

# Hackathon Citizens of Wallonia

## Présentation de TSimulus

Un générateur de séries chronologiques opensource

*10 mars 2018*

*Alexandre Nuttinck*

*alexandre.nuttinck@cetic.be*

# TSimulus, c'est qui?



Mathieu Goeminne

Créateur de TSimulus

Chercheur Expert Big Data

Département SST@CETIC

[mathieu.goeminne@cetic.be](mailto:mathieu.goeminne@cetic.be)

<https://www.linkedin.com/in/mathieu-goeminne-724ab99/>



Alexandre Nuttinck

Contributeur de TSimulus

Chercheur Cloud Computing

Département SST@CETIC

[alexandre.nuttinck@cetic.be](mailto:alexandre.nuttinck@cetic.be)

<https://www.linkedin.com/in/alexandre-nuttinck/>

# Use Case à l'origine de TSimulus

- EAM-SDI: projet de recherche avec Guardis
    - Domaine spécifique de la gestion automatique d'infrastructures.
    - Récolter les données cpus, ventilateurs, ...
    - Faire du monitoring, réactions face aux événements.
    - Problème: absence de données réelles et en temps réelles.
- solution permettant la simulation d'un grand nombre de capteurs remontant de l'information auprès d'un système de collecte et d'analyse

# TSimulus, c'est quoi?

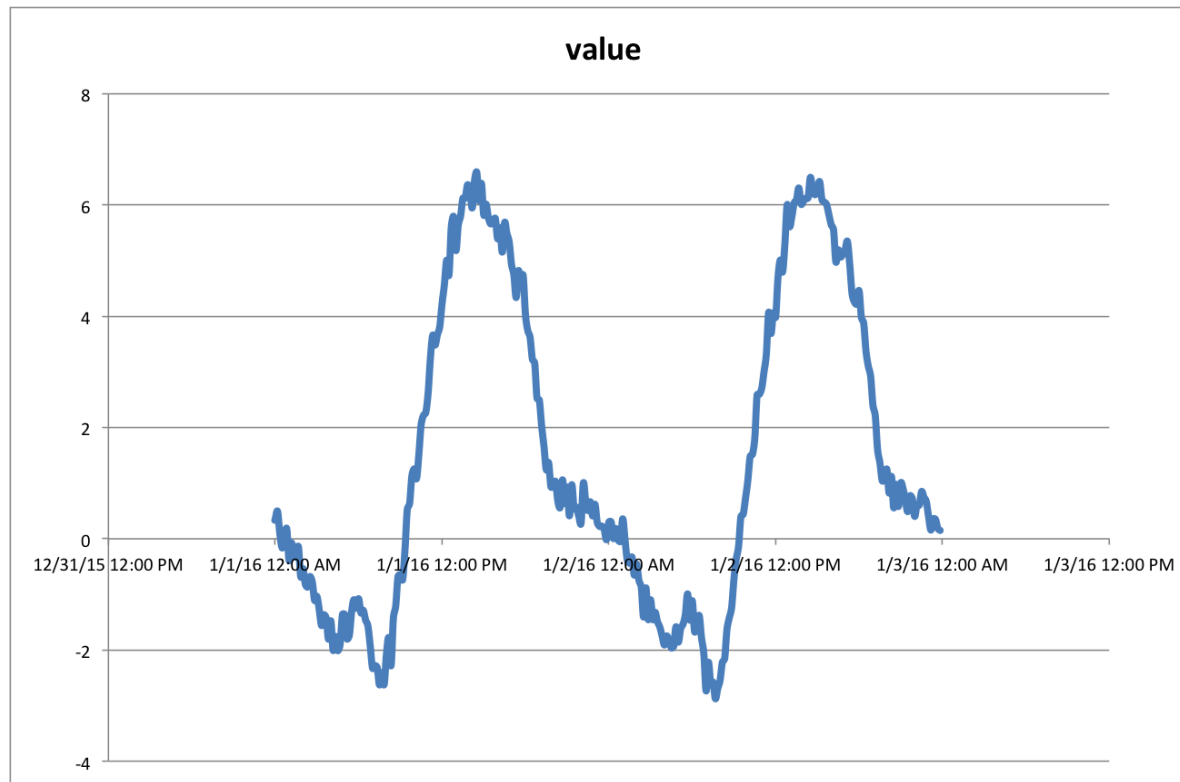
- Un générateur de séries chronologiques aléatoires, mais réalistes.
- Série chronologique?
  - séquence ordonnée de points dans le temps, chacun d'entre eux étant associé à au plus une valeur.
  - timestamp → valeur (typiquement, réel)
  - utilisée dans une grande variété de domaines.
- Aléatoires, mais réalistes?
  - bruit convaincant
  - modèles, patterns

# Fonctionnement

- En entrée: Un DSL pour décrire « l'allure » des séries
  - Cycles (quotidiens, hebdomadaires)
  - Bruit (blanc, ARMA)
  - Combinaisons
  - Conditions

```
{
  "generators": [
    {
      "name": "monthly-basis",
      "type": "monthly",
      "points": {"january": 3.3, "february": 3.7, "march": 6.8, "april": 9.8, "may": 13.6, "june": 18.4, "july": 18.4, "august": 18, "september": 14.9, "october": 11.1, "november": 6.8, "december": 3.3}
    }
  ],
  "exported": [
    {"name": "temperature", "generator": "monthly-basis", "frequency": 3600000}
  ],
  "from": "2016-01-01 00:00:00.000",
  "to": "2017-12-31 23:59:59.999"
}
```

- En sortie: triplets (série, timestamp, valeur)



- La documentation
  - Maintenu sur readthedocs (<http://tsimulus.readthedocs.io>)
  - Introduction, générateurs, exemples, ...

- La bibliothèque
  - Pur Scala (jar téléchargeable depuis Maven Central)
  - Développement sur Github (<https://github.com/cetic/TSimulus>)



- Le CLI
  - Prend en entrée un fichier de configuration et produit du CSV sur la sortie standard.
  - <https://github.com/cetic/tsimulus-cli>

- Le micro-service
  - API akka, envoi d'un fichier de configuration en POST
  - Possible de générer une valeur sur POST ou un ensemble de valeurs
  - CSV en réponse
  - Image Docker disponible: <https://hub.docker.com/r/ceticasbl/tsimulus-ms/>
  - <https://github.com/cetic/tsimulus-ms>

- Souvent, manque de données
  - Simuler la température (Use case présenté dans la démo d'aujourd'hui)
  - Simuler le trafic sur les routes
  - Simuler le nombre de places de parking
  - Simuler le nombre de piétons en circulation dans le centre
  - Simuler les passages des transports en commun
  - ...

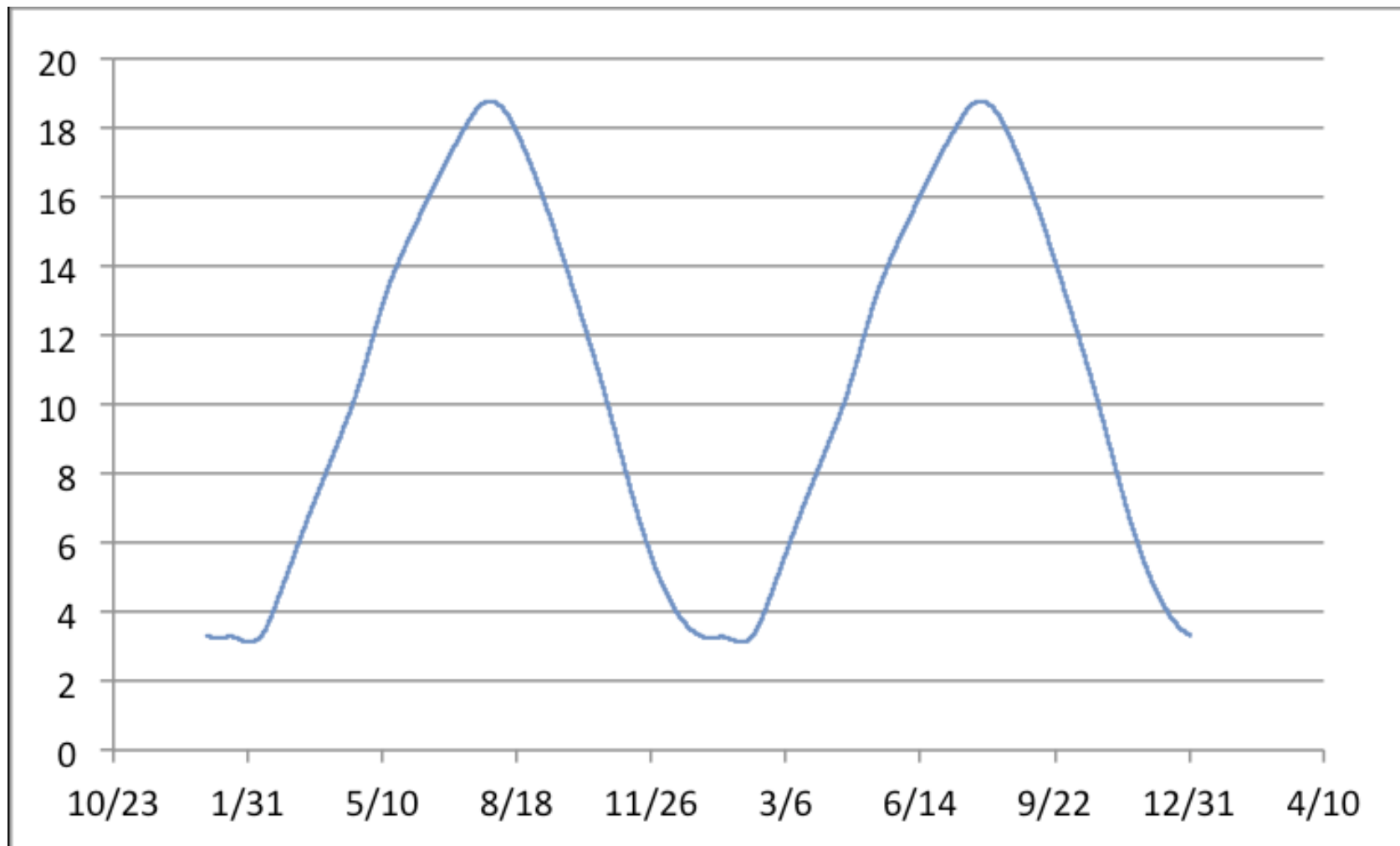
# Démo - Une première génération

- Pour votre première génération de séries chronologiques, nous allons nous baser sur la température extérieure d'Uccle.
- TSimulus Microservice, disponible ici: <http://tsimulus.openshift.cetic.be>
- Fichier de configuration: [https://raw.githubusercontent.com/cetic/tsimulus/master/examples/get\\_started\\_1.json](https://raw.githubusercontent.com/cetic/tsimulus/master/examples/get_started_1.json)
- Faire une requête **POST** sur <http://tsimulus.openshift.cetic.be/generator>

Month	Temperature
January	3.3
February	3.7
March	6.8
April	9.8
May	13.6
June	16.2
July	18.4
Augustus	18.0
September	14.9
October	11.1
November	6.8
December	3.9

Source: [Royal Meteorology Institute of Belgium](#)

# Démo - Une première génération



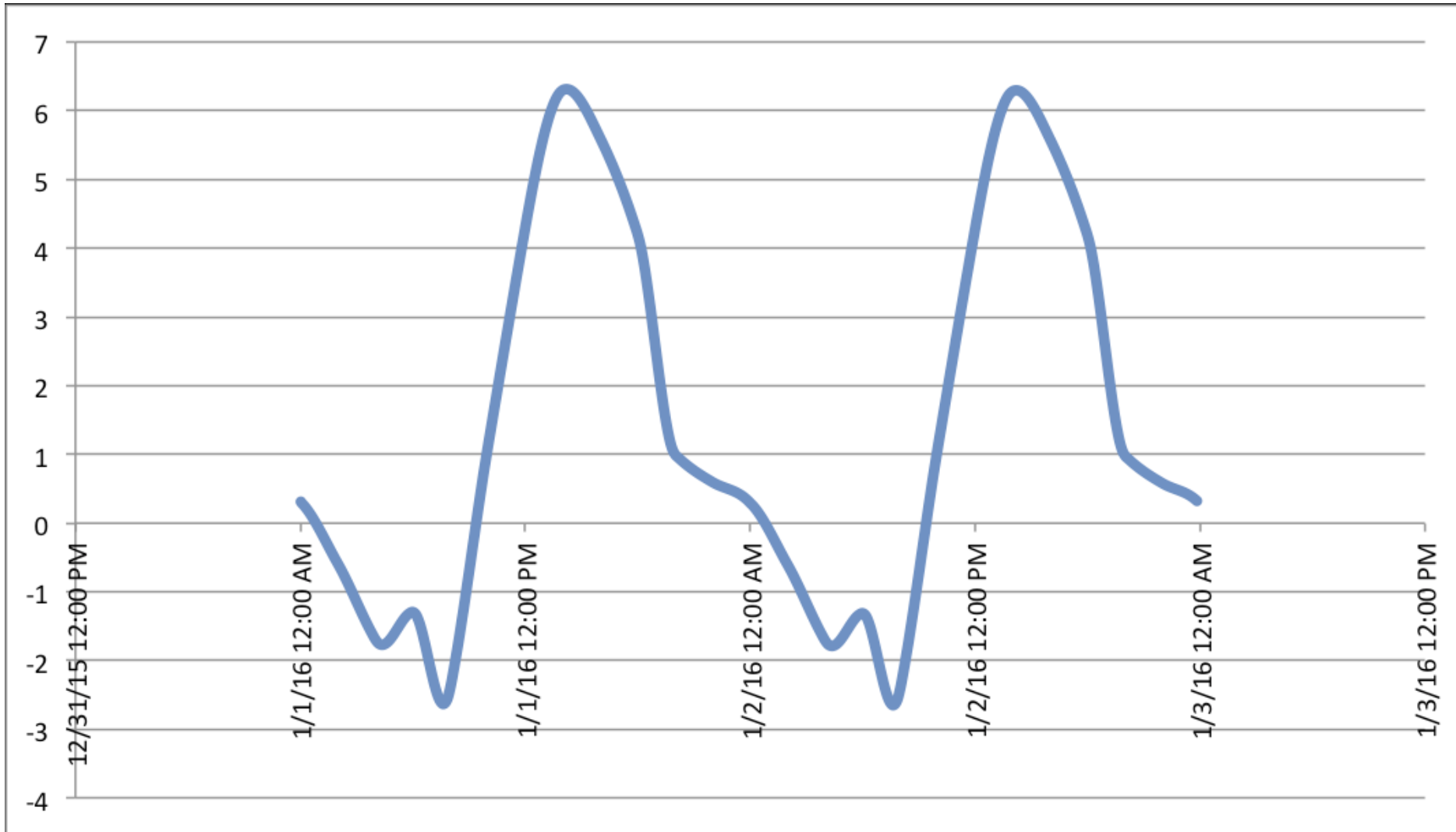
# Démo - Vers un modèle plus réaliste

- Ok ça fonctionne!
- La température varie de façon constante, et vers le milieu du mois, on obtient bien les valeurs que l'on a spécifiées.
- Mais, les valeurs sont bien trop régulières. Ce n'est pas réaliste.
- Les températures sont variables en fonction de l'heure de la journée.

[https://raw.githubusercontent.com/cetic/tstimulus/master/examples/get\\_started\\_2.json](https://raw.githubusercontent.com/cetic/tstimulus/master/examples/get_started_2.json)

Hour	Relative Temperature
00:00	-3.0
02:00	-3.9
04:00	-5.0
06:00	-4.6
08:00	-5.7
10:00	-2.2
12:00	+1.0
14:00	+3.0
16:00	+2.3
18:00	+0.9
20:00	-2.3
22:00	-2.7

# Démo - Vers un modèle plus réaliste

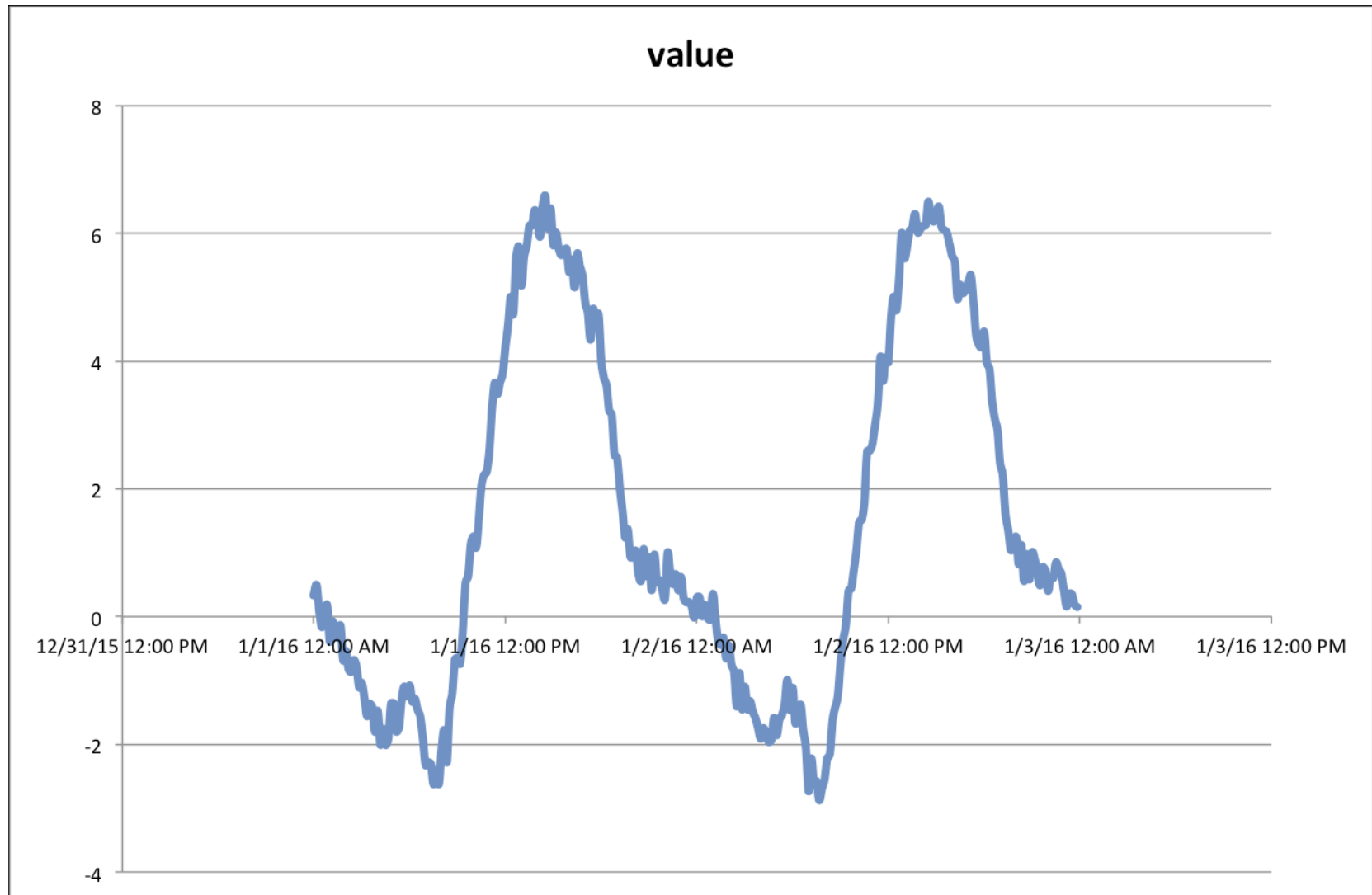


# Démo - Vers un modèle encore plus réaliste

- Pas encore totalement réaliste!
- La temperature varie au cours des années du aux modifications complexes des conditions atmosphériques.
- Pour ce faire, ajoutons un peu de bruit à notre générateur.
- [https://raw.githubusercontent.com/cetic/tsimulus/master/examples/get\\_started\\_3.json](https://raw.githubusercontent.com/cetic/tsimulus/master/examples/get_started_3.json)



# Démo - Vers un modèle encore plus réaliste



Merci!

## Questions?

- Le code source: <https://github.com/cetic/TSimulus>
- Le documentation: <https://github.com/cetic/TSimulus>
- Email: [alexandre.nuttinck@cetic.be](mailto:alexandre.nuttinck@cetic.be)