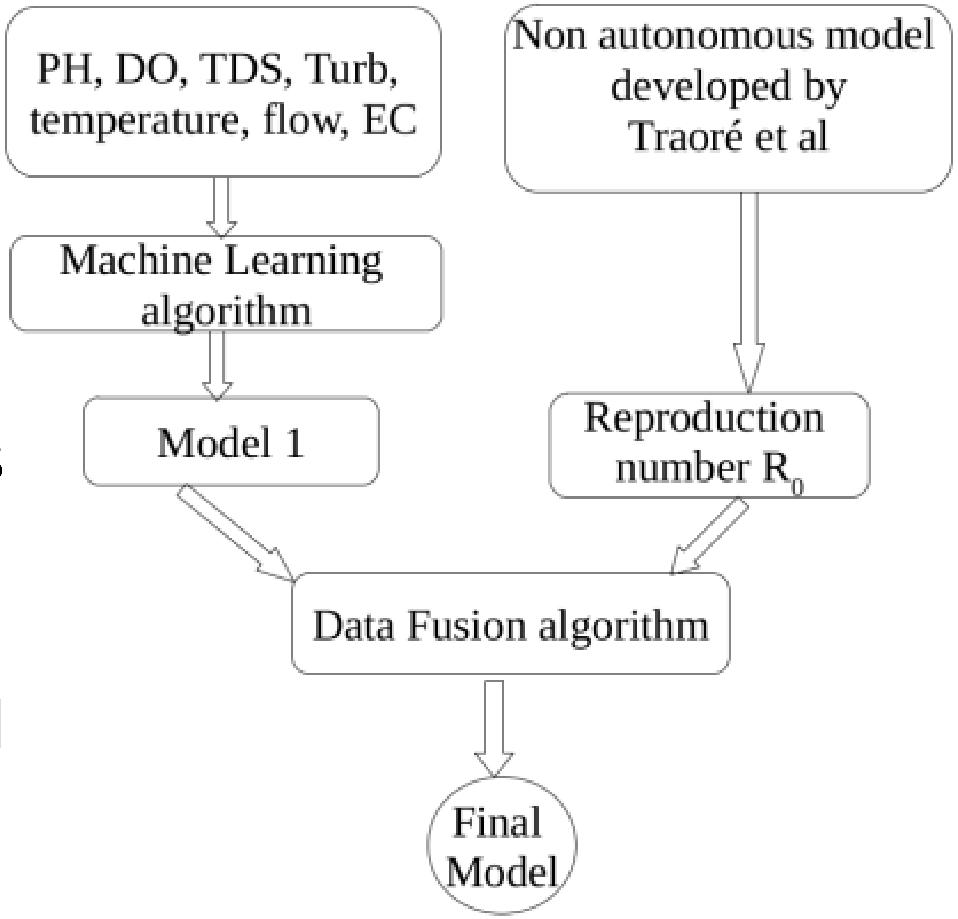
- SVM is able to detect 100% of anomalies according to over all sliding windows
- SVM triggers 0.84% of false alarms wrt a sliding window of 3 hours
- ANN test has wrongly reported as positives several samples according to overall considered sliding window
- RF test exhibits good detection rate but 11.11% of anomalies are not detected wrt a sliding window of 1 hour

Limitations

- Possible skewness in our forecasting approach due to lack of handling outliers or missing values
- Lack of awareness on the dynamic behavior and population densities of snail vectors and parasitic worms
- A single learning algorithm was considered

FuMalMMO: Fusion of Machine learning and Mathematical Models

- Leveraging information coming from different sources for decision making
- Consider machine learning model for real-time water point forecasting
- Epidemiological model to figure out the parasites density evolution and snails
- Use theory of belief functions (evidence theory), which assesses a “subjective judgment” provided by both models, to refine and improve models outputs



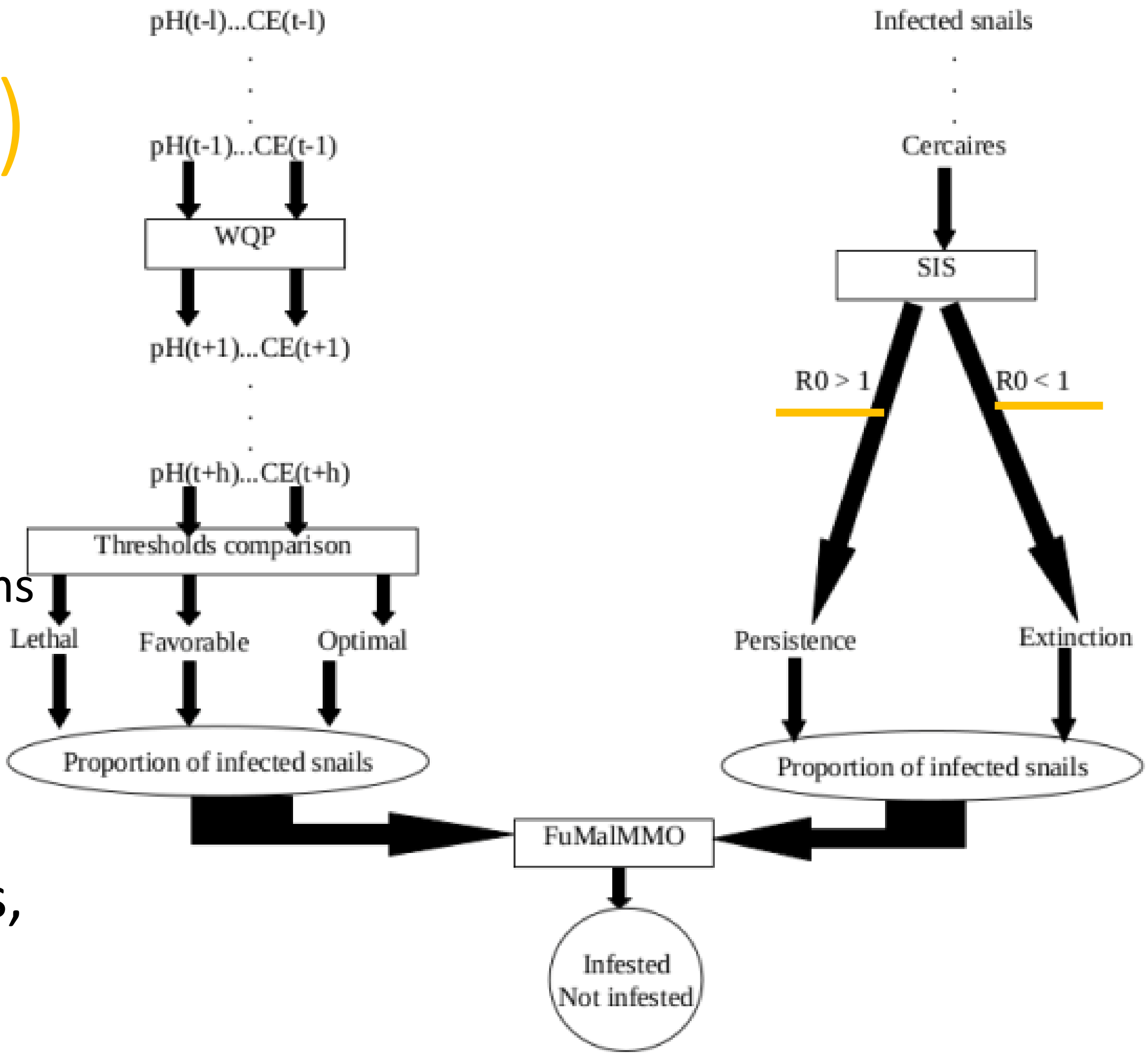
Data fusion approach

$$\begin{bmatrix} & S_1 & S_2 & ... & S_m \\ d_1 & M_1^1(x) & M_2^1(x) & ... & M_m^1(x) \\ d_2 & M_1^2(x) & M_2^2(x) & ... & M_m^2(x) \\ ... & ... & ... & ... & ... \\ d_n & M_1^n(x) & M_2^n(x) & ... & M_m^n(x) \end{bmatrix}$$

- $M_j^i(x)$ depicts information provided by S_j which argues decision d_i about observed parameter « x » like freshwater source status
- Final decision is taken based on $M_j^i(x)$ combination
- $M_j^i(x)$ is represented by a mass function (theory of belief)
- From a mass function, we can derive other functions:
  - Credibility which represents the total mass of belief wrt hypothesis or a set of hypotheses
  - Plausibility which exhibits the maximum belief wrt hypothesis or a set of hypotheses

Epidemiological data (SIS)

- Characterize water points into 3 categories
  - Lethal: parasitic and snails live is impossible
  - Favorable: physicochemical parameters are favorable but do not promote the multiplication of parasitic and snails
  - Optimal: favorable snails breeding's and parasitic worms
- Leverage the "basic reproduction number (**R_0**)" which models dynamic behavior of infected population and governs the system [1]
- A mathematical model with compartments which takes as input the numbers of human populations, parasites (miracidia and cercariae) and intermediate hosts (snails)



[1] Traoré et al., "Global dynamics of a seasonal mathematical model of schistosomiasis transmission with general incidence function", Journal of Biological Systems, 2019

Water Quality Prediction

- A novel data-driven approach, using the Long Short-Term Memory (LSTM) network, a special type of recurrent neural network
  - LSTM is its ability to learn long-term dependencies
- SVR (Support Vector Regression) enables to predict values with a fixed gap
- Wavelet Artificial Neural Network (WANN) is an hybrid technique which employs discrete wavelet transform and artificial neural network for time series forecasting
  - Feed-forward neural network with back-propagation is considered as ANN
- Algorithms were validated and tested by applying criteria such as Root Mean Squared Error (RMSE) (for error reducing) and Mean Absolute Error (MAE) (reduce outliers impact on model's result)

Background on evidence theory

- For each output, of the WQP and SIS, we defined a mass function
- d1 (resp. d2) means how the model believes the water is infested (resp. not infested) ; d(1,2) ignorance about water point status

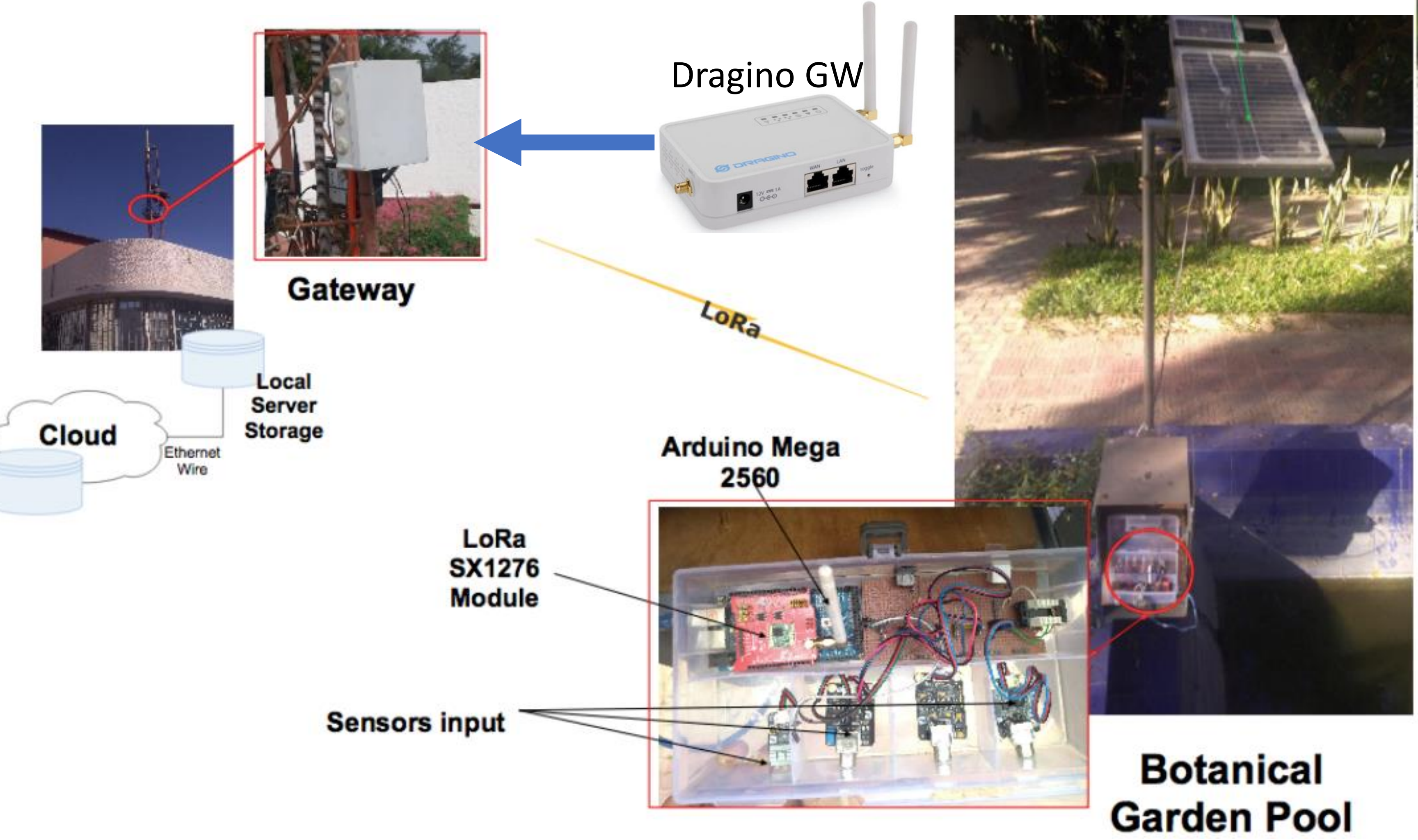
$$\{m_{wqp}(d_1), m_{wqp}(d_2), m_{wqp}(d_{1,2})\} \qquad \{m_{sis}(d_1), m_{sis}(d_2), m_{sis}(d_{1,2})\}$$

- Estimate the mass function, for each output (WPQ and SIS), by considering the infected snails proportion
- Determine a discounting coefficient of the mass functions since discrepancies may exist between predicted and expected values
- Combine mass function and provide final decision based on the maximum plausibility criteria

$$m_{comb}(A) = (m_1 \oplus m_2)(A) = \frac{1}{1-k} \sum_{B_1 \cap B_2 = A} m_1(B_1) m_2(B_2)$$

Gildas et al.,"Schistosomiasis Prevention Undermined by Water Point Forecasting: An Approach Based on Data Fusion Model, [ICRTCIS 2023]

Panamasso river (Burkina Faso) deployment



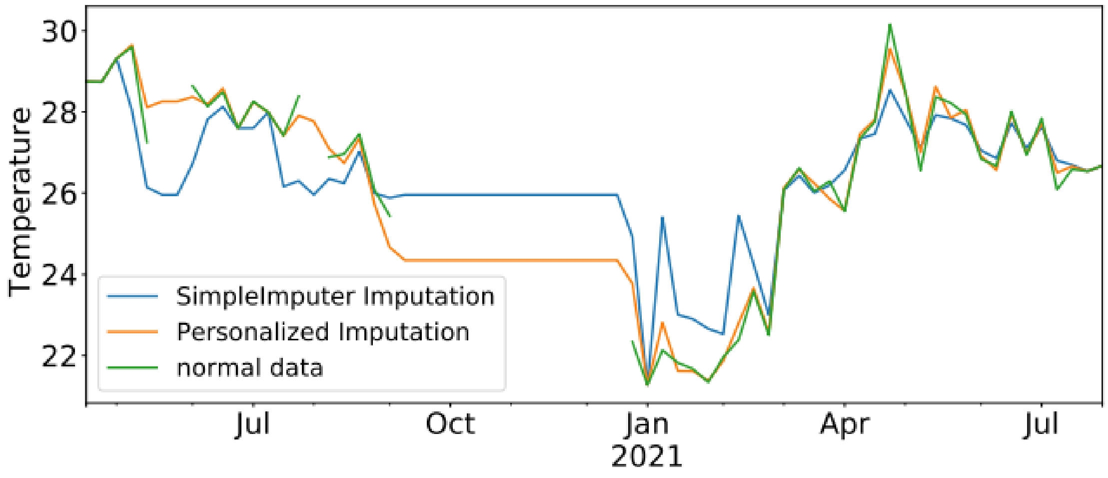
Experimental settings

- Prevalence rate is estimated at 27.47% over 3065 inhabitants
- The malacological study consisted of collecting snails of the genus Biomphalaria
- Epidemiological data collection: cold (December 2016 - February 2017) and hot (March 2017 - May 2017)
- pH, electrical conductivity, temperature, oxygen dissolved are measured, each 10mn, during April 2020 to July 2021
- Sliding windows of 1 to 7 days to figure out previous values as well as predict the future ones
- 3 layers considered for WANN and LSTM
- Number of nodes of layers (input, output) is dynamic through forecast horizon
- Training set : april 2020 to march 2021
- Validation set: march 2021 to April 2021
- Test set: may 2021 to July 2021

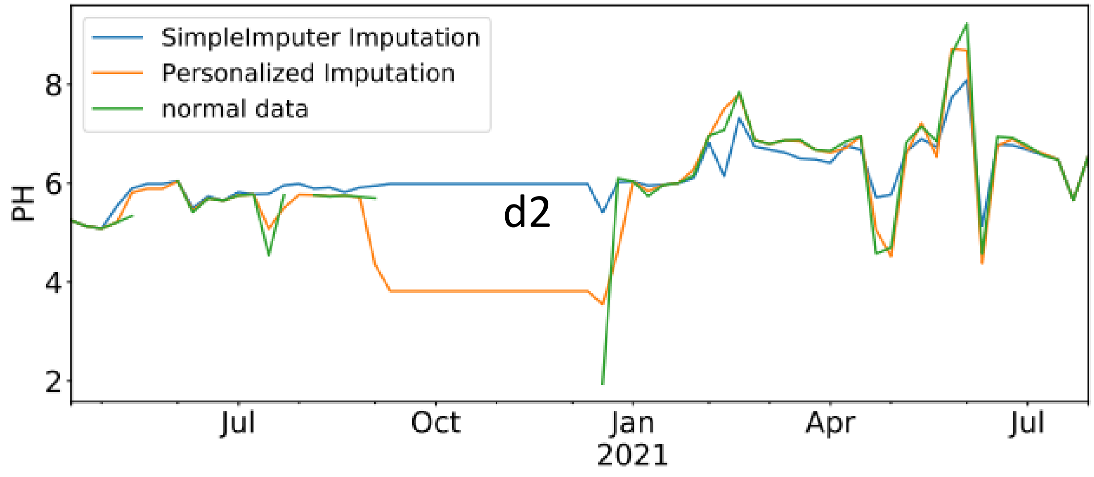
Imputation of missing values

- We tried mean imputation through SimpleImputer class from scikit-learn

- Replace missing values by an average value calculated with the two values separating each missing period in the series



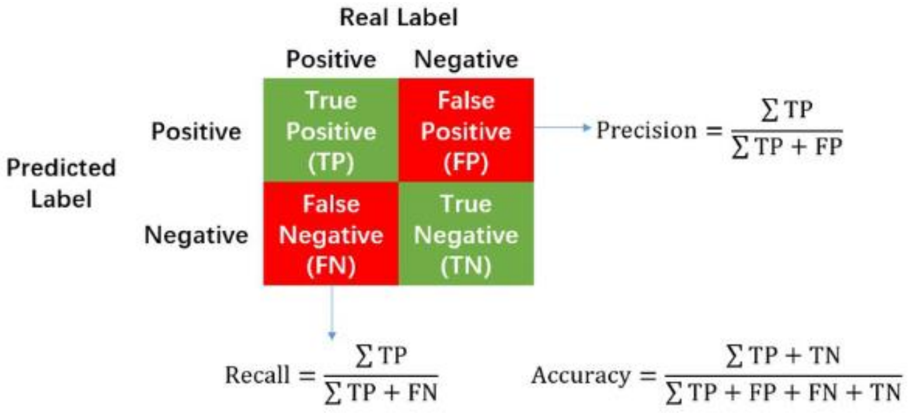
(a) Temperature Data

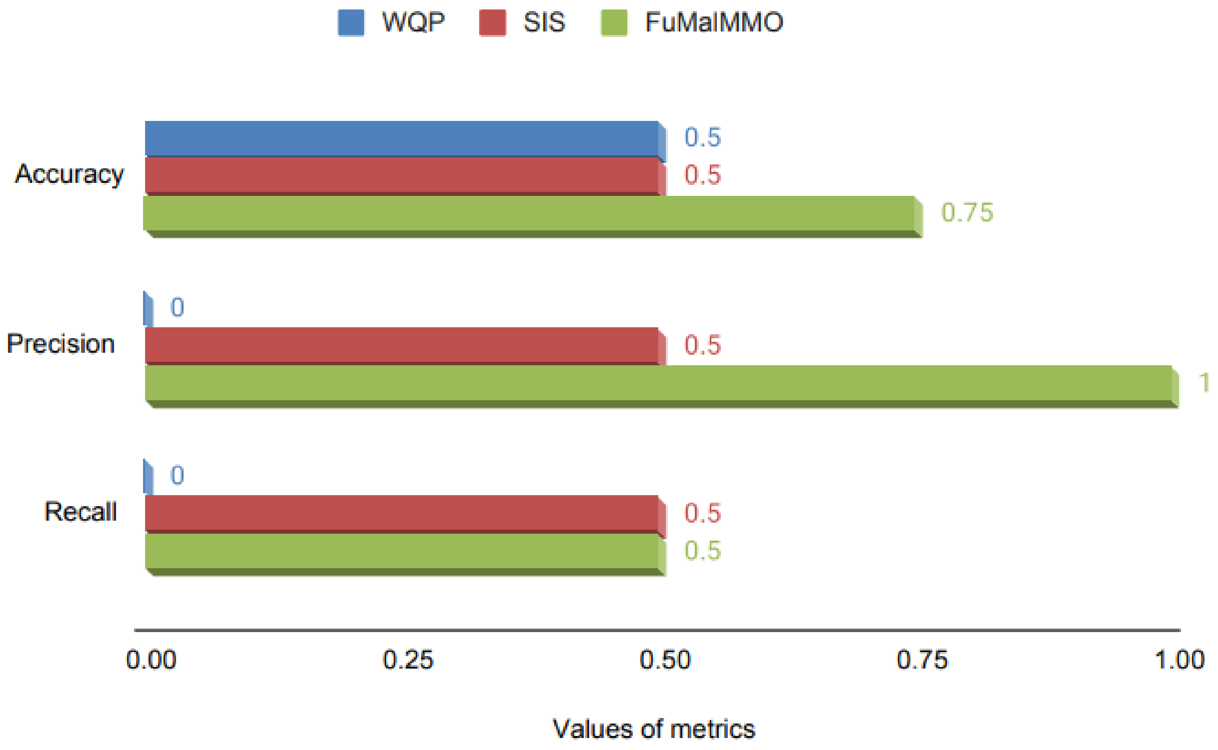


(b) PH Data

Comparison of FuMalMMo against WQP and SIS

- "Accuracy" show how often FuMalMMo classification model is correct overall
- "Precision" counts the proportion of relevant items among the selected items
- "Recall" counts the proportion of relevant items selected from all selectable relevant items
- Each model taken separately has modest performances compared to the data fusion approach
- The fusion approach leads to an efficient model to warn in one day ahead that a water point is likely to be infested with parasites causing schistosomiasis





Learned lessons

- High temperature breaks network devices
- Missing data due to stolen or disconnected sensors
- WANN model outperforms SVR and LTSM based on raw dataset collected at Panamasso, Burkina Faso
- IoT enables to assess the impact of climate change on the dynamic behavior and population densities of snail vectors
- Leverage data fusion to earlier assess whether a given water point harbors infected snails

Future work

- Use DNA biosensors to detect proteolytic enzyme that can be excreted either by miracidium or cercariae
- Generate data cleaning pipelines to check and resolve data quality issues
- Holistic approach will integrate genomic, serology, environmental and behavior approaches
- Deploy a distributed data collection, transmission, storage and visualization infrastructure along Senegal river and other endemic areas

References

- Teegwende Gildas Zougmore, Malo Sadouanouan, Bamba Gueye,"Schistosomiasis prevention undermined by water point forecasting : An approach based on data fusion mode" [2023]
- Teegwende Gildas Zougmore, Bamba Gueye, Malo Sadouanouan," An AI-Based Approach to the Prediction of Water Points Quality Indicators for Schistosomiasis Prevention" [2022]
- Bassirou Kassé, Fiorenantsoa Santatra, Bamba Gueye, Moussa Diallo, Halima Elbiaze,, "IoT based Schistosomiasis Monitoring for More Efficient Disease Prediction and Control Model" [2019]
- Bassirou Kassé, Moussa Diallo, Bamba Gueye, Halima Elbiaze, "SBSD: Towards a Proactive Sensor-Based Shistosomiasis Detection"[2017]