# EXPERIMENTS AND ANALYSIS OF QUALITY- AND ENERGY-AWARE DATA AGGREGATION APPROACHES IN WSNS

Cinzia Cappiello – Fabio A. Schreiber

Politecnico di Milano
Dipartimento di Elettronica ed Informazione

# TECHNOLOGICAL ISSUES WITH SENSOR NODES IN WSNs

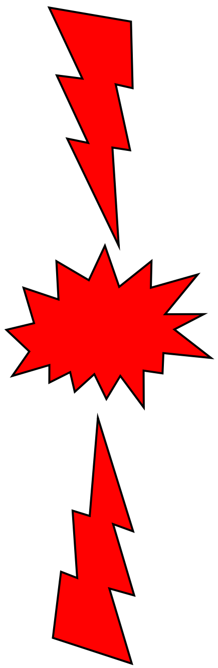


- MEMORY
  - SMALL ALGORITHM FOOTPRINT
  - SMALL DATA BUFFERS
    - **STORE FEW DATA**
    - **FOR A LIMITED TIME SPAN**

TRANSFER DATA ELSEWHERE AS SOON AS POSSIBLE

- POWER
  - BATTERY LIFE IS LIMITED
  - TRANSMISSION IS THE MOST POWER CONSUMING FUNCTION

LIMIT TRANSMISSION FREQUENCY AND DURATION



- **PRESERVE DATA QUALITY**
  - ACCURACY
  - PRECISION
  - TIMELINESS

# DATA QUALITY DIMENSIONS



- ACCURACY

  the degree of conformity of a measured or computed quantity to its actual (true) value ( $|v_{avg}-v_{ref}| < \varepsilon_{acc}$)

- PRECISION

  the degree to which repeated measurement show the same or similar results

  (small variance $1/n * \sum_{n=1}^{N} (v_n - \mu)^2 < \varepsilon_{prec}$ )

- TIMELINESS
  - CURRENCY

    **the time interval from the instant the value was sampled to the instant at which it is sent to the base station**

  - VOLATILITY

    **the amount of time during which data remain valid**

$$Timeliness = max(1 - Currency/Volatility; 0)^s$$

# PREVIOUS WORK



*DATA COMPRESSION IS THE PROCESS OF ENCODING INFORMATION USING FEWER BITS*

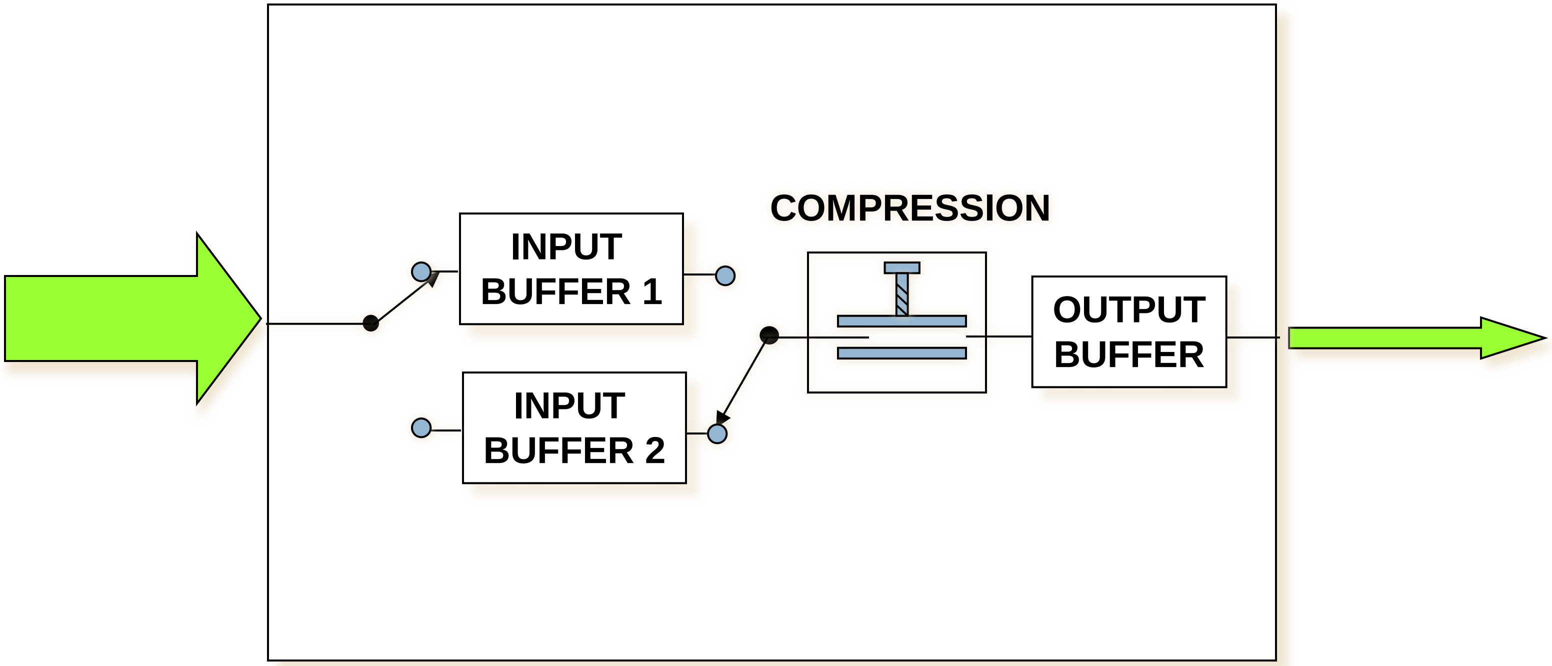
- Traditional data compression algorithms are not feasible for WSNs owning to their size and complexity

- Data compression algorithms proposed for WSN are focused on:
  - Temporal correlation/aggregation
  - Spatial correlation/aggregation

- … and they are characterized by the following limitations:
  - Specialization on specific contexts
  - Utilization on predictive models and thus, limitations on acceptable values
  - Data Quality is scarcely considered
  - Outliers are discarded

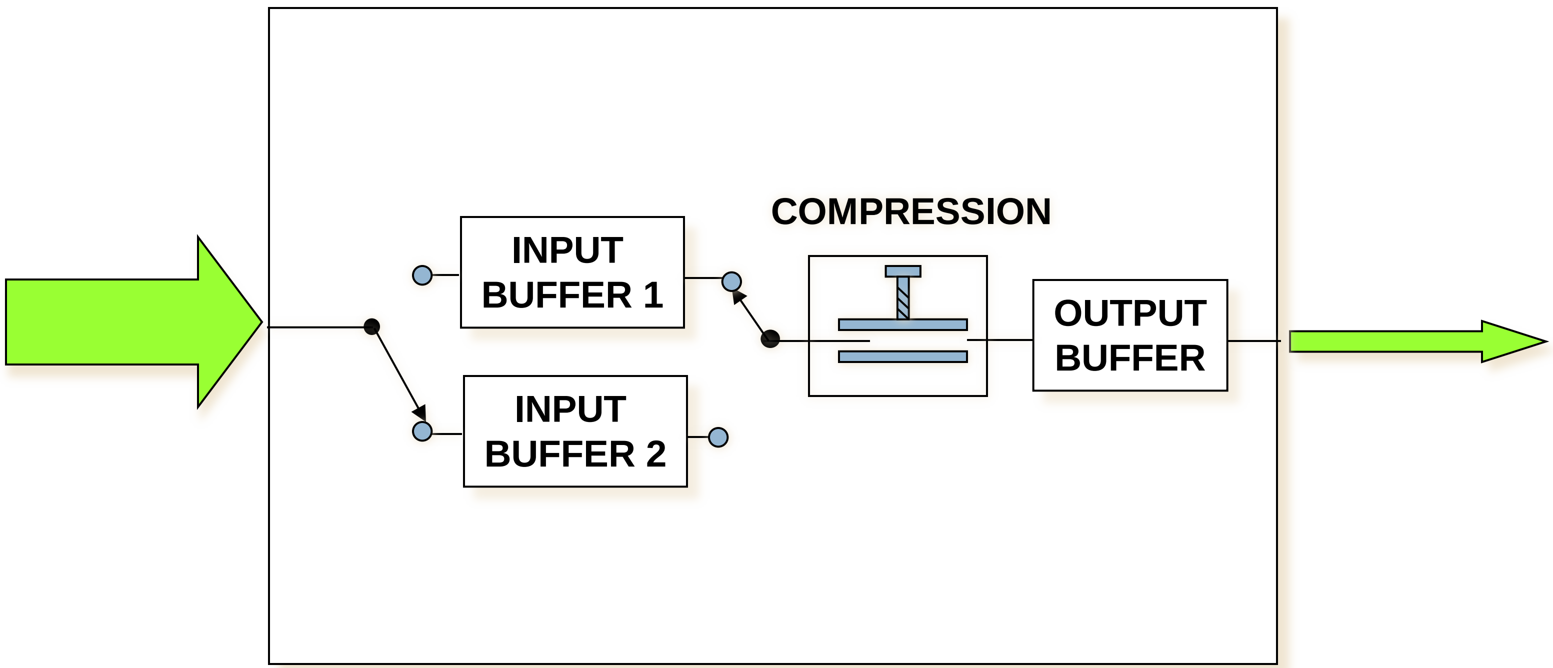
# SYSTEM STRUCTURE





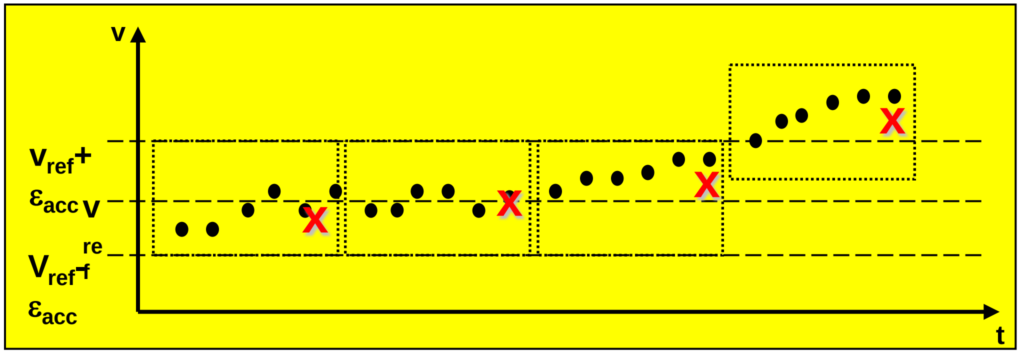
# SYSTEM STRUCTURE





# BASIC PRINCIPLES OF THE PROPOSED AGGREGATION ALGORITHM






- ACCURACY IS REPRESENTED BY THE WINDOW HEIGHT
- VALUES FALLING WITHIN THE WINDOW CAN BE CONSIDERED SIMILAR ENOUGH TO BE FAIRLY REPRESENTED BY THEIR AVERAGE
- VALUES FALLING OUTSIDE THE WINDOW ARE OUTLIERS
- OUTLIERS CAN BE OCCASIONAL OR CONSECUTIVE: IN ANY CASE OUTLIERS INFORMATION MUST BE PRESERVED FOR FURTHER INVESTIGATION

# CONSIDERED CASES



- **EXPECTED TREND**

  THE TREND IS REGULAR

  **VALUES ARE BOTH PRECISE AND ACCURATE**

- **SLOW CHANGE**

  THE TREND SHOWS AN UNPREDICTED BUT LASTING VARIATION

  **VALUES ARE STILL PRECISE, BUT NOT ACCURATE**

- **OSCILLATORY / BURSTY / SUDDEN CHANGE**

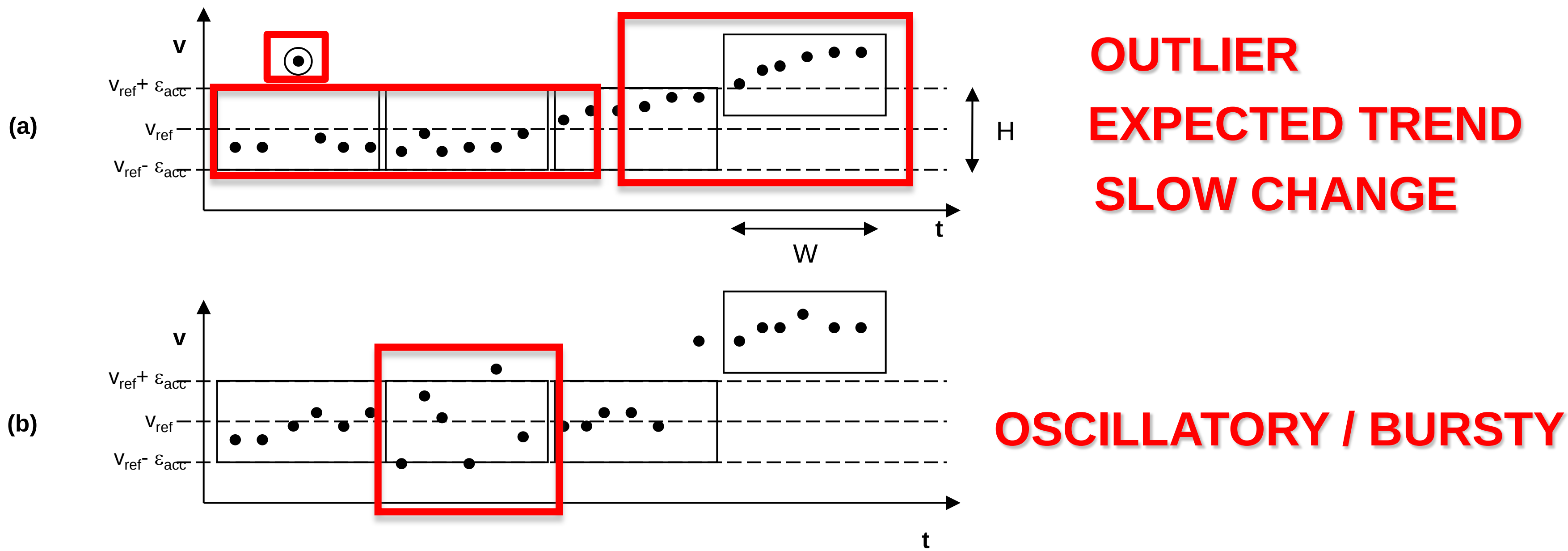
  THE TREND SHOWS DISCONTINUOUS VARIATIONS

  **VALUES ARE NOT PRECISE, BUT THEY CAN BE BOTH ACCURATE OR NOT**

ANY DATA STREAM CAN BE DESCRIBED AS A COMBINATION OF THESE CASES

# CONSIDERED CASES





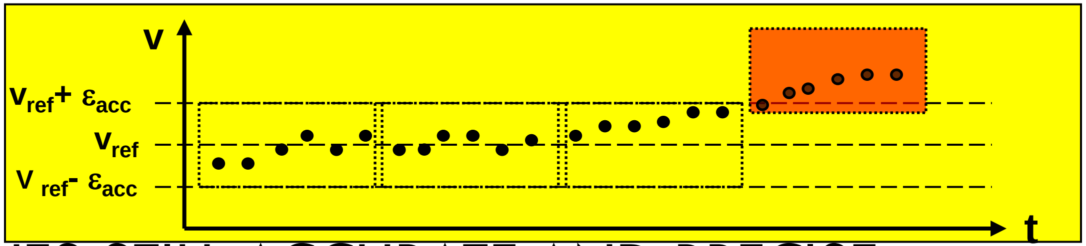

By considering Z aggregate values and J outliers out of a set of N measures, the algorithm is considered efficient if the output is composed by (Z+J) values instead of N where (Z+J)<<N

# OUTLIER DETECTION



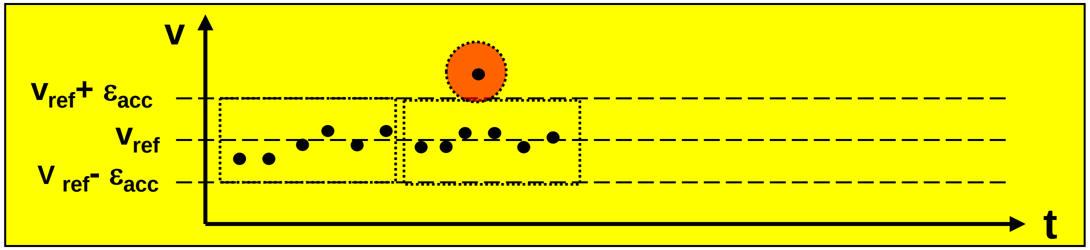
- VALUES NOT ACCURATE BUT PRECISE
  - INCOMING VALUES DON'T MATCH THE REFERENCE VALUE
  - THEIR STANDARD DEVIATION IS SMALL

**CHANGE IN THE BEHAVIOUR**




- VALUES STILL ACCURATE AND PRECISE

**ERROR**



# ALGORITHM BANDWIDTH



- COMPRESSING DATA AMOUNTS TO LOWERING THE BANDWIDTH OF THE MEASUREMENT SYSTEM

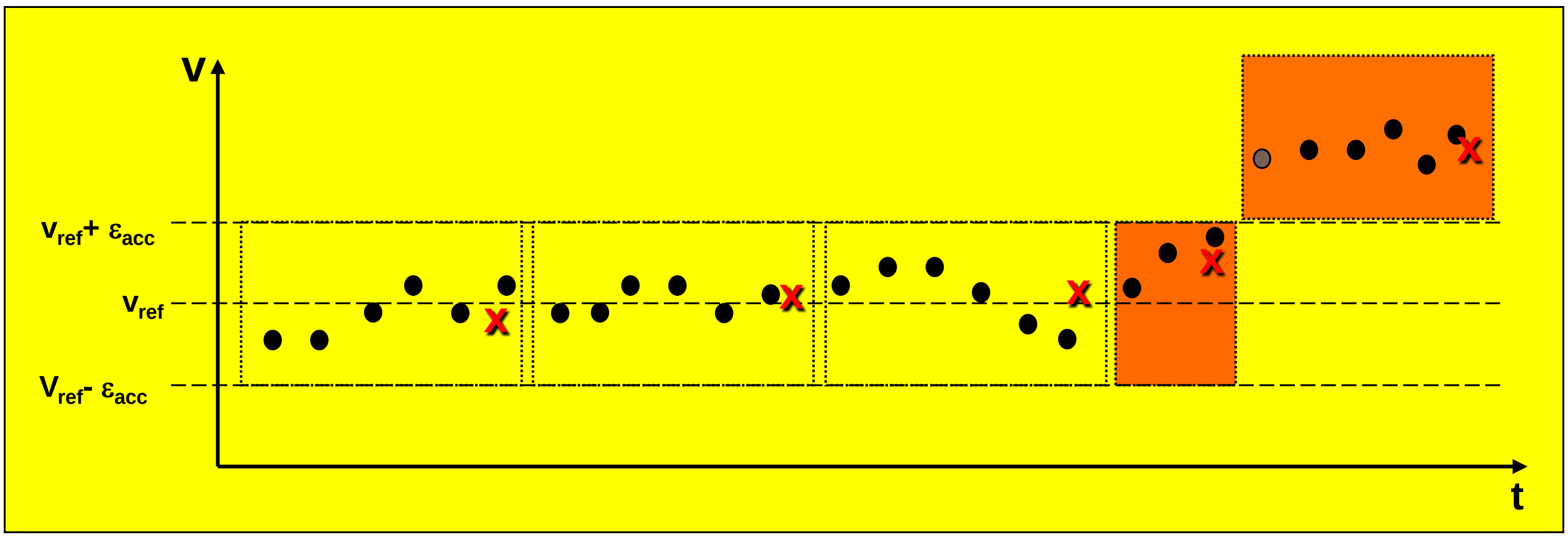
- THE WINDOW WIDTH DETERMINES THE NUMBER OF MEASURED VALUES WHICH ARE AGGREGATED
  - 1 POINT WINDOW → NO COMPRESSION → MAX BANDWIDTH

- THE WINDOW WIDTH ALSO DETERMINES THE TIMELINESS BY WHICH DATA ARE DELIVERED TO THE BASE STATION

# ALGORITHM BANDWIDTH



- THE WINDOW WIDTH MUST BE ADAPTED TO CATCH VERY STEEP/SUDDEN CHANGES



# ALGORITHM INPUT/OUTPUT



- INPUT PARAMETERS
  - TIME SERIES $V = <v_1, v_2, \ldots v_n>$
  - EXPECTED VALUE $v_{ref}$
  - ACCURACY TOLERANCE $\varepsilon_{acc}$
  - PRECISION TOLERANCE $\varepsilon_{prec}$
  - WINDOW WIDTH N
  - CONTINUITY INTERVAL C

- OUTPUT PARAMETERS
  - AGGREGATE VALUES $T = < a_1,t_1 >; < a_2,t_2 >; \ldots < a_z,t_z >$
  - OUTLIERS $O = < o_1,t_1 >; < o_2,t_2 >; \ldots < o_i,t_i >$

ALGORITHM COMPLEXITY

O(N)

ALGORITHM FOOTPRINT

11 KB RAM; 1 KB ROM

# EXPERIMENTAL SET UP

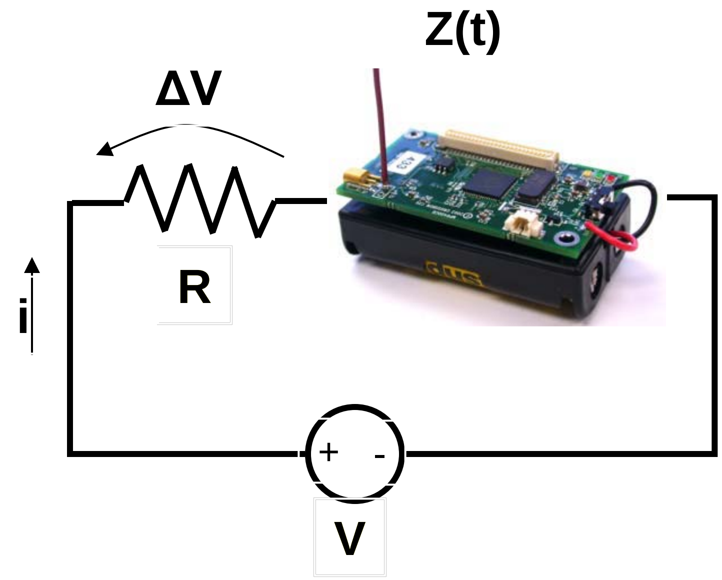


**100 Ώ ≤ $Z_R(t)$ ≤ 1000 Ώ (measured)**

**R = 1 Ώ → R + $Z_R \approx Z_R$**

**0 mA ≤ i ≤ 30 mA (Data sheet)**

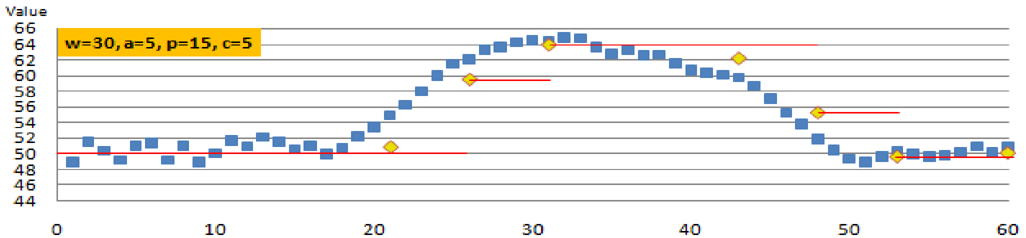
**0 mV ≤ ΔV ≤ 30 mV**



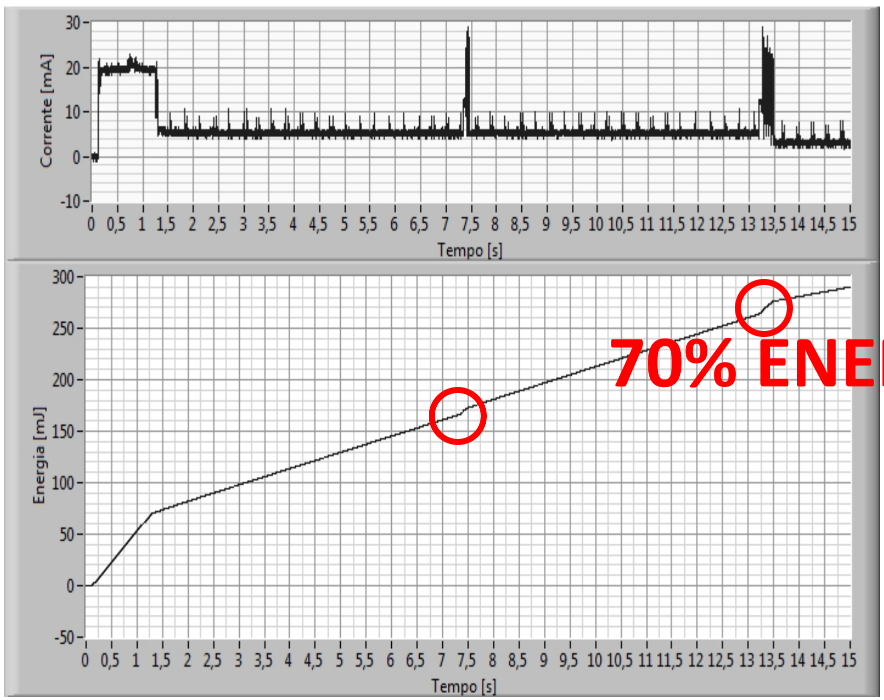

$$E(t) = \int_0^t \left[\frac{v2(t')}{R}\right]^* v1(t')dt'$$

# ALGORITHM BEHAVIOUR

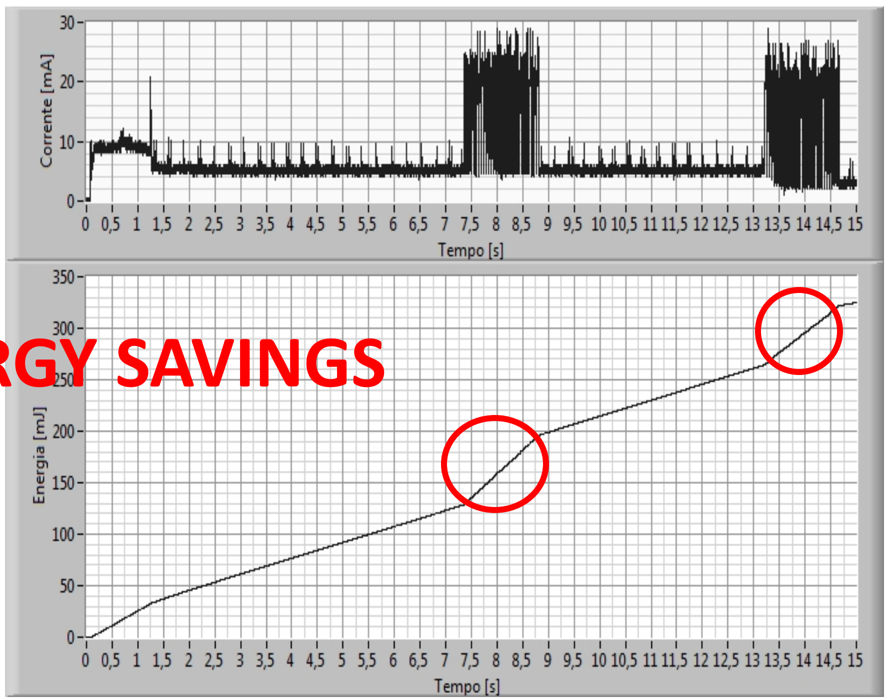





WITH AGGREGATION




WITHOUT AGGREGATION (BYPASS)



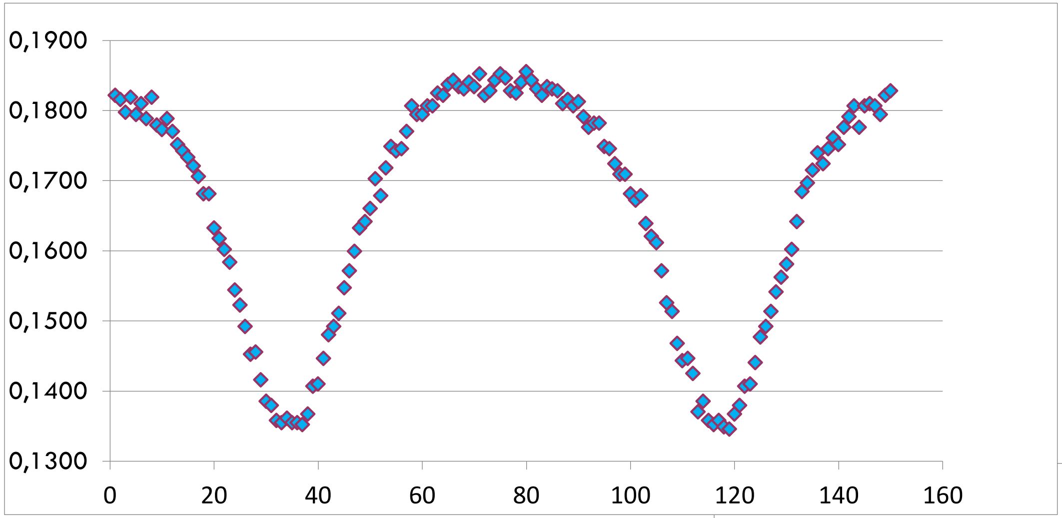

**70% ENERGY SAVINGS**

7 TRANSMITTED VALUES , 30mJ

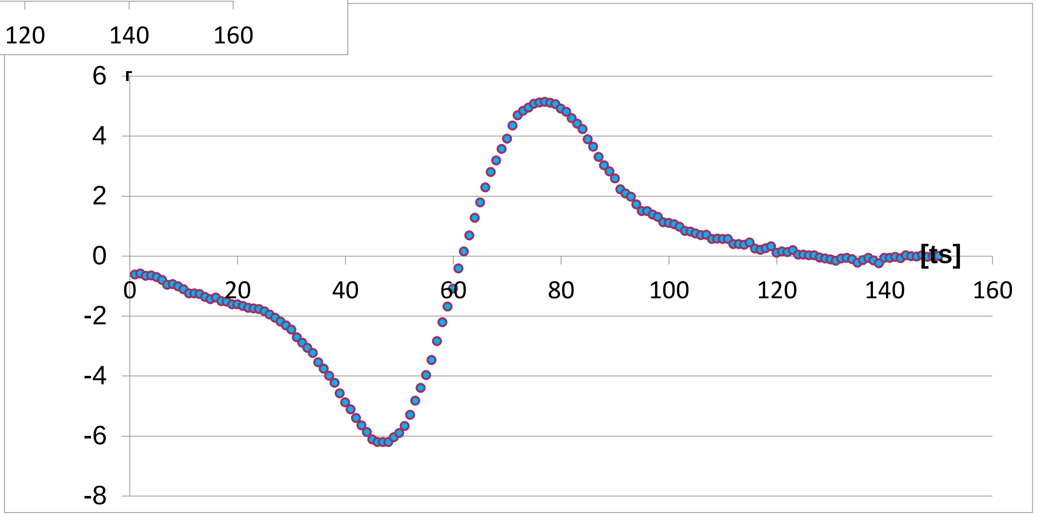
60 TRANSMITTED VALUES , 120mJ

# ALGORITHMS COMPARISON: CONSIDERED SIGNALS






(a) $C_2N_2$ absorption spectrum




$C_2N_2$ absorption spectrum FM (b)

# COMPARISON CRITERIA (1/2)



- The data sets shown in the previous slide have been processed by using the algorithm proposed and two other aggregation algorihms:
  - I. Lazaridis, S. Mehrotra, Capturing Sensor-Generated Time Series with Quality Guarantees, in: ICDE, 2003, pp. 429–439.
  - T. Schoellhammer, E. Osterweil, B. Greenstein, M. Wimbrow, D. Estrin, Lightweight Temporal Compression of Microclimate Datasets, in: LCN, 2004, pp. 516–524.

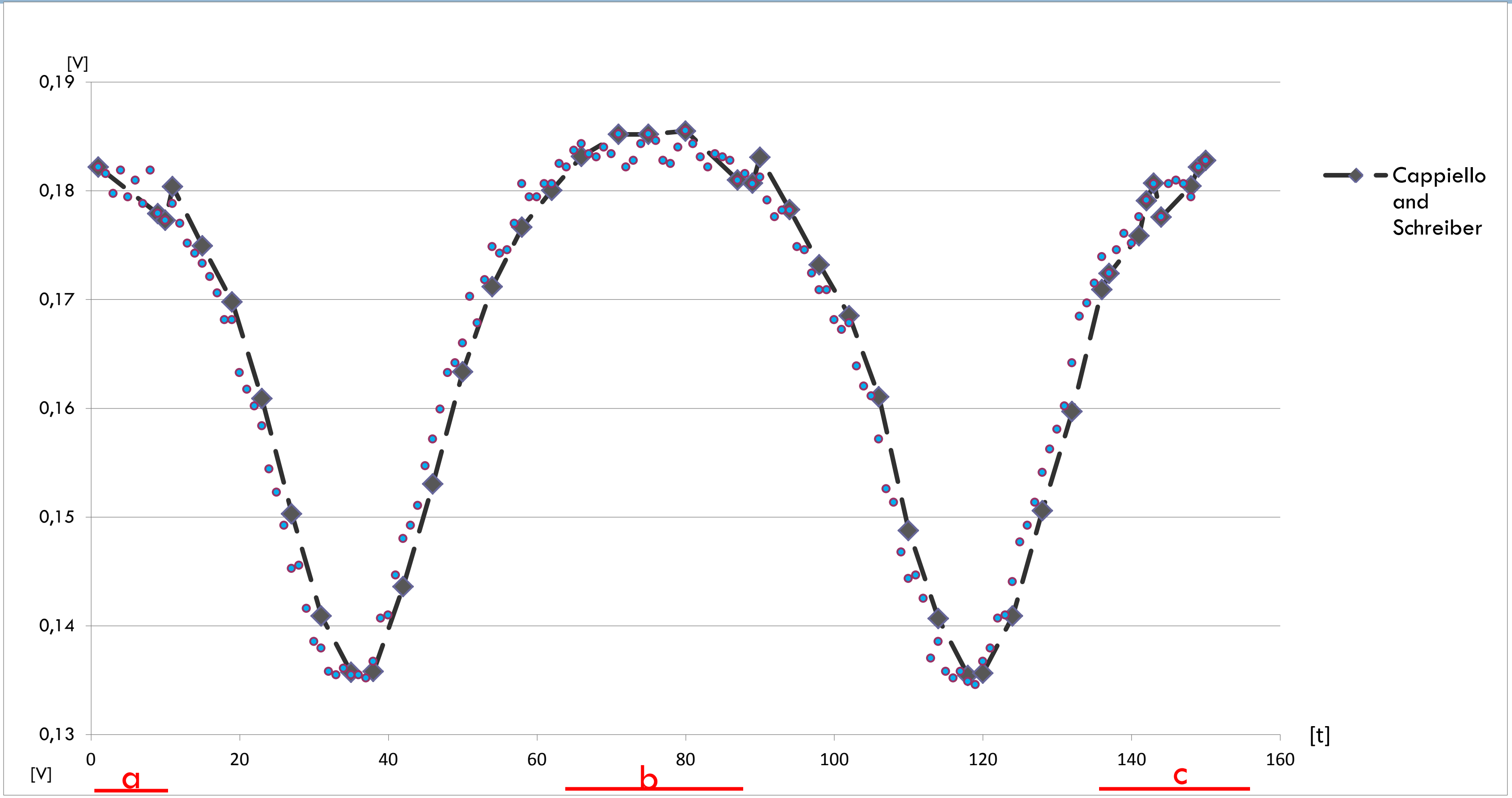
# COMPARISON CRITERIA (2/2)



- The comparison among algorithms have been based on three main criteria:
  - *Compression rate*: the degree with which data have been aggregated.
  - *Energy savings*: the degree with which the aggregation allows sensors to save energy with respect to the case in which all the original values are sent to the base station.
  - *Correctness*: the degree with which the aggregated data allow the base station to retrieve the original trend. Correctness has been evaluated by using the Mean Absolute Error (MAE) and the related Mean Absolute Percentage Error (MA%E).

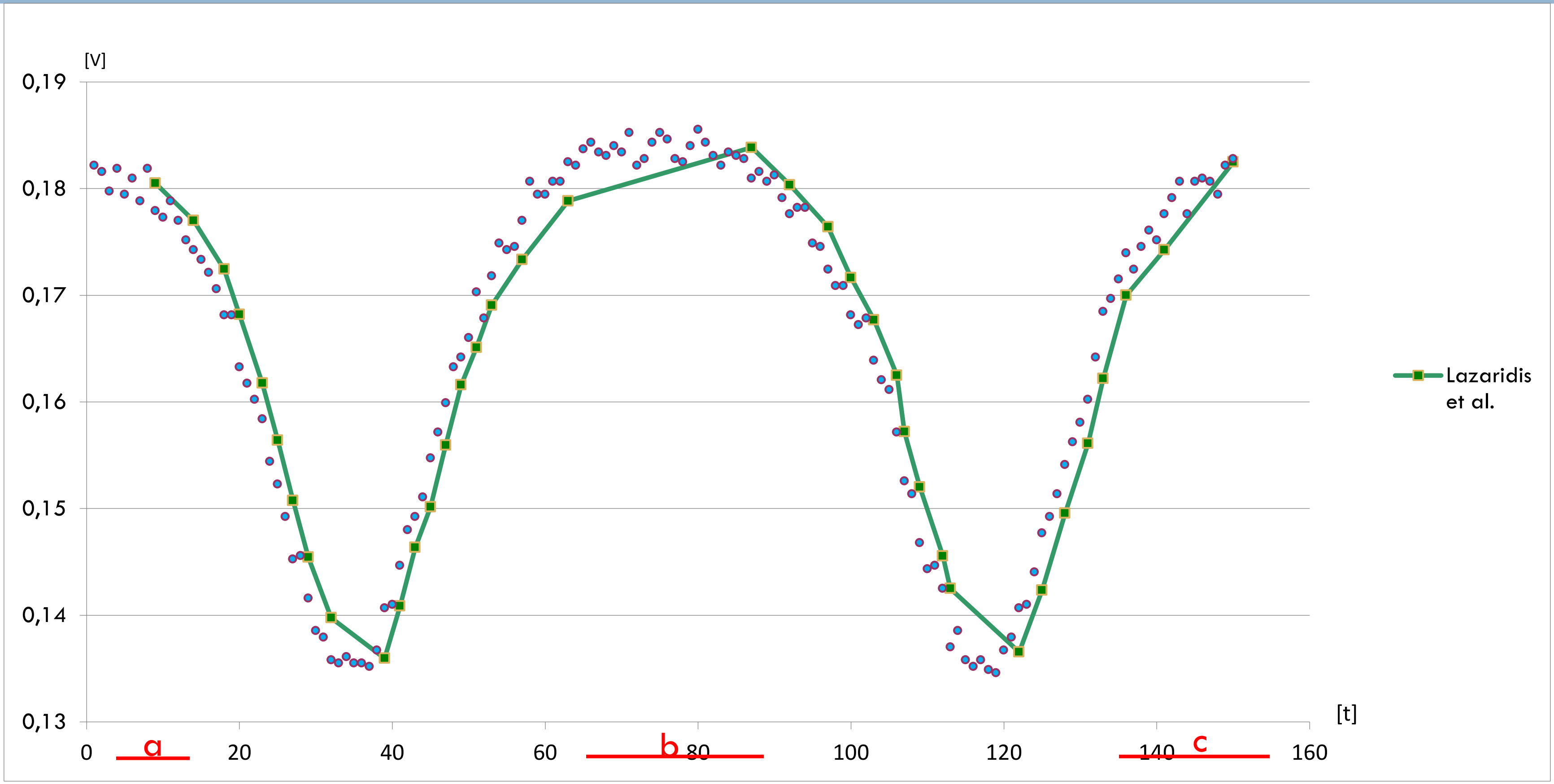
# DATA SET (A) RESULTS





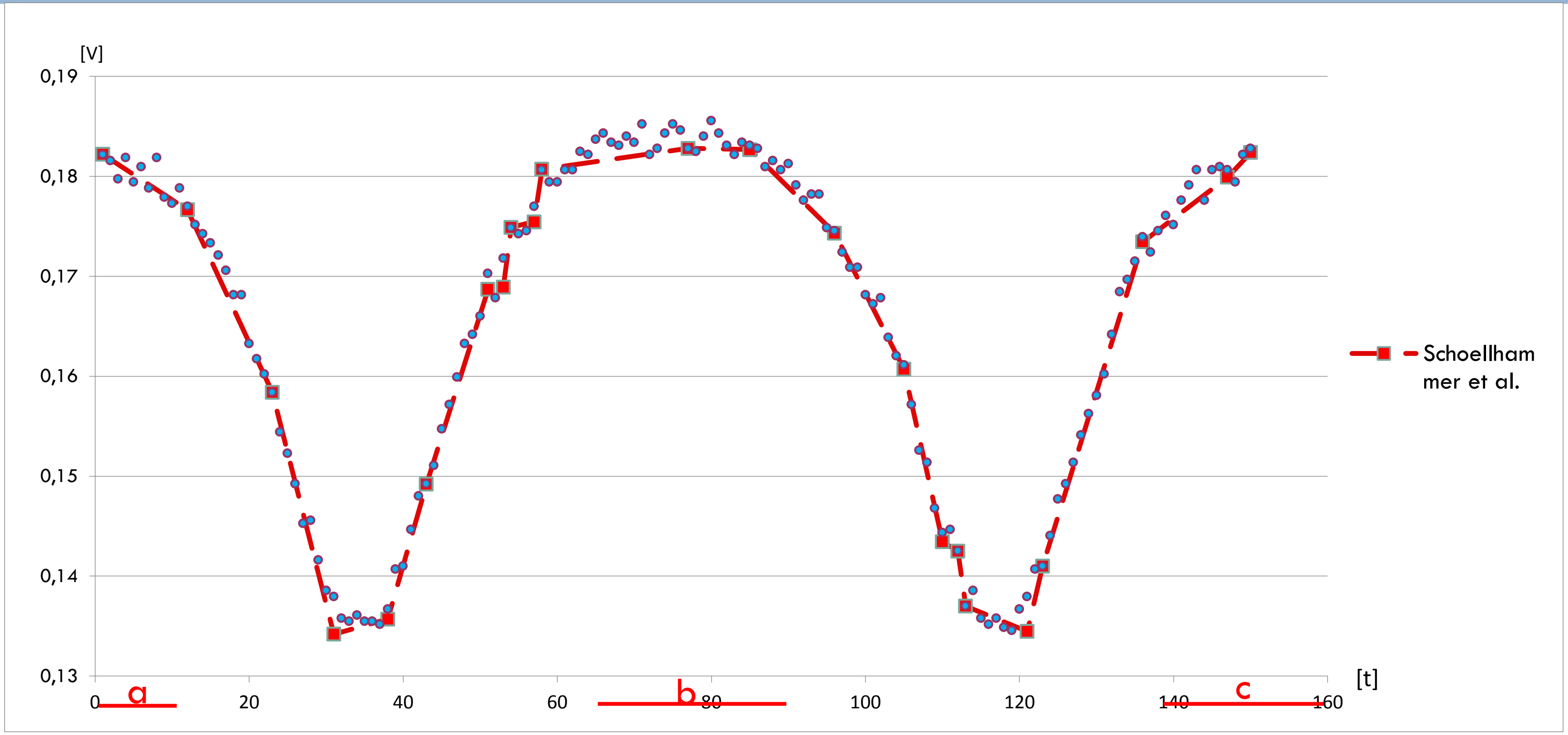
# DATA SET (A) RESULTS





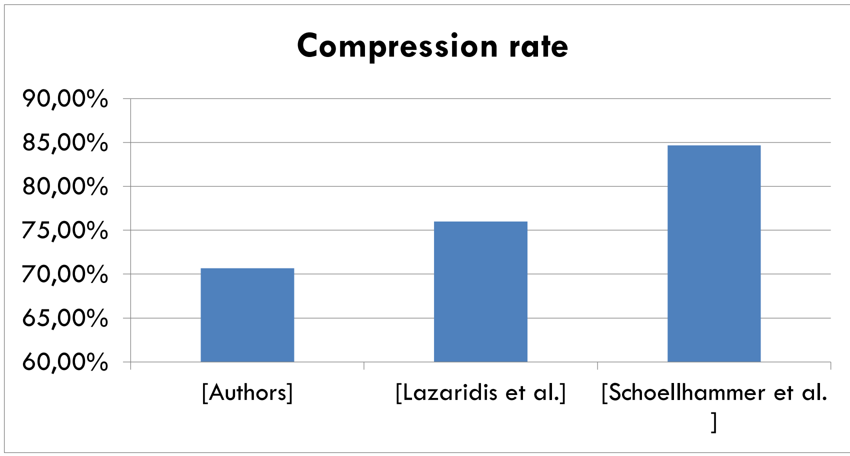
# DATA SET (A) RESULTS

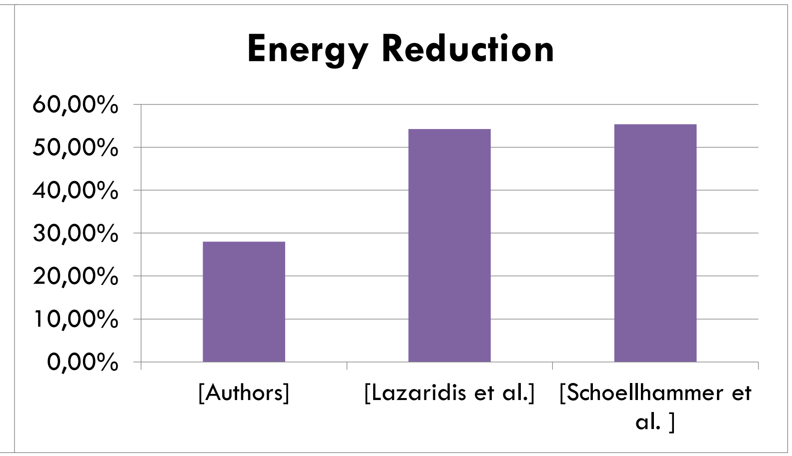




# DATA SET (A) RESULTS





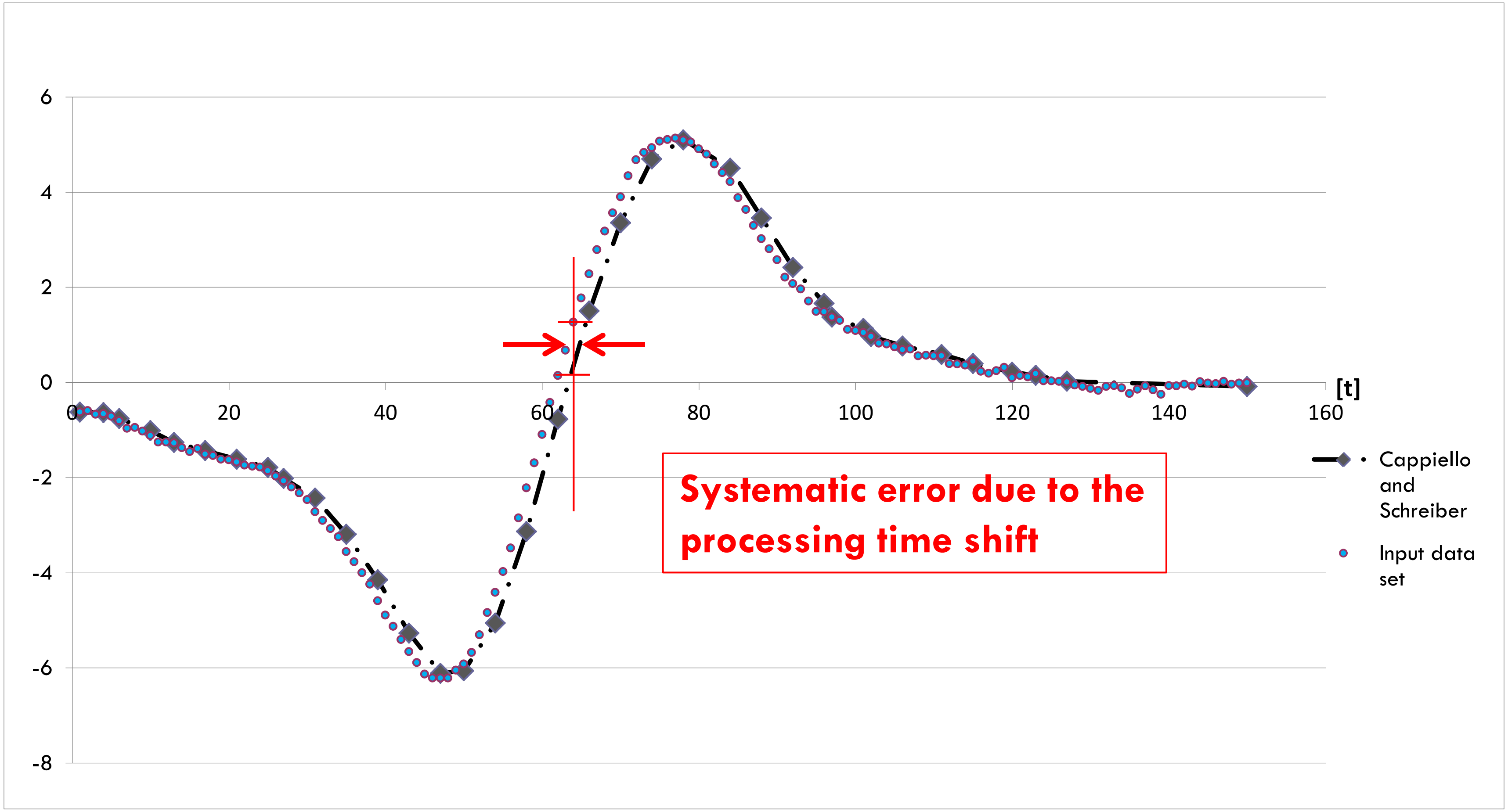




*MAE in case of non linear trends*

| | Authors | Schoellhammer et al. |
|---|---|---|
| (a) | 0.0008 | 0.0009 |
| (b) | 0.0011 | 0.0014 |
| (c) | 0.0008 | 0.0009 |

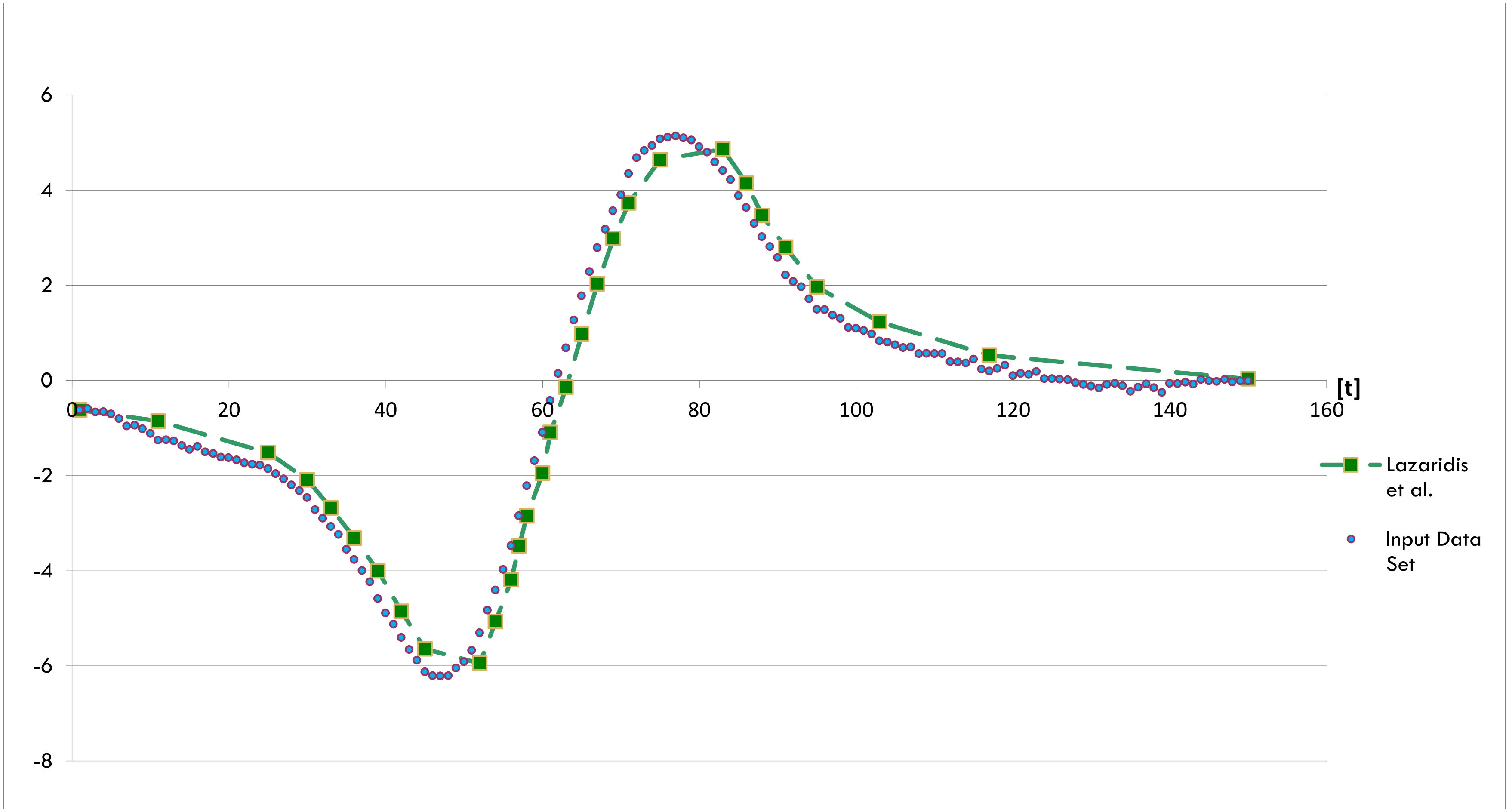
# DATA SET (B) RESULTS





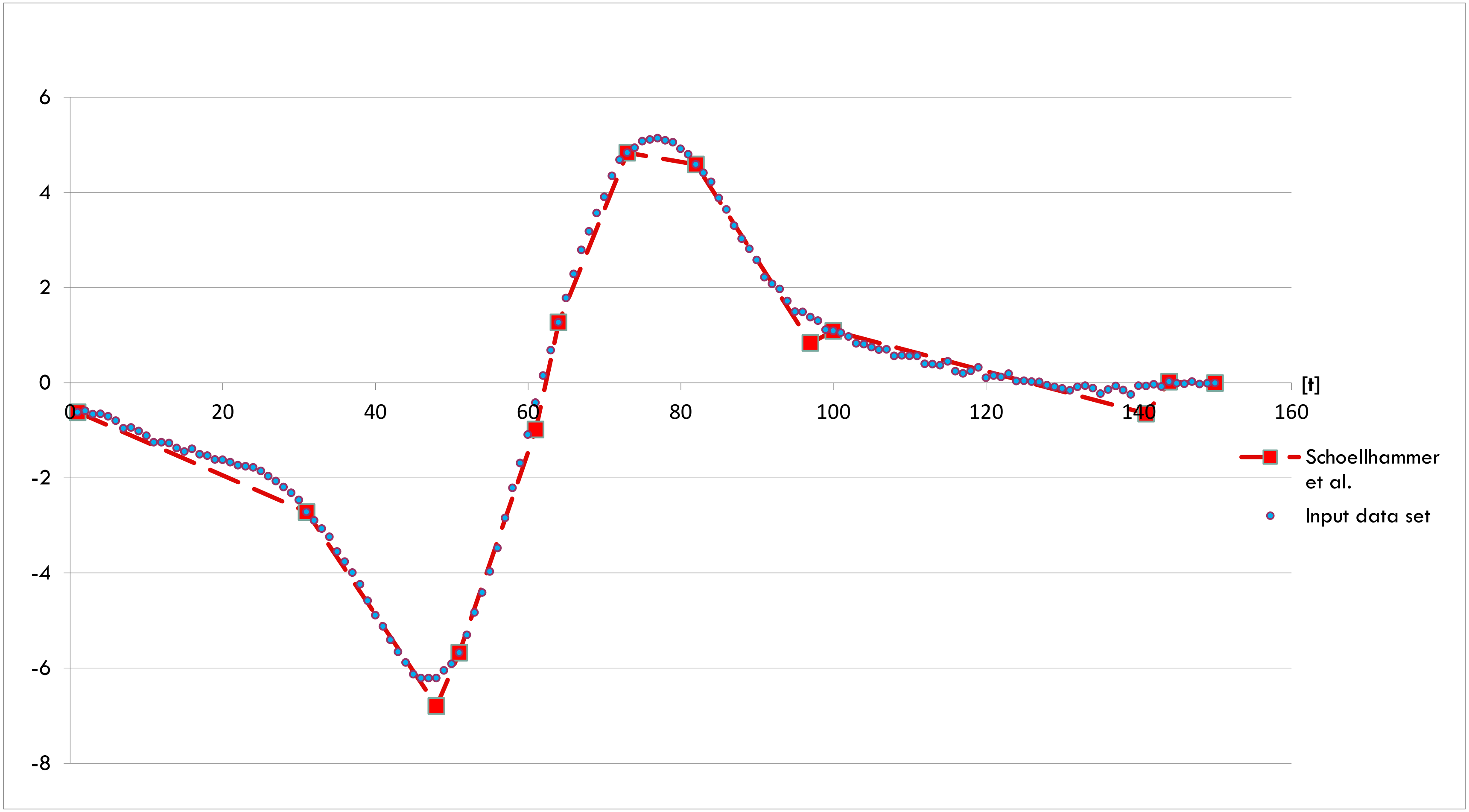
# DATA SET (B) RESULTS





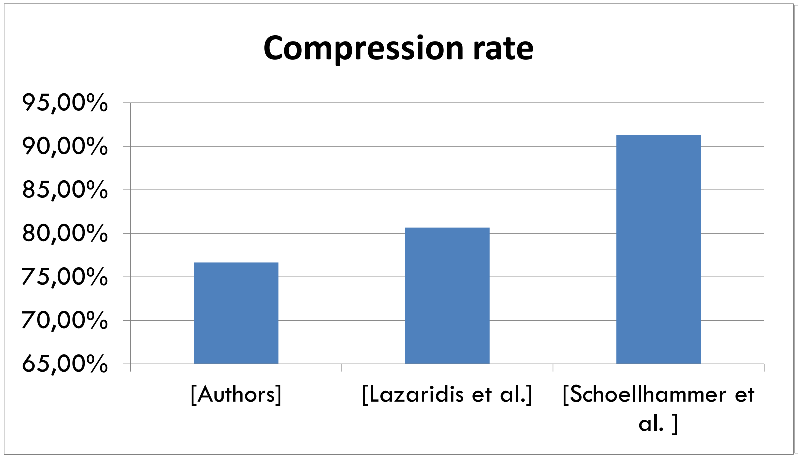
# DATA SET (B) RESULTS

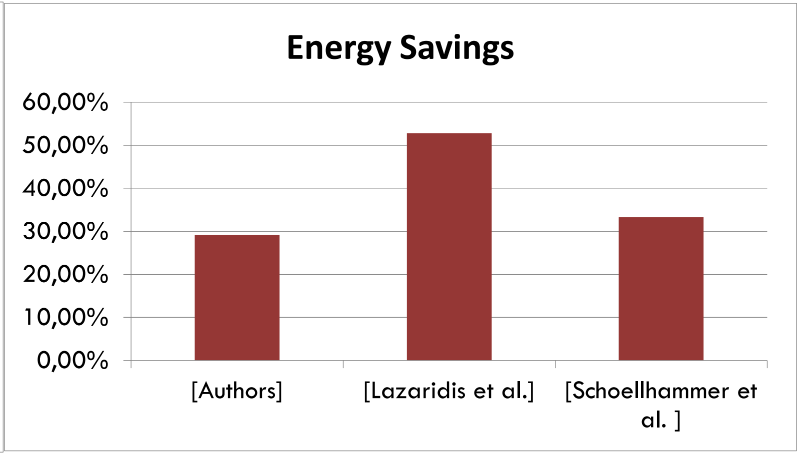




# DATA SET (B) RESULTS





# SUMMARY COMPARISON AND CONCLUSIONS



| | Compression rate/Energy saving | MA%E/Compression rate | MA%E/Energy saving |
|---|---|---|---|
| Authors | 2.5 | 0.64 | **1.6** |
| Lazaridis et al. | **1.55** | 1.77 | 2.77 |
| Schoellhammer et al. | 2.84 | **0.615** | 1.75 |

- No single algorithm is «the best»
- Transmission procedures with packed based protocols can affect the analysis
  - Higher packing factors improve energy efficiency
  - Higher transmission delays negatively affect timeliness
- Adaptable procedures should be used on the basis of
  - The peculiar features of the signals to be processed
  - The quality requirements of the applications

# OTHER COMPRESSION TECHNOLOGIES




**THANK YOU**