

# AMiDST TOOLBOX

Session 2: Introduction to the Amidst Toolbox

**Andrés R. Masegosa**

University of Almeria  
[andres.masegosa@ual.es](mailto:andres.masegosa@ual.es)

# Setting up AMIDST Toolbox

<https://github.com/andresmasegosa/GeiloWinterSchool2018>



## Install AMIDST Toolbox

First, check whether you have installed Java 8:

```
$ java -version
```

If Java 8 is not installed download it from [here](#).

For compiling and running the toolbox you have two options:

1. **IntelliJ IDEA:** You can download it from [here](#).
2. **Maven:** Follow the [official web page](#) for instructions about how to install it.

## Download GeiloWinterSchool2018 code repository

First, download the project code:

```
$ git clone https://github.com/andresmasegosa/GeiloWinterSchool2018.git
```

Enter in the downloaded folder:

```
$ cd GeiloWinterSchool2018/
```

If you have installed maven, you can compile and build the package from command line:

```
$ mvn clean package
```

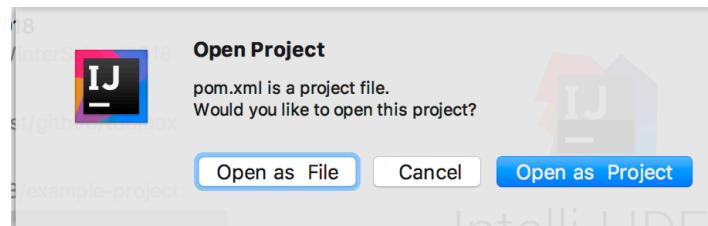
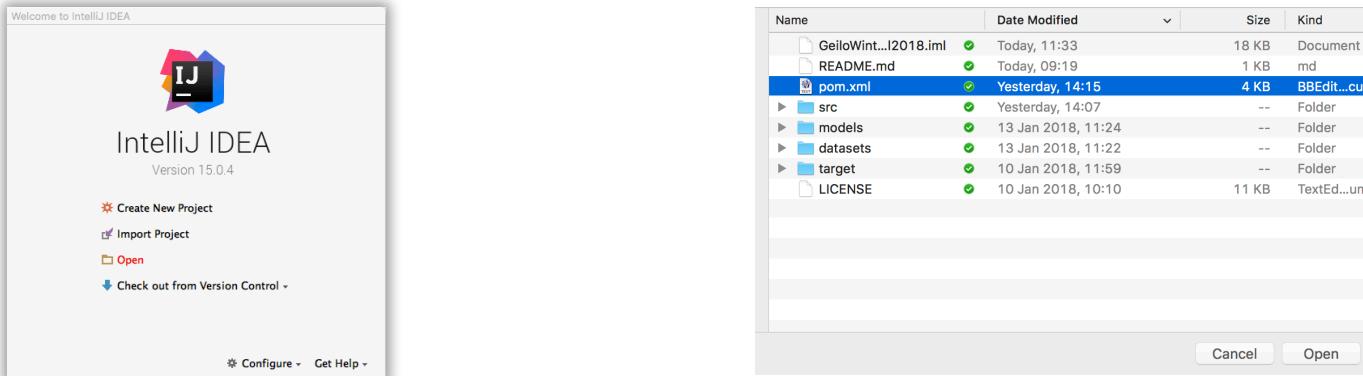
For running any Java file you should type:

```
$ java -cp target/GeiloWinterSchool2018-full.jar winter.Session2.A_GaussianMixture
```



# SETTING UP INTELLIJ IDEA

AMIDST  
TOOLBOX

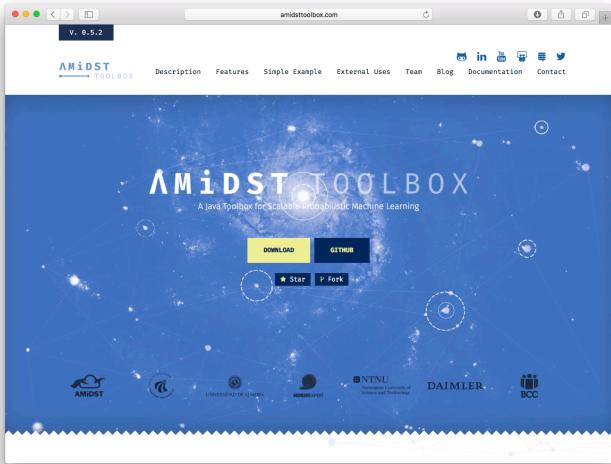


# AMIDST Toolbox

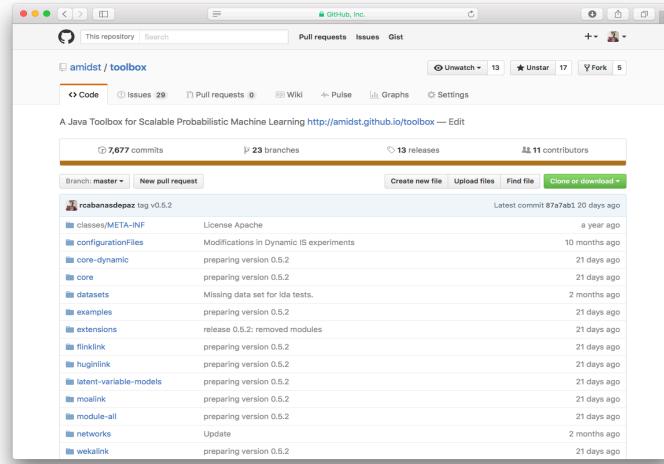


# AMIDST TOOLBOX

AMIDST  
TOOLBOX



[www.amidsttoolbox.com](http://www.amidsttoolbox.com)

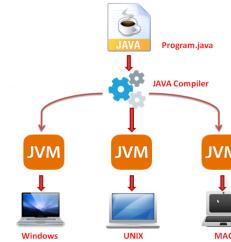


[github.com/amidst/toolbox](https://github.com/amidst/toolbox)



Apache  
License 2.0





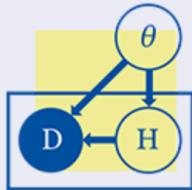
- **Machine Learning on the Java Virtual Machine**
  - Integration with other big data technologies
    - Spark, Flink, Hadoop, Hbase, etc rely on JVM.
  - Many IT systems also rely on the JVM.

## Code-project



## Λ M i D S T TOOLBOX





## Probabilistic Graphical Models

Specify your model using probabilistic graphical models with latent variables and temporal dependencies



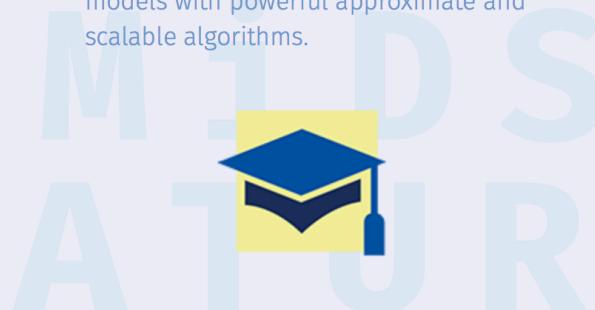
## Large-scale Data

Use your defined models to process massive data sets in a distributed computer cluster using Flink or Spark.

$$P(\theta | \mathbf{D})$$

## Scalable inference

Perform inference on your probabilistic models with powerful approximate and scalable algorithms.



## Extensible

Code your models or algorithms within AMIDST and expand the toolbox functionalities. Flexible toolbox for academics performing their experimentation in machine learning.

01011100

## Data Streams

Update your models when new data is available. This makes our toolbox appropriate for learning from data streams.



## Interoperability

Leverage existing functionalities and algorithms by interfacing to existing software tools such as Hugin, MOA, Weka, R, etc



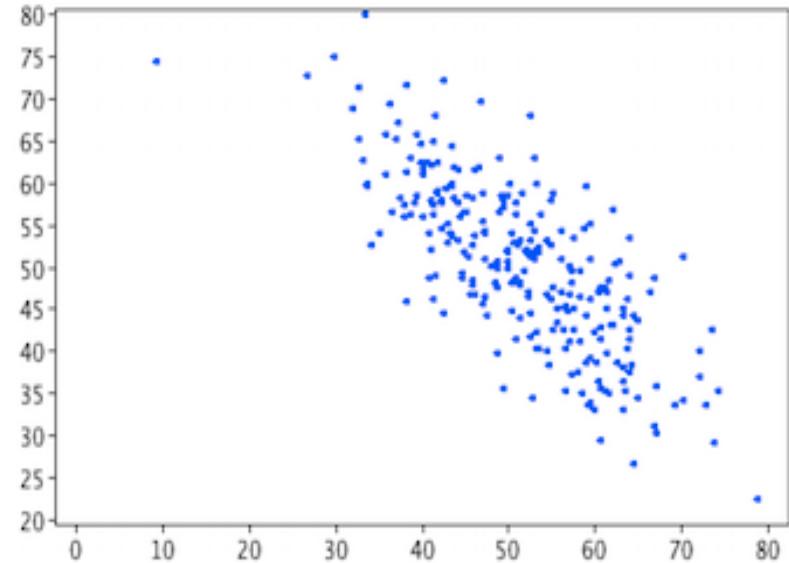
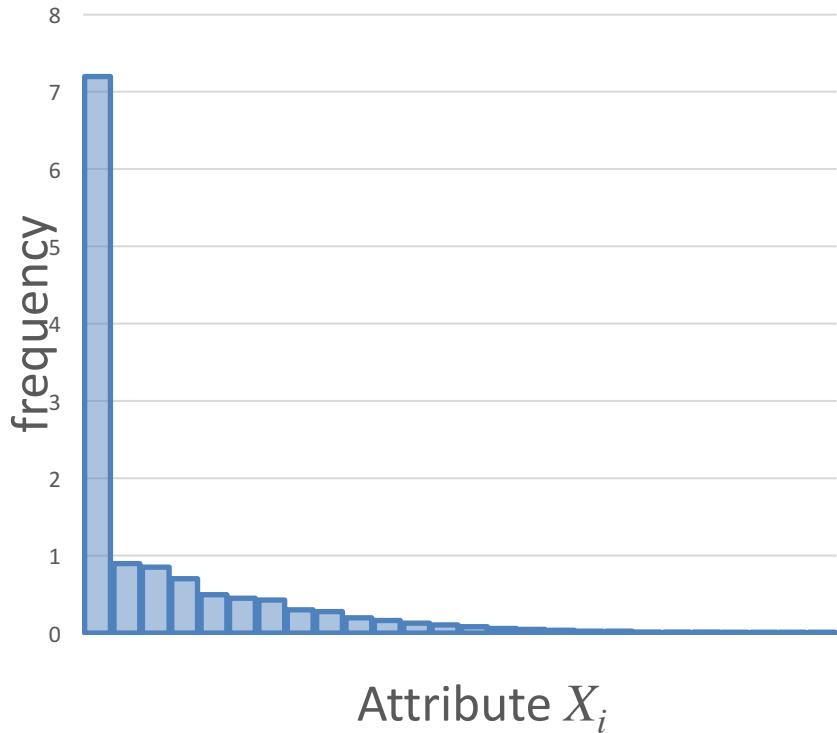
# Use Case I



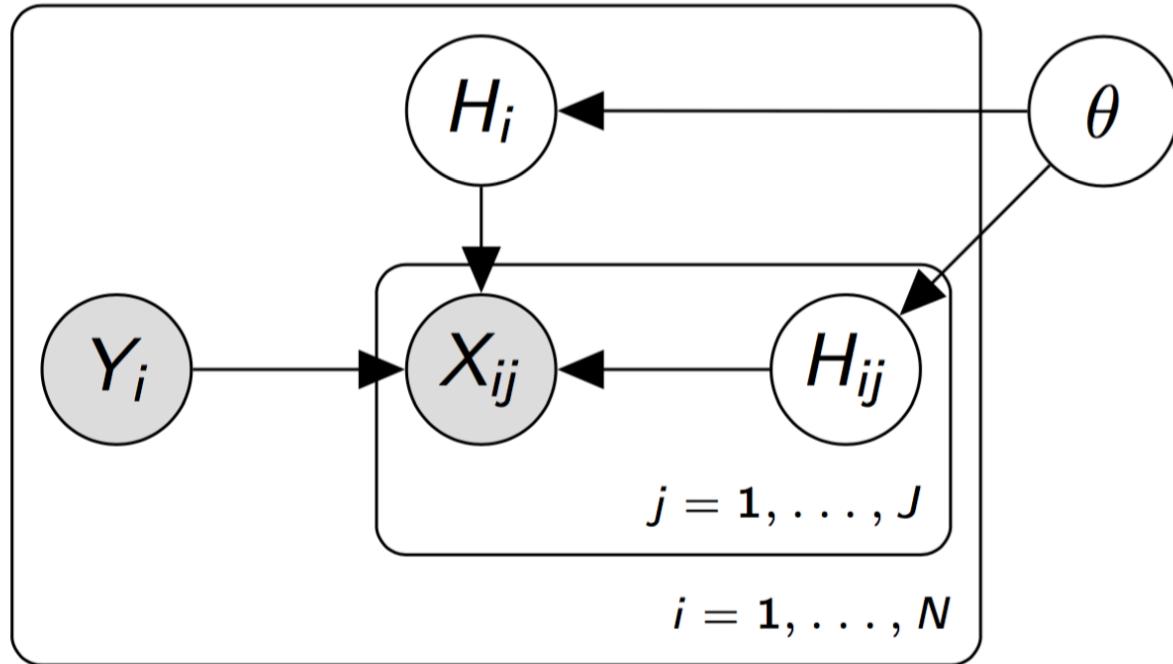


## Predicting Defaulting Clients

Predicts probability a customer will default within 2 years



- Daily data for millions of clients
- Tons of missing data (around 70% of entries).
- Odd distributions.



## Custom Gaussian Mixture Model

$H_{ij}$  defines local mixture

$H_i$  defines a global mixture.



## Predicting Defaulting Clients

- Old BCC's models based on logistic regression got an AUC around 0.8
- AMIDST's models gets an AUC over 0.9
- Model will be in production soon.

.



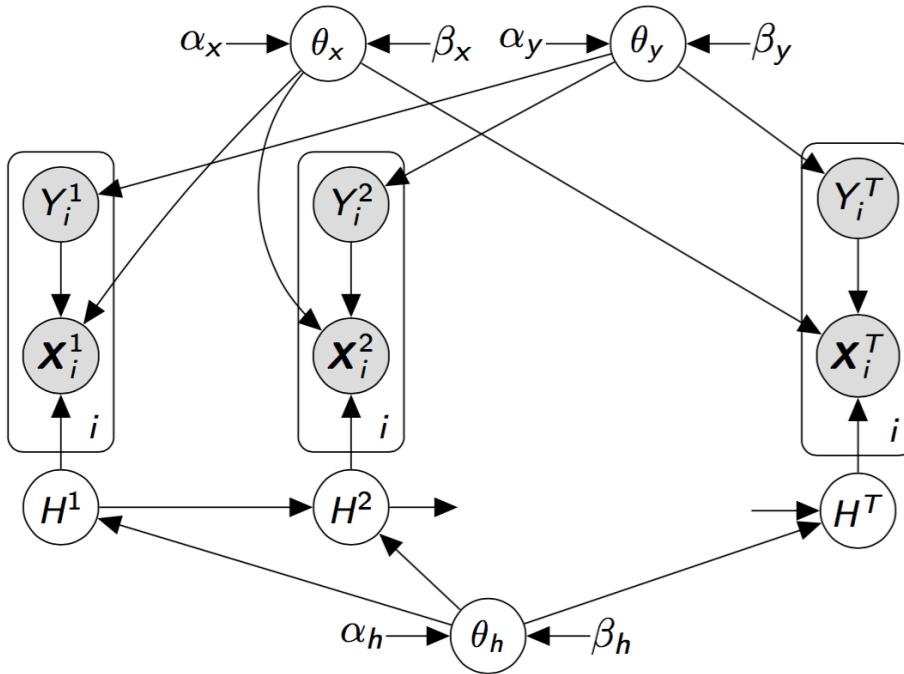
# Use Case II





## Tracking Concept Drift

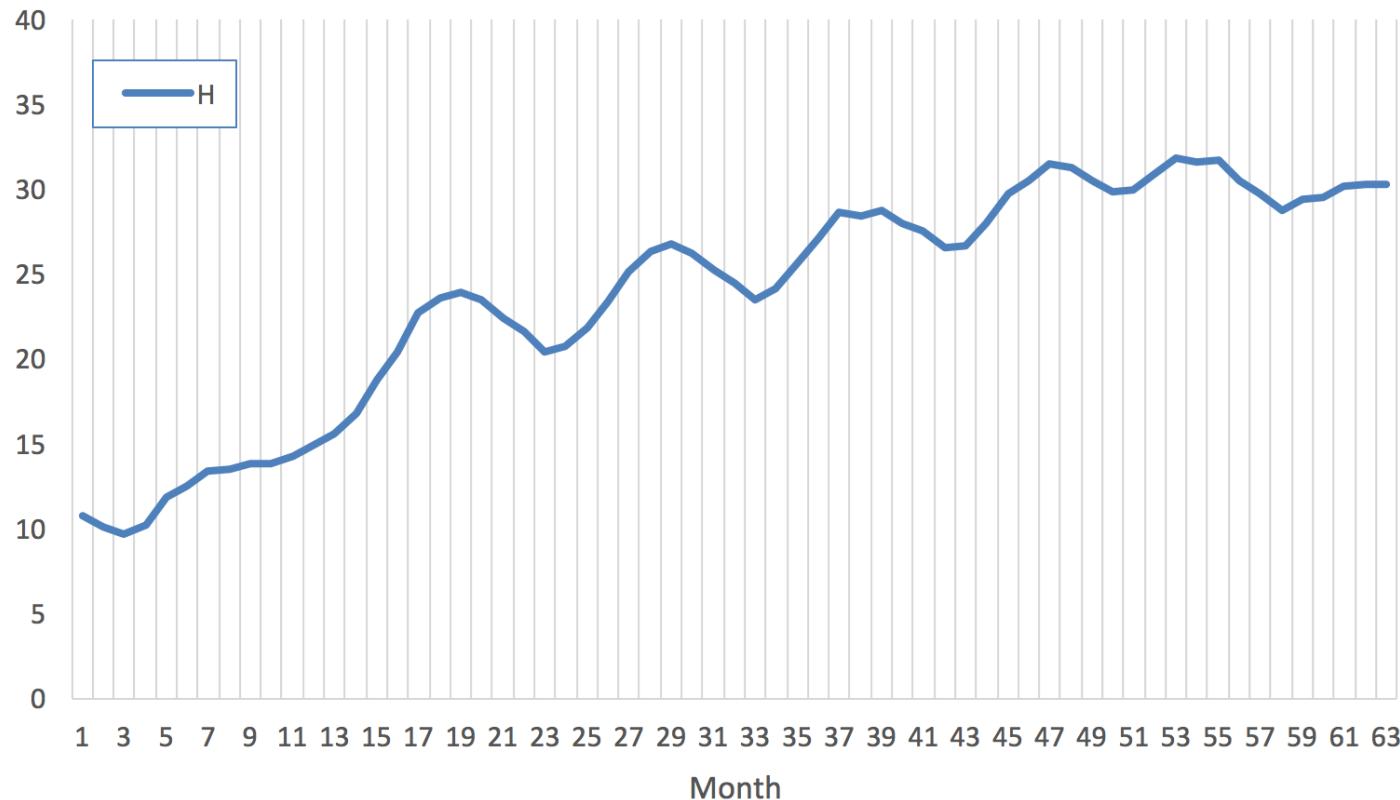
Detects changes in customer profiles during Spanish financial crisis



Latent Variables are used to capture changes in customer profile

# CONCEPT DRIFT DETECTION RESULTS

AMIDST  
TOOLBOX



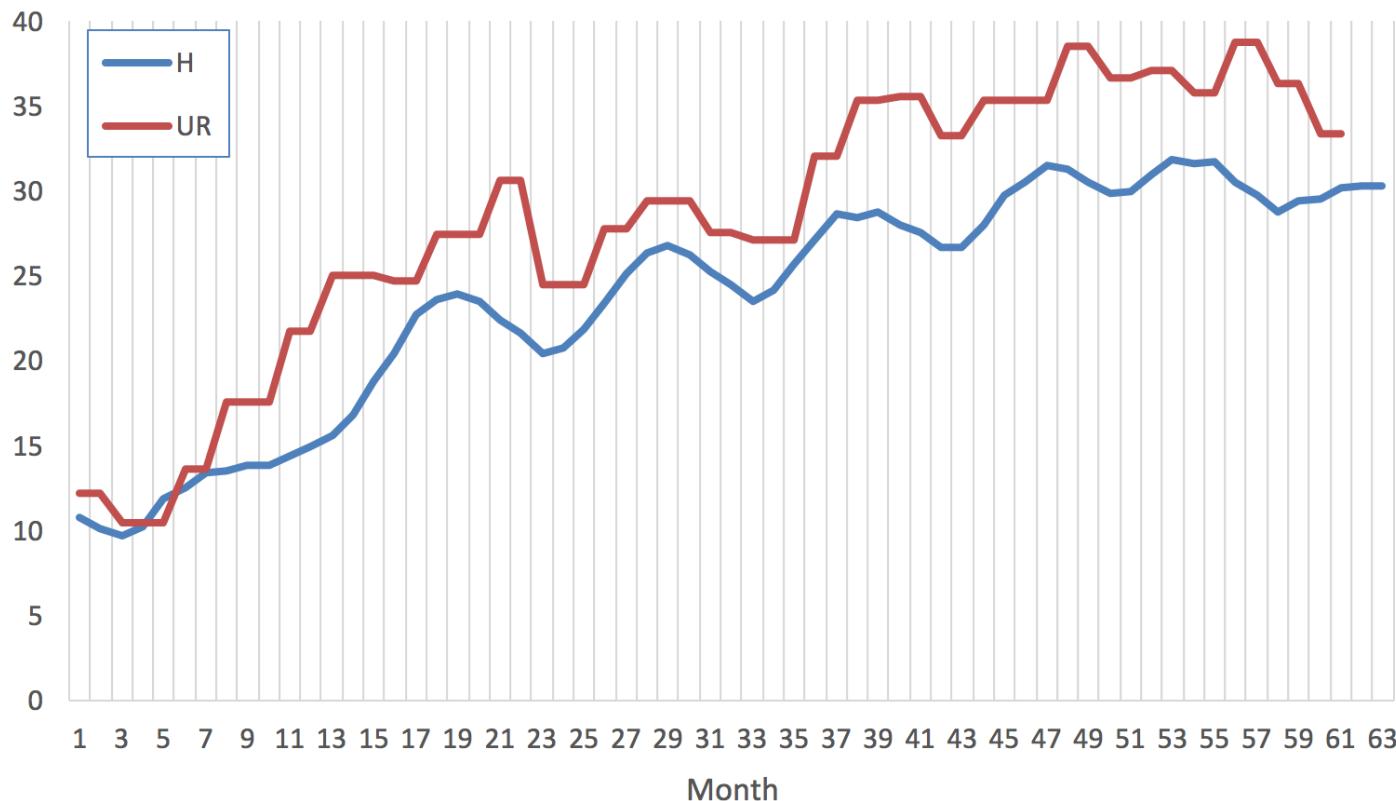
## Latent Variable Captures Concept Drift

Drift Pattern: Seasonal + Global trend



# CONCEPT DRIFT DETECTION RESULTS

AMIDST  
TOOLBOX



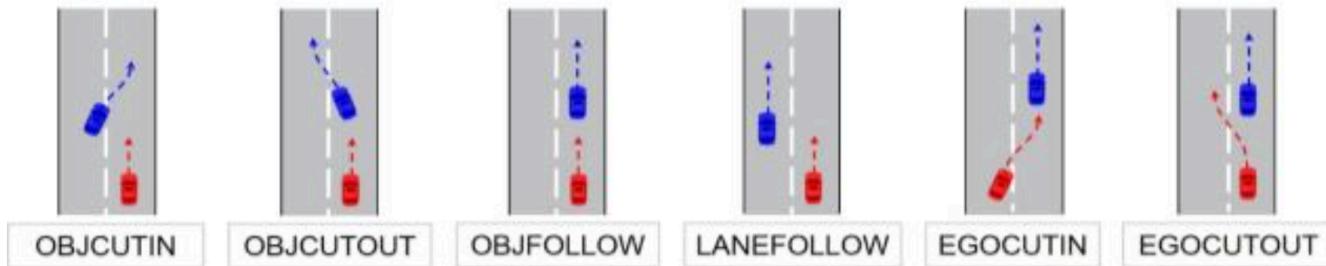
**Unemployment Rate main driver of Concept Drift**

Latent Variable correlates with unemployment rate ( $\rho = 0.961$ )



# Use Case III





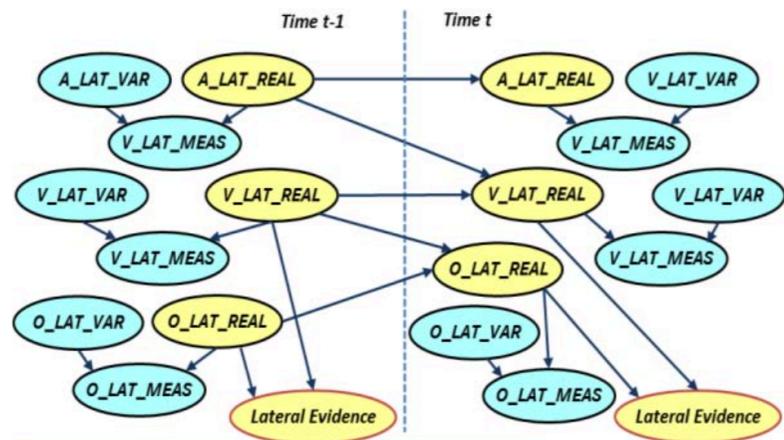
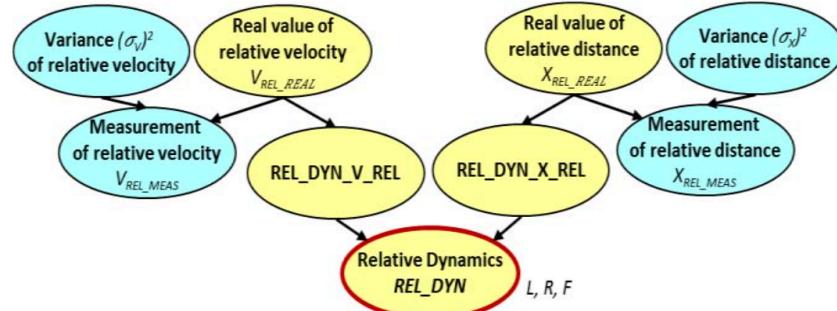
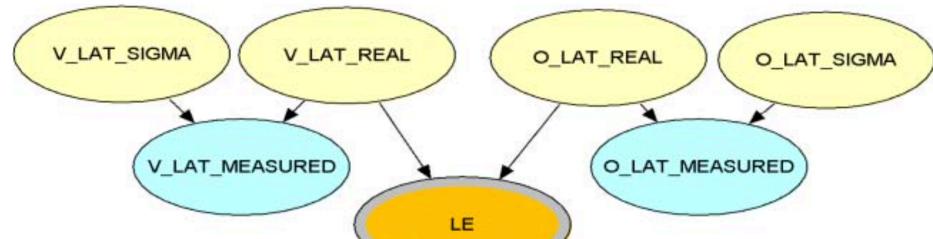
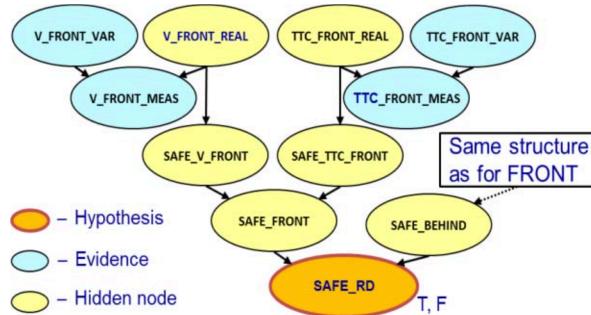
Weidl, Galia, et al. "Early Recognition of Maneuvers in Highway Traffic." *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*. Springer International Publishing, 2015.

## Maneuver Recognition

Early detection of traffic maneuvers changes for intelligent cruise control (and autonomous driving).

# PROBABILISTIC MODEL

AMIDST  
TOOLBOX



Weidl, Galia, et al. "Early Recognition of Maneuvers in Highway Traffic." *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*. Springer International Publishing, 2015.



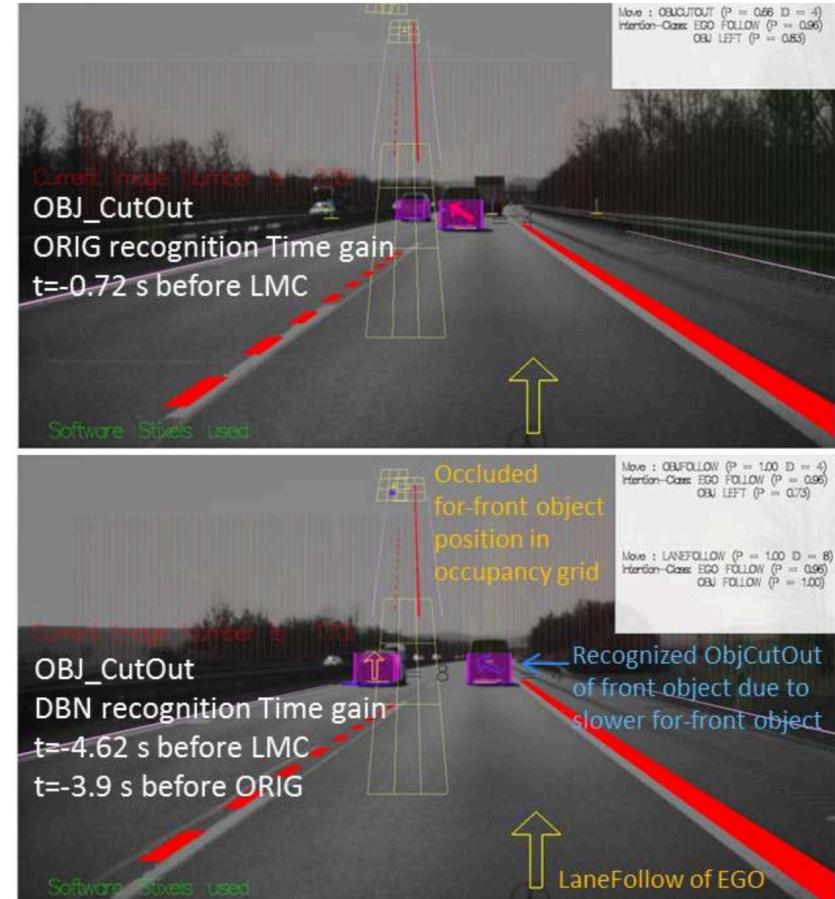
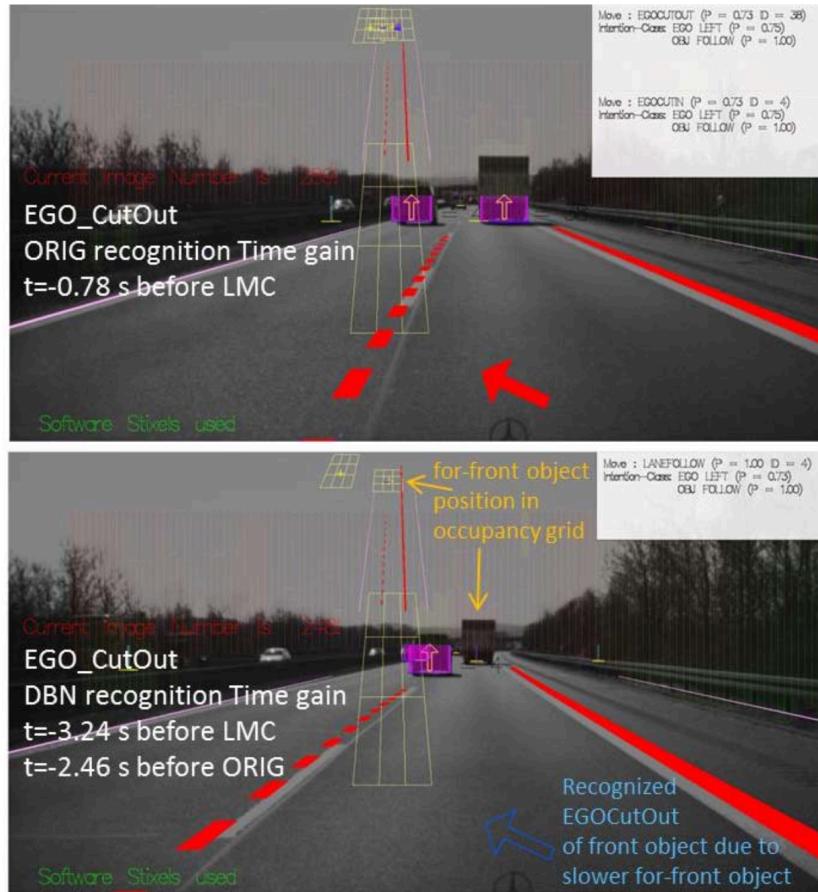
HUGINEXPERT

The leading decision support tool



# PROTOTYPE

AMIDST  
TOOLBOX



Weidl, Galia, et al. "Early Recognition of Maneuvers in Highway Traffic." *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*. Springer International Publishing, 2015.



# Thanks for your attention

www

[www.amidsttoolbox.com](http://www.amidsttoolbox.com)

@

[contact@amidsttoolbox.com](mailto:contact@amidsttoolbox.com)



[@AmidstToolbox](https://twitter.com/AmidstToolbox)

AMIDST  
→ TOOLBOX