POLYTECHNIC OF TURIN

MASTER's Degree in DATA SCIENCE AND ENGINEERING

MASTER's Degree Thesis

TITLE

Supervisors

Prof. LUCA CAGLIERO

Prof. DAMIEN ERNST

Candidate

MAURIZIO VASSALLO

July 2022

Table of Contents

Abbreviations

ANN	Artificial neural network
ANM	Active network management
ANSI	American National Standards Institute
API	Application programming interface
BBM	Bus-branch model
BDEW	German Association of Energy and Water Industries
CNN	Convolution neural network
DER	Distributed energy resources
DES	Distributed energy storage
DNN	Deep neural network
DSO	Distribution system operator
EBM	Element-branch model
FN	False negative
FP	False positive
IEC	International Electrotechnical Commission
FACTS	Flexible alternating current transmission system
GARPUR	Generally accepted reliability principle with uncertainty modelling and through probabilistic risk assessment
GAN	Generative adversial network
GHG	Greenhouse gas
LSTM	Long short term memory
MAE	Mean absolute error

MLP	Multi layer perceptron
PF	Power flow
PV	Photovoltaic
ReLU	Rectified linear unit
RNN	Recurrent neural network
SLP	Standard load profile
Tanh	Hyperbolic tangent function
TN	True negative
TP	True positive
WP	Wind park

Symbols

CO_2	Carbon dioxide
I	Electric current
V	Voltage magnitude
ϕ	Voltage phase angle
p.u.	Per unit
P	Active power
W	Active power unit: Watts
Q	Reactive power
VAR	Reactive power unit: Volt-Ampere reactive
$\mathcal{G}$	Directed graph
$\mathcal{N}$	Set of positive integers representing the bus (or node) in the network
$\mathcal{E}$	Set of directed edges linking buses together
e_{ij}	Directed edge with sending bus i and receiving bus j
$\mathcal{D}$	Set of all devices
$\mathcal{L}$	Set of all loads
r.m.s.	Root mean square

i.i.d.	Independently and identically distributed

Chapter 1

Changes

1.1 11/03

- Added 'Changes' chapter (to be removed in the final version)
- Added 'Problem statement' section **??**
- Added 'MV Oberrhein' section 5.1
- Minor changes

1.2 18/03

- Added 'Aim of the thesis' section 2.1
- Added 'Power system reliability' section 3.1.3
- Minor changes

1.3 25/03

- Added 'N-1 reliability criterion' sub section 3.1.3
- Modified network elements' description part 3.1.1
- Added 'Power Flow' sub section 3.1.2
- Modified 'Problem formulation' chapter 4
- Added 'Simbench dataset' 5.1.1 and 'Time series' 5.1.2 sub sections
- Minor changes

1.4 01/04

- Changed chapter 4 title to 'Problem analysis'
- Added 'Network topology' section ??
- Modified 'Problem statement' section 4.1
- Added 'Solving methodology' section 4.2
- Added 'Pandapower' section 3.2
- Modified 'MV Oberrhein' section 5.1
- Minor changes

1.4.1 Questions

- Evaluation of the model issue 5.1.5

1.5 04/04

- Modified 'Problem analysis' chapter 4

1.6 08/04

- Modified 'Problem analysis' chapter 4
- Modified 'Project implementation' chapter 5
- Minor changes

1.6.1 Questions

- GAN for missing values 5.1.3
- Choosing how many missing values 5.1.3

1.7 14/04

- Modified 'Problem analysis' chapter 4
- Modified 'Project implementation' chapter 5
- Minor changes

1.8 22/04

- Modified 'Problem analysis' chapter 4
- Modified 'Project implementation' chapter 5
- Added 'Machine learning overview' section 3.3

1.9 22/04

- Modified 'Problem analysis' chapter 4
- Modified 'Project implementation' chapter 5
- Modified 'Machine learning overview' section 3.3
- Minor changes

Chapter 2

Introduction

The rapid growth of the world population and the limited ability to supply non-renewable energy is leading to a rise of power demand, especially in developing countries. The current energy demand requires an intensive usage of fossil energy causing environmental pollution, as well as climate change.

Indeed, the increase of global temperature and the worsening of the air quality are posing a real problem for the environment. The changes observed in Earth's climate are primarily driven by human activities, particularly fossil fuel burning. The biggest disadvantage of fossil fuels is that during the process of combustion in addition to produce energy, greenhouse gases (GHG) are emitted [1].
Normally the solar energy from the sun would hit the earth and then part of this energy bounce back to the space, but these gases trap the heat from the sun in the atmosphere, increasing the earth temperatures.

In order to improve the situation, the 2015 Paris Agreement set an ambition to limit global warming to well below 2 °C above pre-industrial levels and pursue efforts to limit it to 1.5 °C - in part by pursuing net carbon neutrality by 2050. The substantial reduction of global greenhouse gas emissions (including CO_2) will limit the increase of global temperature [2].
Countries were asked to go through a process of decarbonization: the reduction of carbon dioxide emissions through the use of low carbon power sources.
These sources convert the energy coming through naturals elements (sun, wind, geothermal heat) in another form of energy, electricity for example, with low or no waste products such as CO_2 or other chemical pollutants.

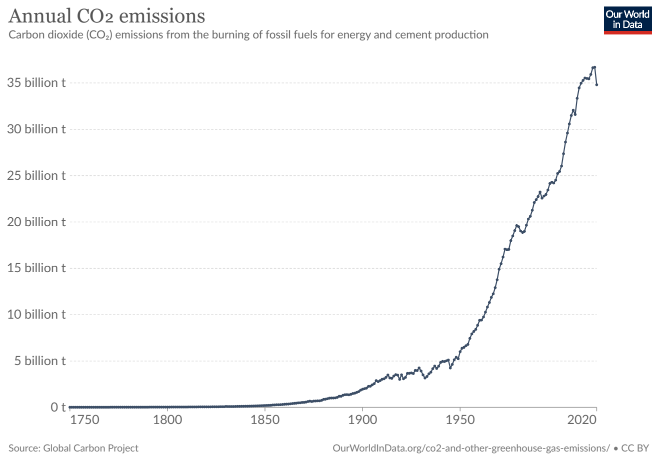

Figure 2.1: World CO_2 production over the years [3]

Thanks to this emerging trend of decarbonization, more and more renewable power energy devices are introduced in the distribution networks [4]. High penetration of these renewables devices bring in some technical complications for the distribution of power and voltage in the grids.
The networks, that have been designed around the conventional centralized energy production, have to adapt to the new generators in the system. It is possible to say that the distribution networks are moving from unidirectional power flow (from the distribution system to the consumers) to a bidirectional power flow (in this case the consumers are also producers and the exceed energy can be transported from the consumers to the distribution system. They are also known as prosumers [5]). This switch from unidirectional to bidirectional power flow requires a smarter system that can handle in an efficient way the generation and distribution of voltage.

In the literature, this smarter way to control a distribution system is known as active network management (ANM) and it refers to the design of control schemes that modulate the generators, the loads, and the distributed energy storages (DESs), as well as other elements like switches, connected to the grid.

2.1 Aim of the thesis

The aim of this thesis is to exploit data-driven approaches to forecast the future distribution of voltages based on historical measurements in a medium-voltage distribution system using machine learning techniques, in particular deep learning models. Predicting over voltages problems of lines would allow avoiding possible consequences related to these issues.

2.2 Thesis outline

Chapter 3

Background

The voltage control problem has been studied for years, but it only comes under the spotlight in the last years for the increasing number of distributed resources introduced in the networks.

There are two voltage problems that can raise on a network: under and over voltage problems.
Generally, electronic devices have defined voltage limits they work well with. The voltage is not always constant during time, but it fluctuates. These fluctuations can be large, and the voltage can drop below the device's minimum allowed voltage limit, in this case there would be an under voltage problem, or increase above its maximum allowed limit, in this case there would be an over voltage problem.

The introduction of more and more DER devices in the networks increases the number of voltages problems, in particular over voltage problems. These devices generate electric power and when this power is greater than the energy consumed, the extra energy is emitted back in the network.

For this reason, it is important to control the voltage in an electrical power system for a regular operation of the electrical equipment. It can prevent damages such as overheating of devices and lines, reduce transmission losses and maintain the ability of the system to last and avoid voltage collapse. Over voltages other than shorten the lifetime of equipment have a negative impact on the stable operation of both supply-side devices and demand-side appliances.

It is therefore required to take some actions to avoid these over voltages situations. In particular, it is useful and sometimes needed to reduce the output of renewable generators to prevent possible voltage problems. This means that the output power of these generators is less from what they could otherwise have produced given the available resources. Such prevention is often referred to as the process of **curtailment**. Said generation curtailment, along with storage and transmission losses, constitute the principal sources of energy loss that could be minimised with smart control system like active network management ANM [6].

Controlling the voltage in an active way has many interesting properties:

- It is a combination of local and global problem: the voltage at each node is influenced by the powers of all other nodes, but the impact depends on the distance between them.
- It is a constrained optimization problem with many constraints, for example to keep the voltage in a given range, and the objective is to minimise the total power loss.
- Voltage control has a relatively large tolerance, and there are no severe consequences if the control fails to meet the requirements for short periods of time. [7]
- It is a hierarchical problem where the information available can be represented as a pyramid: much information is available at the top of the pyramid (distribution stations and substations) and it decreases at the base of the pyramid (houses, factories) mainly due to the absence of many sensors.

3.1 Power system

3.1.1 Description of a power system

A power system is a complex infrastructure that produces and distribute electricity to different consumers. A power system consists of generation, transmission and distribution system and each of them has a different function.

In the traditional power system, electricity is generated in large, centralised power plants. The electricity is then transferred to the loads using the transmission and distribution networks. Transmission substations are located near the power plants, their main function is to increase the voltage level to high and extra-high voltages levels. The reason for transmitting power at high and extra-high voltage levels is to increase efficiency. The lower current accompanying the high voltage transmission allows for the use of thinner, lighter-weight cables. This reduces the cost in the tower and electrical line construction. In Belgium, high and extra-high voltages refer to voltage magnitudes $30kV \leq |V| < 380kV$ for the high voltage and $|V| \geq 380kV$ for the extra high voltage [8].

Large industrial complexes and factories that require a substantial amount of power often utilise medium supply voltages. The high voltage coming from the transmission lines is sent to the primary substation, this can supply step-down power to secondary substations or to single buildings. Secondary substations can have transformers to further step down the power, and they are generally located in areas that can serve one or more buildings. Medium voltages refer to voltage magnitude $0.4kV \leq |V| < 30kV$.
Then the medium supply power is step down again to a low voltage and sent to the domestic household or home appliances power supply. Low voltages refer to voltage magnitude $|V| < 0.4kV$.

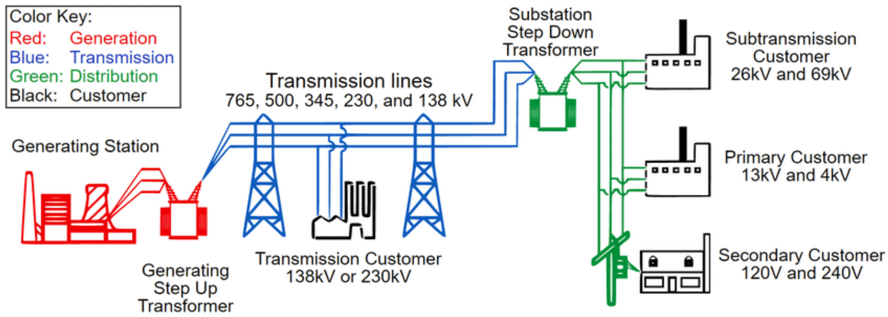

Figure 3.1: Power network distribution [9]

A power system is usually made up of the following main elements:

- **Generator**. These generate energy, converting a form of energy into electricity. In general, electricity is produced when a magnet is moved near a wire to create a steady flow of electrons.
  So, many generators produce energy using turbines: a fluid spin the generator's blades, producing electricity. This fluid, either water or air, can derive from natural sources: hydroelectric, wind or geothermal turbines, or generated by combustion of some fuel, for example coal, natural gas, oil or nuclear source. There are other generators that do not need a turbine to generate electricity, for example solar panels.

- **Lines**. These transport the power from where energy is generated to where it is consumed. One of the main issues about transportation lines is insulation. There are different types of lines: overhead cables, they use air to insulate the bare conductors or underground cables, for these cables particular attention must be taken to insulate them from other conductors and from the earth (ground). Also, the material used must be resistant to damages, corrosion, and it must avoid that the water is being absorbed.

- **Transformers**. Transformers are used to interlink systems operating at different voltages. These can increase the voltage magnitude near a generator power plant or decrease it near the consumptions facilities.
  Changing the voltage magnitude allows reducing the power loss due to transportation. One of the main causes of power loss is the Joule effect: some part of the energy transmitted is converted in heat generated by the current flowing through a conductor. This power lost is given by the equation $P = VI$, where V is the voltage and I is the current, so decreasing the voltage reduces the energy loss.

- **Switchgear**. In an electricity supply, it is necessary to disconnect equipment from the network quickly if a fault occurs to avoid damage on the elements of the network, or to disconnect some points of the network to avoid excessive losses or too high or low voltages.
  Switchgear is a broad term that describes a wide variety of switching devices that fulfil the need of controlling, protecting, and isolating power systems. Among these switching devices the most common are: *circuit breaker*, during an electrical fault, a circuit breaker will detect the anomaly and interrupt the power flow, effectively limiting damage to the system; *switch* is an electrical component that can disconnect or connect the conducting path in an electrical circuit, interrupting the electric current or diverting it from one conductor to another; *recloser* similar to the circuit breaker but used in high voltage networks, these devices handle trouble temporary occurrences such as lightning, windblown tree branches or wires, birds, or rodents damaging the wires.

- **Loads** are electric components that consume the electric power generated by power plants. The type of loads can be divided base on the consumption in:

  Domestic loads, the domestic loads mainly consist of lights, fan, refrigerator, air conditioners, mixer, grinder, heater, ovens, small pumping, motor, etc. The domestic loads consume very little power.

  Commercial loads, the commercial loads mainly consist of lightning, fans, heating, air conditioning and many other electrical appliances used in establishments such as markets, restaurants, shops. This type of load occurs for more hours during the day as compared to the domestic load.

  Industrial loads, the industrial loads refer from a small-scale industry, to a heavy industry. It includes all electrical loads used in industries along with the employed machinery. Industrial loads may be connected during the whole day [10].

- **Buses** are nodes where a line or several lines are connected and may also include several components such as loads and generators.

3.1.2 Power flow

An important procedure in power system networks is to perform a numerical analysis to determine the electrical state of the network, starting from some parameters that are known. This analysis is called power flow (PF).
The objective of a power flow study is to calculate the voltages, magnitude and angle, for a given bus, load, generation device. After this information is known for all elements, line flows and losses can be calculated.

Buses

The power flow gives information about the steady state of the entire system such as voltage, active, reactive power and lines' loading.
Each bus is associated with four quantities: voltage magnitude V, phase angle ϕ, real power P and reactive power Q. Depending on the quantity that have been specified, buses in the power system are classified into the following three different types:

- **Slack bus**. It is taken as reference where the magnitude and phase angle of the voltage are specified. Slack bus magnitude considers 1 p.u. and phase angle 0 degrees. This bus provides the additional real and reactive power to supply the transmission losses, since there are unknown until the final solution is obtained.

- **Load buses or PQ bus**. At these buses, the real and reactive powers are specified. The magnitude and phase angle of the bus voltage are unknown until the final solution is obtained.

- **Voltage controlled buses or PV bus**. At these buses, the real power and voltage magnitude are specified. The phase angles of the voltages and the reactive power are unknown until the final solution is obtained. The limits on the value of reactive power are also specified.

Solution techniques

Defining and solving the power flow equations are the main tasks in load flow analysis.
The definition of the power flow equations is based on Ohm's Law, which is the relationship between voltages and currents. For a network, it can be expressed in matrix notation as follows:

$$\mathbf{Y} \times \mathbf{V} = \mathbf{I}$$

$$\begin{bmatrix} Y_{1,1} & Y_{1,2} & \cdots & Y_{1,N-1} & Y_{1,N} \\ Y_{2,1} & Y_{2,2} & \cdots & Y_{2,N-1} & Y_{2,N} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ Y_{N-1,1} & Y_{N-1,2} & \cdots & Y_{N-1,N-1} & Y_{N-1,N} \\ Y_{N,1} & Y_{N,2} & \cdots & Y_{N,N-1} & Y_{N,N} \end{bmatrix} \times \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_{N-1} \\ V_N \end{bmatrix} = \begin{bmatrix} I_1 \\ I_2 \\ \vdots \\ I_{N-1} \\ I_N \end{bmatrix}$$

Where:

- **Y** is the bus admittance matrix
- **V** is an array of bus voltages
- **I** is an array of bus current injections (positive value when generation, and negative value when load)

The power flow formulation is based on the application of Kirchhoff's laws to meshed electric networks. The basic concept is that the sum of all flows into each and every node should be equal to zero.
The flows are in complex form, they consist of real and reactive components, or Ws and VARs. That means that if there are n nodes, then there are n complex equations. The resulting system of equations involves non-linear relationships, making the calculations not easy. Solution methods for this system of equations are primarily iterative with the objective of reducing the sum of flows in all nodes to some acceptably small value known as the mismatch tolerance.

All these iterative methods follow the same basic concept: they assume starting values for the dependent variables, primarily voltage at nominal voltage magnitude (i.e. 1 p.u.) and zero phase angle; compute new values for those voltages using the nodal network equation or a numerical approximation and repeat until the convergence criteria are met.

The solution has to satisfy some network constrains, in particular:

- Active and reactive power balance: the sum of the power injections (that can be positive or negative) at each bus must be equal to 0. This, as said, results from the Kirchhoff's laws.
- Voltage limits: the voltage magnitude at each bus and the voltage phase difference between two directly connected buses are bounded by some specific values to maintain the system safe.

- Thermal limits on transmission lines: the flow in each transmission line is limited due to the thermal limit of the conductors.
- Generators' active and reactive power limits: the generating units have generally a minimum and maximum level of output power.
- Generator ramping limits: the output power of a generating unit can not be instantaneously increased or decreased. The operator must take into account the ramping limits of the generators.

Convergence

The PF is a non-linear and non-convex numerical analysis, with a large number of constraints (both equality and inequality constraints) and variables (that can be both continuous and discrete). It is therefore a hard problem, whose cost of finding a solution can increase exponentially, particularly with the increasing size of the network. Moreover, there is no guarantee to find the global optimum.
When a solution exists, and it is reached, it is said that the network has converged. Convergence is the state when all nodes have met the mismatch tolerance.
The main power flow solution methods are:

- Gauss-Seidel method updates the voltage one node at a time until all nodes are within the mismatch tolerance.
- Newton-Raphson method uses a first order expansion of the power flow equations to approach convergence. Generally faster than the Gauss-Seidel method and able to converge to small tolerances. However, the method is prone to the phenomenon of **divergence**, when mismatches increase instead of decrease from iteration to iteration. This occurs when the solution vector exits outside the feasible solution space at any point during the algorithm. Once outside feasible space, the solution gradient tends to further increase mismatches, leading to solutions that "blow-up" in the numerical sense. This method requires calculating the first order approximation matrix (known as the Jacobian).
  Several variations on the Newton-Raphson are in use, including:

  Fast Decoupled: separates the loosely linked real and reactive components of the power flow equations in order to speed up solution.

  Fixed Newton: does not update the first order approximation matrix every iteration to reduce computational burden.

  Non-divergent power flow: applies a reduction to the Jacobian multiplier whenever the solution appears to exit feasible space. In certain situations, this may prevent divergence, or at least stop it before blow-up.

- Interior-Point Newton method forces the solution inside feasible space to avoid divergence. The interior point method uses a second order expansion of the power flow equations as a basis for its algorithm. The method is more computationally intensive than either the Gauss-Seidel or Newton-Raphson, but is less susceptible to numerical divergence.

Divergence

Divergence is the condition of the power network when the numerical solution can not be found any more due to some possible issues:

- the power system is going to “blow-up.”
- the power system is in voltage collapse.
- the power system is unstable.
- the initial conditions defined were bad or poor.
- some issues related to software or input data.

Divergence of the power flow solution has traditionally been associated with the singularity of the Jacobian matrix. Since some methods require an inverse of the Jacobian as part of its solution algorithm, singularity of the Jacobian means division by zero [11].

3.1.3 Power system reliability

Reliability of a power system is an important factor concerning the quality of energy supply.
Power reliability can be defined as the degree to which the performance of the elements in a system results in electricity being delivered to customers within accepted standards and in the desired amount [12].
Reliability indices typically consider such aspects as:

- the number of customers;
- the connected loads;
- the duration of the interruption measured in seconds, minutes, hours, or days;
- the amount of power interrupted;
- and the frequency of interruptions.

These factors depend on variable such as reliability of individual items of equipment, circuit length and loading, network configuration, distribution automation, and available transfer capacity [10].

For reliability purposes, it is important to know the maximum voltage that can be transferred with transmission lines to meet the anticipated load demand. It is also important to know the levels of power through various transmission lines under certain contingency outage conditions to maintain the continuity of service. Knowledge of power flows and voltage levels under normal operating conditions are necessary in order to determine fault currents and the ensuing consequences on the stability of the system [11].

There exists some standards about power system reliability.
The International Electrotechnical Commission (IEC), IEC TS 62749 [13], that states that the energy suppliers and facility managers need to verify the conformity of the energy supplied to:

- maximum limits
- statistical limits over a week or a year

Under normal operating condition some values must be verified:

- during each period of one week 95% of the 10 minutes mean r.m.s. values of the supply power voltage shall be within the range of $\pm 10\%$ p.u. and
- all 10 minutes mean r.m.s. values of the supply voltage shall be within the range $+10\% - 15\%$ p.u.

The American National Standards Institute (ANSI), C84.1-2016 [14], voltage standards for service voltage limits, for example, are classified as Range A and Range B limits. The voltage between 0.950 p.u. and 1.050 p.u. of nominal voltage lies under Range A, and the voltage between 0.917 p.u. and 1.058 p.u. of nominal voltage for 240 V service voltage lies under Range B. Note that the voltage can be within Range B for only a short duration and frequency, and thus corrective measures are necessary to constrict.

Reliability criteria

The goal of a distribution system operator (DSO) is to ensure a reliable system. Unfortunately, a completely reliable electricity supply is not feasible to obtain since it comes at an infinite cost. So, network operators need to determine an acceptable reliability level, by balancing the costs and benefits, where acceptable reliability level means that all the elements in a network have an acceptable voltage range.

The European GARPUR project (**G**enerally **A**ccepted **R**eliability **P**rinciple with **U**ncertainty modelling and through probabilistic **R**isk assessment) developed reliability management approaches and criteria. One of these criteria used by system operators is the N-1 criterion.

The basic principle of N-1 security in network planning states that if a component, for example a transformer or circuit, should fail or be shut down in a network operating at the maximum levels of transmission and supply, the network security must still be guaranteed. This means that the safety of the system is guaranteed and the spreading of the failure is avoided.
It is possible that there may be another contingency before restoring the network after the fail of one element, this criterion is known as N-1-1 criterion.

With the increasing of network complexity more than one element may fault, for this reason there exists other levels of reliability, like the N-2 criteria. In this case, even if in the network two components fail, the network security is guaranteed.
This N-2 criteria requires much more computational power since, the system operator must calculate what happens to the network for any combination of two fault elements. So, the problem becomes a combination problem, where the possible combination are given by: $\binom{N}{2}$, with N the number of elements in the network.

In general, the calculation can be extended to any generic k elements, but the complexity of the problem increases with the value of k. Indeed, the possible combination in a N-k contingency are: $\binom{N}{k}$, with $2 < k < N$

3.2 Pandapower

This thesis project will be developed with the help of Pandapower.
Pandapower is a Python based power system analysis tool aimed at automation of static and quasi-static analysis and optimization of power systems [15].
Pandapower is a powerful tool that allows to easily create a model for any power network using customizable predefined data structures, it can solve the PF problems, perform the state estimates, topological graph searches and diagnose the system for possible errors.

3.2.1 Data structure

Pandapower is based on a tabular data structure, where every element is represented by a table that holds all parameters for a specific component. It is possible

to add more information at the data structure, indeed, after the calculation of the power flow, a result table, which contains the element specific results of the different analysis methods, is added to the structure.
The tabular data structure is based on the Python library pandas. It allows storing variables of any data type, so that electrical parameters can be stored together with status variables and meta-data, such as names or descriptions. The tables can be easily expanded and customized by adding new columns without influencing the Pandapower functionality. All inherent pandas methods can be used to efficiently read, write and analyse the network and results data.

A Pandapower network is a Python dictionary that holds all information about the network. Most importantly, it includes element and a result tables for each element type, such as line, transformer, switch, loads. The element table holds all input parameters that are specified by the user, while the result table is used to store the results of the power flow calculation. Input and output parameters are identified by the same index in both tables [15].

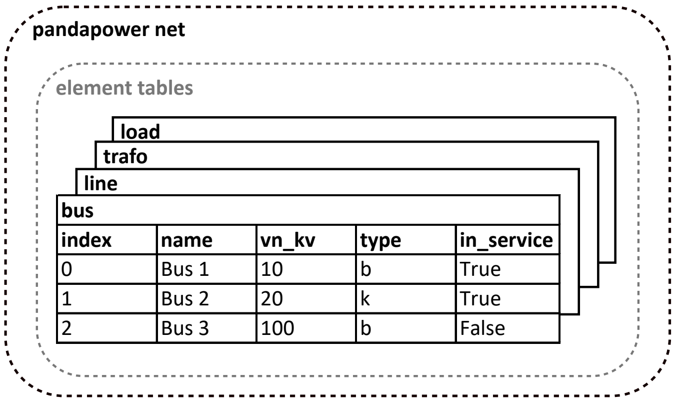

Figure 3.2: Pandas data frame representation of the Pandapower network

3.2.2 Power flow solver

As said, power flow is one of the most important electric analysis function for power system planning. It allows calculating the current flows and voltages in the network.
The Pandapower power flow solver is based on the Newton-Raphson method. The implementation is based on PYPOWER Python library. To solve the PF, the bus constraints include maximum and minimum voltage magnitude, active and reactive power limits can be defined for PV and slack-elements like external grids and generators, but also for PQ-elements, such as loads and static generators.

After running the power flow calculation, new tables are added to the network data frame.

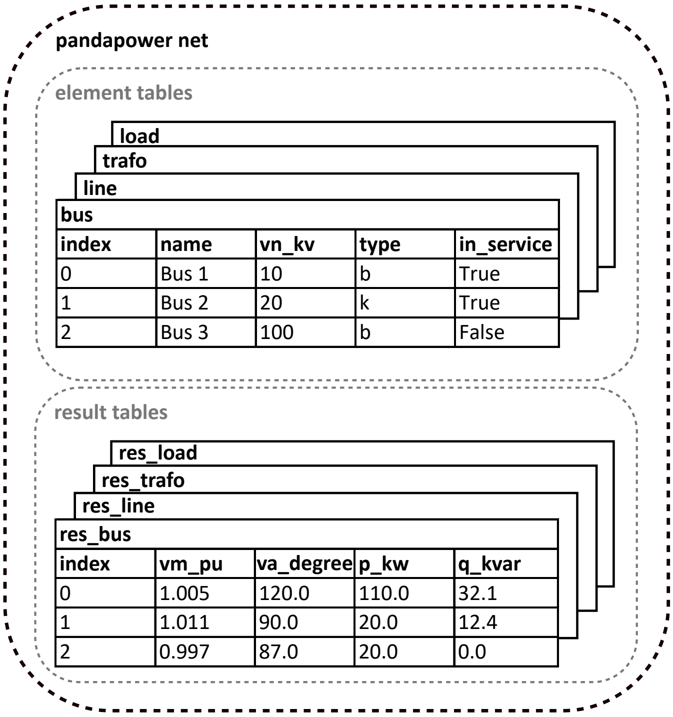

Figure 3.3: Pandas data frame representation of the Pandapower network after the power flow calculation

The power flow calculation on a Pandapower network can fail to converge for a vast variety of reasons, which often makes debugging difficult, annoying and time-consuming. To help with that, the diagnostic function automatically checks Pandapower networks for the most common issues leading to errors. It provides logging output and diagnoses with a controllable level of detail.

3.2.3 Time series

Pandapower allow running time series analysis for a given network. There are two main requirements for time series calculations:

- a Pandapower network
- some time series (in a panda's data frame for example)

To execute the time series calculation, the loads, generators and other elements' active and reactive power time series have to be passed to a controller that will be in charge to change the elements' values according to the time series.

The time series calculation can be run with the command:

```
pandapower.timeseries.run_time_series.run_timeseries(net, ...)
```

this command will start a loop that iterates over every **time_step**. For each step, a control loop is started for each controller by **run_control**. The controller updates the elements' values at each step with the values given in the time series.

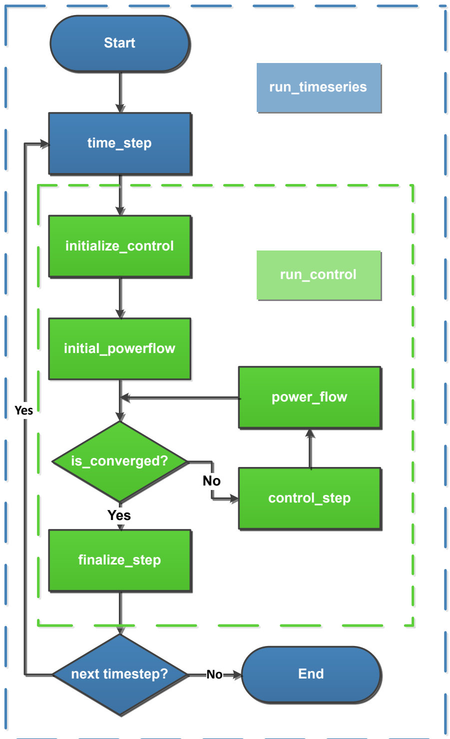

Figure 3.4: Pandapower time series calculation loop [16]

After each step, the elements' values are stored in an output writer object and this allows, after the full calculation is finished, to easily save the values on disk.

3.2.4 Other functionality

Pandapower has some other features:

- Predefined Networks. In addition to creating custom networks through the application programming interface (API), 66 predefined, published test and benchmark networks can be directly accessed through Pandapower. One of these networks, MV Oberrhein, is the one used in this thesis.

- Plotting features. Pandapower comes with extensive plotting features using the Matplotlib library. All Pandapower elements can be translated into different Matplotlib collections that can be customized with respect to shape, size and colour to allow highlighting and create individual network plots. It is also possible to use colour maps to codify information, like the loading of lines or the voltage at buses.
- Converter. Pandapower includes converters in order to export a Pandapower grid as a MATPOWER or PYPOWER casefile or the other way.

3.3 Machine learning overview

Machine learning is a subset of artificial intelligence that trains a machine to learn. In particular machine learning is the study of how a computer algorithm improves its performances at some task through experience or more precisely:

> A computer program is said to learn from experience E with respect to some class of tasks T and performance P, if its performance at tasks in T, as measured by P, improves with experience E [17].

where generally, in a machine learning problem, T is a task too complex to be solved with human written algorithms.

Machine learning differs from the traditional computer science methods. In traditional approaches, algorithms are sets of explicitly programmed instructions, or rules, used by computers to solve a problem. Machine learning algorithms instead allow for computers to train on data inputs and use statistical analysis in order to output values or answers.

image with traditional vs ml approaches

In machine learning, tasks are generally classified into broad categories. These categories are based on how learning is received or how feedback on the learning is given to the system developed. There are three main type of machine learning algorithms: supervised, unsupervised and reinforcement learning.

3.3.1 Supervised learning

In supervised learning, the goal is to learn a function that maps an input X to an output Y based on example input-output pairs and applying this learnt function to predict the output of future unseen data.

More formally, in a supervised learning problem, the goal is to find a function $f : X \to Y$, from a sample data S_n composed by pairs of (input, output) points:

$$S_n = ((x_1, y_1), \dots (x_n, y_n)) \in (X \times Y)^n$$

Typically, $x_i \subset R^n$ and $y_i \subset R$ for regression problems or y_i discrete for classification problems, for example $y_i \in \{0,1\}$ for binary problems.

In the statistical learning settings, an important hypothesis is that the training data is independently and identically distributed (i.i.d.). from a probability distribution function $P(X, Y)$. The goal of the learning is to find a mapping function f that can encode the property of $P(X, Y)$ between the inputs X and the output Y.

Another important concept is to evaluate how well the function f performs, calculating the error or loss between the predicted values f(x) and the actual value y. This error is evaluated with a loss, or cost, function $L : Y \times Y \to R^+$. There are many loss functions depending on the problem and requirements, one example is the mean absolute error (MAE) loss function:

$$L(f(x), y) = \frac{1}{N} \sum_{i=0}^{N} |f(x_i) - y_i|$$

Many supervised learning algorithms consider the minimisation of this loss function as an optimisation problem to find the best predictor among all the possible candidate input-output mappings in the solution space B.

With the loss function L, the definition of risk of the function f, also called generalization error, must be introduced:

$$R(f) = \int L(f(x), y))\, dP(x, y).$$

The objective is to find the function f in B that minimise the generalization error, R(f). Since it is not possible to solve R(f), because of the joint probability distribution P(x,y) is unknown, f inferred from available data set S_n.

3.3.2 Artificial neural networks

Artificial neural network (ANN) is a computational model that consists of several processing elements that receive inputs and deliver outputs based on their predefined activation functions. They have been proved to provide a strong approach to approximate functions in order to solve continuous and discrete problems.

They have been inspired by the biological neural networks that constitute animal brains. For the first time, in *1943* Walter Pitts and Warren McCulloch published a paper with the mathematical modelling of a neural network, taking inspiration from the human biology. They thought a human neuron cell as a threshold logic unit working together with other neurons to build a complex system. This neuron cell collects multiple signals arriving at the dendrites, elaborate them and if the accumulated signal exceeds a certain threshold, an output signal is generated that will be passed on by the axon [18].

Perceptron

The simple unit in an ANN is called perceptron. The perceptron is a mathematical function inspired by biological neurons, where each neuron takes inputs, weighs them separately, sums them and pass the sum through a nonlinear function to produce output. This mathematical function can be written as follows:

$$o(x_1, x_2, ..., x_{n-1}, x_n) = f(\sum_{i=0}^{n} w_i x_i + w_0) \tag{3.1}$$

where n is the number of connected neurons, x_i is the input from the neuron i, w_i is the weight that determines the contribution of input i, and f is a nonlinear function, like for example:

$$f(x) = \begin{cases} 1 \text{ if } x > T \\ -1 \text{ otherwise} \end{cases}$$

with T a real value representing the threshold that x has to surpass for the function to output 1. In the formula 3.1 the threshold T is given by the value w_0.

A single perceptron can be used for classification tasks: it builds a hyperplane that separates the data and output a value between -1 and 1 whether a point is on a side of the hyperplane or on the other side. The perceptron can find a hyperplane in any n-dimensional space as long as this decision boundary exists; this happens if the data points are *linearly separable*.

Multi layer perceptron

As said, perceptrons can express only linear decision boundaries. To solve this problem, it is possible to use more perceptrons to represent more complex decision surfaces.

Multi layer perceptron (MLP) are constructed by many perceptrons. These neurons are organized in layers: there are always at least two layers, input and output layer, and one or more hidden layers; from here the term *Multi* layer perceptron. Each layer is composed by many neurons and each neuron in one layer is connected to all the neurons in the next layer, so the information from the input layer is propagated to hidden layers and then to the output layer. These model are also known as deep neural network (DNN) since the network's hidden layers make the model 'deep'. The output of a layer before being propagated to the next layer pass through a non-linear activation function, for example Rectified Linear Unit (ReLU), Sigmoid function or Tanh. The non-linearity of the activation function is needed since it introduce more complexity to the model and summing operations of many linear layers the output would still be linear, so a deep network would perform similarly to a single layer network.

The main idea behind stacking many layers is that each layer represents a boundary region, that will pass to another layer to represent a more complex boundary region. Using many layers, like a chain, it is possible to represent very complex decision boundaries. It is possible to write this sequence of operations as follows:

$$f(x) = f^{(n)}(f^{(\dots)}(f^{(1)}(x))) \tag{3.2}$$

where n is the number of hidden layer, $f^{(n)}(x)$ is the boundary representation at the last hidden layer before the output layer, $f^{(1)}(x)$ is the decision boundary representation at the first hidden layer after the input layer and $f(x)$ is the mapping function that can solve the problem. Equation 3.2 can be viewed as a chain, where the output of the first decision boundary is propagated to the next decision boundary function up to the final hidden layer and then to the output layer to get the predictions.

Convolutional neural network

Convolutional neural networks (CNNs) are a particular type of artificial neural network that process data with a grid-like topology. These kinds of networks are usually used with images, considering an image as a 2D matrix of pixels or for with time series, considering a time series as a 1D structure.

The term convolutional comes from the usage of a mathematical operation called convolution. Convolution is a linear operation that involves two functions:

$$(x * w)(t) \stackrel{def}{=} \int_{-\infty}^{\infty} x(\tau) \cdot w(t-\tau) d\tau \tag{3.3}$$

where $*$ is the sign for the convolution and $\cdot$ is the sign for the dot product.
The output of this linear operation, given by the input x and the weights w (also

called **kernel** or **filter**) is referred to as **feature map**.

The main idea is to *convolve* the input, let's take as reference an image of size n, with a filter of size f. The filter is applied to an area of the image, and the dot product between the that portion of the input image and the and the filter is computed. Then the filter is shifted to the next portion of the input image, and this way the dot product is calculated for the full width and height of the input. Since the convolution is a linear operation and since the result of multiple linear operations are still linear, the output of the convolution must go through a non-linear activation function, usually ReLU in the case of CNNs.

The output of a convolutional layer is passed to a pooling function. This pooling function aggregates the output of the convolutional layer at a certain location with a value that represents statistically their values. This allows to reduce the data dimensionality, to shorten the training time and reduce overfitting.
Usually, the most common pooling function are max pooling, which takes the max value of the window, or average pooling, which averages the values of the window.

With these series of operation, a single convolutional block can extract some important feature from the input data. Generally, many convolutional blocks are stacked together so that each of the next block can represent more and more complex and specific features.
The output of the last block is flatten and passed to one or more fully connected layers to get the final prediction (in case of a classification task).

CNNs have shown to perform very well on some task, especially image classification, and their main advantages are lower number of weights compared with a MLP network and the ability to automatically learn how to extract important features.

Recurrent neural network

Recurrent neural networks (RNNs) are a particular type of ANN that work with time series data or data that involves sequences. They use the output of network units at time t as the input of the other units at time $t + 1$.

Generally RNNs are represented as follows:

image of a rnn x -> rnn -> y

where x is the input RNN is the recurrent network and y is the output of the model. The arrow coming out and back in in the RNN block is what the *recurrent* refers to: after receiving an input, at time x the RNN computes some operation and an hidden state, h_t, is saved and used for the next input, at time $t+1$.

Mathematically, this process can be represented with the following formula:

$$h_t = f_W(h_{t-1}, x_t) \tag{3.4}$$

where f_W is a function that takes as input the hidden state of the previous time step h_{t-1} and the input at the current time step x_t and it outputs the hidden state at the current time step h_t. At the next time step the state h_t would be passed to with the next input x_{t+1} to f_W and so on until all the input time steps are consumed.

An important thing to notice is that the function f_w depends on some weights W, and these weights are share for every time step of the computation. For example, the function f_W can be represented as:

$$f_W = tanh(W_h \cdot h_{t-1} + W_x \cdot x_t) \tag{3.5}$$

where *tanh* is the hyperbolic tangent function, W_h is the matrix of weights that multiplies the hidden state h_{t-1} and W_x is the matrix of weights that multiplies the hidden state x_t.

A popular type of RNN is the Long Short-Term Memory (LSTM) networks. They were designed to handle the long time dependency of the input. The main difference between a simple RNN and a LSTM network is the complexity of the hidden block: while in a RNN there is only a Tanh function, in a LSTM there usually is the Tanh function and as well some Sigmoid functions.
This more complex model allows the network to keep in memory (as a hidden state) or forget some information that are considered as not too relevant [19].

3.3.3 Metrics

An important part of a machine learning task is to evaluate whether a model performs well or not. There are many ways to evaluate a model, these are commonly referred to as evaluation metrics.

The evaluation metrics differs from task to task, for example there are some specific evaluation metrics for classification and there are other metrics for regression problems. Since this thesis will focus on a classification task, only the classification evaluation metrics will be presented here.

Before talking about the different evaluation metrics, some terms have to be introduced. During classification, there are four outcomes that can occur:

- **True positive (TP)**: when the result of a test tells that a subject belongs to a particular class, and it actually belongs to that class.
- **True negative (TN)**: when the result of a test tells that a subject does not belong to a particular class, and it actually does not belong to that class.
- **False positive (FP)**: when the result of a test tells that a subject belongs to a particular class, but it actually does not belong to that class.
- **False negative (FN)**: when the result of a test tells that a subject does not belong to a particular class, but it actually belongs to that class.

These different outcomes can be represented in a confusion matrix.

add confusion matrix image

When a test is wrong (either FP or FN) a misclassification occurs. The evaluation metrics try to quantify how well a model performs, elaborating how many miss classification were done.

Accuracy

Accuracy measures how often the model classifies correctly. Accuracy is defined as the ratio between the number of correct classification and the total number of predictions.

$$accuracy = \frac{\text{correct predictions}}{\text{total predictions}} = \frac{TP + TN}{TP + TN + FP + FN}$$

Recall

Recall, also called sensitivity or true positive rate, gives how well the model was able to correctly classify the subject of a specific class. Accuracy is defined as the ratio between the number of true positive and the number of subject that were predicted to belong to that class.

$$recall = \frac{TP}{TP + FN}$$

Precision

Recall Precision explains how many of the correctly identified as positive out of all the subject predicted to belong to a specific class. Precision is defined as the number of true positives divided by the number of predicted positives.

$$precision = \frac{TP}{TP + FP}$$

F1 score

It is often convenient to combine recall and precision in a single evaluation metric. It is defined by the harmonic average of the recall and precision.

$$F1score = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

3.3.4 Unbalanced dataset

Chapter 4

Problem analysis

The principle of active network management ANM is to address congestion and voltage issues via short-term decision making policies [20].
ANM creates a smarter network infrastructure providing automated control of various components in the network and provides the information needed to ensure that every device performs in an optimal manner. This automated control allows grid companies to avoid reinforcing the network with expensive upgrades, so reducing the costs. For example, in case of energy generation from the renewable devices higher than what a particular line can handle, a grid company, to avoid congestions and possible overvoltages, has three main options:

- Replace the existing line with a line that can handle a higher voltage. This usually means to replace the existing line with a cable with a larger cross-sectional area.
- Add another parallel line.
- Use ANM.

The first two solution require some infrastructure investment that can be expensive and troublesome, especially in the case of overhead or underground lines.
The solution with ANM does not require construction cost for the grid company; to keep the network working, in this case, the output of the renewable devices can be curtailed to reduce lines' overloading.

In these references, generally, ANM maintains the system within operational limits by relying on the curtailment of the generator devices, PVs, WPs and other DER devices.
Curtailment of renewable energy may be seen as counter-intuitive on the environmental point of view, and it may be considered as last option. Indeed, this process

can slow down the switch to clean energy, because of the lost of the curtailed energy.

In this mindset, ANM could also be used to control flexible loads and reduce the curtailment effects. These flexible loads, also known as virtual batteries, such as water heaters, air conditioning systems, electric vehicles, can be controlled to be turned on if the energy production is higher than the energy consumption so to avoid curtailment on the generators [21].
Another way to reduce the energy curtailment is to use Flexible Alternating Current Transmission System (FACTS) devices. They offer some level of power flow control and enhance the transfer capability over the existing network. This flexibility can be utilized for congestion mitigation and renewable energy integration. Particularly, FACTS devices allow controlling all parameters that determine active and reactive power transmission: nodal voltages magnitudes and angles and line reactance. These devices replace the mechanical switches with semiconductor switches allowed much faster response times. One problem with these devices is the cost a system operator should sustain to implement them in the network [22].

4.1 Problem statement

This thesis will focus on ANM and the problem faced by a DSO to maintain the network within its operational limits. In particular, the system operator will evaluate whether in a given moment there will be a voltage problem, so that it would be possible to proceed with some actions, like applying curtailment to generator devices, to maintain the voltage inside a safe range.

The DSO knows some information about the network:

- The network topology: the number of buses, loads and generators, the lines' length, the distance between the connected buses, and the distance between each load and generator from the external grid. Moreover, the impedance of the lines is known.
- The active and reactive power of some loads at each time step. Some values may not available mainly for two reasons: *a)* for that particular load there are no measurements at all due to privacy reasons or for the lack of sensors or *b)* communication problem so the sensor recording was lost. *remove?*
- The type of DER device connected to each bus. If the DER device is directly connected to the medium voltage grid, the active power and reactive power are considered known.

This data is used to calculate the power flow of the network and obtain more information like the voltage magnitude at each bus, lines' loading and other values.

This calculation can be performed with any power system analysis tool, for this thesis Pandapower will be used.

The DSO will consider the behaviour of the network over a set of discrete time steps $t \in \{1,2,...,T-1,T\}$ of length Δt, with, $T \in \mathbb{N}$, the last time step of the time series' horizon. A time step is considered as a snapshot of the system at one particular point in time.

The DSO can have access to some information, let's define the information domain as $\mathbf{I}$. This information can be divided in static information like the network topology, and the observation O_t collected at every time step t like the loads and generators' active and reactive power and the buses' voltage magnitude, lines' loading and other information. The information at time step I_t can be defined as:

$$I_t \in \mathbf{I}, \text{with}\, t \in \{1,2,...,T-1,T\}$$
$$I_t = (\text{static information}, O_1, O_2, ...)$$

In general power networks are not static, since the operators can modify their topology, for example changing the connections due to some incidents on the lines. In this thesis, the network topology is considered as static, so that no changes are applied on the network the during the whole time series' horizon.

The operator will predict if the system, in some given $t+n$ future time steps, will be in a critical condition. The network is considered in a critical condition if any of its elements is in an unsafe situation, for example an over voltage (or under voltage) condition. Let define $\mathbf{C}$ as the list of critical situation and C_t the critical situation at time step t, with $C_t \in \{0,1\}$.
For predicting whether the system will be in a critical situation or not, the DSO will consider the history of the system only for h preceding steps, with $h \in \mathbb{N}^+$.

Let also define the function $f_\mathcal{N}$ that given the information from a particular network can elaborate this information and output some values, that represent the forecasting of a critical situation.
It is possible to summarize all the information under the relationship:

$$\hat{\mathbf{C}} = f_\mathcal{N}(\mathbf{I}) \tag{4.1}$$

where $\hat{\mathbf{C}}$ represents the forecasted values of the system given the information $\mathbf{I}$. A single instance $\hat{C}_{t+i}$ ($\hat{C}_{t+i} \in \{0,1\}$) states if the system is critical ($\hat{C}_{t+i} = 1$) or not ($\hat{C}_{t+i} = 0$), with $i \in \{1,2,\ldots,n-1,n\}$ and $n \in \mathbb{N}^+$.

The main objective is to find a good mapping function $f_\mathcal{N}$ such that the actual critical values $\mathbf{C}$ and forecasted values $\hat{\mathbf{C}}$ are as close as possible.

4.2 Solving methodology

The goal of this thesis is to find a practical function $f_{\mathcal{N}}$. This function, that can be considered as a pipeline, mainly consist of some steps:

- Solve the partial-observability problem, generating realistic time series or filling the missing measurements for the different elements. These time series are needed to run the power flow calculation. *Remove?*

- Process the information to build a dataset, consisting of $\{\mathbf{x},\mathbf{y}\}$ couples given by the information of the network $\mathbf{I}$ and the critical states $\mathbf{C}$.

- Use the generated dataset to train a classifier that can forecast whether the network will be in a critical situation.

4.2.1 Partial-observability problem

As mentioned, the network has only a partial-observability mainly due to *a)* values not available at all or *b)* few missing values.

Missing values problems can be solved filling these measurements with some techniques: balancing the flow of the network or using generative models in the case of *a)* or using the values from the adjacent time steps, averages or more complex methods like for example interpolation in the case of *b)*.

These methods will allow complementing the network time series and having sufficient data to perform the power flow calculation.

4.2.2 Processing

Given all the information about the network, only a subset will be used to train the classifier, in particular the information given from the power flow calculation, that can be the voltage magnitude at each bus or the lines' loading.

This subset of information is divided in windows of length h for the history time series, that correspond to the inputs $\mathbf{x}$s and in outputs $\mathbf{y}$s, of length n of future time steps, that correspond to the labels.

4.2.3 Forecasting

It is common in time series forecasting problems to use artificial neural networks (ANNs) to find a solution, thanks to their capacity to learn an approximate mapping function from the input space to the output space. In this case, the ANN will take as input the information from the network, and it will output binary values, stating if there will be or not a critical situation in the network.

Given the input $\mathbf{x}$, the ANN must output some binary values stating whether the system is safe or in a critical situation at the time steps $t + n$.

This database, given by the couple of elements $\{\mathbf{x},\mathbf{y}\}$, will be used with supervised learning techniques that may extract a forecasting function in order to solve the problem.

Chapter 5

Project implementation

The solving methodology described will be applied to a specific test case.

5.1 MV Oberrhein

The network used for these experiments is the MV Oberrhein network from Pandapower. MV Oberrhein is a real distribution located at Upper Rhine (German: Oberrhein), Germany. This network is a generic 20 kV power system serviced by two 25 MVA HV/MV transformer stations. The network supplies 141 HV/MV substations and 6 MV loads through four MV feeders. The network layout is meshed, but the network is operated as a radial network with some open sectioning points, i.e., redundant lines/cables. This is common in real network: they are meshed but they operate in a radial way.
The grid also includes geographical information of lines and buses and is assembled from openly available data.

A representation of the network can is presented in 5.1. The blue dots represent the buses where load and/or generators are connected and the yellow squares represent the HV/MV substations.

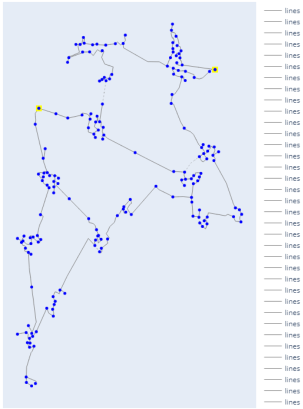

Figure 5.1: MV Oberrhein network. *add feeders division(?)*

To simplify the situation, the network can be divided in 2 independent parts [ref].

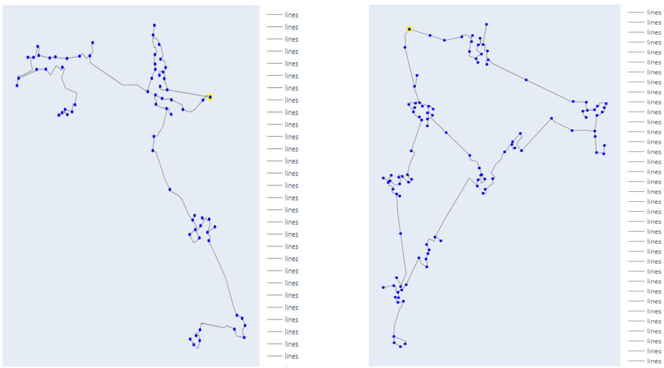

Figure 5.2: MV Oberrhein network. Used network: middle one
TODO add letters

The model consist of: one external grid, one transformer, 70 buses, 61 loads and 60 renewable generators.

5.1.1 Simbench database

The time series dataset used is taken from the Simbench database. This database refers to some real distribution networks in Germany in the year 2016. SimBench includes multiple time series for one year with 15 min resolution for load, generation and storage units. All time series came as active and reactive power. The time series were grouped by element type, reducing the total number of required time series to a reasonable number, while retaining the possibility to model individual nominal power [23]. All active power values are normalized to the maximum active power value.

Power utilities commonly use generic load profiles to group commercial customers with similar load shapes into categories or standard load profiles (SLPs). The most commonly used profiles set is developed by the German Association of Energy and Water Industries (BDEW). It comprises eleven aggregated profiles, one for residential consumers (H0), three for agricultural (L0-L3), and seven for commercial consumers with different opening hours (G0-G6). They are differentiated into workdays, Saturdays and Sundays as well as three seasonal categories winter, summer, and transitional. The set also includes two profiles for street lightning (B0) and band load (G7).
The generation time series for photovoltaics (PVs), wind energy and biomass generated for the SimBench dataset are created using the agent-based simulation tool for optimized grid expansion planning SIMONA. SIMONA's power plant models receive real weather data of Germany from the German Weather Service in 2011 for Wind and 2012 for PV time series as input data.
For 2011 and 2012 generation data, the time axis is adjusted to 2016 by shifting days so that they correspond to the nearest weekday[24].

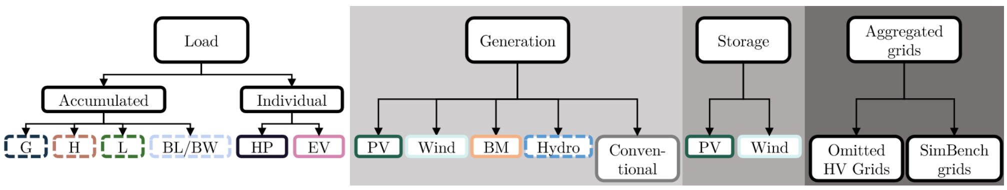

Figure 5.3: Overview of the SimBench time series type

The load time series were distinguished between real measured accumulated, highlighted with a dash, and simulated individual consumers, marked with a solid frame in Figure 5.3.

5.1.2 Time series

Some time series from the Simbench database are taken to adapt to the number of loads and DERs of the MV Oberrhein network in consideration.

As said, each element (load or generator) falls under a specific profile type that represent the consumption or generation over time.

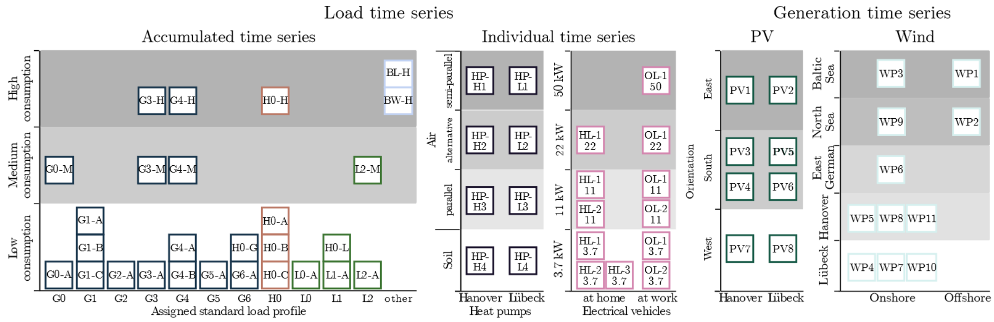

Figure 5.4: Load and generations profiles

In this case, the profile's types for loads and DER are chosen so that every profile type is present to have a network as close as possible to a real network. The profiles' distribution in the Pandapower network is as follows:

Load elements by type: {'G3-A': 2, 'H0-L': 3, 'G2-A': 5, 'G3-M': 2, 'G5-A': 2, 'L1-A': 3, 'L0-A': 2, 'H0-C': 6, 'H0-G': 3, 'G1-B': 2, 'G0-A': 4, 'L2-A': 2, 'L2-M': 3, 'G1-C': 2, 'G6-A': 3, 'G0-M': 3, 'G1-A': 3, 'G4-B': 2, 'H0-B': 3, 'G4-A': 3, 'H0-A': 3}

RES elements by type: {'WP4': 6, 'PV5': 8, 'PV8': 8, 'PV1': 6, 'PV7': 6, 'PV3': 7, 'PV6': 6, 'PV4': 7, 'WP7': 6}

---

where the letters stand for: commercial enterprises (G), households (H), agricultural holdings (L) and industrial companies (BL/BW)'; with last letters –A to C indicating low consumption, -M medium consumption, and -H high consumption customers.
For the DES device there are photovoltaics PVs and wind parks WPs. It is possible to notice a bigger presence of PVs over WPs.

The loads and DERs are chosen so that different profile types are present.

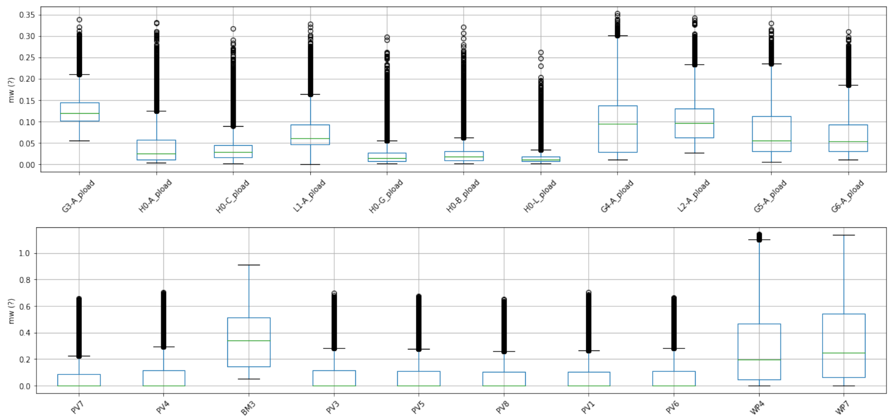

Figure 5.5: Box plots of load and generation for each profile type

Since the profiles are similar for every element of the same type, some noise is added to increase randomness. In particular, the noise added is a scaling factor in the range [0.85,1.15]. The scaling factor allows avoiding negative values in case of a value lower than 1; subtraction may result in a negative value of reactive power for a particular load.

5.1.3 Partial observability

To simulate the partial observability problem, some time series are considered as unknown: for just some sparse time steps (missing values) or not available at all.

In particular, x_1% values are randomly selected and set as *NAN* and x_2% time series of some devices are completely deleted.
Q: Any realistic value for x_1 and x_2?

These missing values are reconstructed as follows:

- for the *NAN* case, it is possible to fill the missing values with the preceding available value. This method, even if simple, is efficient for some reasons: the Δt considered is not too large, so it is plausible to expect that loads and generators' values do not change rapidly; it is easy to compute with respect to more complex methods; using only past values allows not to wait for the future values, waiting for the future values may not be acceptable after a deployment of the model.

- For the case of time series not available at all, it is possible to use balancing flow techniques or generative models like for example GAN.

Q: in this case, isn't a bit overkilling using such method when the load and generator's profile are equal/similar?

5.1.4 Processing

For this case, the information used to build the dataset is the voltage magnitude at each buses. The time series will be divided in windows of length, h for the inputs $\mathbf{x}$ and length n for the outputs $\mathbf{y}$.

It is possible to define $\mathbf{x}$ as follows:

$$\mathbf{x} = \begin{bmatrix} V^1_{t-h+1} & V^1_{t-h+2} & \cdots & V^1_{t-1} & V^1_t \\ V^2_{t-h+1} & V^2_{t-h+2} & \cdots & V^2_{t-1} & V^2_t \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ V^{n-1}_{t-h+1} & V^{n-1}_{t-h+2} & \cdots & V^{n-1}_{t-1} & V^{n-1}_t \\ V^n_{t-h+1} & V^n_{t-h+2} & \cdots & V^n_{t-1} & V^n_t \end{bmatrix} \tag{5.1}$$

where V^i_t is the voltage level V of bus i at time stamp t.

Instead, $\mathbf{y}$ can be defined as previously mentioned:

$$\mathbf{y} = [C_{t+1}, C_{t+2}, \ldots, C_{t+n-1}, C_{t+n}] \tag{5.2}$$

where C_t is the condition of the system at time stamp t, with $C_t = 1$ if the network is in a critical situation or $C_t = 0$ if it is in a normal situation.

A time step is considered critical when the voltage of at least one of the buses is out of the boundaries $V_i < v^{\min}$ (under voltage) or $V_i > v^{\max}$ (over voltage), where $v^{\min}$ is the minimum acceptable voltage, 0.95 p.u., and $v^{\max}$ is the maximum one, 1.05 p.u..

5.1.5 No partial-observability problem

As mentioned the time series were normalized so it is possible to use some scaling factors to generate different case. The high generation case refers to scaling factors, as follows:

scale_factor_load = 1
scale_factor_sgen = 1.5

These values are chosen to fit the load and generation profile with the peak capacity of the MV Oberrhein network.

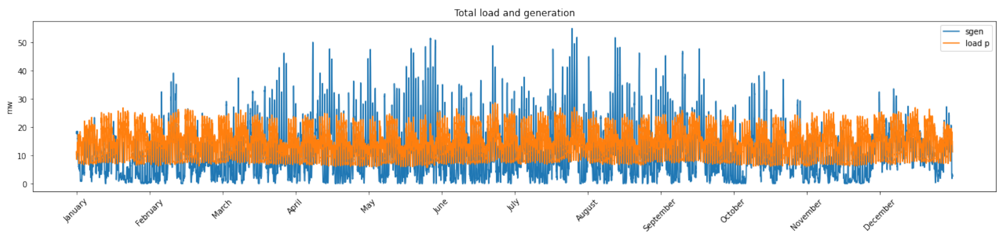

Figure 5.6: Sum of load energy consumption and energy generation over the considered year

The consumption values are as a typical MV network ([25]), while the generation is higher. This case can represent a future power network when the number and the performance of DER devices increase, so to have a generation of power higher than the demand; especially during the summer period when the consumption is lower and the generation is higher.
Such situation is critical for a network since the surplus energy increase the voltages at the buses that may experience problems or over voltages.

To test whether an over voltage situation occurs or not, the PF is calculated using the Pandapower solver.
It is possible to see that there are some overvoltage problems.
The problems are highlighted in the following plot.

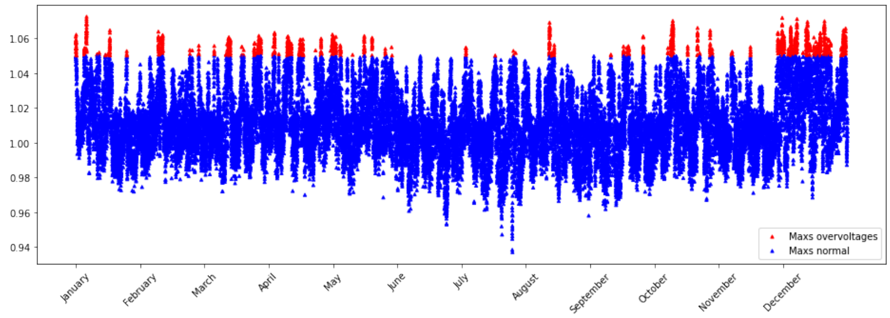

Figure 5.7: In this plot, the maximum and minimum values found for each time step and plotted, so for each time step there are two values representing the highest and lowest value among each bus. The colours, blue and red, represent the normal or critical condition

The full dataset is divided in train, validation and test set with the following proportions: 0.7, 0.2, 0.1 respectively.

The dataset is highly imbalanced since the number of times the network is in a critical situation are lower than the number of times the network is under normal situation. This is what usually happens in a power system: networks perform well most of the time and the critical conditions are few.

Number of critical situations: 1284, over 35040 time steps, ratio: 3.7% *realistic?*
Number of critical instants in Train set: 646, ratio: 2.6%
Number of critical instants in Val set: 193, ratio: 2.8%
Number of critical instants in Test set: 445, ratio: 12.7%

The data input shape is as follows:

Inputs shape (batch size, time steps, features): (512, 16, 69)

Three main ANN are tested:

- MLP with two hidden layers of size 192 and 64.
- CNN with one convolutional layer and one dense layer with 128 neurons.
- RNN with one LSTM as recurrent unit and one dense layer with 128 neurons.

All models have as last output layer a fully connected layer whose output size is just one neuron with a Sigmoid activation function.

The test are perform considering a temporal window h of 16 time steps (4 hours in the past), while the output window is just one time step in the future (*can be changed to any number*):

Labels shape (batch size, time steps): (512, 1)

input |V|, h = 16 (4 h)

	Test Accuracy	Test Precision	Test Recall
MLP	0.92	0.83	0.45
CNN	0.89	0.55	0.93
RNN	0.87	0.64	0.98

input |V|, h = 4 (1 h)

	Test Accuracy	Test Precision	Test Recall
MLP	0.94	0.72	0.93
CNN	0.90	0.56	0.99
RNN	0.95	0.76	0.94

input RES p, h = 16 (4 h)

	Test Accuracy	Test Precision	Test Recall
MLP	0.87	0	0
CNN	0.87	0.49	0.47
RNN	0.87	0	0

input RES p, h = 4 (1 h)

	Test Accuracy	Test Precision	Test Recall
MLP	0.87	0	0
CNN	0.87	0.12	0.002
RNN	0.86	0	0

Any suggestion?
Moreover, what would be an acceptable level of recall/precision/accuracy? I was thinking that since the network can handle a critical situation for some time, if the DSO would curtail some generators, these values should be >0.8 so that it makes sense to trust the model and apply some action.

Chapter 6

Conclusions and further works

Bibliography

[1] Hannah Ritchie and Max Roser. «CO and Greenhouse Gas Emissions». In: *Our World in Data* (2020). https://ourworldindata.org/co2-and-other-greenhouse-gas-emissions (cit. on p. 4).

[2] United Nations Framework Convention on Climate Change (UNFCCC). *THE PARIS AGREEMENT*. 2015 (cit. on p. 4).

[3] Max Roser Hannah Ritchie and Pablo Rosado. «CO and Greenhouse Gas Emissions». In: *Our World in Data* (2020). https://ourworldindata.org/co2-and-other-greenhouse-gas-emissions (cit. on p. 5).

[4] Hannah Ritchie and Max Roser. «Energy». In: *Our World in Data* (2020). https://ourworldindata.org/energy (cit. on p. 5).

[5] Axel Gautier, Julien Jacqmin, and Jean-Christophe Poudou. «The prosumers and the grid». In: *Journal of Regulatory Economics* 53.1 (Feb. 2018), pp. 100–126. ISSN: 1573-0468. DOI: 10.1007/s11149-018-9350-5. URL: https://doi.org/10.1007/s11149-018-9350-5 (cit. on p. 5).

[6] Robin Henry and Damien Ernst. «Gym-ANM: Reinforcement learning environments for active network management tasks in electricity distribution systems». In: *Energy and AI* 5 (Sept. 2021), p. 100092. ISSN: 2666-5468. DOI: 10.1016/j.egyai.2021.100092. URL: http://dx.doi.org/10.1016/j.egyai.2021.100092 (cit. on p. 8).

[7] Jianhong Wang, Wangkun Xu, Yunjie Gu, Wenbin Song, and Tim C. Green. *Multi-Agent Reinforcement Learning for Active Voltage Control on Power Distribution Networks*. 2022. arXiv: 2110.14300 [cs.LG] (cit. on p. 8).

[8] Elia. «Our infrastructure». In: (2022). URL: https://www.elia.be/en/infrastructure-and-projects/our-infrastructure (cit. on p. 8).

[9] A Muir and J Lopatto. *Final report on the August 14, 2003 blackout in the United States and Canada : causes and recommendations*. Apr. 2004 (cit. on p. 9).

[10] E. Lakervi and E. J. Holmes. *Electricity distribution network design*. English. IEE power series. Contribution: organisation=svt,FACT1=1. Peter Peregrinus Ltd., 1995 (cit. on pp. 10, 15).

[11] MOHAN Ned. *Power electronics: a first course*. 2012 (cit. on pp. 14, 15).

[12] JD Kueck. «Measurement Practices for Reliability and Power Quality». In: (May 2005). DOI: `10.2172/885961`. URL: `https://www.osti.gov/biblio/885961` (cit. on p. 14).

[13] IEC. «IEC TR 63213 Overview». In: (2020). URL: `https://assets.iec.ch/public/tc85/IEC%20TC85%20Supporting%20document%20on%20IEC%20TR%2063213%20measurement%20applications.pdf` (cit. on p. 15).

[14] ANSI. «American National Standard for Electric Power Systems and Equipment—Voltage Ratings (60 Hz)». In: (2020). URL: `https://webstore.ansi.org/standards/nema/ansic842016` (cit. on p. 15).

[15] Leon Thurner, Alexander Scheidler, Florian Schafer, Jan-Hendrik Menke, Julian Dollichon, Friederike Meier, Steffen Meinecke, and Martin Braun. «Pandapower—An Open-Source Python Tool for Convenient Modeling, Analysis, and Optimization of Electric Power Systems». In: *IEEE Transactions on Power Systems* 33.6 (Nov. 2018), pp. 6510–6521. ISSN: 1558-0679. DOI: `10.1109/tpwrs.2018.2829021`. URL: `http://dx.doi.org/10.1109/TPWRS.2018.2829021` (cit. on pp. 16, 17).

[16] «Timeseries Module Overview». In: (). URL: `https://pandapower.readthedocs.io/en/v2.2.2/timeseries/timeseries_loop.html` (cit. on p. 19).

[17] Tom M. Mitchell. *Machine Learning*. 1997 (cit. on p. 20).

[18] Warren S. McCulloch and Walter Pitts. «A logical calculus of the ideas immanent in nervous activity». In: *The bulletin of mathematical biophysics* 5.4 (Dec. 1943), pp. 115–133. ISSN: 1522-9602. DOI: `10.1007/BF02478259`. URL: `https://doi.org/10.1007/BF02478259` (cit. on p. 22).

[19] Sepp Hochreiter and Jürgen Schmidhuber. «Long Short-term Memory». In: *Neural computation* 9 (Dec. 1997), pp. 1735–80. DOI: `10.1162/neco.1997.9.8.1735` (cit. on p. 25).

[20] Quentin Gemine, Damien Ernst, and Bertrand Cornélusse. «Active Network Management for Electrical Distribution Systems: Problem Formulation and Benchmark». In: (May 2014) (cit. on p. 28).

[21] Quentin Gemine, Damien Ernst, and Bertrand Cornélusse. «Active network management for electrical distribution systems: problem formulation, benchmark, and approximate solution». English. In: *Optimization and Engineering* 18.3 (2016-10). ISSN: 1389-4420. DOI: `10.1007/s11081-016-9339-9`. URL: `http://dx.doi.org/10.1007/s11081-016-9339-9` (cit. on p. 29).

[22] A. Adamczyk, R. Teodorescu, R.N. Mukerjee, and P. Rodriguez. «Overview of FACTS devices for wind power plants directly connected to the transmission network». In: (2010), pp. 3742–3748. DOI: `10.1109/ISIE.2010.5637987` (cit. on p. 29).

[23] Steffen Meinecke, Džanan Sarajlić, Simon Ruben Drauz, Annika Klettke, Lars-Peter Lauven, Christian Rehtanz, Albert Moser, and Martin Braun. «SimBench—A Benchmark Dataset of Electric Power Systems to Compare Innovative Solutions Based on Power Flow Analysis». In: *Energies* 13.12 (2020). ISSN: 1996-1073. DOI: `10.3390/en13123290`. URL: `https://www.mdpi.com/1996-1073/13/12/3290` (cit. on p. 35).

[24] Christian Spalthoff, Džanan Sarajlić, Chris Kittl, Simon Drauz, Tanja Kneiske, Christian Rehtanz, and Martin Braun. «SimBench: Open source time series of power load, storage and generation for the simulation of electrical distribution grids». In: (May 2019) (cit. on p. 35).

[25] Leon Thurner. *Structural Optimizations in Strategic Medium Voltage Power System Planning*. 2018 (cit. on p. 39).