Concept Version June 27

Spark

Capacity Forecast DALI DATA

# GENERAL Introduction

Enexis is regulated regional distribution network operator (DNO) in the Netherlands, responsible for transporting electricity and gas to 2.6 million customers.

Power is delivered via grid components as cables and transformers (for electricity) to residential and industrial connections of our customers. These grid components have to have the needed capacity. This is monitored by Grid Planners and if needed transformers are swapped for heavier ones, or cable connections are strengthened.

For decennia this monitoring was done by mainly looking to historical yearly extremes. However, the increasing growth rate of power demand and supply driven by emerging technologies as electrical vehicles (EV) and photovoltaics (PV), requires shorter monitoring periods and nowadays even forecasts. Otherwise, there is no time for mitigating actions and customers will be out of power.

Fortunately, Enexis started a few years ago with the measurement of its distribution transformer population (in the project “distribution automation light” (DALI)). Power measurements are available that enable monitoring ad hoc by the grid planners.

This project takes the next step by forecasting on that data in autumn if a transformer overloads in spring.

## Heading 2

### To easily apply any text formatting you can see in this outline with just a tap, in the Home tab of the ribbon, take a look at Styles.

### For example, this paragraph uses Heading 3 style.

# Way of working

This project will follow the CRISP-DM process in combination with a scrum like components like sprint review and 3-weekly sprints. The steps of CRISP-DM are depicted below and will function as the main chapters in this report.

CRISP-DM models

Figure . The Crisp-DM model.

# business Understanding

## Business Objectives

### Background

#### [Enexis](https://www.enexisgroep.nl/over-ons/onze-organisatie/) is a regulated DNO with a monopoly position regarding the distribution of electricity and gas. This project focusses on electricity and specifically on the capacity of MV-LV transformers in the distribution grid.

#### For Enexis as a company three aspects are essential for [value creation](https://www.enexisgroep.nl/over-ons/onze-organisatie/strategie-waardecreatie/): **Reliable** and **affordable** delivery of energy and accelerating the energy transition (**sustainability**).

#### With the increase of renewable energy sources (RES) and electrification of transportation (EV) and heating (e.g. heatpumps), the power flows are more volatile than ever witnessed. Figure 2, Figure 3, Figure 4 show the increase of the forementioned technologies. Chart, line chart Description automatically generated

Figure . Solar power in The Netherlands (source: [CBS.nl](https://www.cbs.nl/nl-nl/nieuws/2020/10/productie-groene-elektriciteit-in-stroomversnelling)).

#### EVs in the Netherlands.

Figure . Battery (B), Fuel Cell (FC) and Plugg-in Hybride (PH) Electric Vehicles (EV) in The Netherlands (source: [RVO.nl](https://www.rvo.nl/sites/default/files/2021/03/Elektrisch%20Rijden%20op%20-%20de%20-%20weg%20-%20voertuigen%20en%20laadpunten%20-%20jaaroverzicht%202020.pdf)).

#### Installed power of heat pumps in The Netherlands.

Figure . Installed power of heat pumps in The Netherlands (source: [CBS.nl](https://opendata.cbs.nl/statline/#/CBS/nl/dataset/82380NED/line?dl=55480)).

#### The more volatile and more rapidly increasing power flows require that Grid Planners more often monitor if the minimum and maximum power is still withing the transformer’s capacity. Otherwise, this will cause unsafe operating situations and it can lead to power outages. Grid Planners expect a big increase of the number of transformers that reach their operating capacity in the upcoming years.

#### The more frequent monitoring can be done with the measurement data of DALI. Already 9k (out of ±35k) transformers are equipped with a so-called DALI-box to measure and transmit the 15-minute data such as voltages, currents and powers.

#### Grid Planners are under pressure to arrange that the distribution grid is not a bottleneck for the energy transition. Although there is more data available than ever, also there is more work for the Grid Planners than ever. The Grid Planners want an quick and effective way to use the available data for insight and their planning tools.

#### Currently, there are two big projects using this DALI data: The first is a Dashboard that gives insights on the data quality (completeness, gaps of missing data, etc.). The other one has a monitoring function on safety issues (e.g. left open door of substation) or reaching the capacity limit, but only on historical data. There is another project under construction to fulfill the needs of the Grid Planners, but solely on historical measurement data (no forecasting is done).

#### After a transformer is identified to have too little capacity, it can be swapped for a heavier one relatively easy (if the rest of the installation permits this). But if the rest of installation in the substation needs additional alteration as well, the lead time can be a year. Especially, in those cases a reliable forecast on measurement data would be of great value. Planning, engineering and acquisition require time especially, with lack of resources such as technical personnel. Knowing the number of transformers and which individual transformer has to be swapped, is expected to lower the costs per transformer and increases the work that can be done by our internal service provider (Infra Services) by foreseeing the upcoming work (load).

### Business Objectives

#### Primary objective:

##### More volatile power flows require a monitoring tool that forecasts transformer overloading. The goal is to preventively identify future overloading of transformers.

#### General description based on an example case

##### The project should result in a tool that is able to predict in autumn that a transformer will be overloaded in spring due to EV with a prediction interval. This enables better planning of grid strengthening which prevents overloading of transformers (safety and reliability), foreseeing future work (costs) and enabling the energy transition better (sustainability).

#### Business Value Diagram How the business value is created by the forecast model is shown by Figure 5.Business Flow Down Diagram

Figure . Business Flow Down Diagram for the project.

#### Additional objectives/requirements

##### Grid Planners don’t want another tool to log into/install..

##### The presentation of results of the tool has to be quick (no long waiting times).

##### The tool has to be scalable up to 35k transformers.

##### The tool should be reliable (working itself and forecast should be accurate).

##### Prediction intervals should be available to show the confidence of the forecast.

##### Results should also be available for other tools (e.g. dashboarding via Power BI integration with the under construction project described at 7)

##### There should be a way to sort transformers/forecasts based on their urgency.

##### The measurement/forecast data should be exportable for other planning tools (such as Vision)

##### It would be preferred if forecasts could also be made on transformers that have limited history of data with knowledge of the general population.

##### Forecast horizon should be 6-12 months.

##### The model used can be explained clearly.

#### Organization

##### The project is the graduation project of Bram Vonk for Lead Track of [JADS](https://www.jads.nl/). Besides the business side of Enexis this project will therefore also include stakeholders (regarding support and supervision) from the institute of JADS as depicted in Figure 5.

Figure . Stakeholders directly involved in the project.

##### Involved stakeholders are:

###### Grid Planners: The end users of the tool. Ad hoc, they will be updated/asked for input/their expertise on the field of grid planning and their needs. They assess if the project is a success.

###### Data Engineers: To get the tool into production the ICT guidelines and processes within Enexis have to be respected. Expertise of the Data engineers is crucial in the second half of the project to get things into production/deployment.

###### Management: The direct manager enables Developer/Lead Bram to work two days a week on this project. Together with all other Enexis stakeholders they will be updated at the end of every sprint in the sprint review session.

###### Academic supervision: Jeroen de Mast will be the academic director that monitors progress and the academic level. Every three weeks there will be an one-on-one meeting to discuss progress and issues.

###### Academic support: A PdEng student of JADS is available for support on technical and academic issues. Together with a buddy from the Lead Track periodic meetings (every four weeks) will be planned, but also ad hoc issues will be discussed directly.

### Business Success Criteria

#### The project is a success if there is a tool that is being **used** by the Grid Planners that **accurately** forecasts overloading of transformers.

##### Used 🡪 measure usage / interview Grid Planners.

##### Accurately 🡪 performance assessment on historical data.

#### The Grid Planners will assess the eventual success of the project.

## Situation Assessment

### Inventory of Resources

#### Personnel

##### Lead 🡪 Bram Vonk, available two days a week

##### Grid Planners 🡪 Dirk van den Berg, René Bosch, Wout Maas, Robin van Wijngaarden, available 30 minutes a week.

##### Data Engineers 🡪 to be requested, available 2 hours a week, second half project?

#### Data

##### DALI data available via SnowFlake database within Enexis.

#### Computer Resources

##### Macbook pro, AWS environment, available from Enexis.

#### Software

##### Licensed: AWS software, Gitlab, PyCharm, etc., available via Enexis.

##### Open: Python, PYMC3, , Stan, Docker, etc. available within Enexis.

### Requirements, Assumptions and Constraints

#### Requirements

##### Schedule The project is initially scheduled as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| week | Date | CRISP-DM Steps |  |
| 26 | June 29 | Business Understanding |  |
| 29 | July 20 | Data Understanding | **GO / NO GO** |
| Summer Break | | - | - |
| 36 | unknown | Data Preparation |  |
| 39 | unknown | Modelling |  |
| 42 | unknown | Evaluation | **GO / NO GO on production** |
| 45 | unknown | Deployment |  |

##### Deployment

###### Maintainable: The code has to have docstrings and unit tests.

###### Scalable: The tool has to be able to process 35k transformers.

###### Deployable: The tool has to be able to go in production according to the Enexis standards (Test-Acceptance-Production).

#### Assumptions

##### DALI data is available without huge quality issues during the project

##### Weekly extremes on transformer data is an acceptable aggregation level for capacity planning.

##### Data Engineering has capacity for several hours a week for support between September and December.

##### The computational burden for probabilistic models is no problem regarding the computational power within the AWS environment.

### Risks and Contingencies

#### Availability

##### Lead/developer has just become a father (is technically up to September on parental leave) and bought a house that has to be renovated. This could result in lower availability for this project. Mitigation: 🡪 None.

##### Grid Planners are immensely occupied with the current challenges in the grid. Although not a lot of time is required, it might be that other activities have higher priority than this project. Mitigation: 🡪 Be clear and direct regarding expectations and communications and limit the effort and time for this project for Grid Planners without giving in on quality/input.

##### Data Engineers are also loaded with work and might not have time/resources available. Mitigation: 🡪 It is essential to request capacity in the beginning of the project, although they will be involved only in the second half.

#### Legal

##### There is no use of sensitive data in this project regarding privacy (GDPR) or security. DALI data is allowed to be used. Credentials are not embedded in code and access to data sources is restricted by design.

### Costs and Benefits

#### Costs

##### Data collection: Data is available in an existing database. Only querying is needed, no active additional data collection.

##### Implementation: Open source software is used besides already licensed applications. Only computation power will cost additionally. More details will be available after a first proof of concept.

##### Etc.

#### To be done after more details on

## Data Mining Goals

### Goal

#### The data mining goal is: Forecasting with a prediction interval the transformer loading for all DALI enabled transformers up to 6-12 months ahead. This is a timeseries problem that requires probabilistic modelling.

### Success Criteria

#### The model forecasts prediction intervals.

#### The working of the model can be explained clearly.

#### The prediction intervals ranges are acceptable to the Grid Planners.

#### The model can use prior information of the rest of the population if historic measurements are missing.

#### The computational burden is acceptable.

## Project Plan

### Plan to achieve data mining goals. Steps:

#### Create aggregated data collection:

##### Input:

###### Queried data from the Snowflake database

###### Aggregated extremes per DALI Box, transformer phase, week.

##### Output:

###### Saved parquet on S3, partitioned by DALI Box, transformer phase.

#### Initial probabilistic model for proof of concept:

##### Input:

###### Data of ten DALI boxes with long history.

##### Output:

###### Initial probabilistic model with prediction interval.

#### Addition of population prior to probabilistic model:

##### Input:

###### Data of five DALI boxes with long history.

###### Data of five DALI boxes with short history.

##### Output:

###### Base model with prior.

#### Visualization of results:

##### Input:

###### Base model

###### Data of ten transformers.

##### Output:

###### Web interface for displaying model results.

#### Scale up:

##### Input:

###### Base model

###### Data all DALI Boxes

###### Visualization

##### Output:

###### Deployment ready code.

#### Deployment:

##### Input:

###### All data, models and code

##### Output:

###### Tested, Accepted code into Production.

### Initial Assessment of Tools and Techniques.

#### To be worked out in detail:

#### The grid planners want to know how likely / uncertain the forecast is. This can be done by displaying prediction intervals such as often used in weather forecasts as show in Figure 6. Weather forecast with a prediction interval.

Figure . Prediction interval used for a weather forecast (source: [KNMI.nl](https://www.knmi.nl/nederland-nu/weer/waarschuwingen-en-verwachtingen/weer-en-klimaatpluim)).

#### To be able to display the uncertainty of the predictions with a prediction interval a probabilistic model will be used (with PYMC3, STAN, or Edward via Python).

#### To explain the model intuitively a model that is based on timeseries decomposition can be used in a generalized additive model (GAM). Such a model enables the decomposition of daily, weekly or yearly seasonality and trends (see Figure 7), which helps to explain the model to the general public. Chart Description automatically generated

Figure . Decomposition of a timeseries (top) into, respectively, a trend, yearly, and weekly seasonality (source: [RitchieVink.com](https://www.ritchievink.com/blog/2018/10/09/build-facebooks-prophet-in-pymc3-bayesian-time-series-analyis-with-generalized-additive-models/)).

#### Even if transformers are missing a long history of measurements a forecast would be valuable. With an [Bayesian modelling approach](https://minimizeregret.com/post/2019/04/16/modeling-short-time-series-with-prior-knowledge/) it might be possible to use prior knowledge on seasonality from the rest of the population to make accurate forecasts as depicted in Figure 8. Forecasting with prior based on other data.

Figure . Forecast with seasonality based on other data (source: [minimizeregret.com](https://minimizeregret.com/post/2019/04/16/modeling-short-time-series-with-prior-knowledge/)).

# Data Understanding

## Initial Data Collection

### Data Sources Enexis will is transferring all data into [Snowflake](https://www.snowflake.com/). This includes DALI measurement and meta data and that is already available to query. Two tables will be used for acquiring data:

#### DALI measurement data:

##### Description: 15 minute data (extremes, averages) on voltages, currents and powers per phase and feeder per DALI Box.

##### Documentation:

|  |  |  |  |
| --- | --- | --- | --- |
| **Database** | DB\_CDWH\_P | | |
| **Schema** | CDWH\_4\_BDM | | |
| **Table** | BDM\_DALI\_METINGEN | | |
| **Fields  of interest** | **Name** | **Format** | **Example** |
| BOXID | VARCHAR | ESD.000240-2 |
| CHANNELID | VARCHAR | register://electricity/0/activepower/l1?avg=15 |
| WAARDE | DOUBLE | -3.408062 |
| DATUMTIJD | TIMESTAMPTZ | 2021-05-12 07:45:00.000000000 |

#### DALI metadata:

##### Description: Metadata of the DALI Box such as initial installation date and transformer limits.

##### Documentation:

|  |  |  |  |
| --- | --- | --- | --- |
| **Database** | DB\_CDWH\_P | | |
| **Schema** | CDWH\_5\_INF\_MEETDATA\_E View: DIM\_DALI\_BOX\_VW | | |
| **View** | DIM\_DALI\_BOX\_VW | | |
| **Fields  of interest** | **Name** | **Format** | **Example** |
| BOXID | VARCHAR | ESD.000240-2 |
| BOX\_BEDRIJFSSTATUS | VARCHAR | in bedrijf |
| IN\_BEDRIJFSNAME\_DATUMTIJD | TIMESTAMPTZ | <null> |
| STATIONS\_FUNCTIEPLAATS\_ID | VARCHAR | MS-116855 |
| VESTIGING | VARCHAR | Breda |
| VERMOGEN\_NOMINAAL | NUMBER | 400 |

### Quick Analysis

#### Completeness A quick completeness analysis on present 15 minute values per week (4\*24\*7=772 🡪 100% completeness) was done on DALI Boxes of region Breda. In Figure 9 one can see the results, where on the horizontal axis the weeks are depicted and on the vertical axis the different DALI BOXES ordered by first available data. Completeness of DALI data in region Breda.

Figure . Completeness of DALI data per DALI Box and week of the region Breda.

#### There are clearly issues in February 2021 and later in May as well. This has been brought under attention of the responsible collogues and it can be expected to be solved in the following months.

#### Overlap between data and metadata

#### At the moment of writing the DALI measurements consist of 10,992 IDs. Only 10,092 IDs have also metadata (meaning that nominal power is present and the transformer/measurement box is in operation).

# Data Preparation

#### 