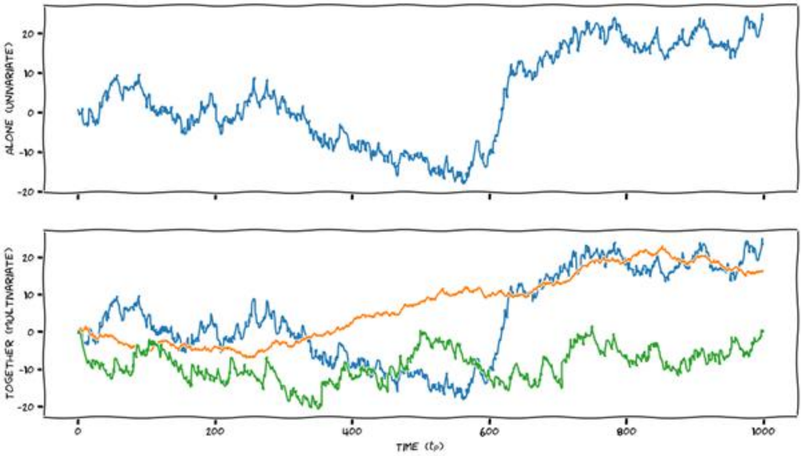
Generating Time Series with Conditional GANs

Time Series. Introduction

- Definition: A sequence of data points indexed in time order.

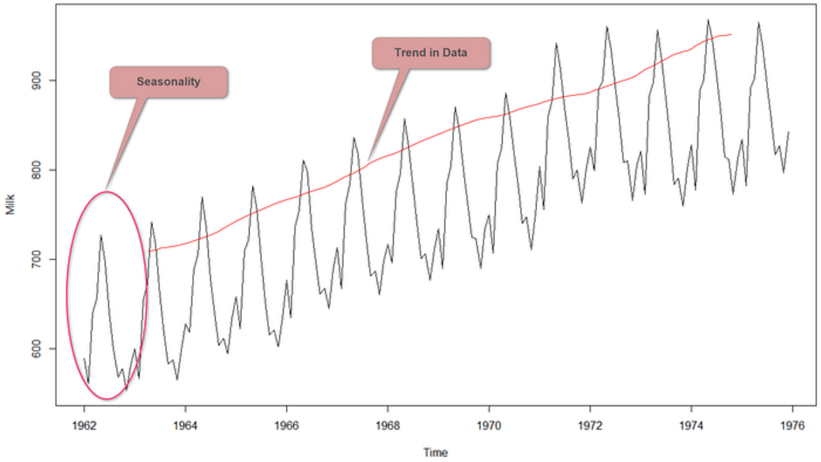
- Characteristics:
  - Observations are dependent on time.
  - Data can be univariate or multivariate.



- Examples:
  - Stock prices
  - Weather data
  - Sensor readings

Time Series. Introduction

- Components of Time Series
  - Trend: Long-term direction (e.g., increasing sales over years).
  - Seasonality: Regular, repeating patterns (e.g., daily, monthly).
  - Cyclicity: Non-fixed, repeating patterns (e.g., business cycles).
  - Noise: Random variations not explained by other components.



Time Series. Introduction

- Applications of Time Series
  - Finance: Predicting stock prices, interest rates.
  - Healthcare: Monitoring patient vitals over time.
  - Marketing: Forecasting sales or customer behavior.
  - Engineering: Predictive maintenance using sensor data.

Time Series. Challenges

- Handling missing data.
- Dealing with seasonality and trends.
- Non-stationarity and noise.
- Scalability for large datasets.

Time Series with GANs

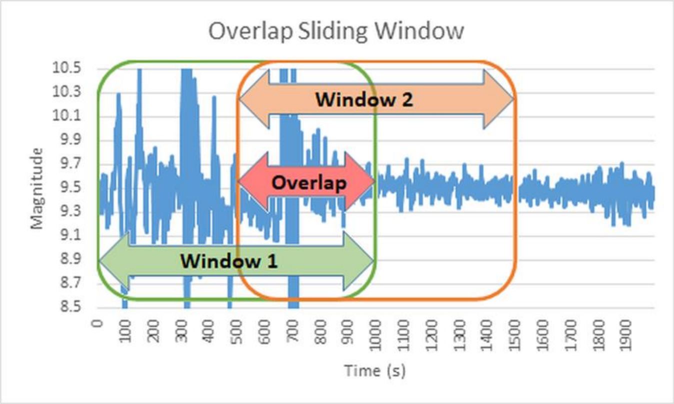
A time series can be represented as collection of vectors that are defined according to a time window size and a selection procedure

For example:

- Time window size 8
- Overlapping sliding window procedure

5.0	6.0	5.0	5.0	5.0	11.0	13.0	7.0

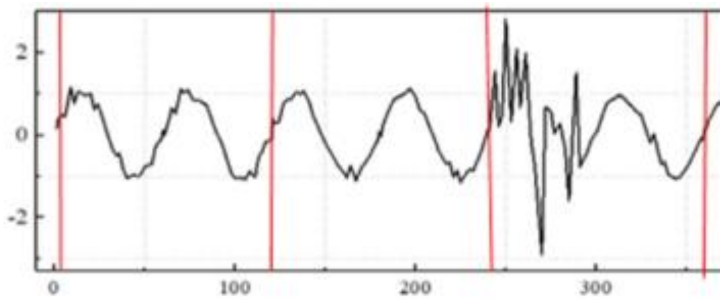
6.0	5.0	5.0	5.0	11.0	13.0	7.0	6.0



date	measure	ID_estacion	geometry	poll
2019-01-02 08:00:00	5.0	Colon	POINT (570970 6149046)	NO2
2019-01-02 09:00:00	6.0	Colon	POINT (570970 6149046)	NO2
2019-01-02 10:00:00	5.0	Colon	POINT (570970 6149046)	NO2
2019-01-02 13:00:00	5.0	Colon	POINT (570970 6149046)	NO2
2019-01-02 14:00:00	5.0	Colon	POINT (570970 6149046)	NO2
2019-01-02 15:00:00	11.0	Colon	POINT (570970 6149046)	NO2
2019-01-02 16:00:00	13.0	Colon	POINT (570970 6149046)	NO2
2019-01-02 17:00:00	7.0	Colon	POINT (570970 6149046)	NO2
2019-01-02 18:00:00	6.0	Colon	POINT (570970 6149046)	NO2
2019-01-02 19:00:00	6.0	Colon	POINT (570970 6149046)	NO2
2019-01-02 23:00:00	5.0	Colon	POINT (570970 6149046)	NO2
2019-01-03 20:00:00	18.0	Colon	POINT (570970 6149046)	NO2
2019-01-03 21:00:00	7.0	Colon	POINT (570970 6149046)	NO2
2019-01-04 21:00:00	72.0	Colon	POINT (570970 6149046)	NO2
2019-01-04 22:00:00	38.0	Colon	POINT (570970 6149046)	NO2
2019-01-04 23:00:00	9.0	Colon	POINT (570970 6149046)	NO2
2019-01-05 22:00:00	7.0	Colon	POINT (570970 6149046)	NO2
2019-02-03 00:00:00	5.0	Colon	POINT (570970 6149046)	NO2

Time Series with GANs

A time series can be represented as collection of vectors that are defined according to a time window size and a selection procedure



For example:

- Time window size 8
- Non-overlapping sliding window procedure

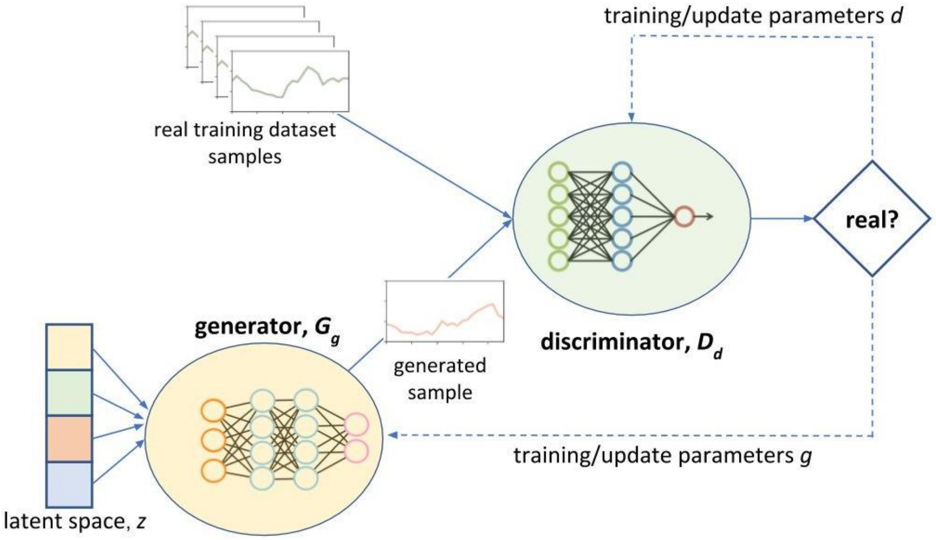
5.0	6.0	5.0	5.0	5.0	11.0	13.0	7.0

6.0	6.0	5.0	18.0	7.0	72.0	38.0	9.0

date	measure	ID_estacion	geometry	poll
2019-01-02 08:00:00	5.0	Colon	POINT (570970 6149046)	NO2
2019-01-02 09:00:00	6.0	Colon	POINT (570970 6149046)	NO2
2019-01-02 10:00:00	5.0	Colon	POINT (570970 6149046)	NO2
2019-01-02 13:00:00	5.0	Colon	POINT (570970 6149046)	NO2
2019-01-02 14:00:00	5.0	Colon	POINT (570970 6149046)	NO2
2019-01-02 15:00:00	11.0	Colon	POINT (570970 6149046)	NO2
2019-01-02 16:00:00	13.0	Colon	POINT (570970 6149046)	NO2
2019-01-02 17:00:00	7.0	Colon	POINT (570970 6149046)	NO2
2019-01-02 18:00:00	6.0	Colon	POINT (570970 6149046)	NO2
2019-01-02 19:00:00	6.0	Colon	POINT (570970 6149046)	NO2
2019-01-02 23:00:00	5.0	Colon	POINT (570970 6149046)	NO2
2019-01-03 20:00:00	18.0	Colon	POINT (570970 6149046)	NO2
2019-01-03 21:00:00	7.0	Colon	POINT (570970 6149046)	NO2
2019-01-04 21:00:00	72.0	Colon	POINT (570970 6149046)	NO2
2019-01-04 22:00:00	38.0	Colon	POINT (570970 6149046)	NO2
2019-01-04 23:00:00	9.0	Colon	POINT (570970 6149046)	NO2
2019-01-05 22:00:00	7.0	Colon	POINT (570970 6149046)	NO2

Non-supervised time series generation

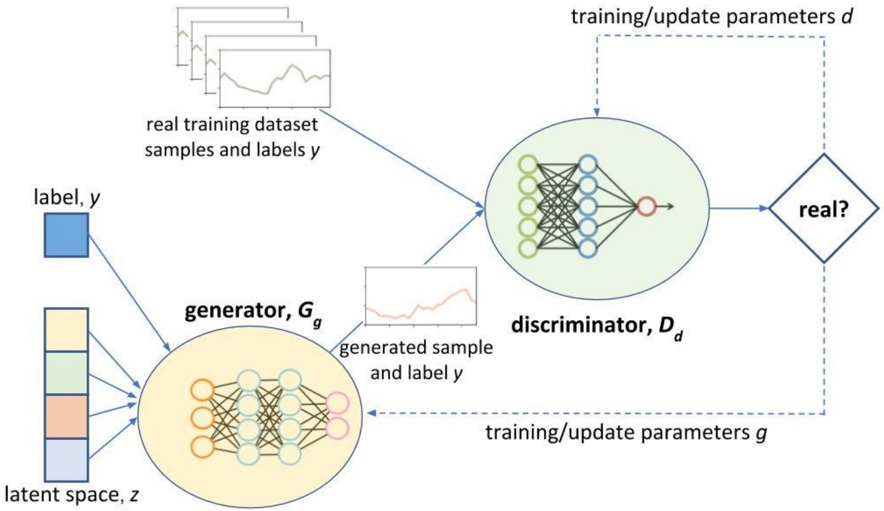
The idea is to use GANs to create the vectors that represent the time series



How can we control the generation?

Conditional GANs for time series generation

In order to improve the generation of the time series (i.e., vectors) some feature/characteristic of the window to be generated to control the generation



Examples. Generating pollution data time series

Two problems (papers):

Time series definition: Average NO_2 concentration in one hour during one day

Data sample: Vector of 24 NO_2 concentration measures

Conditional generative adversarial networks to model urban outdoor air pollution.

- https://arxiv.org/abs/2010.02244

Generative adversarial networks to model air pollution under uncertainty.

- http://ceur-ws.org/Vol-2858/short11.pdf

Example. Generating pollution data time series

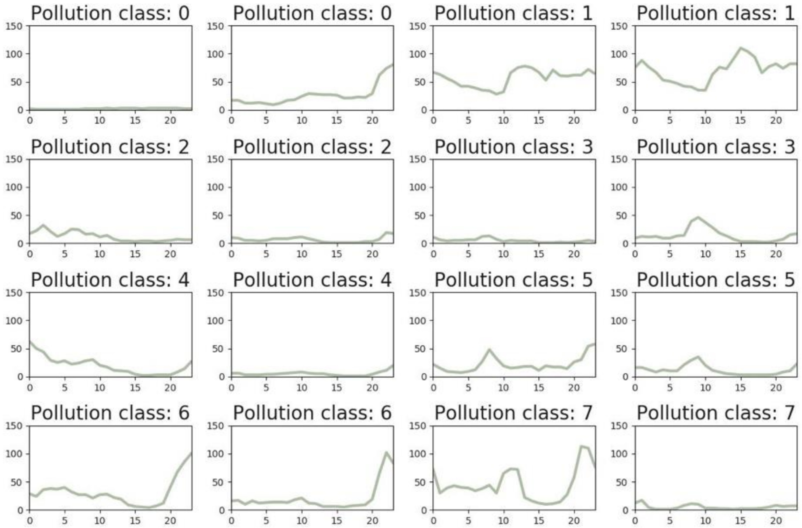
Conditional generative adversarial networks to model urban outdoor air pollution. https://arxiv.org/abs/2010.02244

Time series definition: Average NO_2 concentration in one hour during one day

Data sample: Vector of 24 NO_2 measures

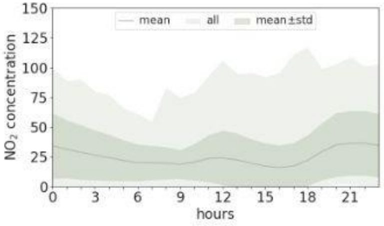
The behaviour of the pollution is affected by the type of day (week-day/weekend) and the season

season	type of day	class	number of samples
winter	weekend	0	439
winter	working day	1	1082
spring	weekend	2	439
spring	working day	3	1119
summer	weekend	4	445
summer	working day	5	1116
autumn	weekend	6	420
autumn	working day	7	1045

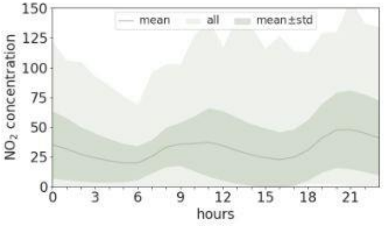


Example. Generating pollution data time series

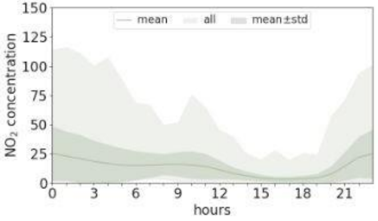
Real data distribution



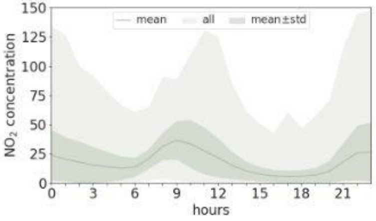
a1) Class 0: winter-weekends.



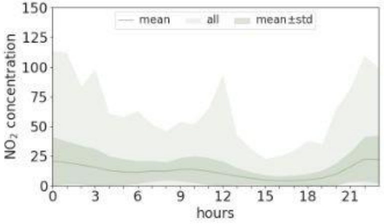
a2) Class 1: winter-working.



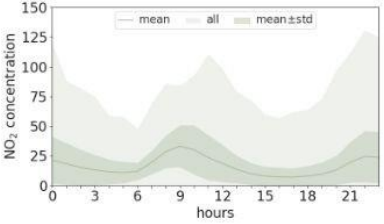
c1) Class 4: summer-weekends.



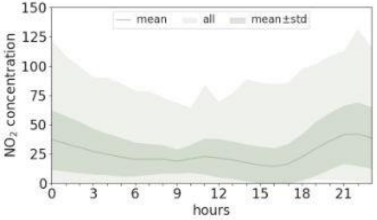
c2) Class 5: summer-working.



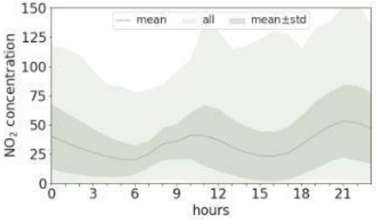
b1) Class 2: spring-weekends.



b2) Class 3: spring-working.



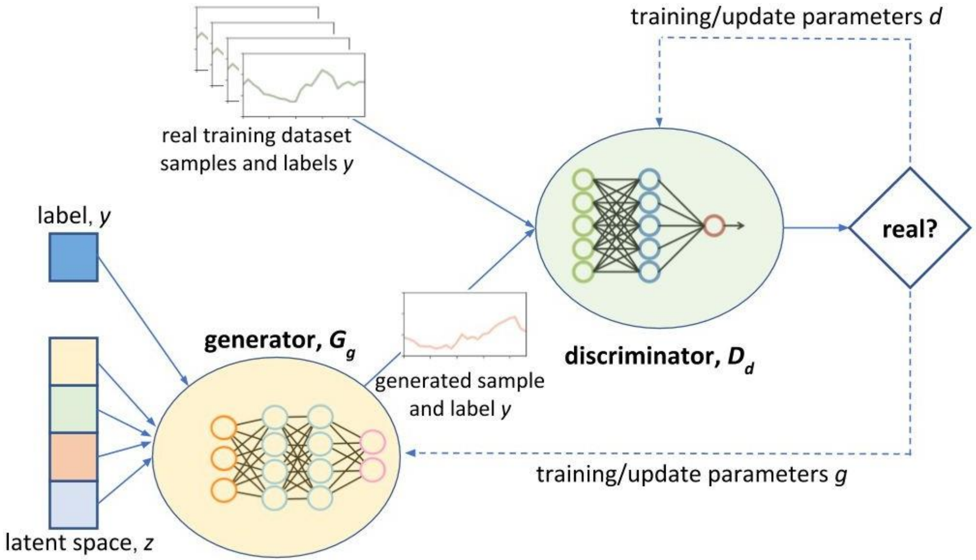
d1) Class 6: autumn-weekends.



d2) Class 7: autumn-working.

Example. Generating pollution data time series

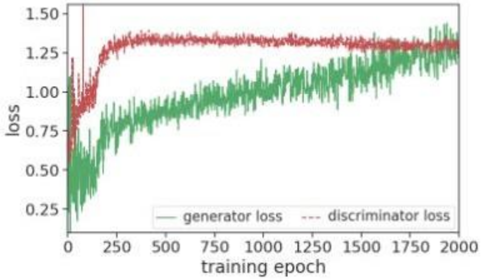
Pollution class = {0, 1, 2, 3, 4, 5, 6, 7}



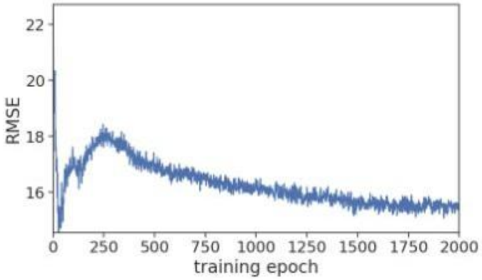
https://colab.research.google.com/drive/1kKpneBxuLyYRoEKDAVYgYYhUh3bdyCNK?usp=sharing

Example. Generating pollution data time series

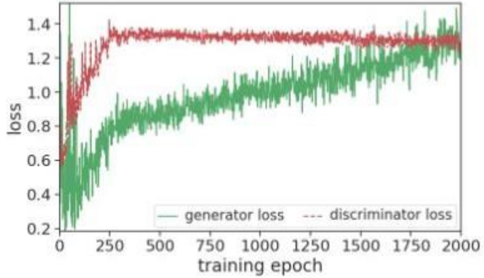
The quality of the generated data was evaluated according to the root mean squared error (RMSE) between the fake samples produced and the time series that represents the mean



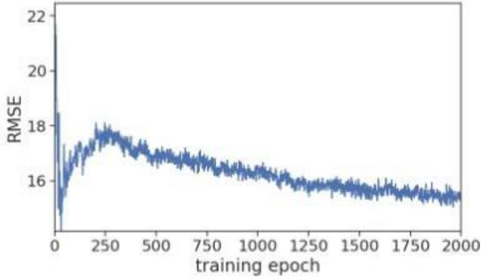
a1) *fake-1* loss values.



a2) *fake-1* RMSE.



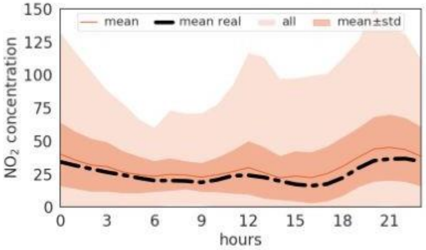
b1) *fake-5* loss values.



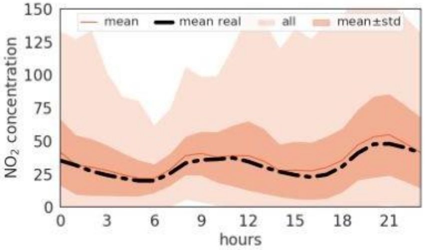
b2) *fake-5* RMSE.

Example. Generating pollution data time series

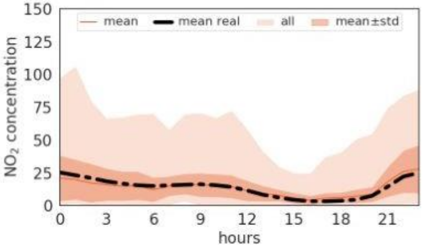
Generated data



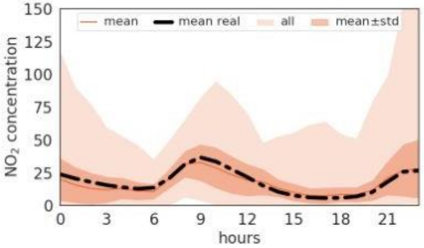
a1) Class 0: winter-weekends.



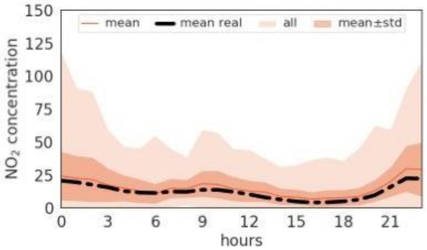
a2) Class 1: winter-working.



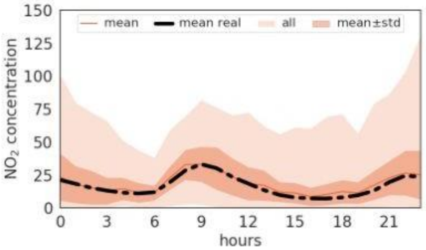
c1) Class 4: summer-weekends.



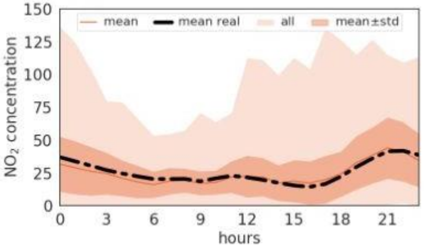
c2) Class 5: summer-working.



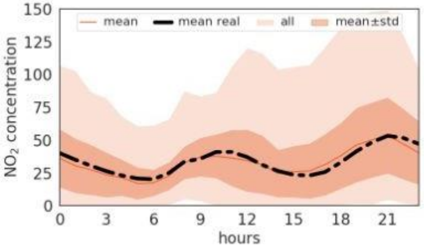
b1) Class 2: spring-weekends.



b2) Class 3: spring-working.



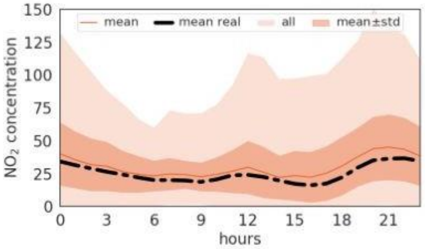
d1) Class 6: autumn-weekends.



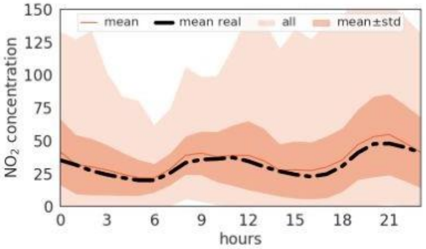
d2) Class 7: autumn-working.

Example. Generating pollution data time series

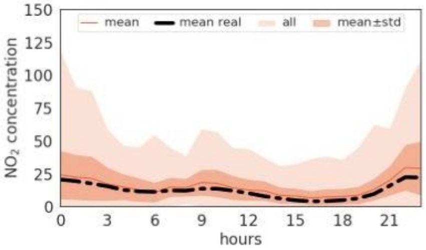
Generated data vs. Real data



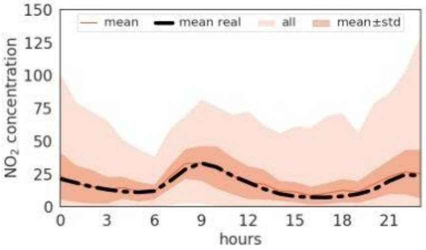
a1) Class 0: winter-weekends.



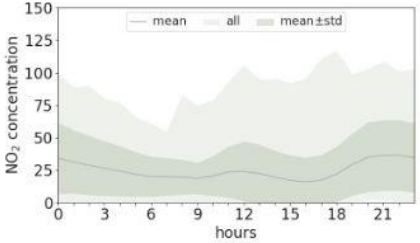
a2) Class 1: winter-working.



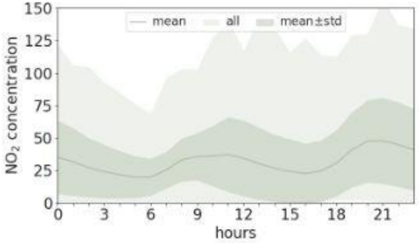
b1) Class 2: spring-weekends.



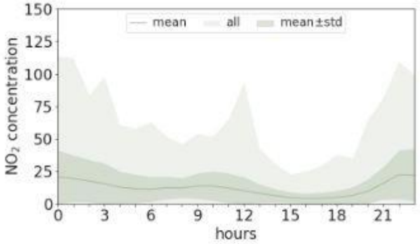
b2) Class 3: spring-working.



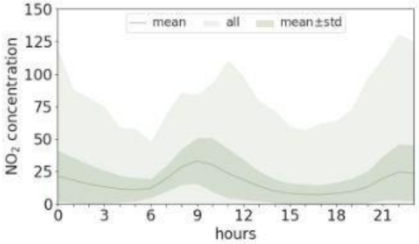
a1) Class 0: winter-weekends.



a2) Class 1: winter-working.



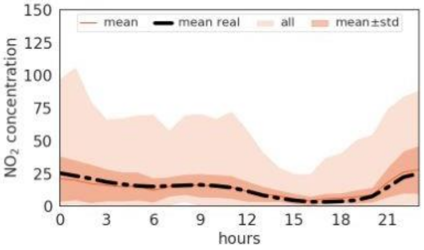
b1) Class 2: spring-weekends.



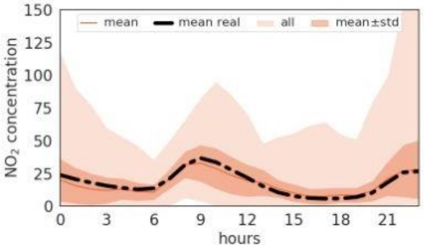
b2) Class 3: spring-working.

Example. Generating pollution data time series

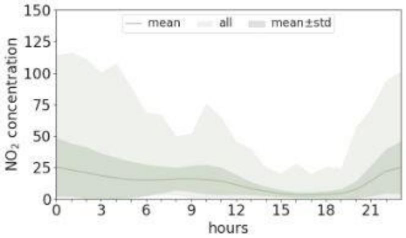
Generated data vs. Real data



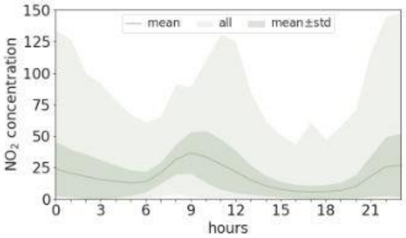
c1) Class 4: summer-weekends.



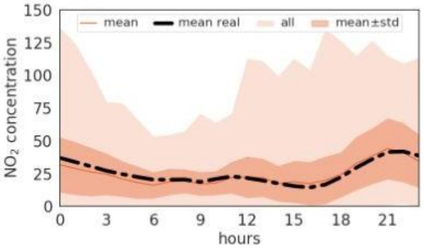
c2) Class 5: summer-working.



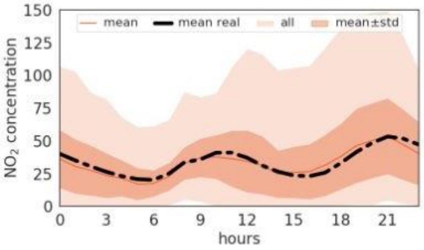
c1) Class 4: summer-weekends.



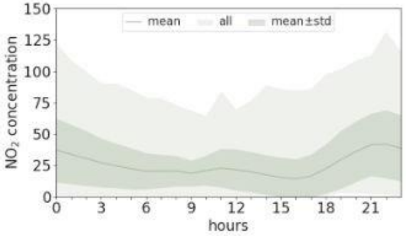
c2) Class 5: summer-working.



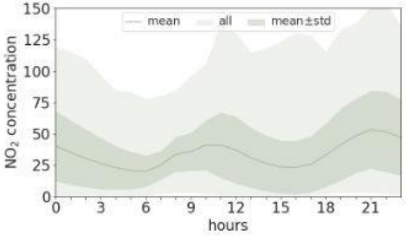
d1) Class 6: autumn-weekends.



d2) Class 7: autumn-working.



d1) Class 6: autumn-weekends.



d2) Class 7: autumn-working.

Generating Time Series with Conditional GANs

Questions?