Post-Processing in Wireless Sensor Networks: Benchmarking Sensor Trace Files

Evangelos Theodoridis
theodori@cti.gr
RESEARCH ACADEMIC COMPUTER TECHNOLOGY INSTITUTE
GREECE

ACM
PE-WASUN 2010

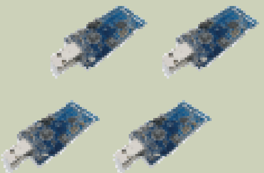
Bodrum. Turkey
Oct(17-21)/
2010

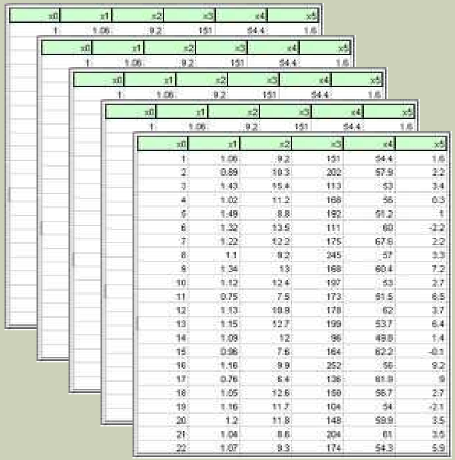
OVERVIEW

- Motivation
- Wisebed /WiseML
- Datasets
- Dataset Classification Metrics
- Evaluations
- Discussion

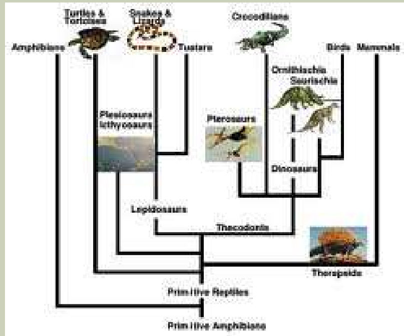
MOTIVATION

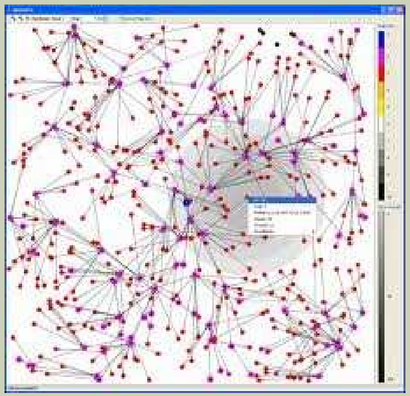
- Dataset Evaluation
- Complexity Classes for Data Aggregation





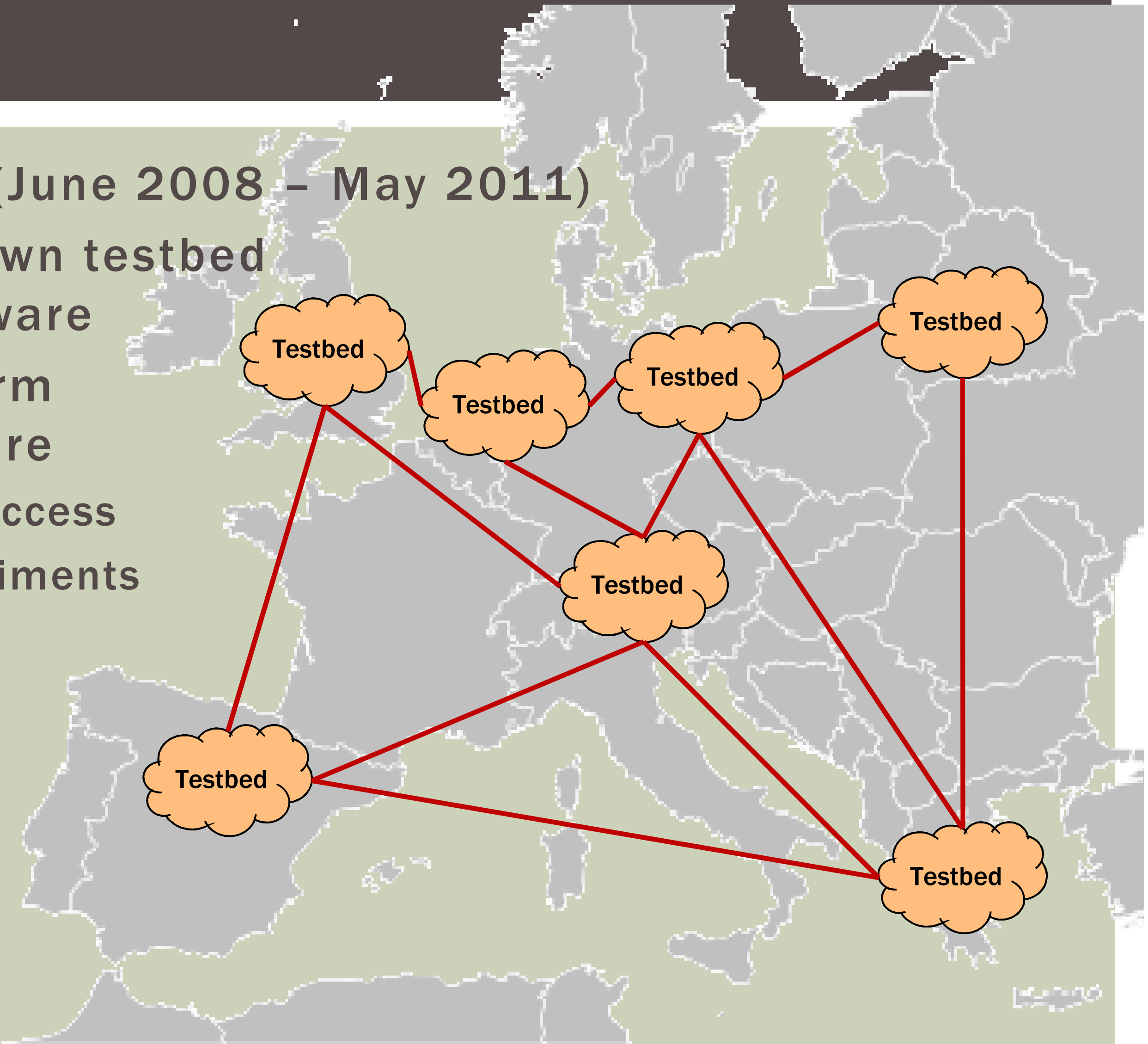
x0	x1	x2	x3	x4	x5
1	1.06	9.2	151	54.4	1.6
2	0.89	10.3	202	57.9	2.2
3	1.43	15.4	113	53	3.4
4	1.02	11.2	168	56	0.3
5	1.49	8.8	192	51.2	1
6	1.32	13.5	111	60	-2.2
7	1.22	12.2	175	67.6	2.2
8	1.1	9.2	245	57	3.3
9	1.34	13	168	60.4	7.2
10	1.12	12.4	197	53	2.7
11	0.75	7.5	173	51.5	6.5
12	1.13	10.9	178	62	3.7
13	1.15	12.7	199	53.7	6.4
14	1.09	12	96	49.8	1.4
15	0.96	7.6	164	62.2	-0.1
16	1.16	9.9	252	56	9.2
17	0.76	6.4	136	61.9	9
18	1.05	12.6	150	56.7	2.7
19	1.16	11.7	104	54	-2.1
20	1.2	11.8	148	59.9	3.5
21	1.04	8.6	204	61	3.5
22	1.07	9.3	174	54.3	5.9



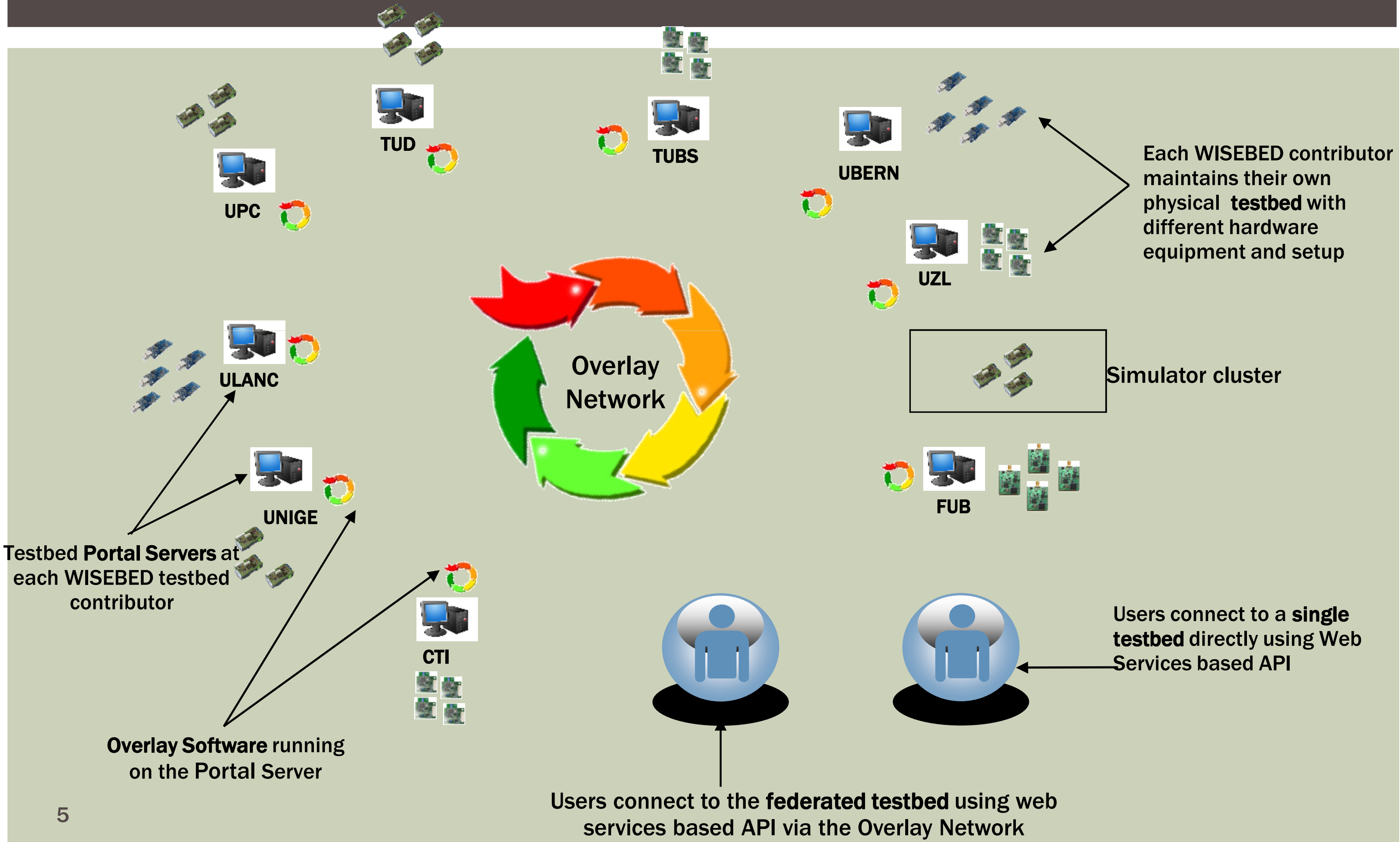


WIRELESS SENSOR NETWORK TESTBED (WISEBED)

- FP7 STREP in FIRE (June 2008 – May 2011)
- Each partner runs own testbed with different hardware
- <u>Requirement</u>: uniform testbed infrastructure
  - Standardization of access
  - Description of experiments and results



WISEBED STRUCTURE



WISEBED APPROACH

- **Goal: Enable evaluation and testing of WSN systems & applications at large scale with great flexibility**
- Approach
  - **Deploy large numbers of heterogeneous sensor nodes**
    - Different types of sensors, topologies and environments
    - Evaluation and test of algorithms and protocols at large scale
  - **Interconnect individual test-beds via the Internet**
    - Remote access to reduce the need for local, private testbeds
    - Define standards for federation of individual WSN testbeds
  - **Mix physicality, virtuality and simulation**
    - Offer interactive, secure, remote access to on-demand "virtual testbed" per-user instances
    - Each virtual testbed can combine (portions of) physical testbeds with virtual connectivity/topology, and simulation environments.
  - **Offer value-added software services**
    - Library of useful algorithms, mechanisms and protocols (Wiselib)
    - Software development kit supporting heterogeneity and dynamic reconfigurability
  - **Support management of data traces**
    - Collection of measurement data and input for simulations → repeatability of experiments
    - Visualization of data traces; performance evaluation

WISEML

```
<wiseml version="1.0" xmlns="http://wisebed.eu/ns/wiseml/1.0">
 <setup>
. . .
</setup>

<scemarios>
. . .
</scemario>

<traces>
. . .
</traces>

</wiseml>
```

```
<node>
        <position>
                <x>1.0</x>
                <y>2</y>
                <z>1.879</z>
        </position>
        <gateway>true</gateway>
        <programDetails>blink.tnode</programDetails>
        <nodeType>TNode v4</nodeType>
        <capability>
                <name>urn:wisebed:node:capability:temperature</name>
                <datatype>decimal</datatype>
                <unit>kelvin</unit>
                <default>273</default>
        </capability>
            . . .
</node>
<link>
        <encrypted>false</encrypted>
        <virtual>false</virtual>
        <rssi datatype="decimal" unit="dBm" default="-90" />
        <capability>
            <name> urn:wisebed:node:capability:lqi</name>
            <datatype>integer</datatype>
            <unit>percentage</unit>
            <default>100</default>
            </capability>
        </capability>
...
</link>
```

WISEML

```
<wiseml version="1.0" xmlns="http://wisebed.eu/ns/wiseml/1.0">
 <setup>
...
</setup>

<scemarios>
...
</scemario>

<traces>
...
</traces>

</wiseml>
```

```
<scenario id="...">
        <timestamp>0</timestamp>
        <enableNode id="..." />
        <disableNode id="..." />
        <enableLink source="..." target="..." />
        <disableLink source="..." target="..." />
        <node id="...">
                <position>
                        <x>0</x>
                        <y>1</y>
                        <z>2</z>
                        <phi>0</phi>
                        <theta>1</theta>
                </position>
                <data key="lqi">50</data>
        </node>
</scenario>
```

WISEML

```
<wiseml version="1.0" xmlns="http://wisebed.eu/ns/wiseml/1.0">
 <setup>
. . .
</setup>

<scemarios>
. . .
</scemario>

<traces>
. . .
</traces>

</wiseml>
```

```
<trace id="experiment_CTI_uniqueID_23453323">
	[...]
	<timestamp>3605.164612</timestamp>
	<node id="urn:wisebed:node:CTI:9">
		<position>
			<x>85</x>
			<y>80</y>
			<z>52</z>
		</position>
		<data key="textOutput">Temperature 2 15</data>
		<data key="textOutput">Light 1 202</data>
	</node>
	<timestamp>3685.164612</timestamp>
	<node id="urn:wisebed:node:ubern:3">
		<data key="textOutput">Light 2 480</data>
		<data key="textOutput">Light 1 223</data>
	</node>
	[...]
</trace>
```

DATA AGGREGATION

- Minimize redundant information
- Topology, aggregation function, etc.
- Accuracy, completeness, latency, message overhead, etc.
- Aggregation Classes
  - Tree Aggregation/Cluster Aggregation
  - Gossip Aggregation
  - Hybrid Aggregation
- Efficacy:
  - Topology changes
  - Correlation of measures :{spatial, temporal, semantic}

DATA AGGREGATION-DISTINCT COUNT

- Probabilistic Counting with Stochastic Averaging (PCSA)
  - Fajolet-Martin :
  - Values →hash→ [0 … 2^L-1]
  - Estimation=$2^{(i-1)}$/0.77351, i – lowest order bit not set
- LogLog Counting Algorithm
  - Durand – Flajolet:
  - LoglogN synopsis size and highest bit set is used
  - 0.78/√k → 1.30/√k
- Time Decaying Sketches
  - Cormode et al.
  - Time decaying function

- Synopsis Diffusion : distribution/aggregation phase
  - Gibbons et al.
  - Synopsis generation
  - Synopsis fusion
  - Synopsis evaluation

DATASET CLASSIFICATION

- time[0] to time[T], $N_{installed}$, window W timestamps (or time)
- Network Based Classification: time[i]
  - Number of active nodes-Network cohesion
  - P_w , P average periods of transmission
  - n_1 = # nodes with P_w<P
  - n_2 = # nodes with P_w~=P
  - n_3 = # nodes silent in W
  - $N_{installed}=n_1+n_2+n_3$

$$n = n_1 + n_2 + \begin{cases} 0, \text{if } \; n_1 \leq \frac{1}{2}n_2 \\ \frac{2}{3}n_3, \text{if } \; n_1 \leq n_2 \\ \frac{n_1}{n_1+n_2}n_3, \text{if } \; n_1 \geq n_2 \end{cases}$$

DATASET CLASSIFICATION

- Measurement-based Classification
- Spatial Correlation:
  - Cluster measurements (K-means 3-5 cluster)
  - Spatial Correlation

$$Scat(c) = \frac{1}{c}\sum_{i=1}^{c}\frac{\|\sigma(C_i)\|}{\|\sigma(X)\|}$$

$$\sigma(X)[p] = \frac{1}{n}\sum_{i=1}^{n}((x_i[p] - \overline{x}[p])^2$$

$$\overline{x}[p] = \frac{1}{n}\sum_{i=1}^{n} x_i[p],$$

- Temporal Correlation:

$$cor_l = \frac{1}{w_1 - l}\sum_{i=1+l}^{w_1}\frac{(m_i \cdot m_{i-l} - \overline{y}_i \cdot \overline{y}_{i-l})}{\mathrm{var}(m)}$$

$$\overline{y}_i = \frac{1}{w_1 - l}\sum_{i=1+l}^{w_1} m_i, \quad and \quad \overline{y}_{i-l} = \frac{1}{N-l}\sum_{i=1}^{N-l} m_i$$

DATASET CLASSIFICATION

$$C_{CL} = \begin{cases} high, \text{if } \sum_{\forall i} \frac{N_{installed} - C_i}{N_{installed}} \leq 0.25 \\ avg, \text{if } \sum_{\forall i} \frac{N_{installed} - C_i}{N_{installed}} \leq 0.75 \\ low, \text{if } \sum_{\forall i} \frac{N_{installed} - C_i}{N_{installed}} \leq 1 \end{cases}$$

$$S_{CL} = \begin{cases} low, \text{if } \frac{\sum_{\forall i} S_i}{T} \leq 1 \\ avg, \text{if } \frac{\sum_{\forall i} S_i}{T} \leq 2 \\ high, else \end{cases}$$

$$T_{CL} = \begin{cases} low, \text{if } \frac{\sum_{\forall i} T_i}{T} \leq 1 \\ avg, \text{if } \frac{\sum_{\forall i} T_i}{T} \leq 5 \\ high, else \end{cases}$$

DATASETS

- LOFAR-agro (Delft University of Technology): precision agriculture potato field
  - Temperature and relative humidity every 10 mins
  - Statistics: neighbors, communication links, battery power routing topology, failures
  - Over 1 million measurements
- Greenhouse (University of Thessaly): Glass greenhouse
  - Grid layout
  - Temperature, humidity, Total solar radiation, photosynthetically active radiation, soil moisture
  - 91656 measurements
- Husbandry (Alexander Fleming Inst.) husbandry of mice
  - Temperature, humidity, ammonia, man-made noise
  - Static and mobile nodes
  - 117454 measurements

DATASETS

- Citysense project: nodes on light poles and public building
  - Weather conditions and air pollutants
  - sub-minute sampling to multi-minute sampling
  - Since 2007 over 31 million measurements
- Webdust (RACTI): environmental measurements
  - Indoor and outdoor nodes
  - light, temperature, humidity, acceleration, magnetic field levels and barometric pressure
  - Over 2.5 million measurements

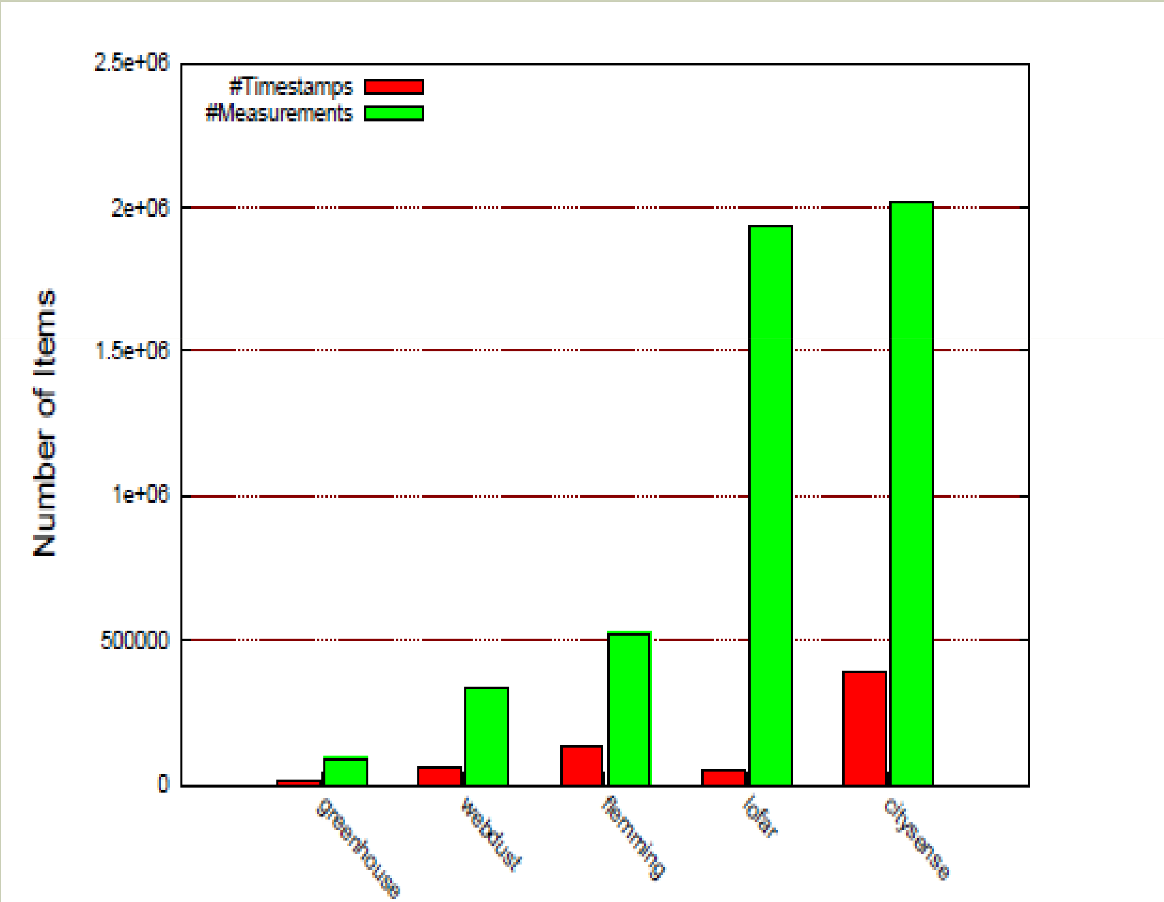
Dataset	*WSN Cohesion*	*Spatial Correlation*	*Temporal Correlation*	Overall *Complexity Level*
Lofar-agro	avg	low	avg	avg
Greenhouse	low	avg	avg	avg
Flemming	high	avg	high	high
Citysense	low	high	high	high
Webdust	high	low	high	high

EXPERIMENTAL EVALUATION

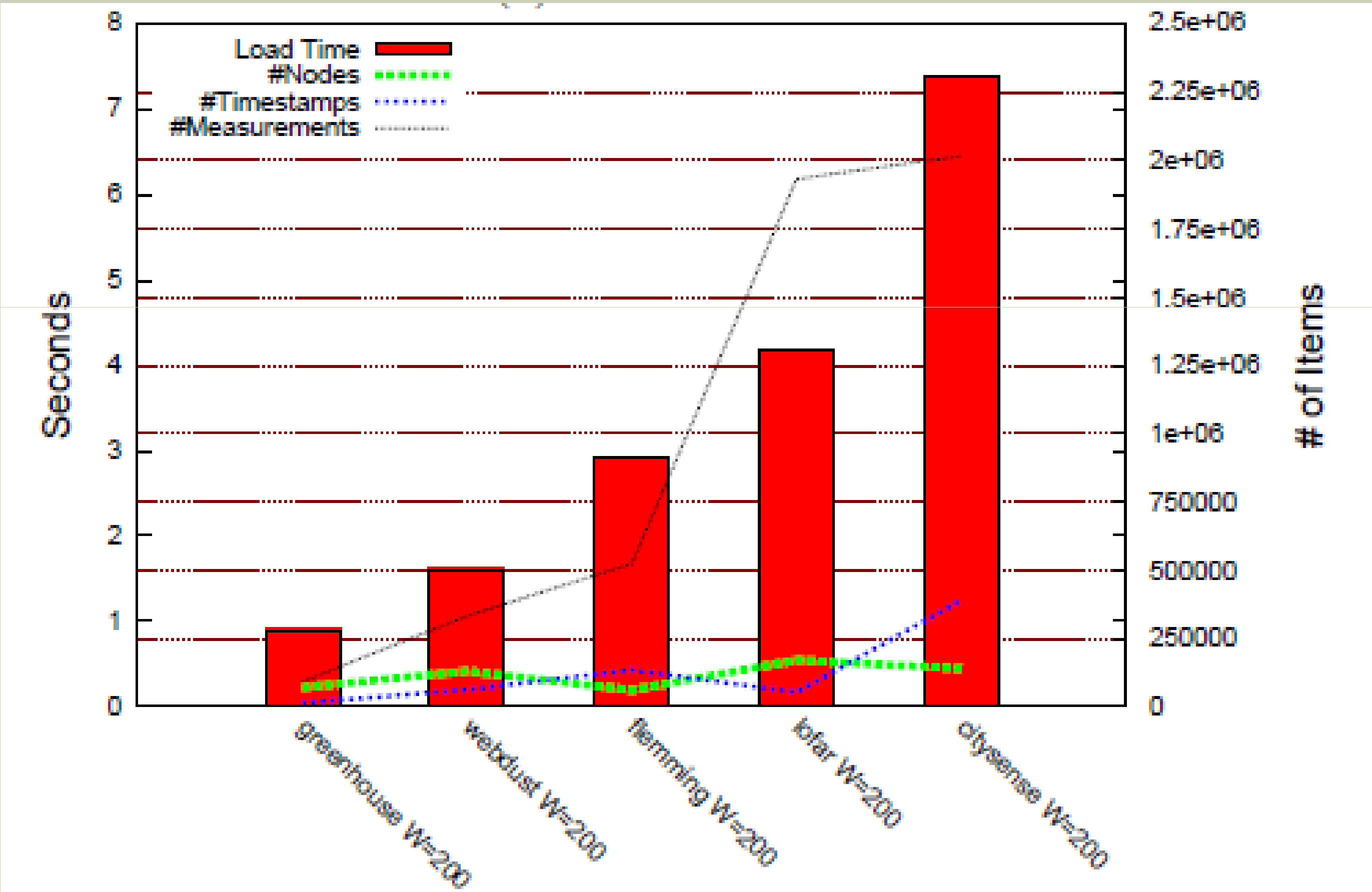
- Simulation:
  - Synopsis Diffusion: W=200,500 time stamps
  - Form a binary tree
  - Each node: generate synopsis for its values and merge with incoming synopses. Then push synopsis to ancestor and to a few other nodes.
- Implementation:
  - Java SE 6, DCore 3Ghz, 4 Gbytes main memory, Ubuntu 10.04 64 bit.
  - WiseML parser based on JIBX library
  - MD4 message digest function by GNU-Crypto library

Dataset	W	#T	#M	#Nodes	#Types	L-Time	P-Time	M-Mem
greenhouse	200	13906	91656	7	7	0.899	4.356	534
	500	13906	91656	7	7	0.899	5.176	535
webdust	200	61283	332742	33	13	1.611	43.656	875
	500	61283	332742	33	13	1.611	52.456	884
flemming	200	133901	523866	5	6	2.916	22.057	1940
	500	133901	523866	5	6	2.916	36.042	2023
lofar	200	51749	1932845	47	17	4.195	155.014	1158
	500	51749	1932845	47	17	4.195	224.467	1317
citysense	200	390953	2015447	13	14	7.384	142.494	1512
	500	390953	2015447	13	14	7.384	216.368	1516

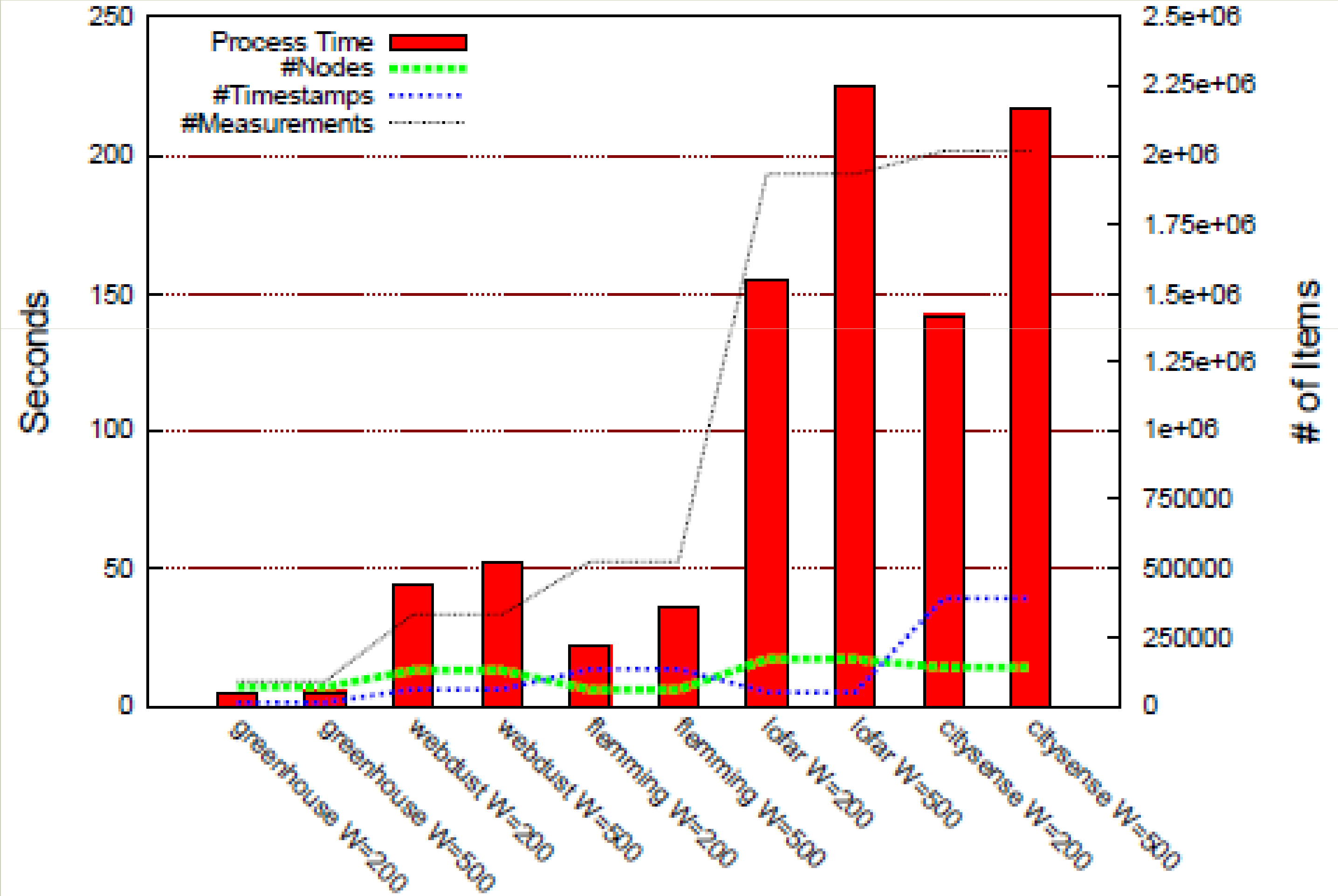
EXPERIMENTAL EVALUATION



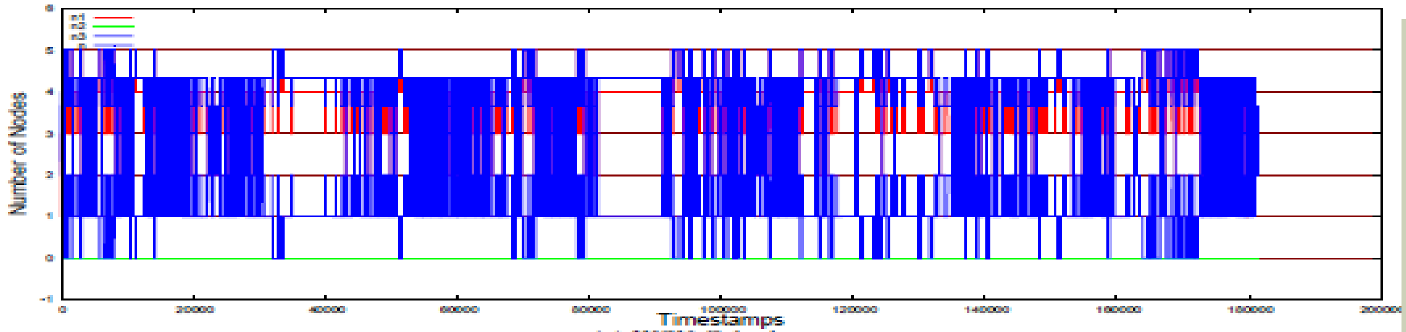
EXPERIMENTAL EVALUATION



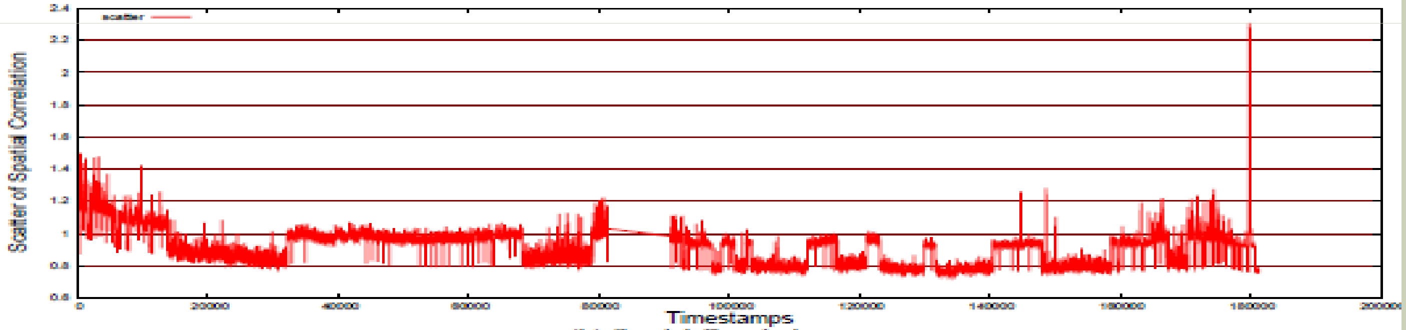
EXPERIMENTAL EVALUATION



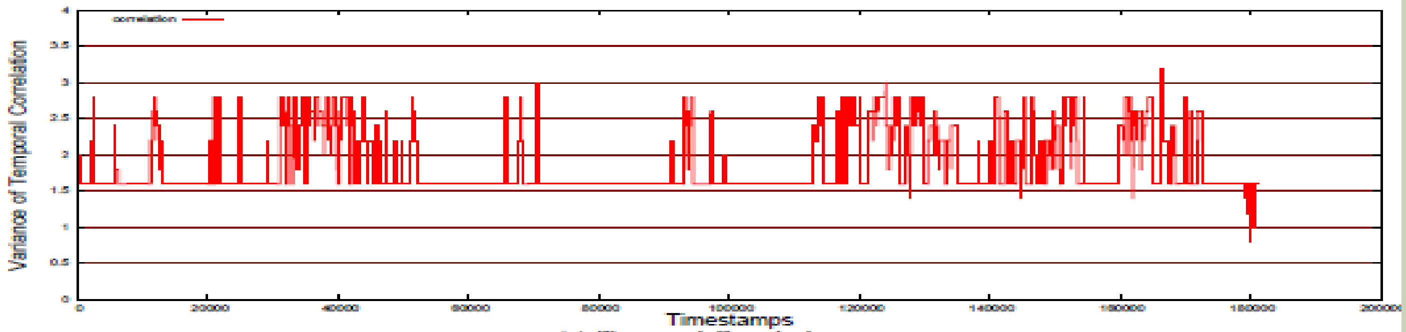
HUSBANDRY DATASET



(a) WSN Cohesion

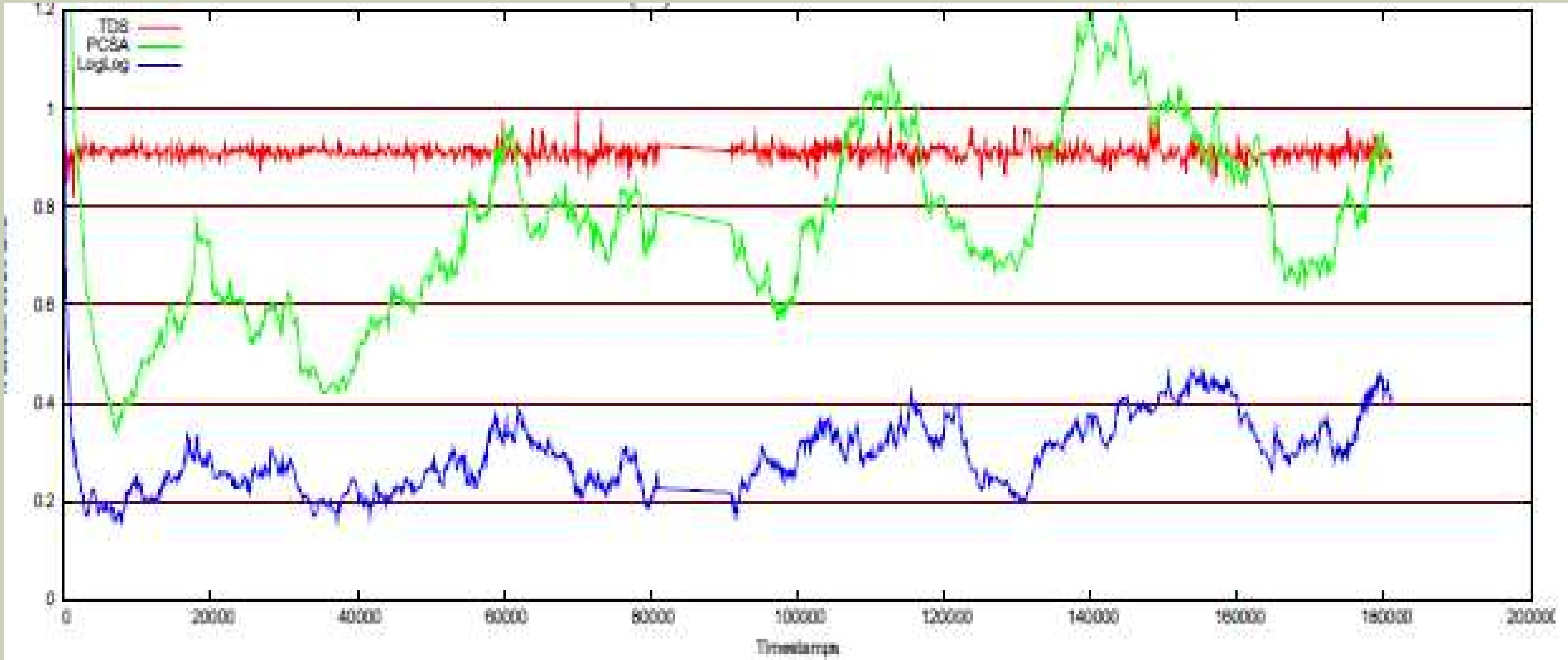


(b) Spatial Correlation

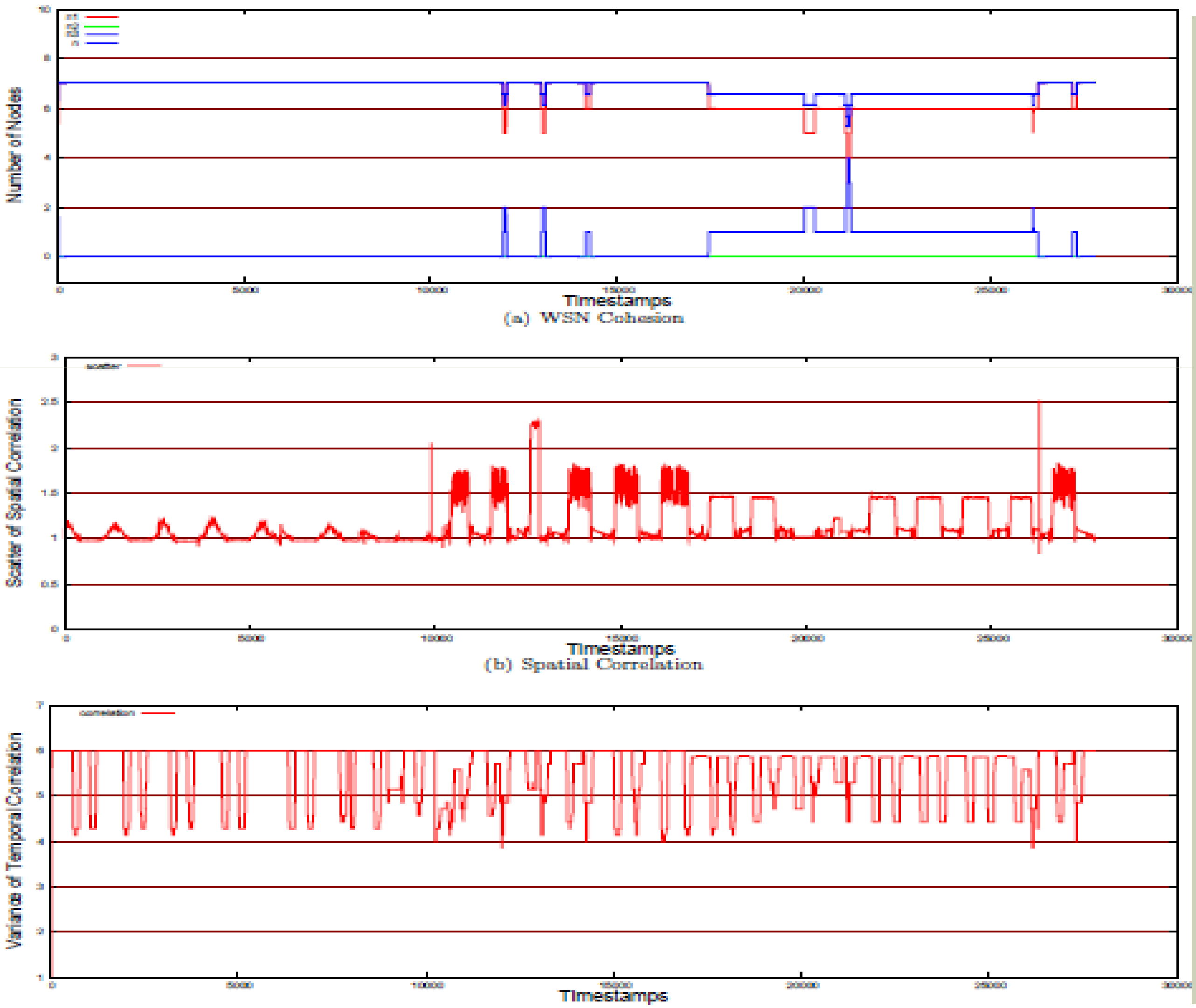


(c) Temporal Correlation

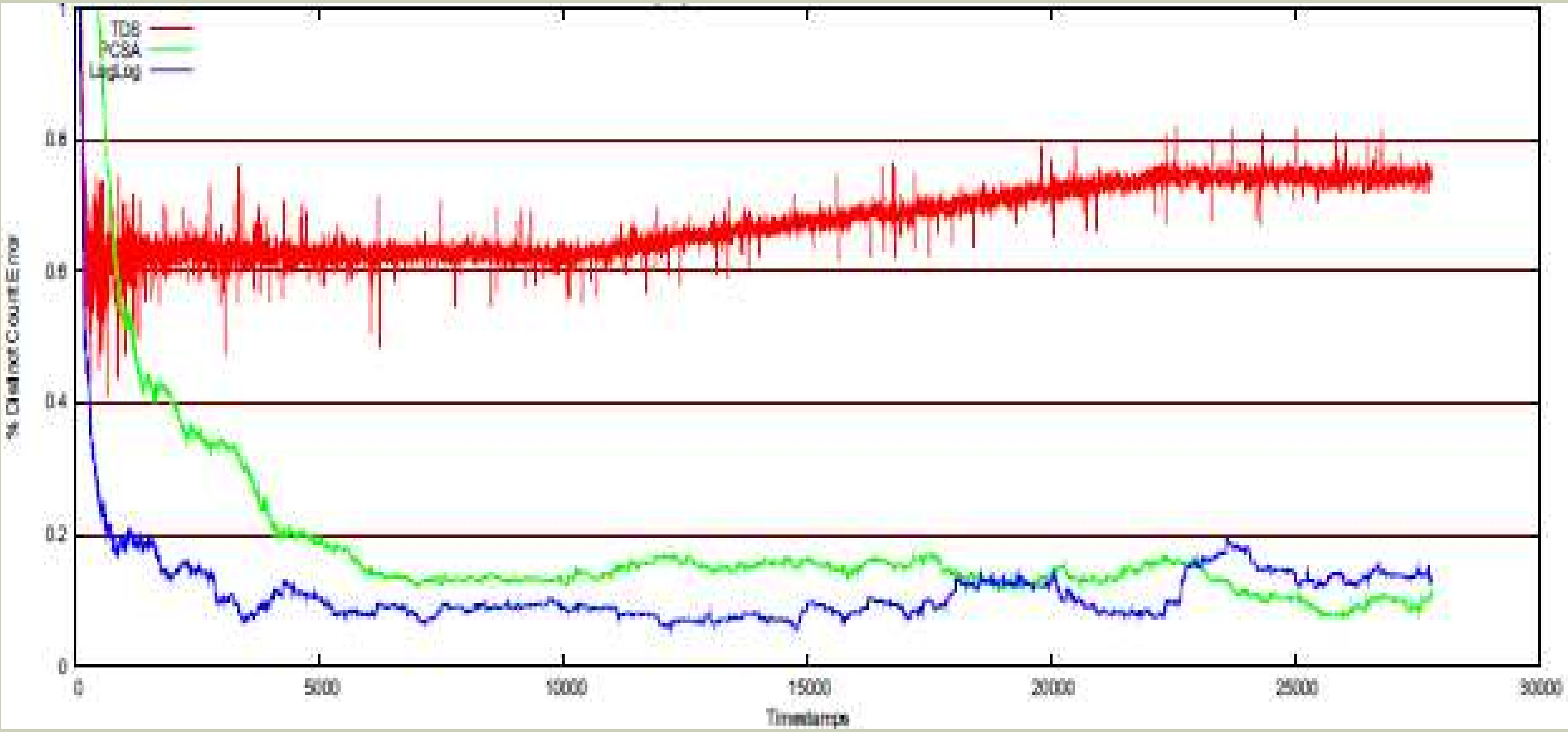
HUSBANDRY DATASET



GREENHOUSE DATASET



GREENHOUSE DATASET



CONCLUSIONS

- Classify datasets into benchmarks for data aggregation Algorithms

- Fine tune metrics
- Different classes of algorithms
- Handling of large WISEML files