



5<sup>th</sup> International Workshop on Semantic Sensor Networks at  
ISWC2012

# Deriving Semantic Sensor Metadata from Raw Measurements

[Jean-Paul Calbimonte](#)<sup>1</sup>, Hoyoung Jeung<sup>2</sup>,  
Zhixian Yan<sup>3</sup>, Oscar Corcho<sup>1</sup>, Karl Aberer<sup>3</sup>

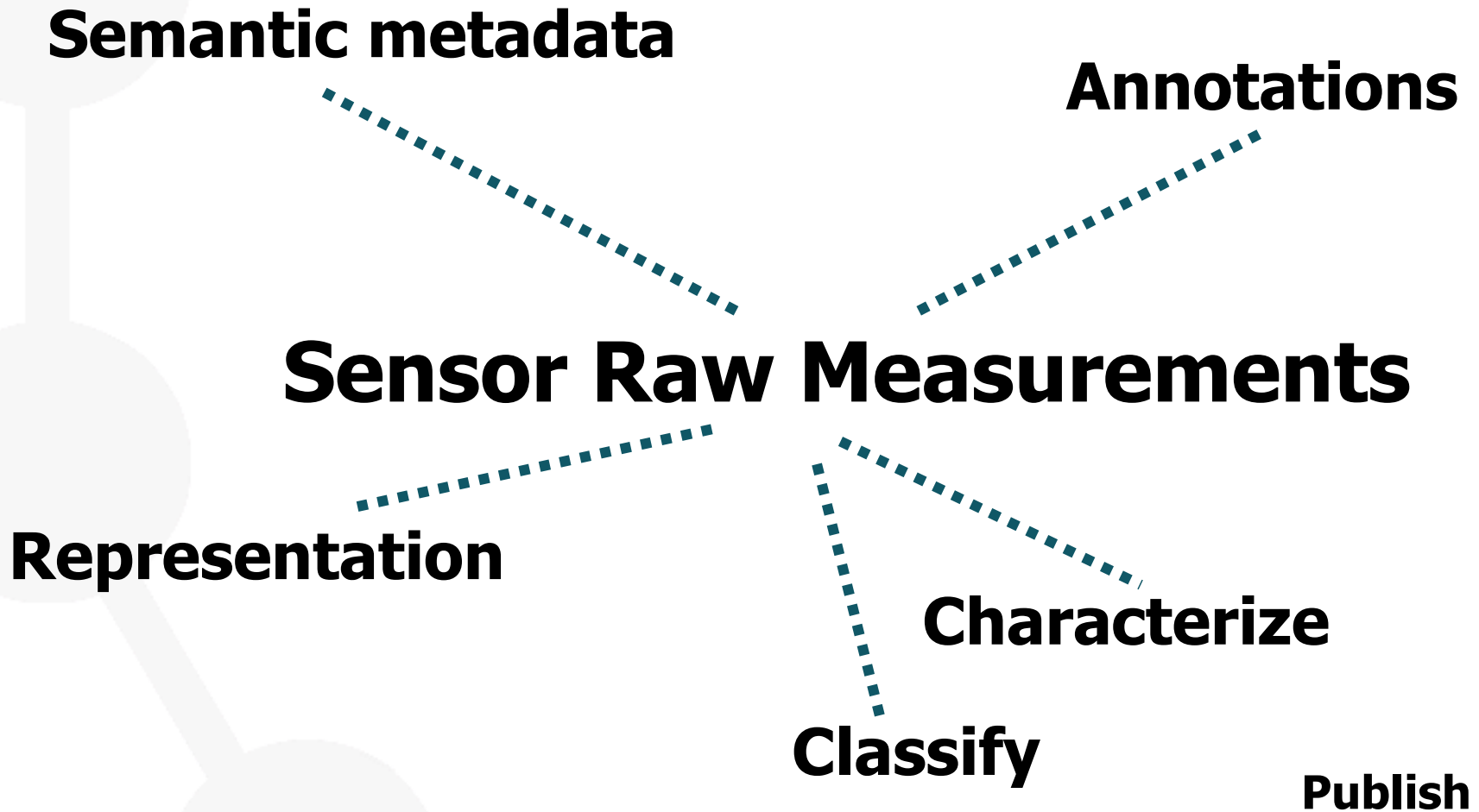
<sup>1</sup>Ontology Engineering Group, Universidad Politécnica de Madrid

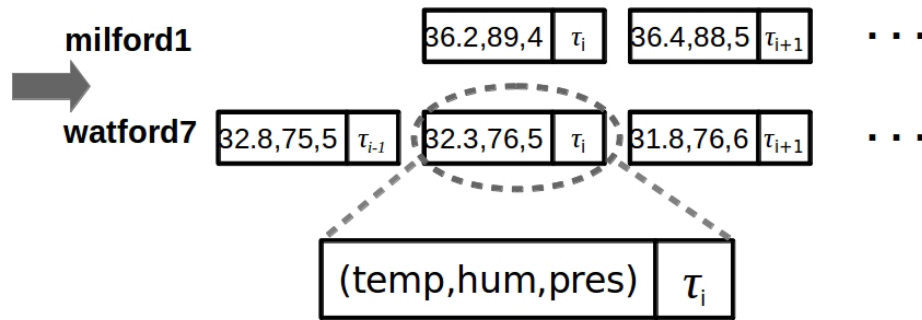
<sup>2</sup>SAP Research, Birsbane

<sup>3</sup>LSIR, EPFL Ecole Polytechnique Fédérale de Lausanne

[jp.calbimonte@upm.es](mailto:jp.calbimonte@upm.es)





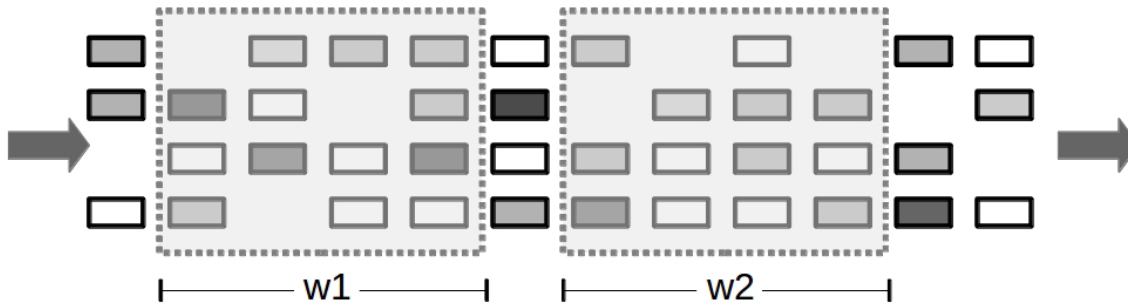


Data streams

Continuous evaluation

Timestamped tuples

Time windows



Do we know **what** we are **sensing**?



Weather Sensors



Sensor Dataset



GPS Sensors



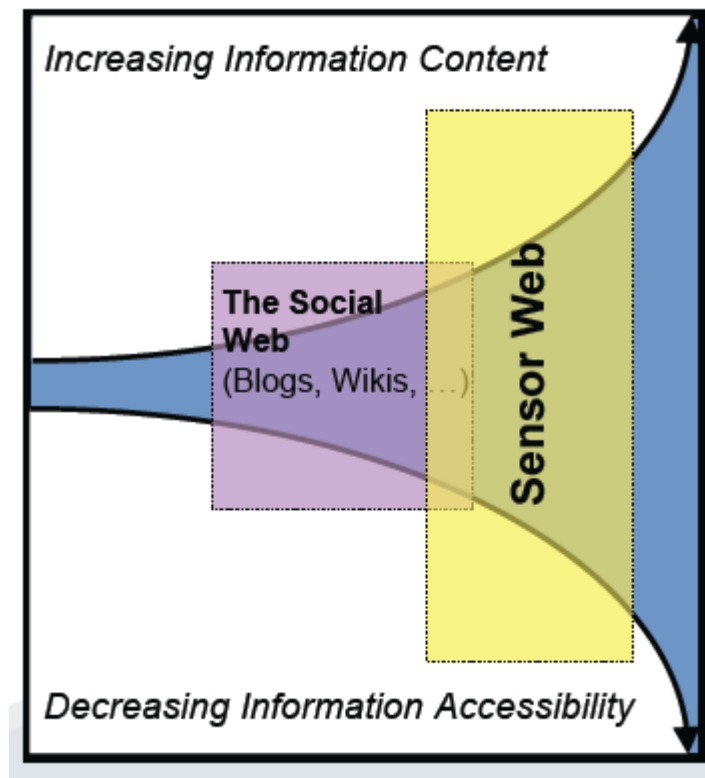
Satellite Sensors



RTMS (Remote Traffic Microwave Sensor) radar  
Camera Sensors

Source: H Patni, C Henson, A Sheth

## Universal, web-based access to sensor data



Source: Adapted from Alan Smeaton's invited talk at ESWC2009

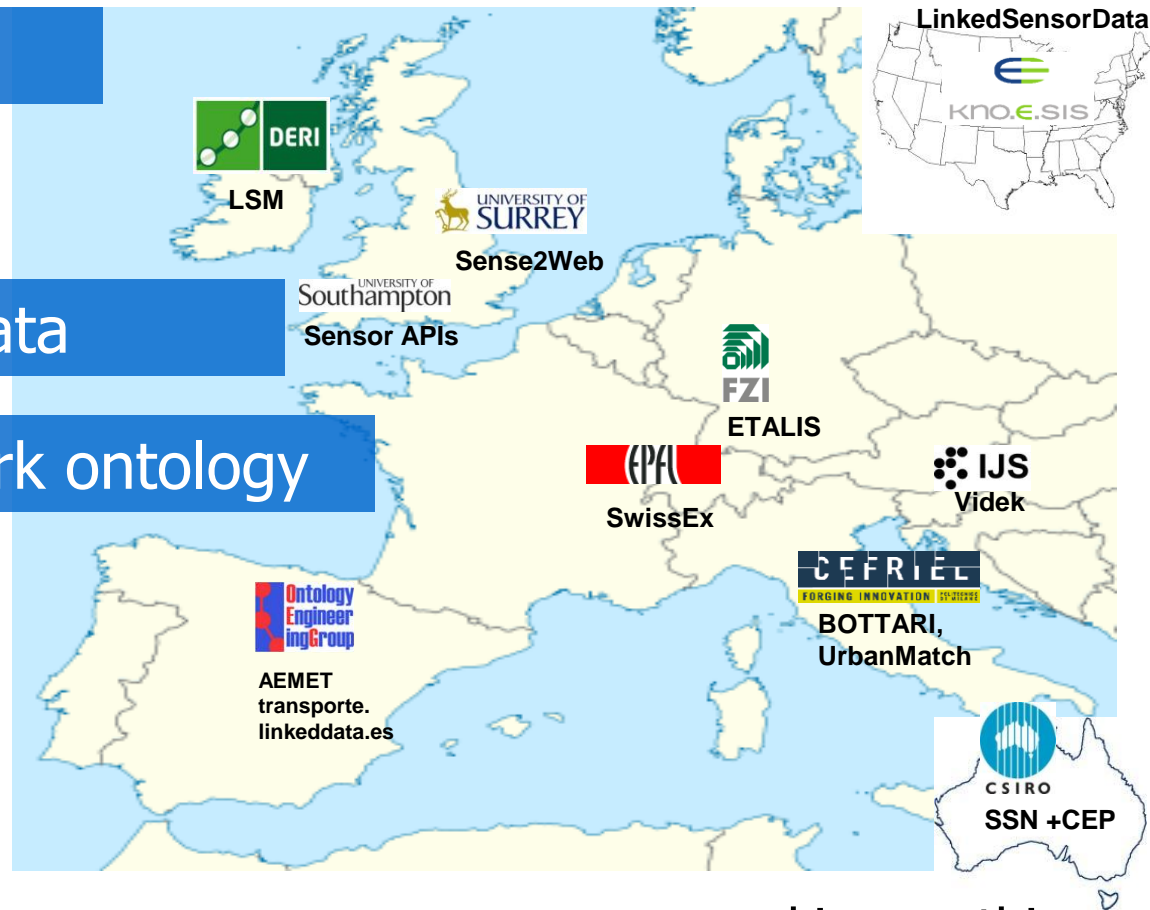
“too much (**streaming**) data but not enough (tools to gain and derive) **knowledge**”\*

Sensor data publishing

Linked Data

Semantic sensor metadata

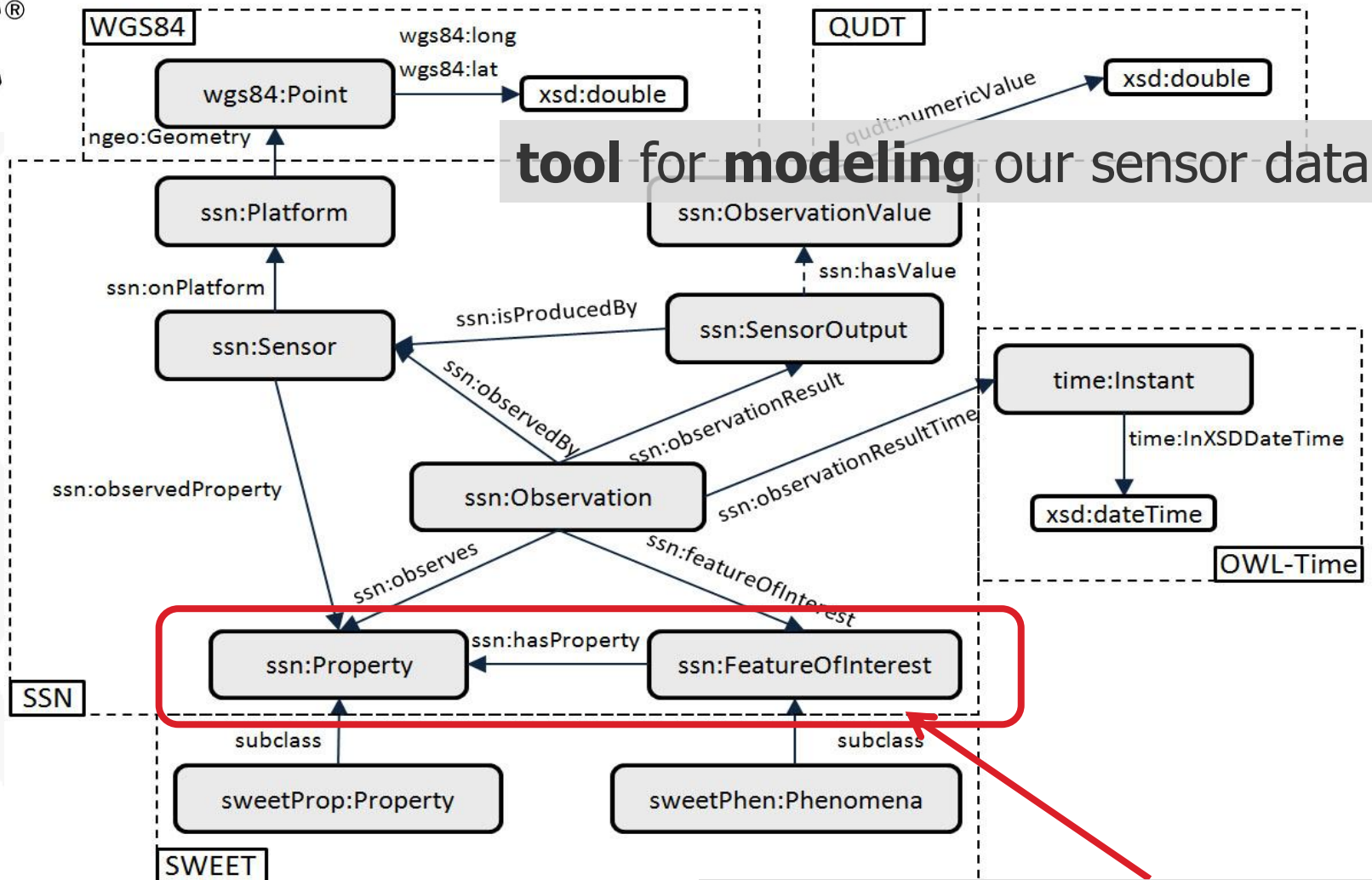
Semantic Sensor Network ontology



...many many more working on this

\* Sheth et al. 2008, Semantic Sensor Web

# SSN Ontology with other ontologies



tool for modeling our sensor data

~what we are observing

swissex:Sensor1

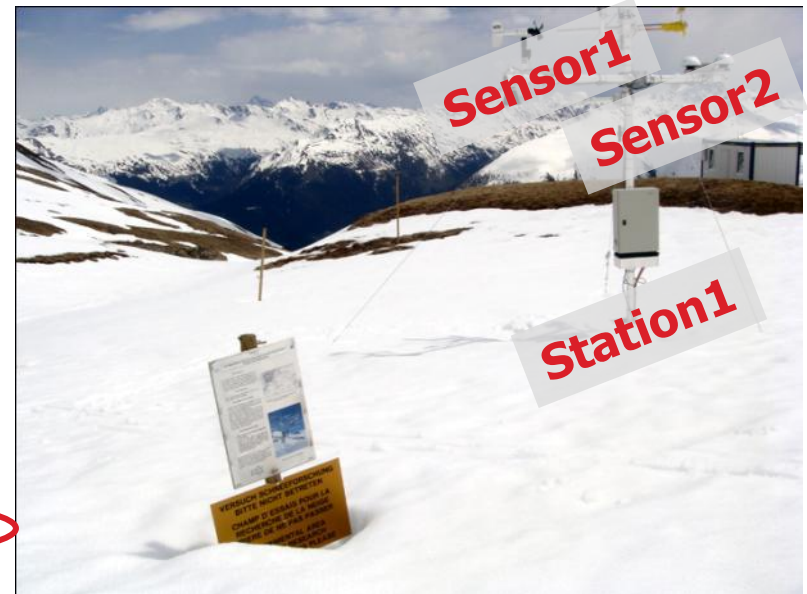
rdf:type ssn:Sensor;  
ssn:onPlatform swissex:Station1;  
ssn:observes cf-property:wind\_speed.

swissex:Sensor2

rdf:type ssn:Sensor;  
ssn:onPlatform swissex:Station1;  
ssn:observes cf-property:air\_temperature.

swissex:Station1

:hasGeometry [ rdf:type wgs84:Point;  
wgs84:lat "46.8037166";  
wgs84:long "9.7780305" ].





swissex:WindSpeedObservation1

rdf:type ssn:Observation;

ssn:featureOfInterest cf-feature:wind;

ssn:observedProperty cf-property:wind\_speed;

ssn:observationResult [rdf:type ssn:SensorOutput;

ssn:hasValue [qudt:numericValue "6.245"^^xsd:double]];

ssn:observationResultTime [time:inXSDDatatime "2011-10-26T21:32:52"];

ssn:observedBy swissex:Sensor1 ;



WindSpeed : 6.245  
At: 2011-10-26T21:32:52

**GSN Server :: GSN**

HOME DATA MAP FULLMAP

Auto-refresh every : 1min refresh close all

**wan4** 03/08/2009 11:10:00 +0200

Real-Time	Addressing	Structure	Description	Download
record 7920.0				
relative_humidity 74.44				
air_temperature 8.326				
wind_speed_northerly_component 8.13				
wind_speed_westerly_component -10.28				
wind_speed_vertical_component -10.04				
snow_height -198.3				

**wannengrat\_unterhalb\_felsen** 03/06/2010 02:06:00 +0200

Real-Time	Addressing	Structure	Description	Download
wind_direction null				
wind_speed null				
solar_rad null				
snow_water_content_1 null				
snow_water_content_2 null				
snow_temp_1 null				
snow_temp_2 null				
air_humid null				
air_temp null				
air_temp_ir null				
snow_surface_temp_ir null				

Heterogeneity

Reuse this data?

Publish as Linked Sensor Data?

Query with SPARQL-Stream?

## **AuroraWatch test**

Test magnetometer for AuroraWatch UK. Magnetometer is under test, for real data see <http://aurorawatch.lancs.ac.uk/>

Temperature and battery voltage not yet operational.

aurorawatch magnetometer V counts

**LIVE** Updated on Fri, 03 Feb 2012 22:10:56 +0000 | Published by [stevemarple](#)

## **Envi**

Envi

envi currentc...

**LIVE** Updated on Tue, 07 Feb 2012 16:07:34 +0000 | Published by [xemuel](#)

## **test1**

premier test de pachube

test V °C

**LIVE** Updated on Tue, 07 Feb 2012 16:07:34 +0000 | Published by [tuxtof](#)

## **Hallway**

air fridge hallway humidit temperature °C %

**LIVE** Updated on Tue, 07 Feb 2012 16:07:34 +0000 | Published by [zws](#)

## **Arduino Analog Read**

This is a test for the Arduino and Ethernet shield. It is reading values from the connected to the analog input 1.

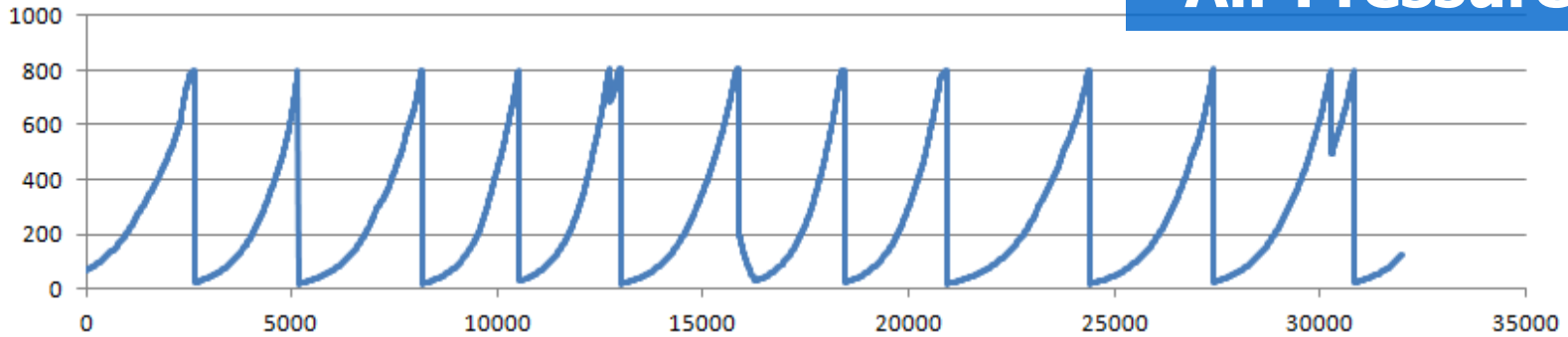
The data is updated every 5 seconds.

Analog Arduino Ethernet Read Shield Uno Wire

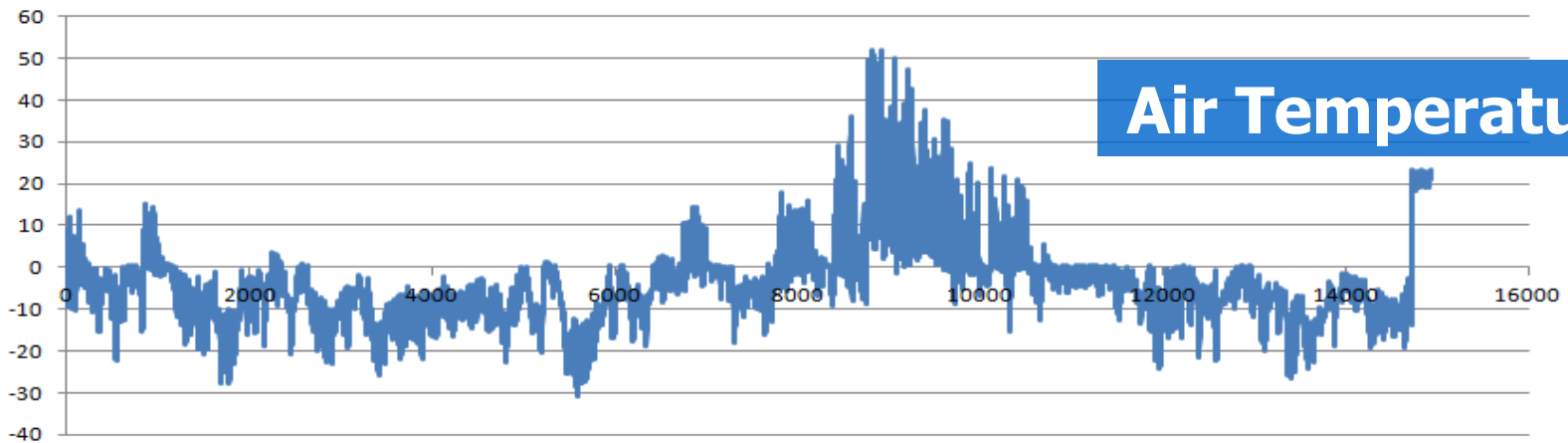
**LIVE** Updated on Tue, 07 Feb 2012 16:07:34 +0000 | Published by [joe95443](#)

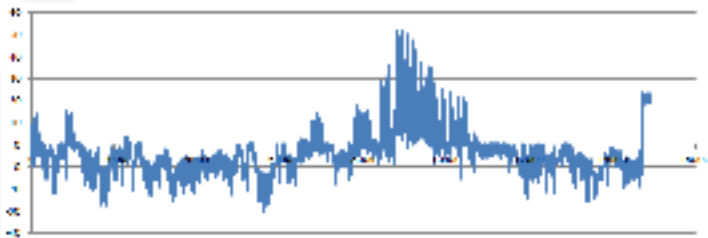
not enough **reliable metadata** about the observations

## Air Pressure?

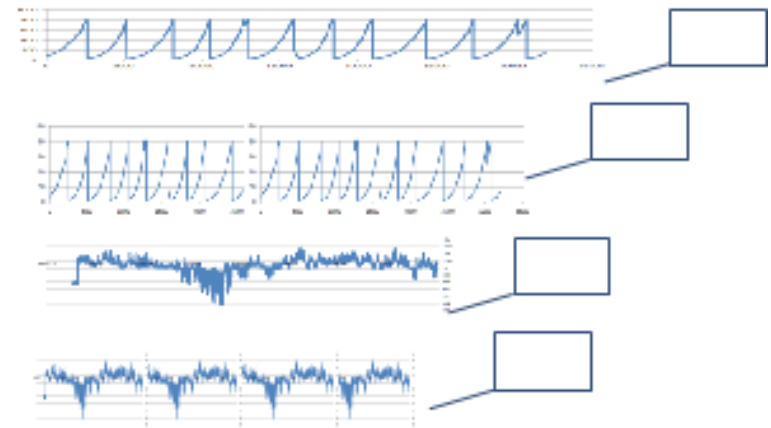


## Air Temperature?





Unclassified input series



Already classified time series

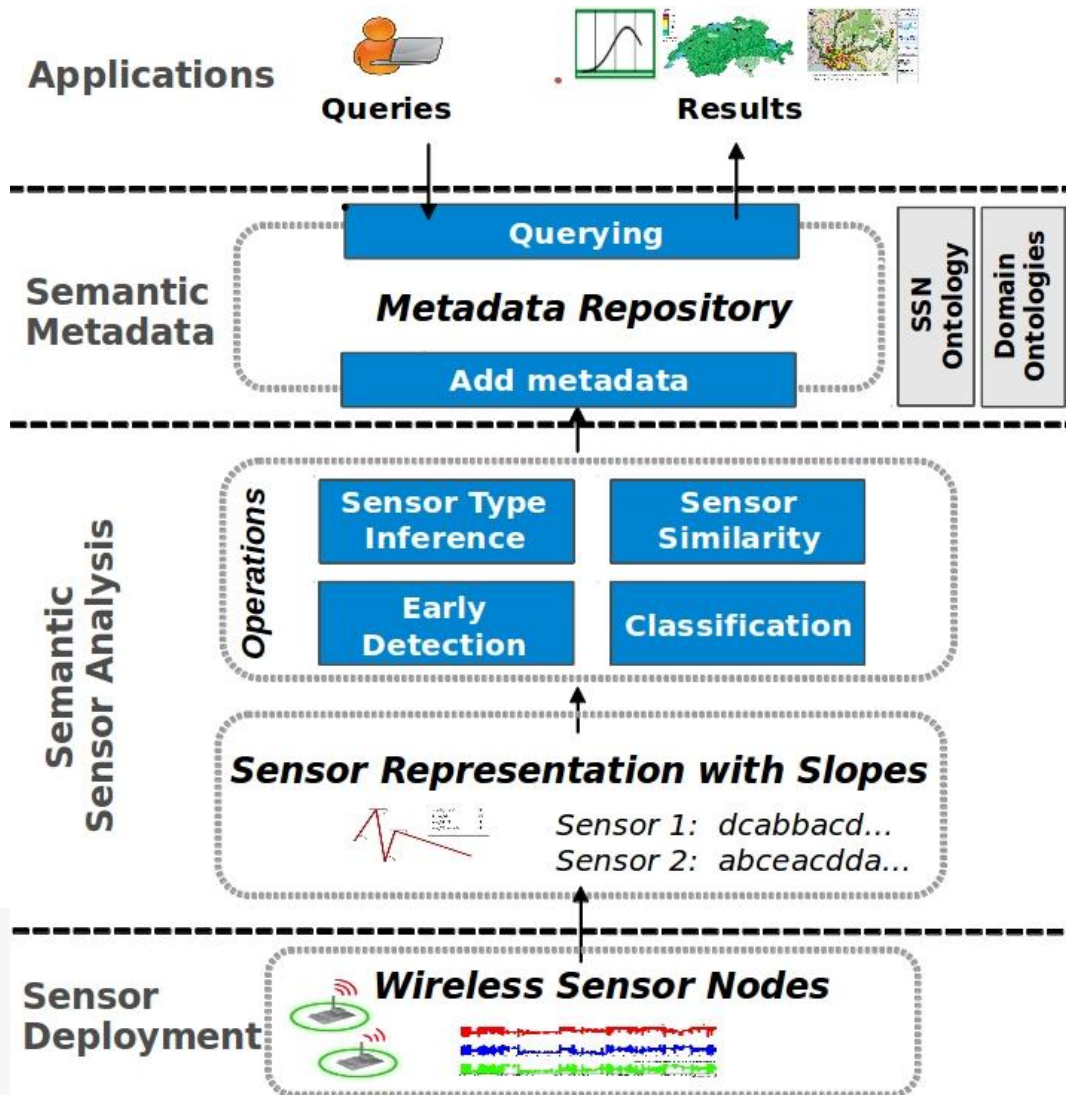
**Representation**

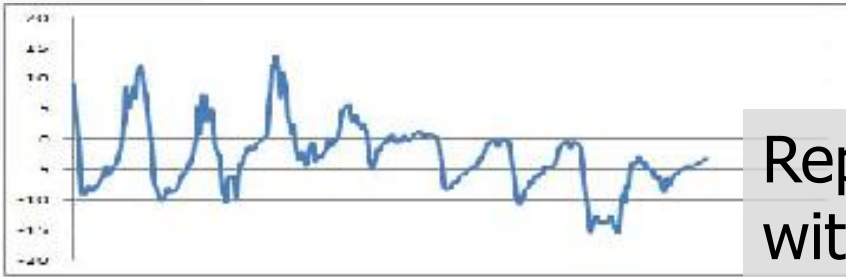
**Classification**

**Metadata**

- Querying time series
  - e.g. find a sub-sequence in a time series database
- Measuring time series similarity
  - e.g. are these time series the same?
- Time series classification
  - e.g. classify heart beat series: normal, murmur, et

# Deriving Semantic Metadata

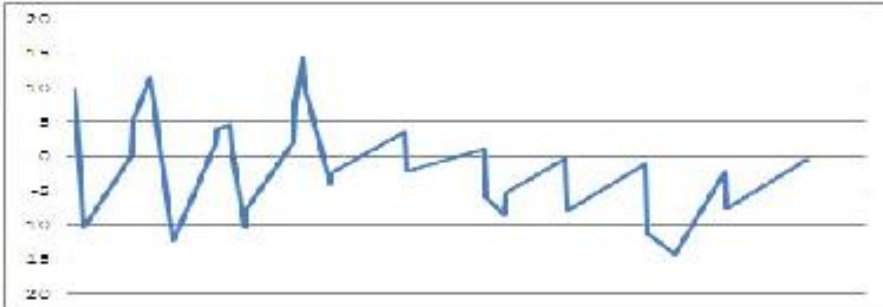




Represent the data **approximating** with fewer linear segments

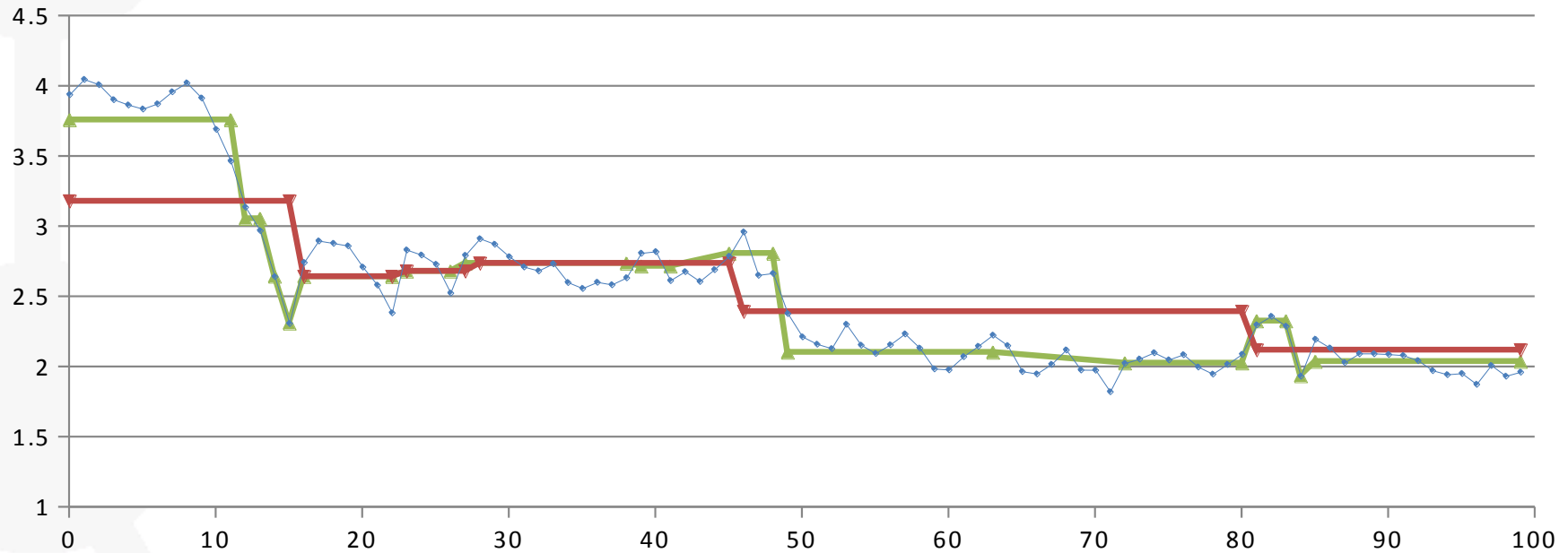


Accuracy vs Numerosity

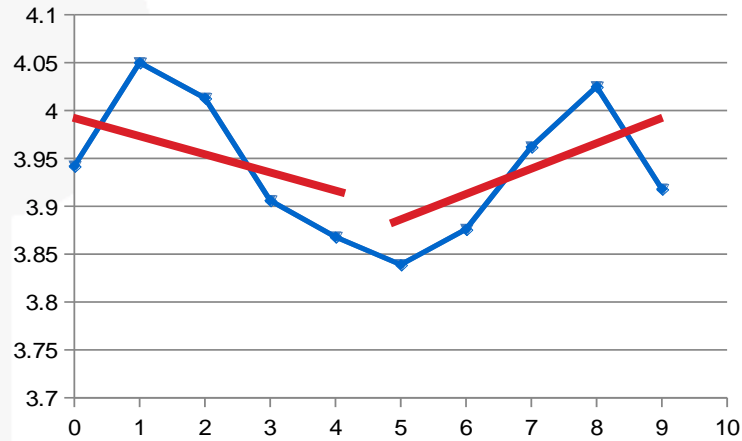




- Use constant segments for a subset of points



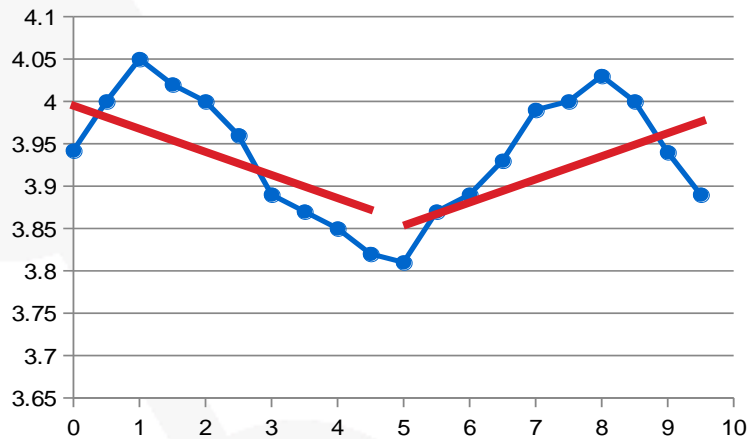
# Piecewise Linear Approximation



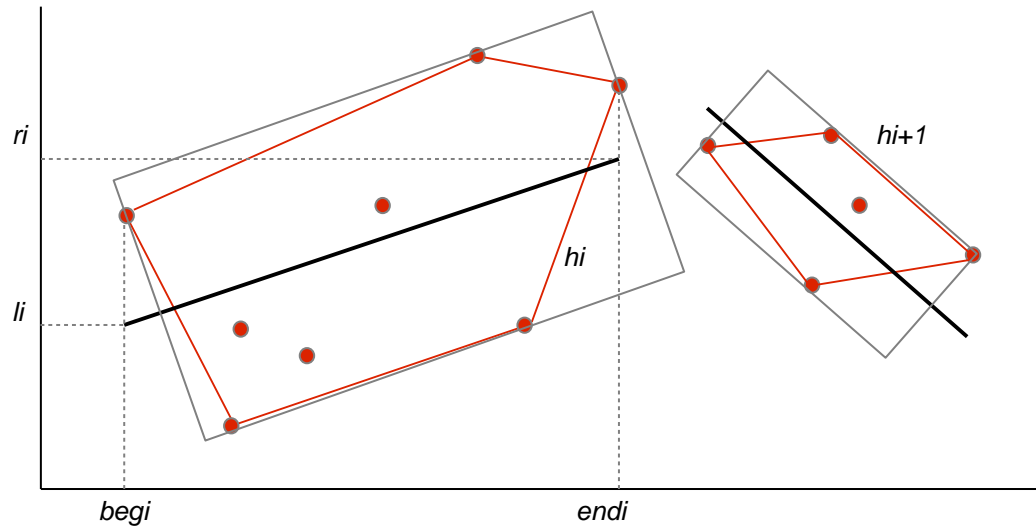
Reflect data trends

Apply with different resolutions

Applicable for different rates

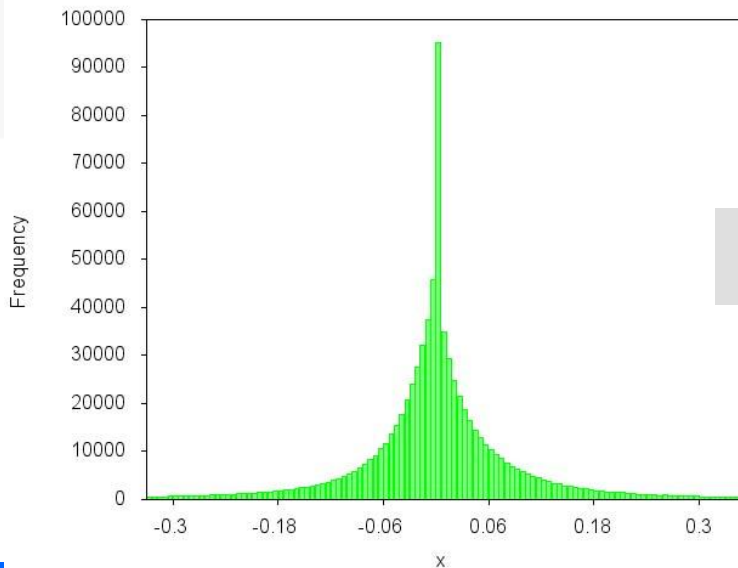
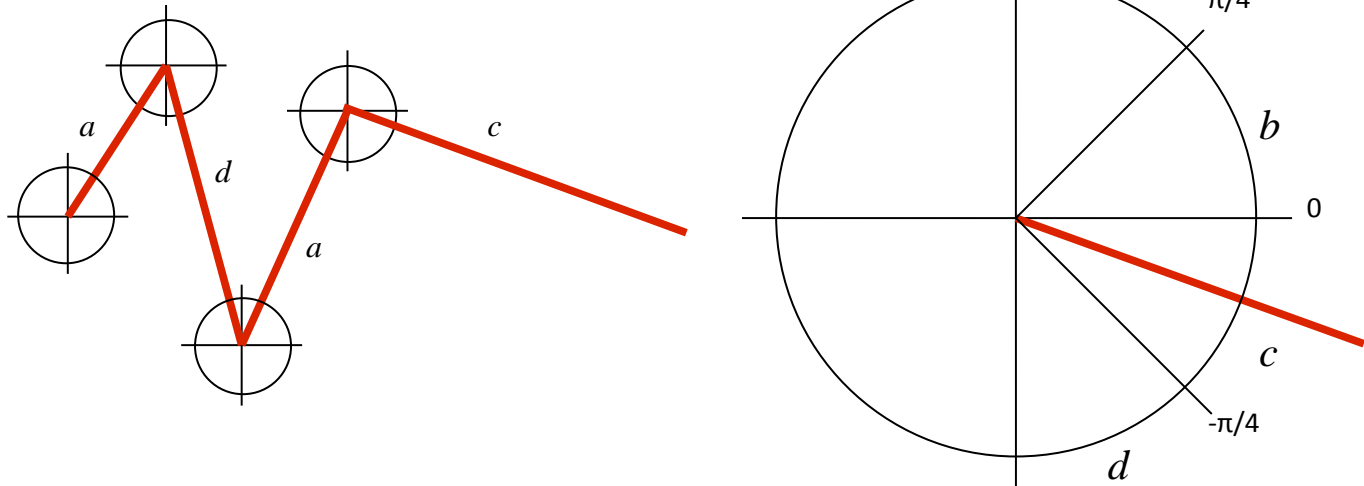


# Piecewise Linear Approximation



**Constructing segments**

We care about the angles



Divide the angle space in sectors

Distribution of angles in training set

ts1	→	adacdaaad	[5a,0b,1c,3d]
ts2	→	adabbaaad	[5a,2b,0c,2d]
ts3	→	adccdaaad	[4a,0b,2c,3d]

**Distance measure**

**Classification**

Linear approximation

Compute distribution of the slopes

K-nearest neighbor classification

Training-Test datasets:

SwissExperiment

AEMET

type	# series	sampling interval			
		< 0.5	$\geq 0.5,$ < 4	$\geq 4,$ < 15	$\geq 15$
humidity	34	1	19	6	8
temperature	78	1	37	23	17
CO <sub>2</sub>	11	11	0	0	0
lysimeter	6	0	0	6	0
moisture	20	0	3	16	1
pressure	4	4	0	0	0
radiation	34	0	14	8	12
snow height	4	0	0	2	2
voltage	16	0	0	11	5
wind direction	35	0	20	8	7
wind speed	46	0	21	13	12

Table 1: Swiss Experiment dataset.

type	# series
humidity	100
temperature	100
battery	81
precipitation	100
soil temperature	81
pressure	98
wind direction	100
wind speed	100
wind direction (max)	100
wind speed (max)	100

Table 3: AEMET dataset. Sampling interval is 10 minutes for all series.

## Confusion matrix

Match results Swissex k=5, 5-fold

test set	ra	mo	te	wd	ws	hu	ly	pr	co	sh	vo	total	fp	tp	fn	p	r
radiation	15		4	7	7						1	34	19	15	0	0.441	1
moisture		10	3	2	1					1	1	20	8	10	2	0.556	0.833
temperature	2	3	56		1	11					2	78	19	56	3	0.747	0.949
wind direction	4		1	25	4							35	9	25	1	0.735	0.962
wind speed			1	4	40	1						46	6	40	0	0.87	1
humidity	1		9		2	21						34	12	21	1	0.636	0.955
lysimeter		2					4					6	2	4	0	0.667	1
pressure								4				4	0	4	0	1	1
co2									10			11	0	10	1	1	0.909
snow height		1	2							1		4	3	1	0	0.25	1
voltage			6		1						9	16	7	9	0	0.563	1
<b>total</b>												<b>288</b>	<b>85</b>	<b>195</b>	<b>8</b>	<b>0.7</b>	<b>0.96</b>

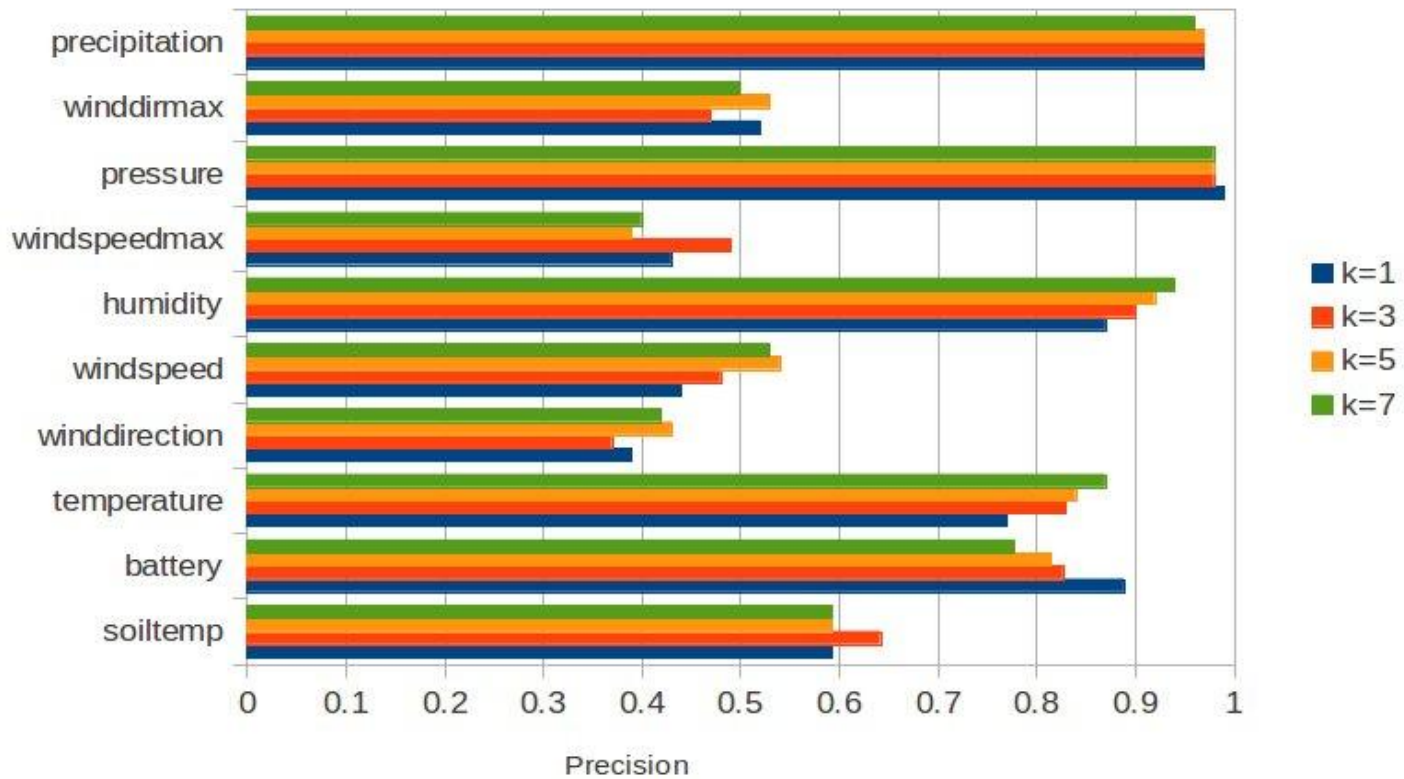
## Confusion matrix

Match results AEMET k=5, 10-fold

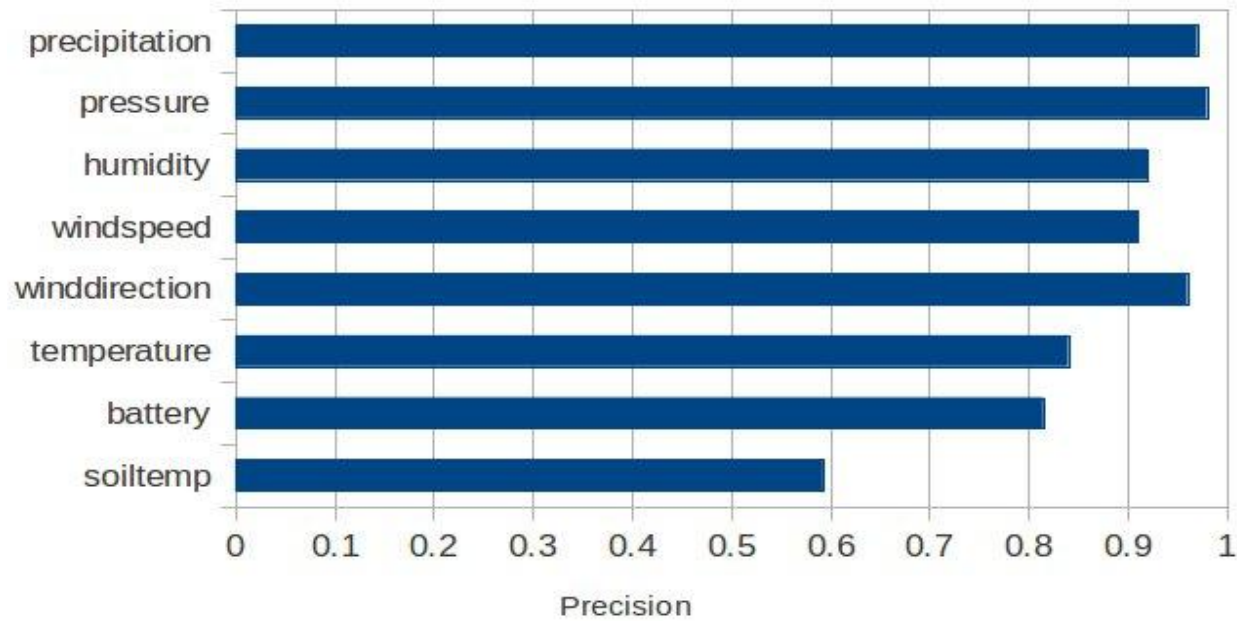
test set	st	ba	te	wd	ws	hu	wsx	pr	wdx	pre	total	fp	tp	fn	p	r
soiltemp	48		23	1	1	2		6			81	33	48	0	0.593	1
battery	2	66		2		1				10	81	15	66	0	0.815	1
temperature	15	1	84								100	16	84	0	0.84	1
winddirection		1		43	3				53		100	57	43	0	0.43	1
windspeed		1	1	2	54	4	37		1		100	46	54	0	0.54	1
humidity	1		1		2	92	2		2		100	8	92	0	0.92	1
windspeedmax		1	1	1	54	3	39		1		100	61	39	0	0.39	1
pressure			2					97			99	2	97	0	0.98	1
winddirmax		1		43	3				53		100	47	53	0	0.53	1
precipitation		2					1			97	100	3	97	0	0.97	1
<b>total</b>											<b>961</b>	<b>288</b>	<b>673</b>	<b>0</b>	<b>0.7</b>	<b>1</b>



## Precision in AEMET for different k



## Precision in AEMET (unified types)



```
cf-property:wind_speed rdf:type dim:VelocityOrSpeed;  
rdfs:label "wind speed";  
ssn:isPropertyOf cf-feature:wind;  
qu:propertyType qu:scalar;  
qu:generalQuantityKind qu:speed.
```

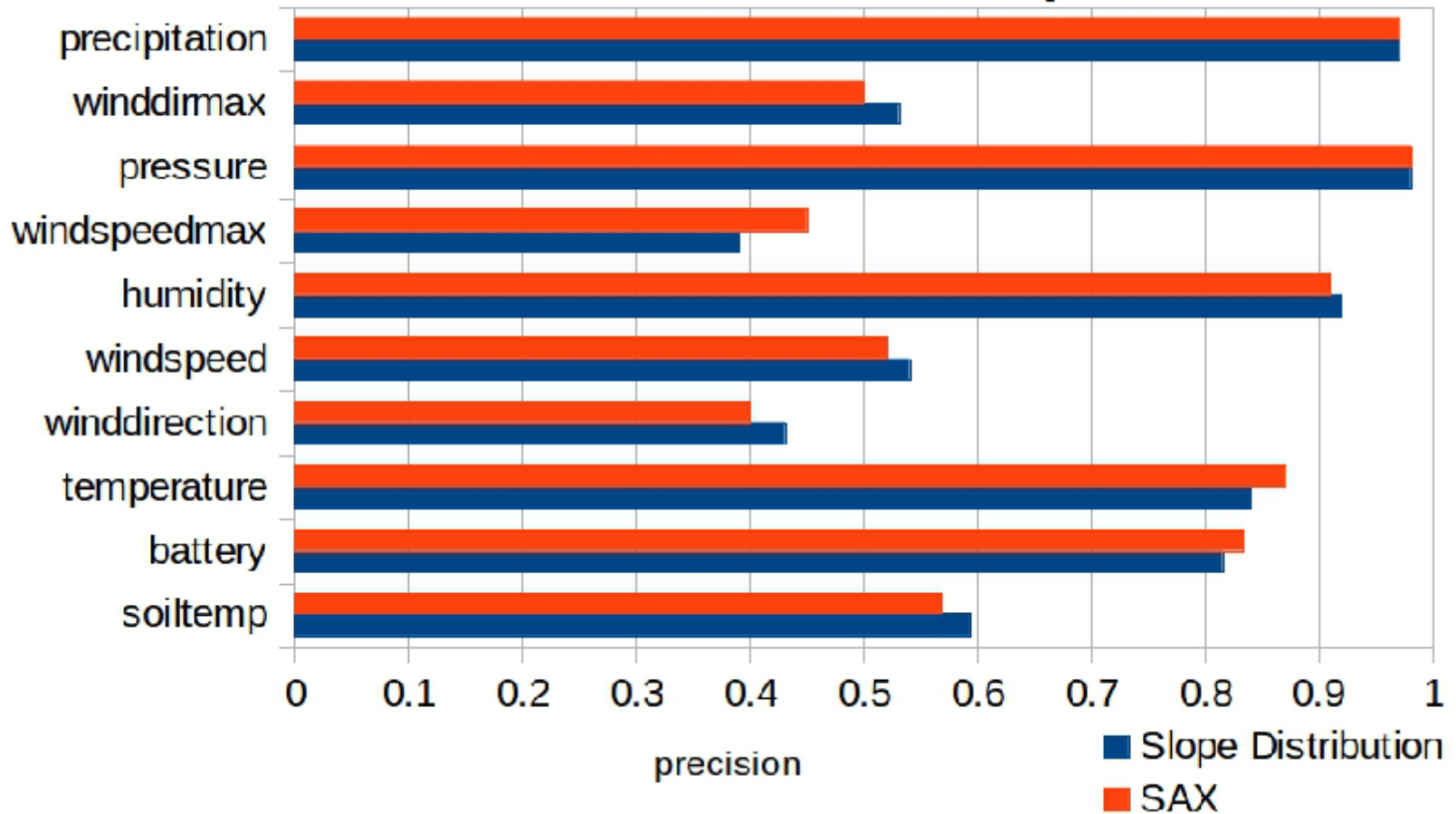
**cf-property:air\_temperature** rdf:type dim:Temperature;  
ssn:isPropertyOf cf-feature:air;  
qu:propertyType qu:scalar;  
qu:generalQuantityKind qu:temperature.

**cf-property:soil\_temperature** rdf:type dim:Temperature;  
ssn:isPropertyOf cf-feature:soil;  
qu:propertyType qu:scalar;  
qu:generalQuantityKind qu:temperature.

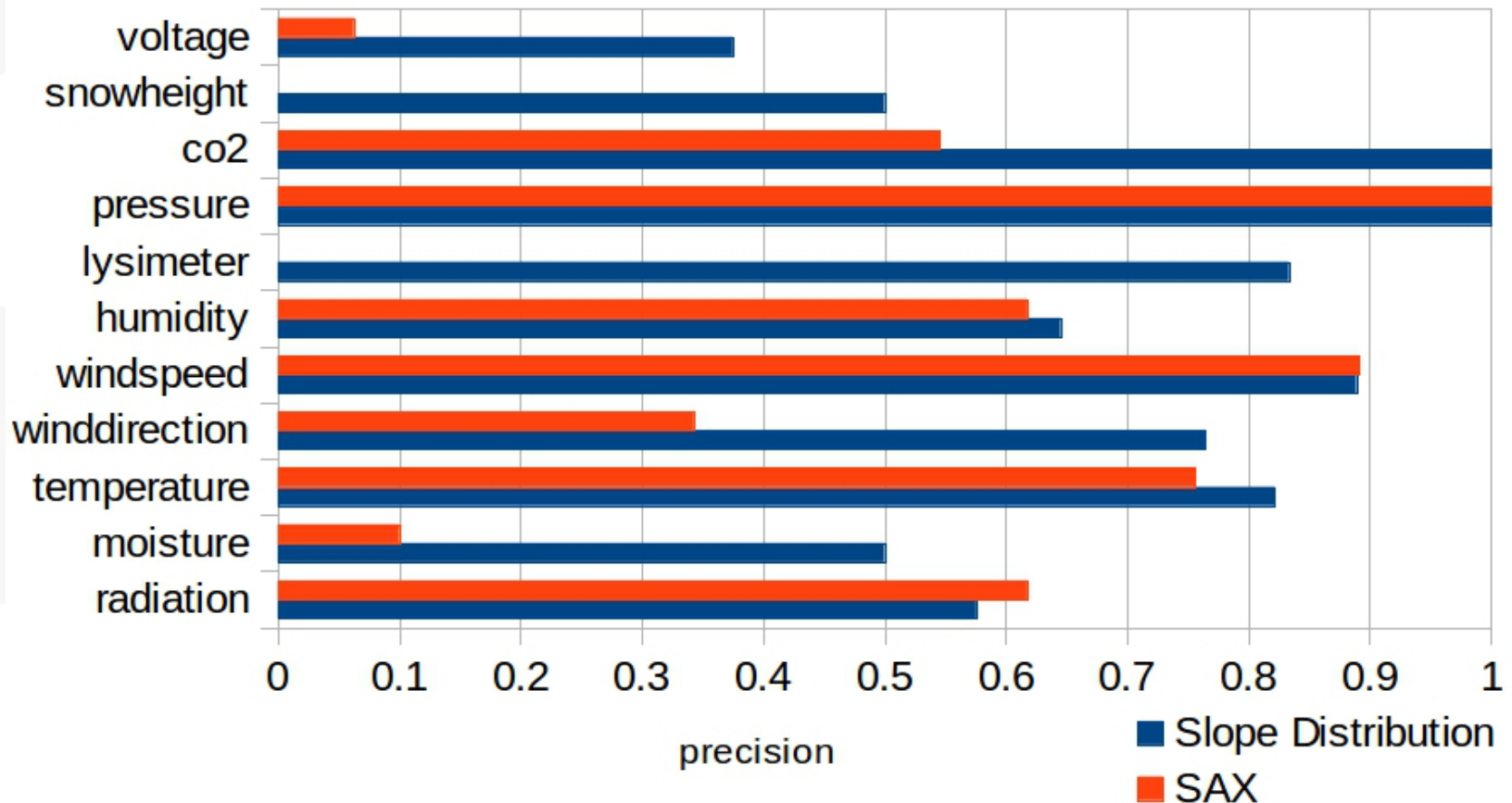
```
SELECT ?sensor
WHERE {
  ?sensor a ssn:Sensor ;
          ssn:observes cf-property:air_temperature .}
```

```
SELECT ?stream ?observedProperty
WHERE {
  ?sensor a ssn:Sensor ;
          ssn:observes ?observedProperty .
  ?stream ssn:isProducedBy ?sensor .
  ?observedProperty qu:generalQuantityKind qu:temperature .}
```

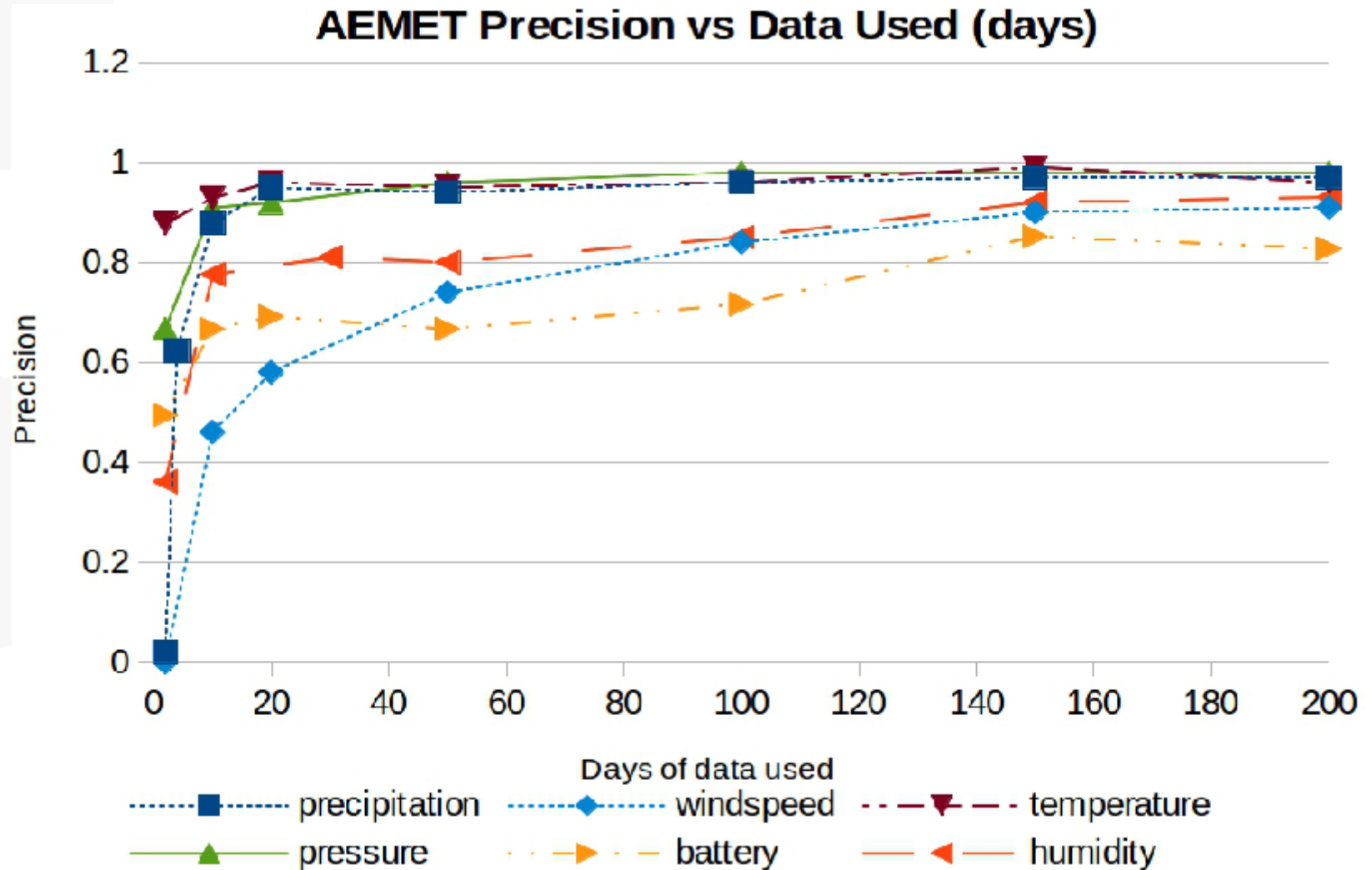
## AEMET Classification Precision: Slope vs SAX



## Swiss Experiment: Classification Precision: Slope vs SAX



# How much data do we need?



## Classify Sensor Data

- Piecewise Linear Representation
- Segment slope distributions
- kNN classification

## Generate Metadata

- Observed properties
- Potentially unknown metadata

## Future work

- Combine with tag disambiguation?
- Use pattern mining for online queries
- Other techniques, shapelets, use other parameters



Questions, please.

**[jp.calbimonte@upm.es](mailto:jp.calbimonte@upm.es)**