# DATA VALIDATION AND RECONSTRUCTION FOR WATER DISTRIBUTION MONITORING SYSTEMS

GREGORIO L. ORTIZ III

JHOANNA RHODETTE I. PEDRASA, PH.D.

UNESCO-JASTIP JOINT SYMPOSIUM ON INTRA-REGIONAL WATER SECURITY AND DISASTER MANAGEMENT

15 NOVEMBER 2017

SEDA HOTEL, QUEZON CITY

# OUTLINE

- Introduction
- Distributed Data Validation & Reconstruction
- Research Design & Results
- Conclusion

# INTRODUCTION

- Water Distribution System (WDS)
  - Provide sufficient supply of water to customers at an adequate pressure level [1]
  - Need operational data such as flow and pressure [1]
  - Network modeling, leakage management, resource planning [2]
- Water Distribution Monitoring System (WDMS) [2]
  - Flow meters and pressure sensors
  - Communication via wireless systems such as GSM/GPRS



[1] Vairavamoorthy, K., Tsegaye, S., Mutikanga, H., and Grimshaw, F. Design of water distribution systems. In Water Distribution Systems, pages 171-192. Thomas Telford Ltd, 2011.
[2] Boxall, J., Machell, J., Dewis, N., Gedman, K., and Saul, A. Operation, maintenance and performance. In Water Distribution Systems, pages 193-226. Thomas Telford Ltd, 2011.

# ERRORS IN WDMS

- Substantial fraction of readings are erroneous or missing [3]
- Loss of data can have detrimental effects on the operation [4,5]
- There are methods for addressing data errors, but they are inadequate
  - Dependence on GSM/GPRS network, that is not always reliable
  - Does not support real-time applications [6]
- Need for a resilient method for overcoming sensor faults

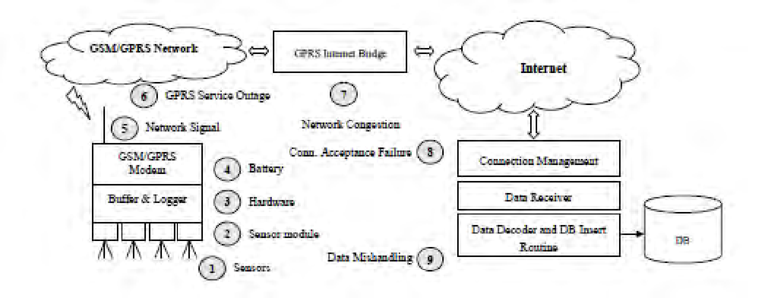



Image taken from: [3] Ediriweera, D. and Marshall, I. Monitoring water distribution systems: understanding and managing sensor networks. Drinking Water Engineering and Science, 3(2):107, 2010.


[[3] Ediriweera, D. and Marshall, I. Monitoring water distribution systems: understanding and managing sensor networks. Drinking Water Engineering and Science, 3(2):107, 2010.
[4] Cugueró-Escofet, M. À., García, D., Quevedo, J., Puig, V., Espin, S., and Roquet, J. A methodology and a software tool for sensor data validation/reconstruction: Application to the Catalonia regional water network. Control Engineering Practice, 49:159172, 2016.
[5] Quevedo, J., Pascual, J., Espin, S., and Roquet, J. Data Validation and Reconstruction for Performance Enhancement and Maintenance of Water Networks. IFAC-PapersOnLine, 49(28):203207, 2016.
[6] Kamal, A. R. M., Bleakley, C., and Dobson, S. Packet-level attestation (PLA):A framework for in-network sensor data reliability. ACM Transactions on Sensor Networks (TOSN), 9(2):19, 2013.

# PROBLEM STATEMENT

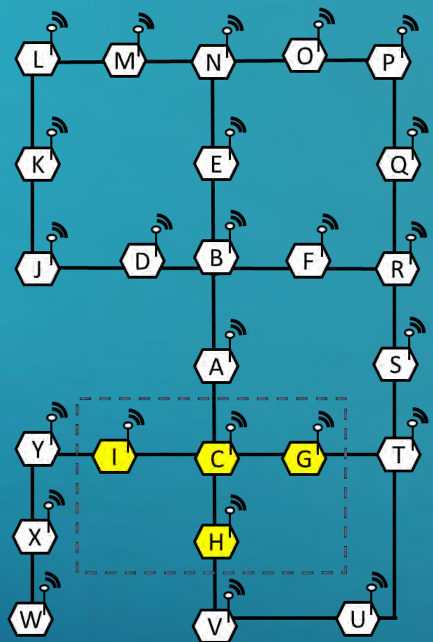
- Aim
  - To provide a solution for validation and reconstruction of measurement data in a water distribution network, that is not dependent on server-side processing
- Research Objectives:
  - Formulation of method for data validation and reconstruction for a water distribution system
    - Distributed
    - Uses principles from conservation of mass and energy
  - Test performance for reduction of error

# DISTRIBUTED DATA VALIDATION & RECONSTRUCTION

- Overview
  - Distributed approach to validating & reconstructing data
  - Shift from centralized communication to wireless sensor networks
  - Sensors communicate with each other to perform validation and replacement functions
  - Validation and replacement of data are performed while data is passed along the network
  - Use Laws of Conservation of Mass and Energy

- Assumptions:
  - Minor losses are insignificant
  - Uniform velocity along a pipe (constant flow and pipe diameter)
  - Pipe is full at all times (EPANET)
  - Sensor node located where it can transmit to/receive from neighbors
  - Prior knowledge of pipe parameters, neighbors
  - Reliable communication system

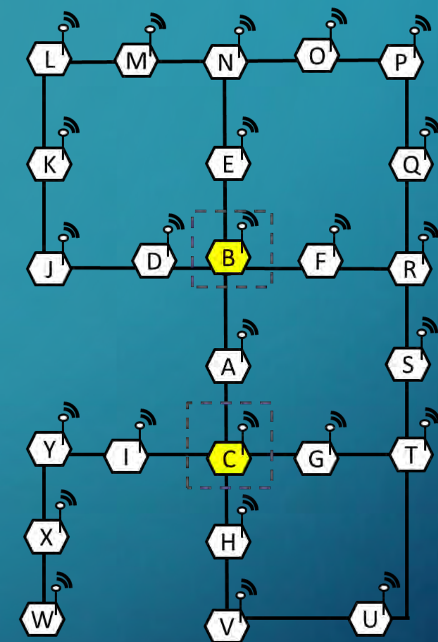
## Mass Balance Computation




$$\boldsymbol{Q_A = Q_D + Q_E + Q_F + D_B}$$

Sum and difference of flows from connected pipes

## Energy Balance Computation

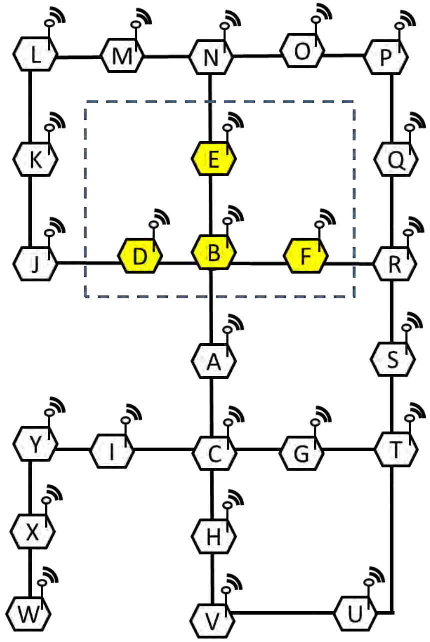


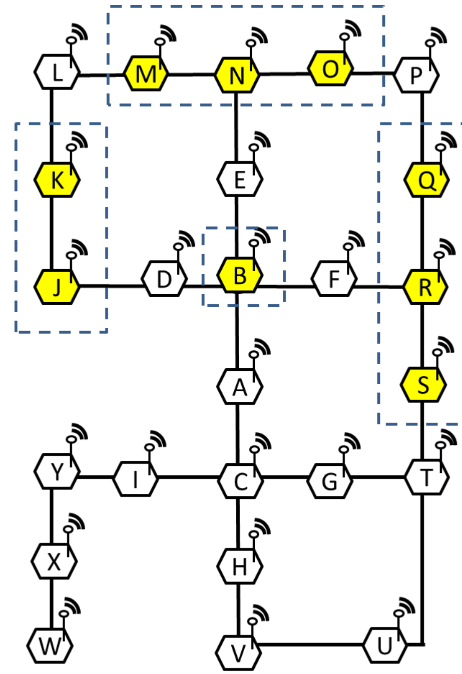

$$\boldsymbol{Q_A} = \sqrt[1.852]{\frac{\left(z_C + \frac{p_C}{\rho g}\right) - \left(z_B + \frac{p_B}{\rho g}\right)}{\left(\frac{L_A}{{C_A}^{1.852}{D_A}^{4.871}}\right)}}$$

Derived from measured values, known constants and pipe parameters

Level 1 MB at Start Junction

$$Q_A = Q_G + Q_H + Q_I + D_C$$



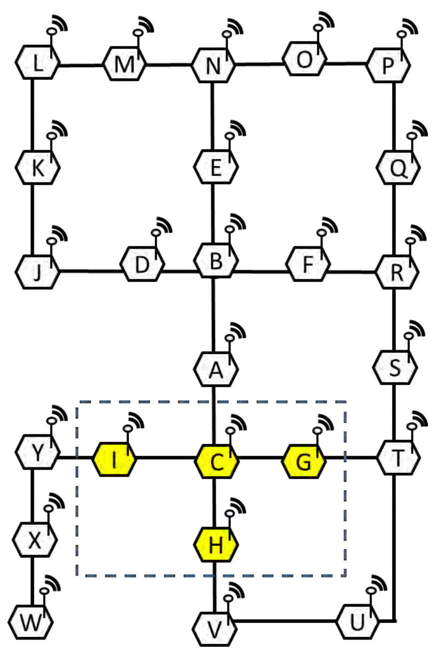



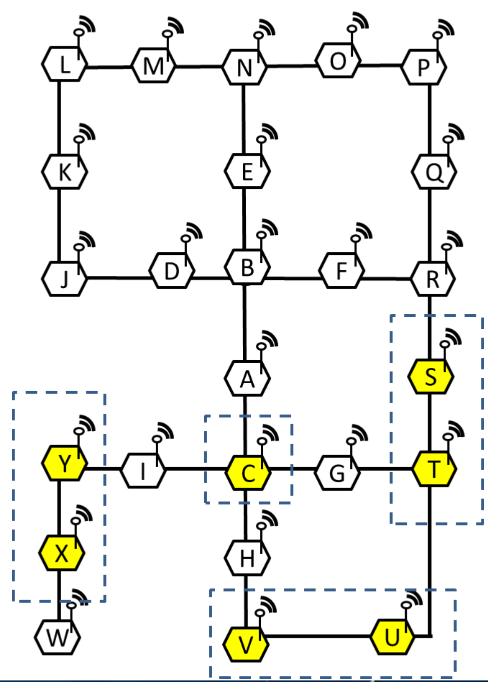

Level 2 MB at Start Junction

$$Q_A = Q_S + Q_U + Q_X + Q_Q + Q_S + D_C + D_T + D_V + D_Y$$

Level 1 MB at End Junction

$$Q_A = Q_D + Q_E + Q_F + D_B$$



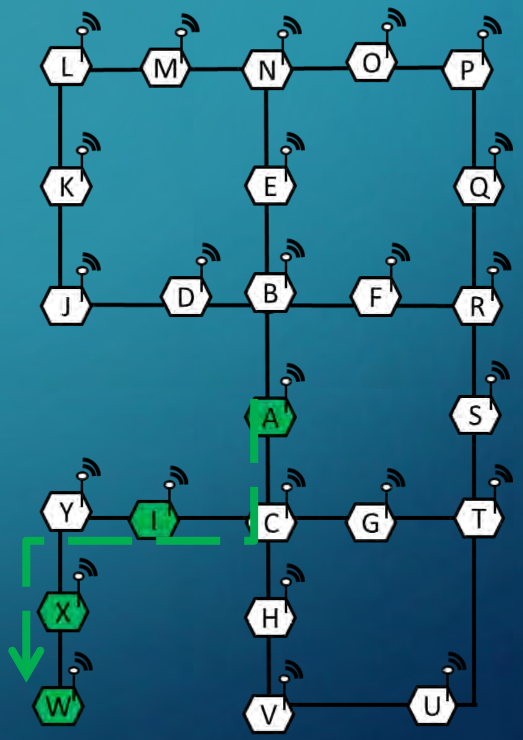




Level 2 MB at Start Junction

$$Q_A = Q_K + Q_M + Q_O + Q_Q + Q_S + D_B + D_J + D_N + D_R$$

# OPERATION

- Sensor node at **A** performs measurement
- **A** validates its data by communicating with nearby sensor nodes and determining a Validation Value
    - **B** and **C** to perform Energy Balance
    - **C, G, H** and **I** (or another set of nodes) to perform Mass Balance
- If **A** is found to be erroneous, it computes for a Replacement Value to replace the measured value
- After validation & reconstruction at **A, A** passes data to next node along the path
- Process repeats until all sensor nodes along path route, validate & if needed, repair, their data



# ALGORITHM

```
% SINK = destination of data, usually server or control point
% PATH = series of pipe sensor nodes towards SINK
% i = order of sensor node along PATH
% PATH(i) = ith sensor node along PATH

thresh = 0.10;

for i = 1 to length(PATH) {
    Qm = getMeasurement(PATH(i));
    Qval = getValidationValue(PATH(i));
    if abs(Qm-Qval) < thresh {
        Qout = Qm;
    }
    else {
        Qrep = getReplacementValue(PATH(i));
        Qout = Qrep;
    }
    transmit(Qout);
}
```

# SIMULATION

- EPANET
  - Generate hydraulic simulation data
- Matlab
  - Introduce Gaussian noise and sensor faults
  - Perform DDVR algorithm
  - Determine performance
  - Selection of Validation and Replacement functions

# TESTS

- Performance of MB and EB methods with/without Gaussian Noise and sensor faults
  - Metric: Percent Error (PE)
- Performance of MB and EB methods as Validation and Replacement functions
  - Factors:
    - Distribution Network (3 different topologies)
    - Percentage of Erroneous Nodes (10%, 30%, 50%)
    - Selection of Erroneous Nodes (3 different sets)
    - Magnitude of Error (-100%, -30%, +30%, +100%)
    - Computation Method (EB, different sets ofMB)
  - Metrics:
    - Root Mean Square Error (RMSE)
    - Mean Absolute Error (MAE)
    - Mean Absolute Percent Error (MAPE)

# PERFORMANCE OF ENERGY AND MASS BALANCE METHODS

## ENERGY BALANCE

- Highly accurate except at some low values of flow
- At low Flow and high R:
  - $h_A$ may become low enough
  - Precision error emerges
- Computing Q using $h_A$ with precision error
  - Computed Q also has precision error

$$Q_A = \sqrt[1.852]{\frac{h_A}{R}}$$

## MASS BALANCE

- Mass Balance susceptible to large magnitudes of Error
- More terms mean more potential sources of error
  - Severity of error increases at higher levels

$$Q_A = Q_D + Q_E + Q_F + D_B$$

$$Q_A = Q_K + Q_M + Q_O + Q_Q + Q_S + D_B + D_J + D_N + D_R$$

# TABLE 1: PERFORMANCE OF ENERGY AND MASS BALANCE AS REPLACEMENT METHOD FOR DDVR USING NEW YORK CITY TUNNEL EPANET NETWORK

| 10 % Erroneous | No Corr | EB | Level 1 MB-SJ | Level 1 MB-EJ | Level 2 MB-SJ | Level 2 MB-EJ |
|---|---|---|---|---|---|---|
| RMSE | 69.47667 | 20.79693 | 84.45383 | 73.60683 | 176.8299 | 174.3713 |
| MAE | 21.97032 | 9.324958 | 46.76075 | 38.3575 | 132.3589 | 129.7671 |
| MAPE | 0.0675 | 0.048706 | 3.272106 | 7.530353 | 13.58119 | 10.90809 |

| 30 % Erroneous | No Corr | EB | Level 1 MB-SJ | Level 1 MB-EJ | Level 2 MB-SJ | Level 2 MB-EJ |
|---|---|---|---|---|---|---|
| RMSE | 177.4398 | 59.65317 | 189.8509 | 263.4825 | 305.65 | 557.6467 |
| MAE | 109.4814 | 22.88697 | 133.3624 | 178.2259 | 220.8358 | 354.7817 |
| MAPE | 0.280833 | 0.061369 | 6.432713 | 3.131472 | 30.47955 | 16.74827 |

| 50 % Erroneous | No Corr | EB | Level 1 MB-SJ | Level 1 MB-EJ | Level 2 MB-SJ | Level 2 MB-EJ |
|---|---|---|---|---|---|---|
| RMSE | 157.612 | 110.2409 | 169.3328 | 320.1242 | 233.6364 | 549.9575 |
| MAE | 88.11808 | 63.25625 | 120.7307 | 215.1157 | 193.992 | 399.6375 |
| MAPE | 0.27 | 0.235725 | 3.924545 | 3.35195 | 38.33617 | 17.214 |

# TABLE 1: PERFORMANCE OF ENERGY AND MASS BALANCE AS VALIDATION METHOD FOR DDVR USING NEW YORK CITY TUNNEL EPANET NETWORK

**10 % Erroneous**

| | No Corr | EB | Level 1 MB-SJ | Level 1 MB-EJ | Level 2 MB-SJ | Level 2 MB-EJ |
|---|---|---|---|---|---|---|
| RMSE | 69.47667 | 18.70722 | 19.69015 | 1.623875 | 19.69201 | 19.59588 |
| MAE | 21.97032 | 5.915625 | 6.698232 | 0.614575 | 7.623533 | 7.3556 |
| MAPE | 0.0675 | 0.039068 | 0.041739 | 0.002148 | 0.044007 | 0.043454 |

**30 % Erroneous**

| | No Corr | EB | Level 1 MB-SJ | Level 1 MB-EJ | Level 2 MB-SJ | Level 2 MB-EJ |
|---|---|---|---|---|---|---|
| RMSE | 177.4398 | 59.2605 | 48.86413 | 59.61017 | 21.6222 | 59.64383 |
| MAE | 109.4814 | 19.86318 | 18.17813 | 22.58828 | 9.94215 | 22.7593 |
| MAPE | 0.280833 | 0.052604 | 0.049533 | 0.059729 | 0.029219 | 0.060821 |

**50 % Erroneous**

| | No Corr | EB | Level 1 MB-SJ | Level 1 MB-EJ | Level 2 MB-SJ | Level 2 MB-EJ |
|---|---|---|---|---|---|---|
| RMSE | 157.612 | 110.0866 | 69.98183 | 106.895 | 87.60317 | 110.2265 |
| MAE | 88.11808 | 61.18892 | 29.53928 | 59.08683 | 43.97087 | 62.83083 |
| MAPE | 0.27 | 0.230102 | 0.151159 | 0.20476 | 0.19498 | 0.234775 |

# PERFORMANCE OF ENERGY AND MASS BALANCE FOR DDVR

- Complementary performance of MB and EB methods
- Best performance when MB is used as Validation function, while EB is used as Replacement Function

# ALGORITHM

```
% SINK = destination of data, usually server or control point
% PATH = series of pipe sensor nodes towards SINK
% i = order of sensor node along PATH
% PATH(i) = ith sensor node along PATH

thresh = 0.10;

for i = 1 to length(PATH) {
    Qm = getMeasurement(PATH(i));
    Qval = getMassBalanceValue(PATH(i));
    if abs(Qm-Qval) < thresh {
        Qout = Qm;
    }
    else {
        Qrep = getEnergyBalanceValue(PATH(i));
        Qout = Qm;
    }
    transmit(Qout);
}
```

# CONCLUSION & FUTURE WORK

- Distributed Data Validation and Reconstruction as a viable alternative to non-automated or centralized methods
  - Utilize complementary performance of energy and mass balance methods
- Potential of wide adoption of wireless sensor networks in water distribution monitoring systems
  - Research on communication aspect and hardware implementation

# DATA VALIDATION AND RECONSTRUCTION FOR WATER DISTRIBUTION MONITORING SYSTEMS

GREGORIO L. ORTIZ III

JHOANNA RHODETTE I. PEDRASA, PH.D.

UNESCO-JASTIP JOINT SYMPOSIUM ON INTRA-REGIONAL WATER SECURITY AND DISASTER MANAGEMENT

15 NOVEMBER 2017

SEDA HOTEL, QUEZON CITY

# REFERENCES


[1] Vairavamoorthy, K., Tsegaye, S., Mutikanga, H., and Grimshaw, F. Design of water distribution systems. In Water Distribution Systems, pages 171-192. Thomas Telford Ltd, 2011.

[2] Boxall, J., Machell, J., Dewis, N., Gedman, K., and Saul, A. Operation, maintenance and performance. In Water Distribution Systems, pages 193-226. Thomas Telford Ltd, 2011.

[3] Ediriweera, D. and Marshall, I. Monitoring water distribution systems: understanding and managing sensor networks. Drinking Water Engineering and Science, 3(2):107, 2010.

[4] Cugueró-Escofet, M. À., García, D., Quevedo, J., Puig, V., Espin, S., and Roquet, J. A methodology and a software tool for sensor data validation/reconstruction: Application to the Catalonia regional water network. Control Engineering Practice, 49:159172, 2016.

[5] Quevedo, J., Pascual, J., Espin, S., and Roquet, J. Data Validation and Reconstruction for Performance Enhancement and Maintenance of Water Networks. IFAC-PapersOnLine, 49(28):203207, 2016.

[6] Kamal, A. R. M., Bleakley, C., and Dobson, S. Packet-level attestation (PLA):A framework for in-network sensor data reliability. ACM Transactions on Sensor Networks (TOSN), 9(2):19, 2013.