

# Big Data with ADAMS

What the heck is ADAMS?





# What is ADAMS?

---

- Java, GPLv3
- Data mining: MOA, WEKA, MEKA, R
- Spreadsheets and databases
- Image and video processing
- Visualizations (plots, GIS)
- Scripting via Jython and Groovy
- ...

# Flow

---

- Operators are called “actors”
- Actors arranged in tree, no connections
- Actor “handlers” nest other actors
  - e.g., sequence of actors
- Control actors control data flow
  - e.g., branch, tee, if-then-else, switch
- Input/output defines
  - standalone , source , transformer , sink 

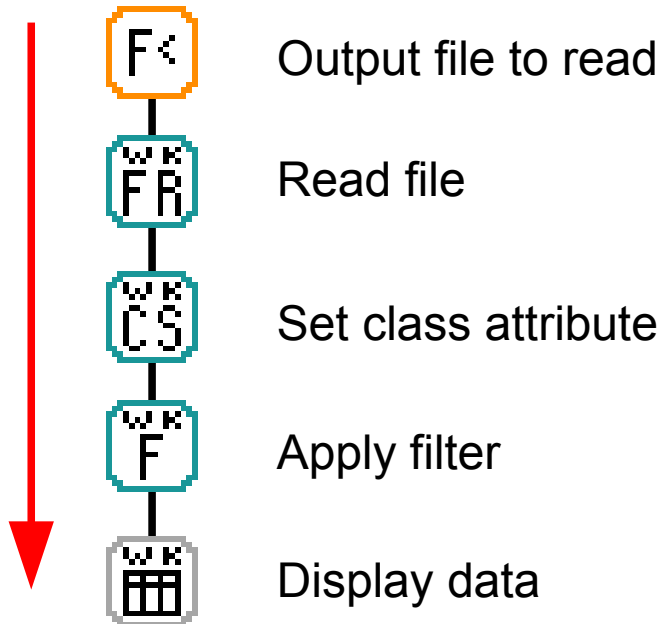
# Flow (2)

---

- Tree only supports 1-to-n connections
- Simulating n-to-m semantics
  - Containers
  - Variables
  - Internal storage
  - Callable actors

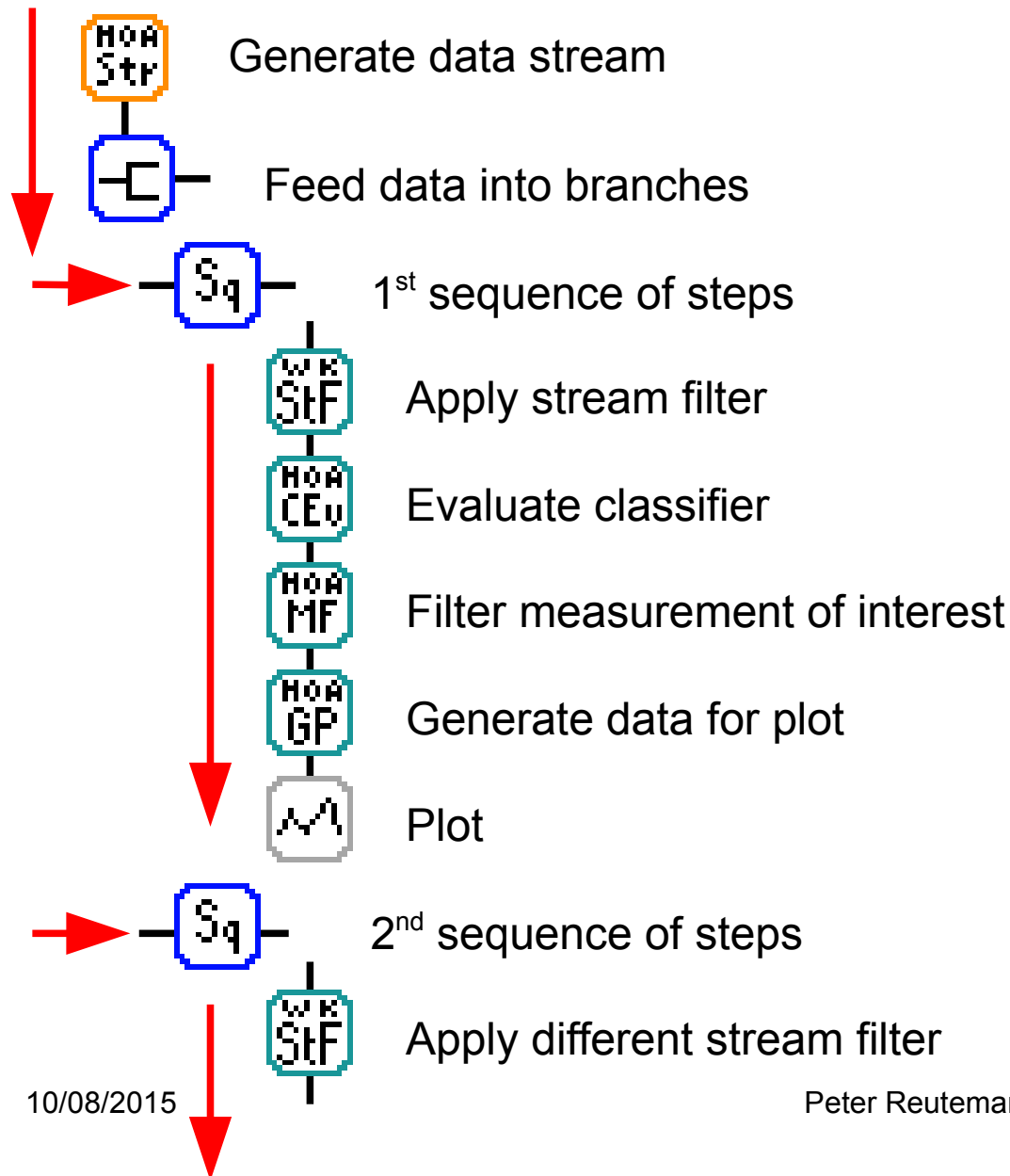
# Examples

**FI** Execute nested actors one after the other



Load dataset,  
apply filter and  
display dataset

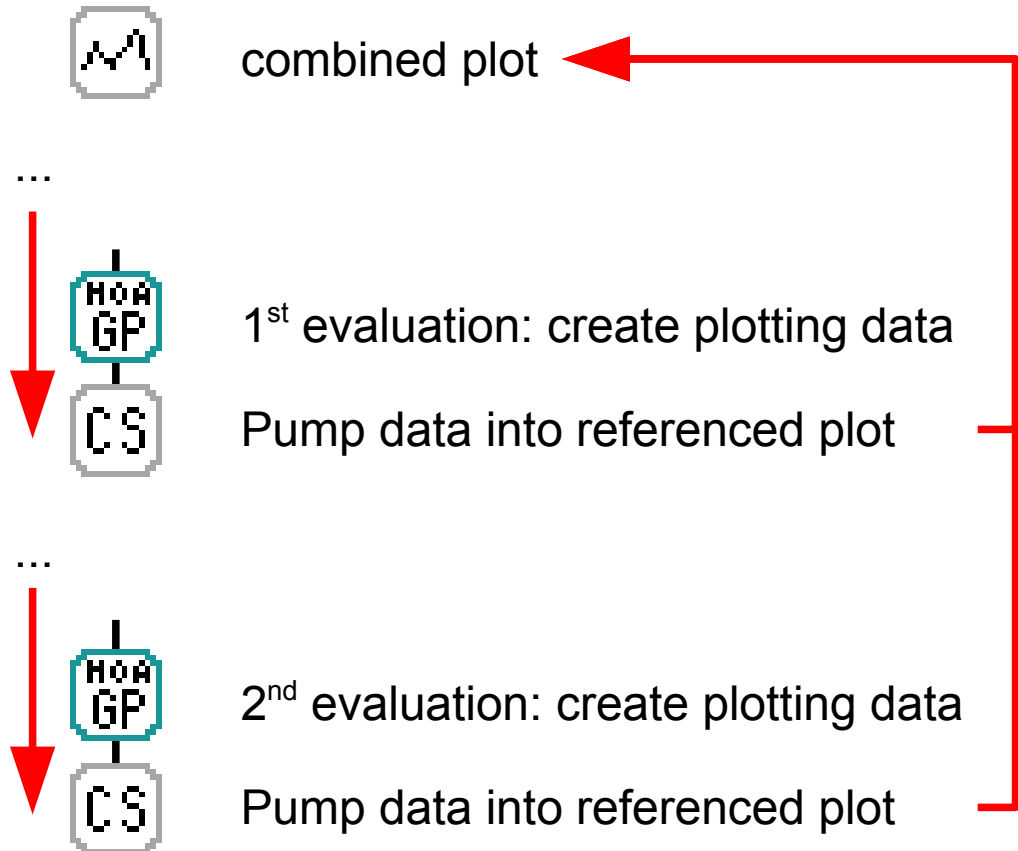
# Examples (2)



Filter data stream in two separate branches with different filters, evaluate classifier and plot metric

# Examples (3)

**CA** groups actors accessible via their name (“callable actors”)



Generate combined plot of two evaluations by using “callable actors” functionality

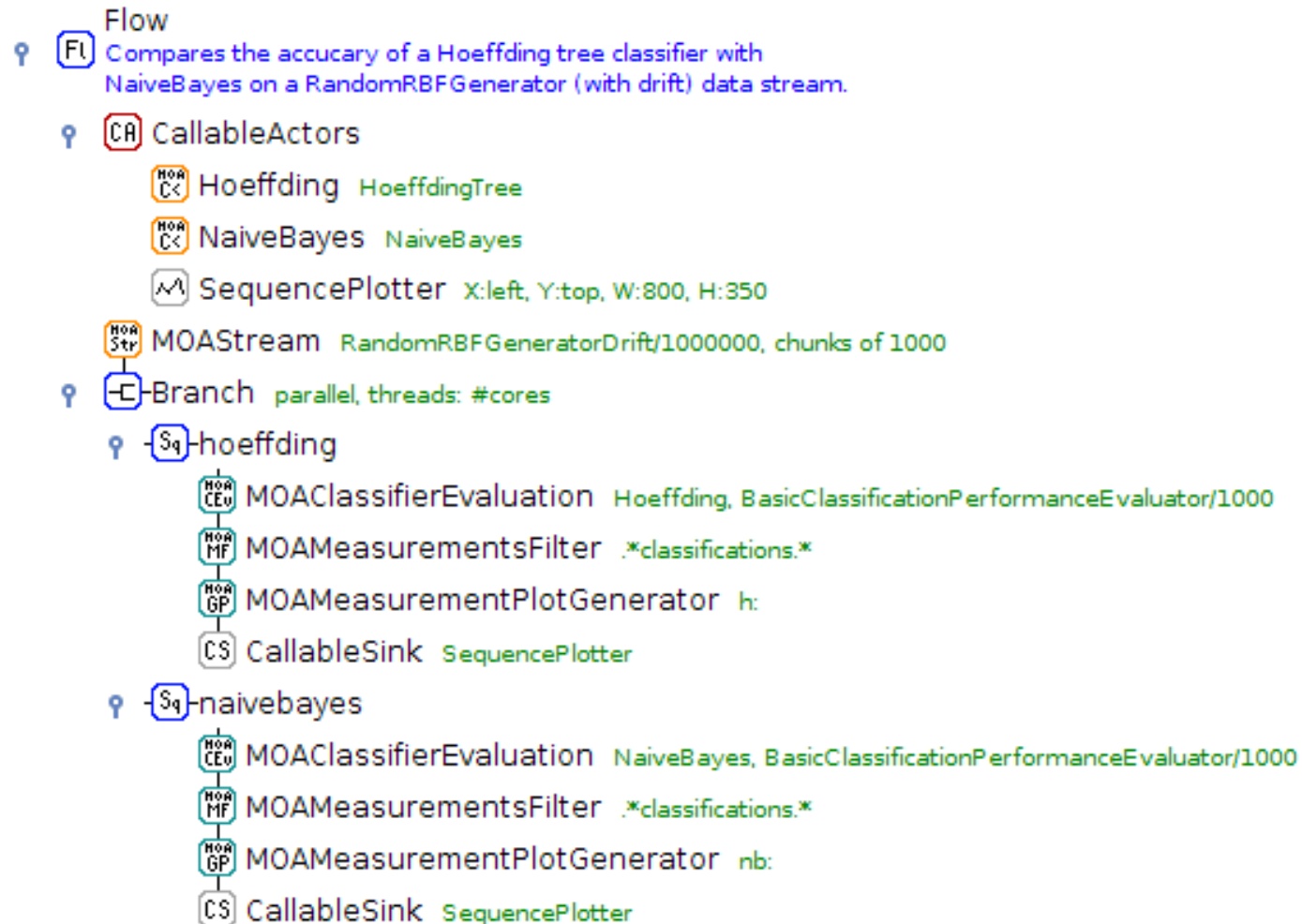
# Research (demos)

---

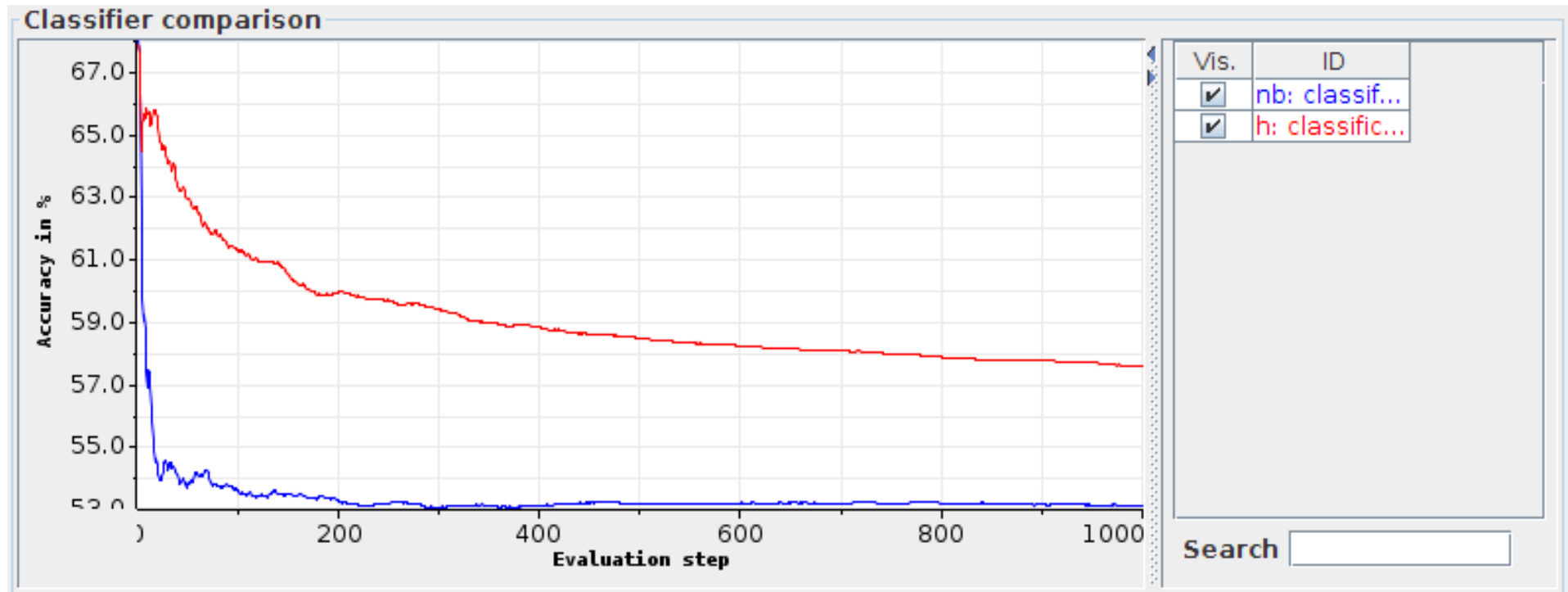
- Compare two MOA classifiers (drift)
- Compare MOA classifier on different streams
- MOA cluster visualization
- Track mouse in video



# MOA - Drift

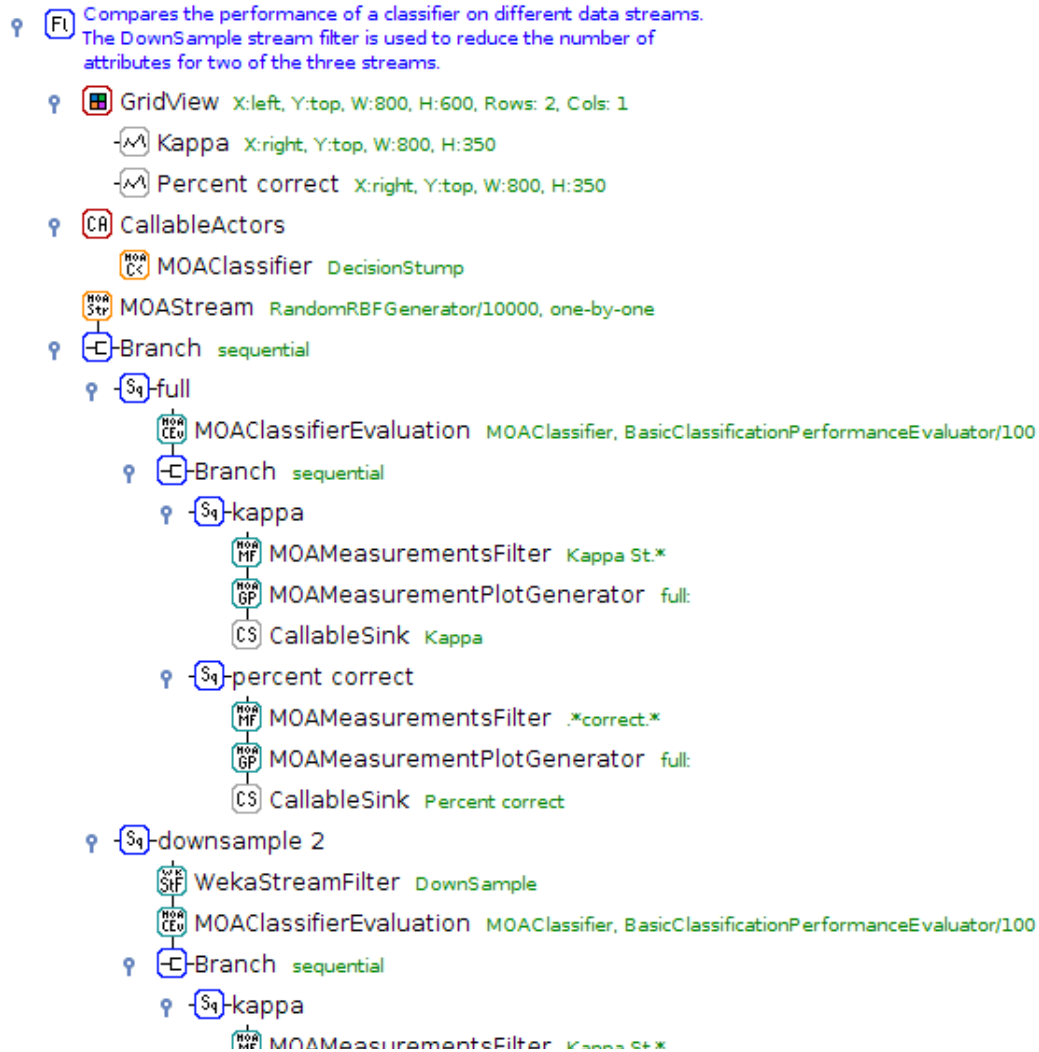


# MOA - Drift

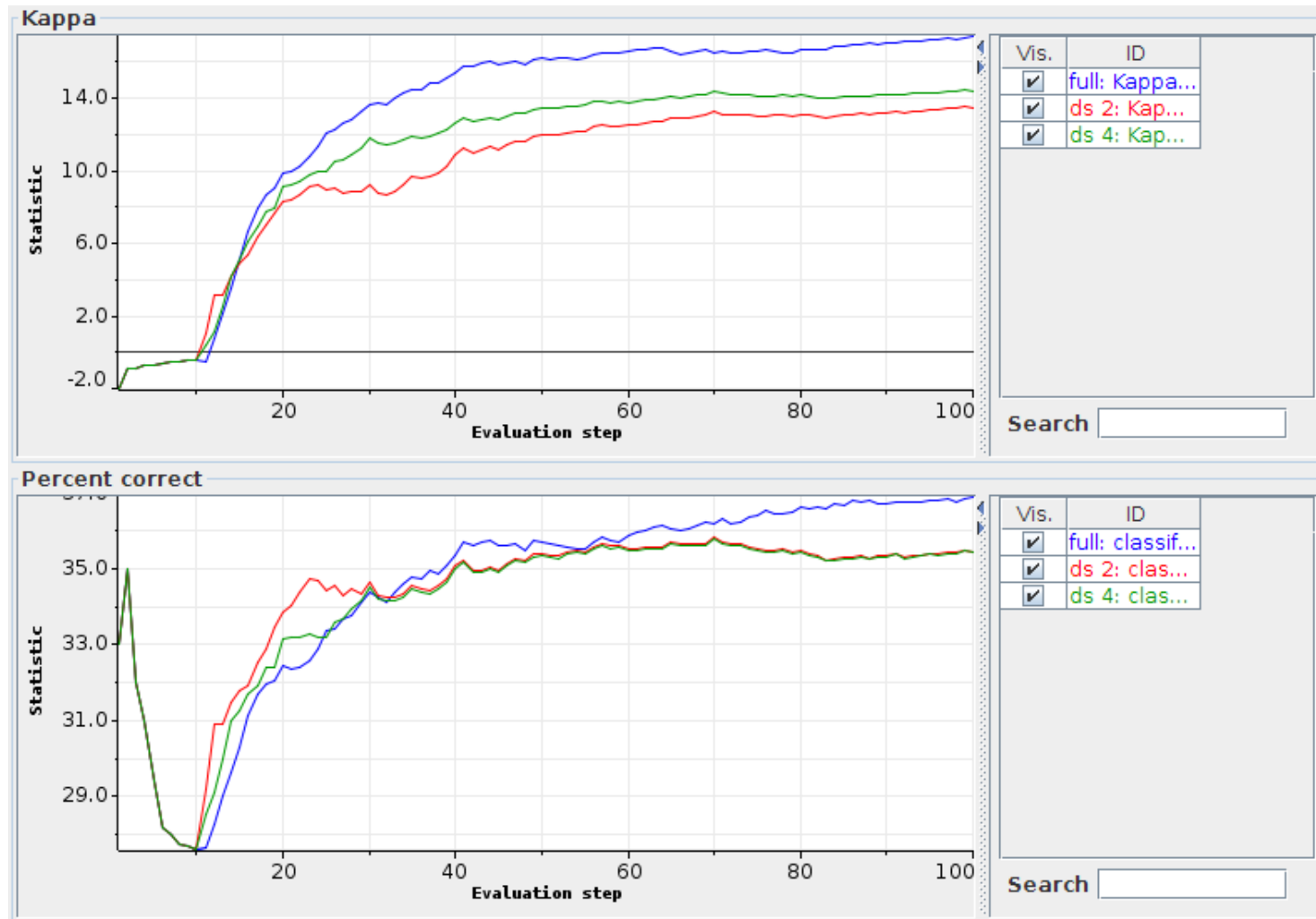


# MOA - different streams

Flow

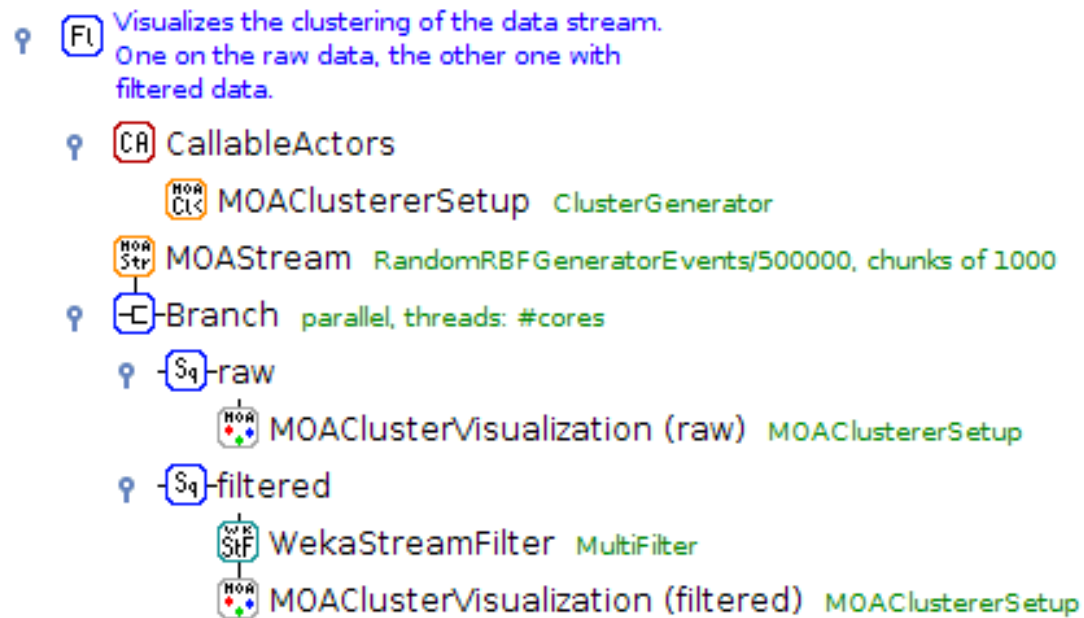


# MOA - different streams

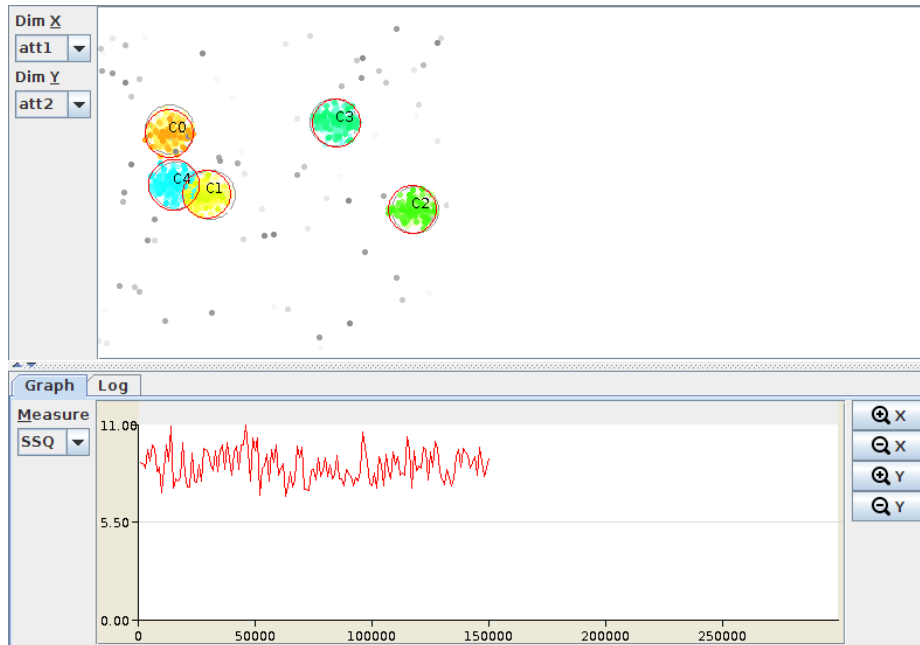


# MOA - Cluster visualization

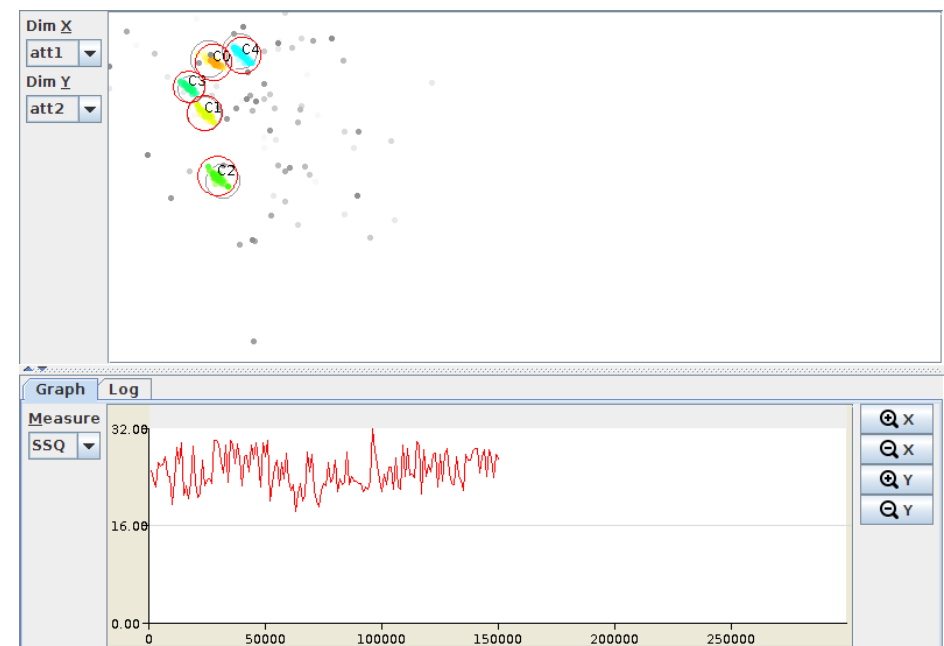
Flow



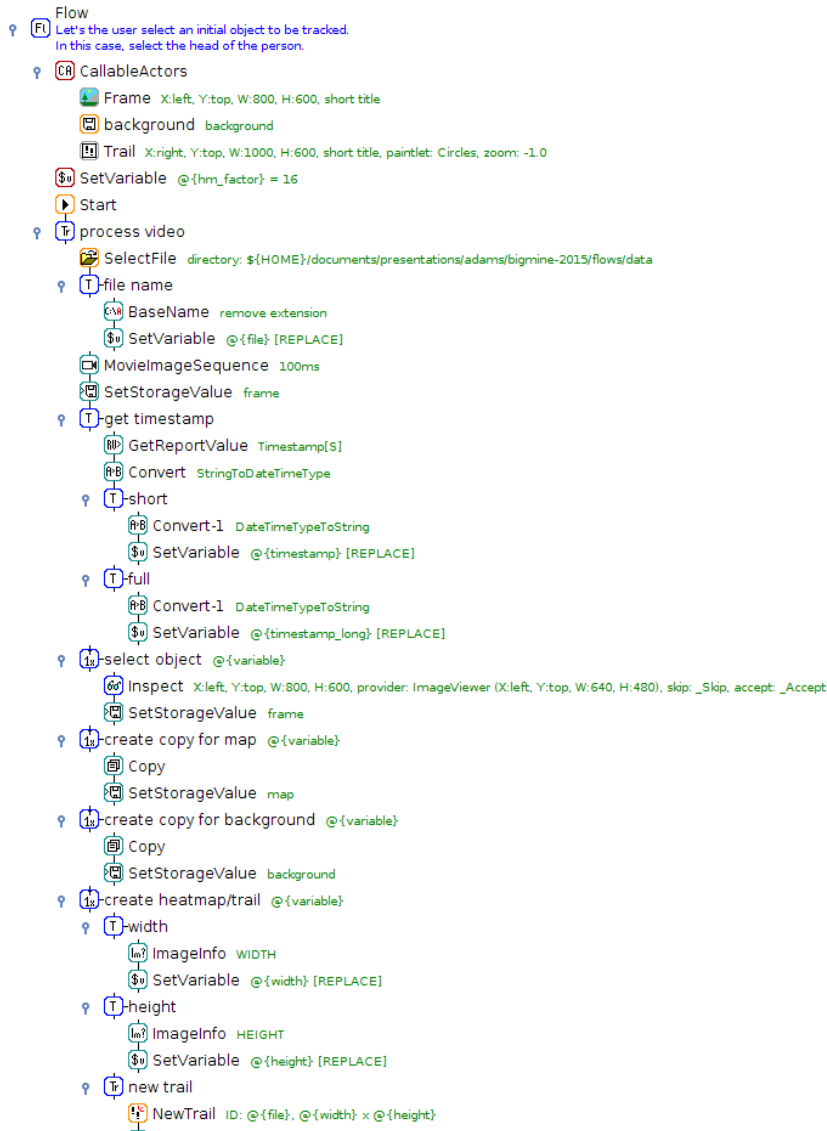
# MOA - Cluster visualization



Stream 2




# Track mouse



# Track mouse

View Trail



Name	Type	Value
Trail.Height	N	720.0
Trail.Width	N	1280.0

X: 474 Y: 22 Zoom: 51.7%

Data Log

Search



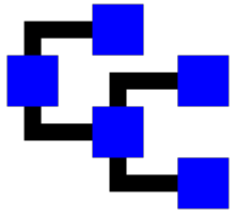
# Industry

---

- BLGG - environmental lab in NL
- Spectral analysis
  - XRF: 10,000, MIR: 2,000, NIR: 1,500
- In operation since 2006
- Predictive modelling: soil, plant (~250 models)
- 1,000 to 3,000 samples per day
- Savings due to less wet chemistry
  - USD 18 million to USD 33 million per year

# Interested?

---



<https://adams.cms.waikato.ac.nz/>

@TheAdamsFlow