LAMARR
INSTITUTE FOR MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE

Hands-on-Session : Time Series Generation using GAN

Dr. Amal Saadallah

Partner institutions:

technische universität dortmund

Fraunhofer IAIS

UNIVERSITÄT BONN

Fraunhofer IML

Institutionally funded by:

Bundesministerium für Bildung und Forschung

Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen

Outline

- Time Series Data
- Time Series Analysis Techniques
- Concept Drift in Time Series
- Applications of Time Series Analysis in Astroparticle Physics
- Time Series Data Generation

Time Series Data

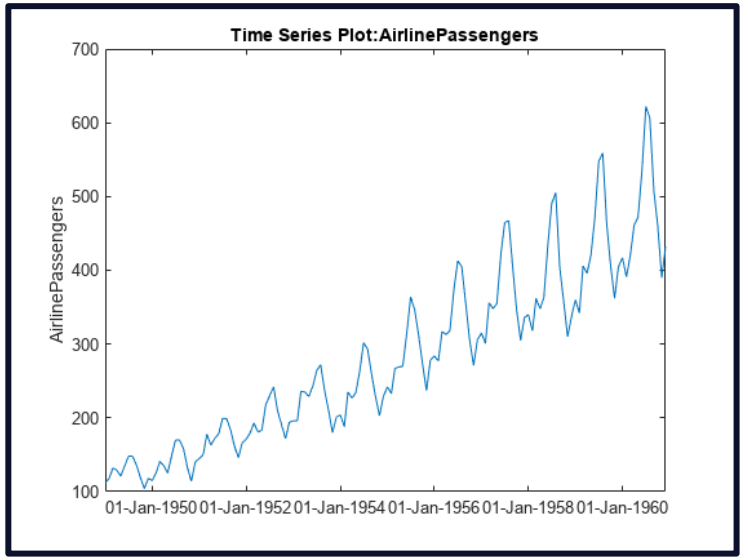
Definition

- **General Definition:** “A time series is a collection of observations made **sequentially through time**, whose dynamics is often characterized by short/long period fluctuations (**seasonality and cycles**) and/or long period direction (**trend**)”

- **Examples:** Stock prices, weather data, sensor readings, economic indicators.

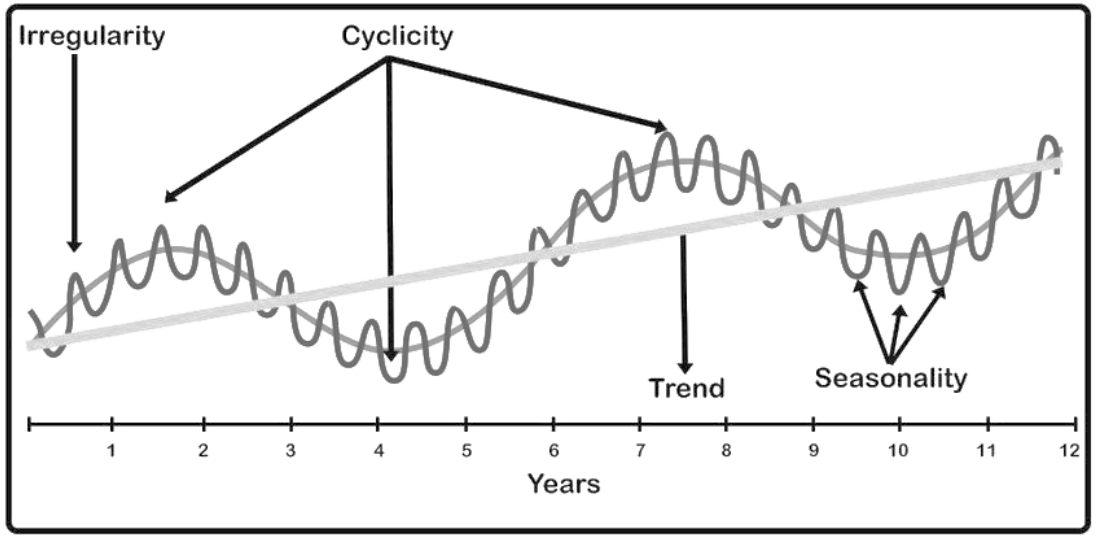
- Observations of a time series X recorded from 1 to T may be denoted by $X_{1:T} = x_1, x_2, \ldots, x_t, \ldots, x_T$ since data are usually collected at discrete points in time.

Observation at time t



Main Components of Time Series Data

- **Trend:** The long-term movement or direction in the data.
- **Seasonality:** Regular pattern or cycle in the data that repeats at specific intervals.
- **Cyclic Patterns:** Fluctuations in the data that occur at irregular intervals, typically influenced by external factors.
- **Noise or Irregular Patterns:** Random variations or irregularities in the data.



Time Series Data Importance

- **Predictive Analysis:** Helps in forecasting future events based on historical data (e.g., sales forecasting, weather prediction).
- **Anomaly Detection:** Identifying unusual patterns that do not conform to expected behavior (e.g., fraud detection, fault detection).
- **Trend Analysis:** Understanding long-term movement or pattern in data (e.g., market analysis, economic research).

Time Series Analysis Techniques

- **Descriptive Analysis:** Identifying patterns, trends, and seasonal effects.

  Example Techniques: Moving average, exponential smoothing.

- **Predictive Analysis:** Forecasting future values.

  Example Techniques: ARIMA, Lstm, CNN.

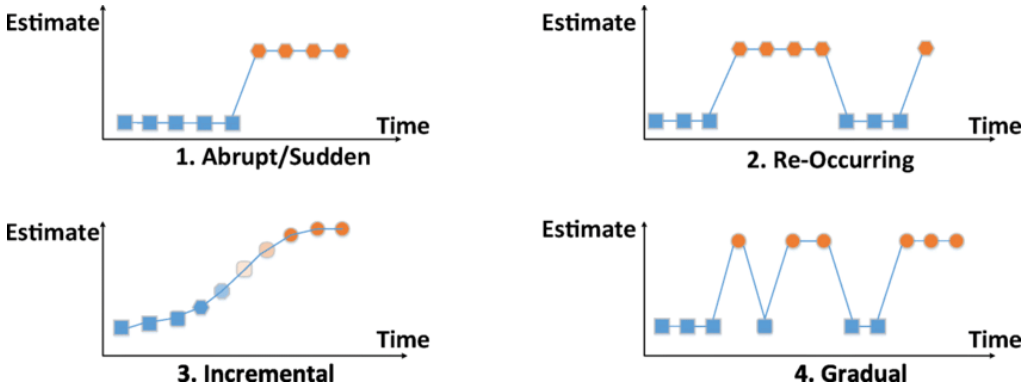
- **Inferential Analysis:** Understanding underlying factors.

  Example Techniques: Granger causality, cross-correlation.

Concept Drift in Time Series

- **Definition:** Concept drift refers to the change in the statistical properties of the target variable over time.

- **Types of Concept Drift:**
  - ✓ ***Sudden Drift:*** Abrupt change.
  - ✓ ***Incremental Drift:*** Incrementally change over time.
  - ✓ ***Recurring Drift:*** Repeated changes over time.
  - ✓ **Gradual Drift:** Gradually replace the old concept over time



- **Impact on Modelling:** Models trained on historical data may become outdated as the data distribution changes. They might loose their accuracy.

Handling Concept Drift

- **Model Retraining:** Periodically updating the model with new data.
- **Ensemble Methods:** Combining multiple models to adapt to new data.
- **Online Learning:** Continuously updating the model as new data arrives.
- **Drift Detection Methods:** Algorithms to detect and adapt to concept drift (e.g., ADWIN, DDM).

Applications of Time Series Analysis in Astroparticle Physics

- **Monitoring Cosmic Events:** Analyzing time series data from telescopes and detectors.
- **Predicting Particle Interactions:** Using historical data to predict future particle interactions.
- **Anomaly Detection:** Identifying unusual cosmic events or particle behaviors.
- **Case Study:** Analysis of gamma-ray bursts, neutrino detection time series.

Time Series Data Generation

Importance:

- *Synthetic Data:* Creating data for training models when real data is scarce.
- *Scenario Testing:* Simulating different scenarios to test model robustness.

Methods:

- *Simulation Models:* Generating data based on mathematical models.
- *Generative Models:* Using machine learning models like GANs to create synthetic time series.

Time Series Data Generation

Generative Adversarial Networks (GANs) for Time Series

- **Overview:** GANs consist of a generator and a discriminator that compete against each other.
- **Process:** The generator creates synthetic data while the discriminator tries to distinguish between real and fake data.
- **Advantages:** Ability to generate realistic and diverse time series data.

Link to the notebook

https://colab.research.google.com/drive/1fvLPhstrnDXUGkGay5CnVonQjgtueiLK?usp=sharing