Towards ALDI++

Protocols of the results of the weekly meetings

# 1st Weekly (March 8th)

Matias and I discussed his implementation work on ALDI today. As part of this, we identified three preliminary ALDI adaptations that we want to test:

* Adding a pre-processing step for ALDI (basically to filter long periods of 0s).
* Building an ensemble of discord detectors
* Classifying “too good” motifs as discords
* Look further back in time when calculating MP values

# 2nd Weekly (March 15th)

**What was done last week:**

* Framework for running experiments with ALDI or potential enhancements to ALDI
* Outsourcing of functionalities (data import, transformation, and metrics) into own classes for better scalability and experimentation.
* Adaptation of ALDI by Matias - Better handling of zero counts
* Enabling ALDI to run on GPUs (computation time reduced by half). A better way of computation performance comparison with other DL methods that mainly use GPU.

**Discussion and ToDos for the future:**

* Possible ALDI adaptations
  + Adding a pre-processing step for ALDI: Could long streaks of 0s be treated as basically missing values? The implementation of Matrix Profile gives different results if values are treated as 0 or as NaN: long streaks of 0 could affect discord or be treated as motifs, NaNs are ignored.
  + Building an ensemble of discord detectors: One at the building level, one at the site level, one at the building type level, etc.
  + Classifying “too good” motifs as discords
  + Look further back in time when calculating MP values
  + Introduction of a threshold value: Minimum number of buildings, within a site, to perform the KS test. E.g.,: What happens if none but one single building has data for a specific period of time?
  + Identify discords also at the building level (not only within the site): How to combine both discords? KS test (site-wise) LOGIC OR discord in individual building MP time series. That way, the date becomes a discord for the building if at least one discord is detected.
  + Find alternatives for the KS test
* Extending the framework for experiments to apply to all sites: Run original ALDI across all ASHRAE GEP datasets and have accuracy metrics (ground truth = rank1 team labeled file) as a baseline. Any further ALDI improvements will be benchmarked against this.

# 

# 3rd Weekly (March 22th)

**What was done last week:**

* ALDI adaptions:
  + Application IQR-Filtering
    - Procedure: Filter the p's that are lower than p\_value (fixed at 0.01). Calculate the 25% quantile (iqr25) on this subset. Use this iqr25 to determine if a day is a discord day or not.
    - Result: Accuracy improvements, notable reduction of false discord days
  + Application of the KS test at the building level
    - Procedure: Adjustment of the parameters for the KS test - KS(a, b). a = MP-value at day d and building bdg; b= all MP-values at day d
    - Result: Accuracy improvements
  + Zero count adaption
    - Procedure: Replace 0-consumption with NaN. (It is assumed that 0 consumption is not a real measurement, at least for energy meters).
    - Result: No improvement? -.-
* Framework implementation:
  + Introduction of a new export function in ALDI, which makes comparison with the basic truth easier
  + Introduction of further evaluation metrics
  + Comparison options across different ALDI implementations
  + Experiment framework to test an ALDI implementation on all sites

**Discussion and ToDos for the future:**

* Preliminary remark:
  + In the course of this week, we want to primarily deal with the search for suitable alternatives for the KS test
  + Baseline calculations of ALDI-vanilla and ALDI-IQR performance on all meters with the rank1 ground truth file (Kaggle winning team)
* Possible ALDI adaptations
  + Find alternatives for the KS test
  + Building an ensemble of discord detectors: One at the building level, one at the site level, one at the building type level, etc.
    - Combining the individual results in a post-processing (Soft classification, ...).
    - Combining within the ALDI building level adaption:
      * KS test (site-wise) LOGIC OR discord in individual building MP time series. That way, the date becomes a discord for the building if at least one discord is detected.
  + Classifying “too good” motifs as discords
  + Introduction of a threshold value: Minimum number of buildings, within a site, to perform the KS test. E.g.,: What happens if none but one single building has data for a specific period of time?
  + Try different periods used when calculating the MP values
* Finding a new metric that penalizes it harder when ALDI calls all data points discord

# 

# 4th Weekly (March 29th)

**What was done last week:**

* ALDI adaptions:
  + Examine how the p-values and D-values are distributed
    - The D-values are intuitively distributed. That means there is a concentration with low distance values (D-values) and (several) concentrations with higher distance values.
  + Draft possible use of HOT SAX framework (E. Keogh et.al.) in combination with MP (instead of SAX)
    - Outer heuristic (goal: get the interesting subsequences first) can be based on the MP-value. Select the subsequences in descending order according to their MP-value.
    - Inner heuristic (goal: get the subsequences first which are similar to the current subsequence) can be imitated by the MP values and the MP indices. The next subsequence is the next NN of the currently compared subsequence...
    - Caveat: This would currently be some sort of discord detector at the building level.
  + Replacement of the KS test by the Arlington-Darling test.
    - + Test of the following combinations:
      * Arlington-Darling test instead of KS
      * Arlington-Darling test + IQR-Filtering (IQR-Filtering was introduced last week)
      * Arlington-Darling test + KS test
        + (Results of the tests are linked by logical OR)
      * Arlington-Darling test + KS test + IQR Filtering
        + (Results of the tests are linked by logical OR)
    - More about the results at our monthly meeting on Thursday.
* Draft of a simple discord detector based on an IQR method computed on a sliding window. The window slides over the MP values of a corresponding time series.
  + Calculate different IQR thresholds (e.g. based on a month, a season, a year, ...) and classify all data points outside the IQR threshold as discord.
  + Caveat: This would currently be some sort of discord detector at the building level.

**Discussion and ToDos for the future:**

* Main topics covered in the course of this week
  + Baseline calculations of ALDI-vanilla and ALDI-IQR performance on all meters with the rank1 ground truth file (Kaggle winning team)
  + Preparation of the results of the experiments with the Arlington-Darling test
  + Implementation of a Discord detector analyzing the distribution of D-values with a GMM. Different approaches:
    - The first n left-hand (low D values) components of the GMM are not discords
    - The first m right-side (high D values) components of the GMM are discords
    - A way of soft classification of MP-values based on the density of D-values
      * This approach seems to make sense when 2 (or more) gaussians are not overlapping significantly (question remains on how to programatically find this, perhaps gaussians with similar SD). When only one big gaussian is found, it’s either all outliers or all normal days.
  + Some minor tasks (KS-test & supremum, p-value calculation, possible application of Mahalanobis distance)
* Possible ALDI adaptations (carried over from previous weeks)
  + Find alternatives for the KS test (ongoing)
  + Building an ensemble of discord detectors: One at the building level, one at the site level, one at the building type level, etc.
    - Combining the individual results in a post-processing (Soft classification, ...).
    - Combining within the ALDI building level adaption:
      * KS test (site-wise) LOGIC OR discord in individual building MP time series. That way, the date becomes a discord for the building if at least one discord is detected.
  + Classifying “too good” motifs as discords
  + Introduction of a threshold value: Minimum number of buildings, within a site, to perform the KS test. E.g.,: What happens if none but one single building has data for a specific period of time?
  + Try different periods used when calculating the MP values
* Finding a new metric that penalizes it harder when ALDI calls all data points discord

# 5th Weekly (April 5th) - *draft*

**What was done last week:**

* Baseline experiments
* IQR Experiments
* ROC (AUC) Metrics
* Discord classification with GMM components

**Discussion and ToDos for the future:**

* Some minor tasks (KS-test & supremum, p-value calculation, possible application of Mahalanobis distance)
* Possible ALDI adaptations (carried over from previous weeks)
  + Find alternatives for the KS test (ongoing)
  + Building an ensemble of discord detectors: One at the building level, one at the site level, one at the building type level, etc.
    - Combining the individual results in a post-processing (Soft classification, ...).
    - Combining within the ALDI building level adaption:
      * KS test (site-wise) LOGIC OR discord in individual building MP time series. That way, the date becomes a discord for the building if at least one discord is detected.
  + Classifying “too good” motifs as discords
  + Introduction of a threshold value: Minimum number of buildings, within a site, to perform the KS test. E.g.,: What happens if none but one single building has data for a specific period of time?
  + Try different periods used when calculating the MP values
* Finding a new metric that penalizes it harder when ALDI calls all data points discord

# 6th Weekly (April 12th)

**What was done last week:**

* Consideration of PSU (primary space use) in the calculation of discord days
  + Resulting processing pipeline:
    - (1) Calculate MP-values per building
    - (2) Divide the data (MP-values) according to their PSU
    - (3) Running the KS-test per PSU
    - (4) Training of GMMs on the D- or p-values per PSU
    - (5) Determine the discords per PSU (based on the D- or p-values)
* Running structured experiments with four ALDI adaptations
  + ALDI + GMM + absolute determination of the GMM components emitting a Discord
  + ALDI + GMM + relative determination of the GMM components emitting a Discord
  + ALDI + GMM + Consideration PSU + absolute determination of the GMM components emitting a Discord
  + ALDI + GMM + Consideration PSU + absolute determination of the GMM components emitting a Discord
  + Current parameters:
    - Maximum number of GMM components that can be used when training it
    - Number (absolute) or share (relative) of GMM components that do not emit discords
  + Interpretations of results are difficult: diffuse metrics and lack of comparison to vanilla ALDI (to be made)
* Preparation of rank-1 solution from ASHRAE competition to test “new discord detectors”

**Setting of the areas of responsibility within the scope of the master thesis and the paper:**

* We have discussed the scope for Matias and Till again
  + Till will lead the development of the new discord detector (ALDI++). Concurrently he will write his master thesis about it. In connection with this, a comparison with other discord detectors is planned, as well as the effect of ALDI++ on the forecast result.
  + Matias will supervise this master thesis. Following this, Matias will take over the process of preparing the results for a scientific publication.

**Discussion and ToDos for the future:**

* Restructuring of the approach with the consideration of PSUs:
  + New processing pipeline:
    - (1) Calculate MP-values per building
    - (2) Divide the data (MP-values) according to their PSU
    - (3) Running the KS-test per PSU
    - (4) *Merge the values into one data set*
    - (5) Training of one GMM on all D- or p-values
    - (6) Determine the discords (based on the D or p values)
* Combination of different discord detectors (soft (probability, depending on possibility) or hard (logic or))
* Test an automatic way to determine the share of GMM components that do not emit discords:
  + share\_non\_discord\_comp = 1 - max( all\_gaussian\_comp.means() )
  + Background:
    - There exists a distant GMM component (distance is close to 1) - assumption: the GMM components do not cluster at the low values, but are rather distributed - then we want to have only a small number of non-discord GMM components.
    - Reverse: The most distant GMM component has a small distance value (close to 0) - assumption: the GMM components cluster at the low distance values - then we want to have a large number of non-discord GMM components.
* Possible ALDI adaptations (carried over from previous weeks)
  + Find alternatives for the KS test (ongoing)
  + Building an ensemble of discord detectors: One at the building level, one at the site level, one at the building type level, etc.
    - Combining the individual results in a post-processing (Soft classification, ...).
    - Combining within the ALDI building level adaption:
      * KS test (site-wise) LOGIC OR discord in individual building MP time series. That way, the date becomes a discord for the building if at least one discord is detected.
  + Classifying “too good” motifs as discords
  + Introduction of a threshold value: Minimum number of buildings, within a site, to perform the KS test. E.g.: What happens if none but one single building has data for a specific period of time?
  + Try different periods used when calculating the MP values

# 7th Weekly (April 19th)

**What was done last week:**

* Running two structured experiments with PSU consideration and a modified processing pipeline:
  + Modified processing pipeline (working title is “PSU second”):
    - (1) Calculate MP-values per building
    - (2) Divide the data (MP-values) according to their PSU
    - (3) Running the KS-test per PSU
    - (4) *Merge the values into one data set*
    - (5) Training of one GMM on all D- or p-values
    - (6) Determine the discords (based on the D or p values)
  + Experiments:
    - ALDI + GMM + PSU second + absolute determination of the GMM components emitting a Discord
    - ALDI + GMM + PSU second + relative determination of the GMM components emitting a Discord
  + Results: Interpreting the results from the experiments is difficult. The only thing that can be deduced is that we are always slightly better than the classic ALDI algorithm. We hope to achieve better interpretations by applying the forecasting algorithms from the ASHRAE competition.
* Running a structured experiment with an ALDI adaptation
  + Experiment:
    - ALDI + GMM + dynamic weights
      * The number of components that do not emit discords is calculated dynamically using the following formula
      * share\_non\_discord\_comp = 1 - max(all\_gaussian\_comp.means())
  + Results:
    - Interpreting the results from the experiments is difficult. The only thing that can be deduced is that we are always slightly better than the classic ALDI algorithm. We hope to achieve better interpretations by applying the forecasting algorithms from the ASHRAE competition.
* Ongoing tasks
  + Preparation of rank-1 solution from ASHRAE competition to test “new discord detectors”
  + Preparation of a suitable export of the discord labels out of ALDI

**Discussion and ToDos for the future:**

* New ideas to take into account the different PSUs:
  + New processing pipeline (third) to take into account the effects of the weekdays:
    - (1) Calculate MP-values per building
    - (2) Divide the data (MP-values) according to their PSU
    - (3) Running the KS-test per PSU
    - (4) Merge the values into one data set
    - (5) *Training of one GMM per weekday on all D- or p-values*
    - (6) Determine the discords (based on the D or p values)
  + Aggregation of all data from a single PSU across all (or multiple) sites in the ASHRAE dataset. The background of this is to increase the amount of data that is used to calculate the GMM.
    - Note: It May be necessary to convert the timestamps
* Evaluation of the results of ALDI and all ALDI adaptations against the Excel file with the “special days” or vacations
* Possible ALDI adaptations (carried over from previous weeks)
  + Combination of different discord detectors (soft (probability, depending on possibility) or hard (logic or))
  + Find alternatives for the KS test (ongoing)
  + Building an ensemble of discord detectors: One at the building level, one at the site level, one at the building type level, etc.
    - Combining the individual results in a post-processing (Soft classification, ...).
    - Combining within the ALDI building level adaption:
      * KS test (site-wise) LOGIC OR discord in individual building MP time series. That way, the date becomes a discord for the building if at least one discord is detected.
  + Classifying “too good” motifs as discords
  + Introduction of a threshold value: Minimum number of buildings, within a site, to perform the KS test. E.g.: What happens if none but one single building has data for a specific period of time?
  + Try different periods used when calculating the MP values

# 8th Weekly (April 26th)

**What was done last week:**

* Running one structured experiment
  + Evaluation of the results of ALDI and all ALDI adaptations against the Excel file with the “special days” or vacations
* Enable is\_discord export
* Ongoing tasks
  + Preparation of rank-1 solution from ASHRAE competition to test “new discord detectors”

**Discussion and ToDos for the future:**

* New ideas to take into account the different PSUs:
  + New processing pipeline (third) to take into account the effects of the weekdays:
    - (1) Calculate MP-values per building
    - (2) Divide the data (MP-values) according to their PSU
    - (3) Running the KS-test per PSU
    - (4) Merge the values into one data set
    - (5) *Training of one GMM per weekday on all D- or p-values*
    - (6) Determine the discords (based on the D or p values)
  + Aggregation of all data from a single PSU across all (or multiple) sites in the ASHRAE dataset. The background of this is to increase the amount of data that is used to calculate the GMM.
    - Note: It May be necessary to convert the timestamps
* Possible ALDI adaptations (carried over from previous weeks)
  + Combination of different discord detectors (soft (probability, depending on possibility) or hard (logic or))
  + Find alternatives for the KS test (ongoing)
  + Building an ensemble of discord detectors: One at the building level, one at the site level, one at the building type level, etc.
    - Combining the individual results in a post-processing (Soft classification, ...).
    - Combining within the ALDI building level adaption:
      * KS test (site-wise) LOGIC OR discord in individual building MP time series. That way, the date becomes a discord for the building if at least one discord is detected.
  + Classifying “too good” motifs as discords
  + Introduction of a threshold value: Minimum number of buildings, within a site, to perform the KS test. E.g.: What happens if none but one single building has data for a specific period of time?
  + Try different periods used when calculating the MP values

# 9th Weekly (May 3rd)

**What was done last week:**

* Implementation of a first discord detector on building level based on IQR filtering. Essentially, all data points outside the IQR range are outliers or discords. With the possibility of using a sliding window method. The motivation is to combine this building-level discord detector with a site-level discord detector in the next step.
  + Setting options, parameters
    - Modification of the IQR range
      * lower\_bound = (Q1 - scaling\_IQR \* IQR)
      * upper\_bound = (Q3 + scaling\_IQR \* IQR)
    - Sliding Window parameter:
      * Window size
      * Step size
    - Merge function in case of overlap (default: logic or)
* Ongoing tasks
  + Preparation of rank-1 solution from ASHRAE competition to test “new discord detectors”

**Discussion and ToDos for the future:**

* Combination of existing ALDI adaptations (site level) with a discord detector at building level (e.g. the IQR discord detector)
* Possible ALDI adaptations (carried over from previous weeks)
  + New ideas to take into account the different PSUs:
    - New processing pipeline (third) to take into account the effects of the weekdays:
      * (1) Calculate MP-values per building
      * (2) Divide the data (MP-values) according to their PSU
      * (3) Running the KS-test per PSU
      * (4) Merge the values into one data set
      * (5) *Training of one GMM per weekday on all D- or p-values*
      * (6) Determine the discords (based on the D or p values)
    - Aggregation of all data from a single PSU across all (or multiple) sites in the ASHRAE dataset. The background of this is to increase the amount of data that is used to calculate the GMM.
      * Note: It May be necessary to convert the timestamps
  + Combination of different discord detectors (soft (probability, depending on the possibility) or hard (logic or))
  + Find alternatives for the KS test (ongoing)
  + Building an ensemble of discord detectors: One at the building level, one at the site level, one at the building type level, etc.
    - Combining the individual results in a post-processing (Soft classification, ...).
    - Combining within the ALDI building level adaption:
      * KS test (site-wise) LOGIC OR discord in individual building MP time series. That way, the date becomes a discord for the building if at least one discord is detected.
  + Classifying “too good” motifs as discords
  + Introduction of a threshold value: Minimum number of buildings, within a site, to perform the KS test. E.g.: What happens if none but one single building has data for a specific period of time?
  + Try different periods used when calculating the MP values

# 10th Weekly (May 10th)

**What was done last week:**

* Preparation of rank-1 solution from ASHRAE competition to test “new discord detectors” (ongoing)
* Function to create plots visualizing the distribution of labeled D-values at a building level
* Implementation of a new pipeline to account for PSU in ALDI+GMM adaptation:
  + - (1) Calculate MP-values per building
    - (2) Divide the data (MP-values) according to their PSU
    - (3) Running the KS-test per PSU
    - (4) Merge the values into one data set
    - (5) *Training of one GMM per weekday on all D- or p-values*
    - (6) Determine the discords (based on the D or p values)
* Function to export the 'is\_discord' column for subsequent utilization in the prediction algorithm (rank-1 solution from ASHRAE competition)

**Discussion and ToDos for the future:**

* New feature to create plots visualizing the distribution of all labeled D-values at a database level (across all sites)
* Adjusted approaches to compare the results from ALDI (and the previous adaptations) with the true labels (labels from the rank-1 team). Concerns the handling of the different resolutions (temporal as well as spatial).
  + Approach 1:
    - Temporal adjustment:

Adjust ALDI so that hourly discord labels can now be calculated. To do this, the modified process takes into account all MP-values (not just midnight values) and performs KS-tests according to the following scheme:

KS( MP(current monday 01:00 am), MP(all mondays 01:00 am) ),  
KS( MP(current monday 02:00 am), MP(all mondays 02:00 am) ),  
etc..

* + - Spatial adjustments:  
      ALDI provides discord labels at the site level, while the true labels are at a building level. For alignment, the labels calculated by ALDI (at the site level) are copied for each building on the site.
  + Approach 2 (quick and easy solution):
    - Temporal adjustment::  
      Calculate on the true labels (labels from the rank-1 team) some kind of majority decision. E.g. if more than 12 hours of the day are labeled as discord, the day is considered a discord day. Other thresholds are possible.
    - Spatial adjustment:  
      see approach 1 above
* Adaptation of ALDI to apply a supervised learning procedure with the discord labels (labels from the rank-1 team). The goal should be to determine a parameter (e.g. p-value or D-value) appropriately.
* Possible ALDI adaptations (carried over from previous weeks)
  + Combination of existing ALDI adaptations (site level) with a discord detector at building level (e.g. the IQR discord detector)
  + New ideas to take into account the different PSUs:
    - Aggregation of all data from a single PSU across all (or multiple) sites in the ASHRAE dataset. The background of this is to increase the amount of data that is used to calculate the GMM.
      * Note: It May be necessary to convert the timestamps
  + Combination of different discord detectors (soft (probability, depending on the possibility) or hard (logic or))
  + Find alternatives for the KS test (ongoing)
  + Building an ensemble of discord detectors: One at the building level, one at the site level, one at the building type level, etc.
    - Combining the individual results in a post-processing (Soft classification, ...).
    - Combining within the ALDI building level adaption:
      * KS test (site-wise) LOGIC OR discord in individual building MP time series. That way, the date becomes a discord for the building if at least one discord is detected.
  + Classifying “too good” motifs as discords
  + Introduction of a threshold value: Minimum number of buildings, within a site, to perform the KS test. E.g.: What happens if none but one single building has data for a specific period of time?
  + Try different periods used when calculating the MP values

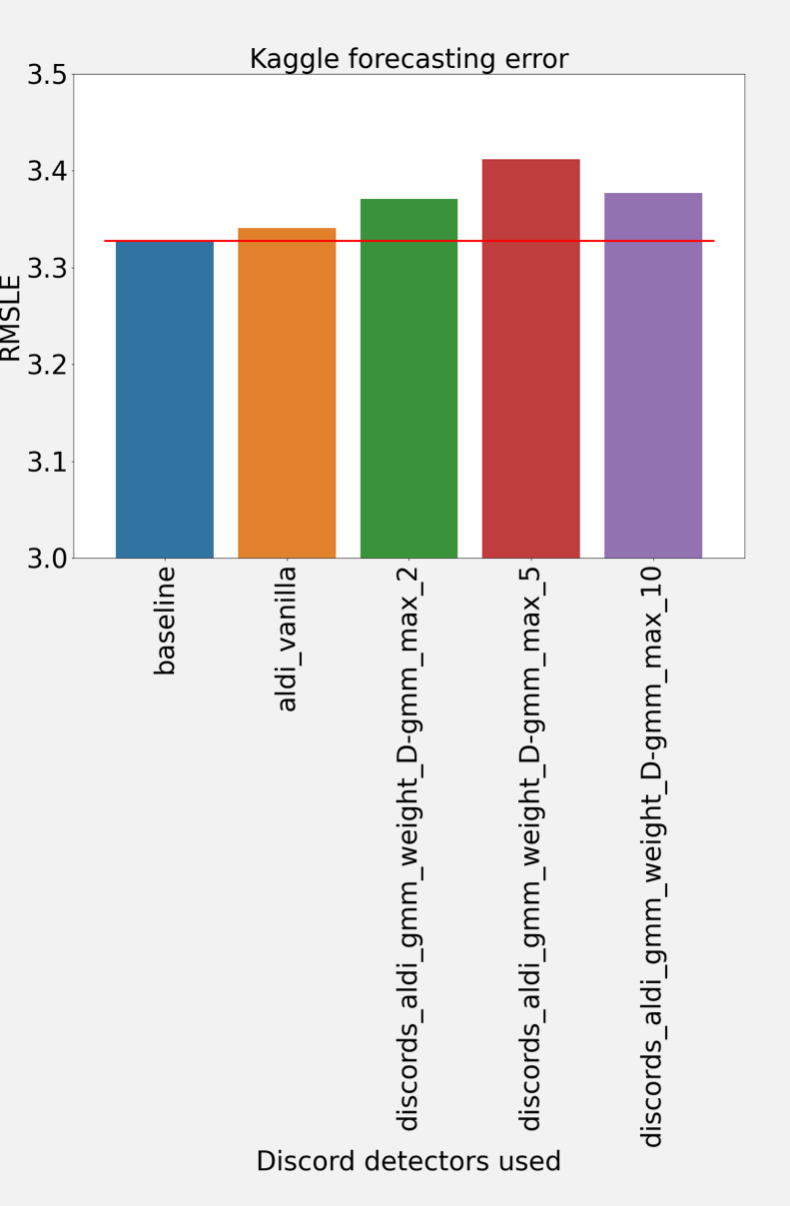
**Future work/ ideas**

* Adaptation of ALDI to apply a supervised learning procedure with the discord labels (labels from the rank-1 team). The goal should be to determine a parameter (e.g. p-value or D-value) appropriately.
* Development of a discord detector (detached from the ALDI methodology) that applies a method from unsupervised or supervised learning to detect discords.
  + Compute clusters on the energy consumption data. Determining which clusters are discords. Assign new energy consumption data to one of the clusters and estimate whether it is a discord or not.

# 11th Weekly (May 17th)

**What was done last week:**

* The implementation of the winning team's prediction algorithm has been completed
  + First experiments were performed with the 'is\_discord' labels from:
    - the "original" (manual labels from the winning team)
    - the labels of vanilla ALDI
    - the labels of ALDI + weighted GMMs + D-value (three different parameterizations)
  + The prediction error of all five experiments is shown in the plot below. Prediction error (RMSLE) was calculated using the Kaggle competition website.



* Running a structured experiments with PSU consideration and a modified processing pipeline (third applied psu-pipeline):
  + Third PSU-pipeline
    - (1) Calculate MP-values per building
    - (2) Divide the data (MP-values) according to their PSU
    - (3) Running the KS-test per PSU
    - (4) Merge the values into one data set
    - (5) *Training of one GMM per weekday on all D- or p-values*
    - (6) Determine the discords (based on the D or p values)
  + Results: The performance is comparable to the other ALDI PSU pipelines (first and second). However, it is noticeable that rather simple GMM configurations (low number of Gaussian bells (e.g. 4)) are sufficient to achieve the best performance so far.
* Implementation of different ways to convert the DataFrame with true labels and an hourly resolution to a DataFrame with daily resolution (equivalent to vanilla ALDI resolution). Aggregation of the 24 labels of a day via:
  + Logic OR
  + Logic AND
  + Majority decision (min 12 discord hours → discord day)
  + Majority decision + noise (min 14 discord hours → discord day)
* Adaptation of vanilla ALDI and all adaptation to an optional hourly processing. (ongoing)

**Discussion and ToDos for the future:**

* Enable all ALDI adaptations to provide hourly results as well.
* Acquire the true prediction results of the energy consumptions to be able to calculate the prediction error (RMSLE) ourselves.
* Possible ALDI adaptations (carried over from previous weeks)
  + New feature to create plots visualizing the distribution of all labeled D-values at a database level (across all sites)
  + Combination of existing ALDI adaptations (site level) with a discord detector at building level (e.g. the IQR discord detector)
  + New ideas to take into account the different PSUs:
    - Aggregation of all data from a single PSU across all (or multiple) sites in the ASHRAE dataset. The background of this is to increase the amount of data that is used to calculate the GMM.
      * Note: It May be necessary to convert the timestamps
  + Combination of different discord detectors (soft (probability, depending on the possibility) or hard (logic or))
  + Find alternatives for the KS test (ongoing)
  + Building an ensemble of discord detectors: One at the building level, one at the site level, one at the building type level, etc.
    - Combining the individual results in a post-processing (Soft classification, ...).
    - Combining within the ALDI building level adaption:
      * KS test (site-wise) LOGIC OR discord in individual building MP time series. That way, the date becomes a discord for the building if at least one discord is detected.
  + Classifying “too good” motifs as discords
  + Introduction of a threshold value: Minimum number of buildings, within a site, to perform the KS test. E.g.: What happens if none but one single building has data for a specific period of time?
  + Try different periods used when calculating the MP values

**Future work/ ideas**

* Adaptation of ALDI to apply a supervised learning procedure with the discord labels (labels from the rank-1 team). The goal should be to determine a parameter (e.g. p-value or D-value) appropriately.
* Development of a discord detector (detached from the ALDI methodology) that applies a method from unsupervised or supervised learning to detect discords.
  + Compute clusters on the energy consumption data. Determining which clusters are discords. Assign new energy consumption data to one of the clusters and estimate whether it is a discord or not.

# 

# 12th Weekly (May 24th)

**What was done last week:**

* Upgrade of all ALDI adaptations, with the possibility of an hourly Discord determination.
  + This skips the existing step that filters out the "midnight MP values". MP values at all hours are used with this option (hourly processing).
  + The KS tests are calculated as follows:
    - KS(Event, Reference)
    - Event: All MP values within a site at a fixed but arbitrary time (hour) (e.g. 01.01.2016 (Mon) 01:00 am)
    - Reference: All MP values within a quarter that occur on the same day of the week and at the same time (e.g. MP values of all Mondays at 01:00 am)
* Tuning of the vanilla ALDI procedure (reduction of runtime from ~30 min to ~8 min).
* Running one structured experiment
  + vanilla ALDI + hourly processing option
  + Results:
    - The experiments with small p-values (0.001, 0.005, 0.01) provided the best discord classifications
    - Compared to the daily version from the vanilla ALDI approach, the hourly version is slightly worse (0.02 in ROC-AUC metric)
      * Note: In order to be able to compare the labels (true and predicted), the hourly labels of ground truth were combined with a logical OR.
* Implementation of a new plot function in ALDI + GMM + dyn weights. It is now possible to plot the ground truth labeled D-values in a histogram at the level of a site.
* Recalculation of the prediction error:
  + Previously, all metering types (including water, ...) were used in the calculation of the prediction and its error. However, ALDI is currently applied only to electricity data.
  + Recalculation of prediction error using only true electrical consumptions and electrical predictions.
  + Results: vanilla ALDI provides an improvement of the prediction result (compared to the predictions calculated with the manually labeled data (ground truth)). The ALDI adaptations also provide better prediction results, but perform worse compared to vanilla ALDI.

**Discussion and ToDos for the future:**

* Running multiple structured experiments: Testing all hourly ALDI variants.
* Visual inspection of the abnormalities in the plots of the D-values
* Applications of a cluster algorithm to differentiate the two classes (Discord and non-Discord).
  + Calculation of the class memberships based on:
    - the D-value
    - the p-value
    - both values (D and p values)
* Identify the time points where the ground truth predicts a discord but the D-values are very high (>0.5) (indicating a strong discord). Subsequent "correction" of the ground truth and calculation of the prediction on the adjusted ground truth.
  + If ALDI's intuition is correct (high D values are caused by a discord), then the prediction error must decrease. This is to be checked.
* Possible ALDI adaptations (carried over from previous weeks)
  + New feature to create plots visualizing the distribution of all labeled D-values at a database level (across all sites)
  + Combination of existing ALDI adaptations (site level) with a discord detector at building level (e.g. the IQR discord detector)
  + New ideas to take into account the different PSUs:
    - Aggregation of all data from a single PSU across all (or multiple) sites in the ASHRAE dataset. The background of this is to increase the amount of data that is used to calculate the GMM.
      * Note: It May be necessary to convert the timestamps
  + Combination of different discord detectors (soft (probability, depending on the possibility) or hard (logic or))
  + Find alternatives for the KS test (ongoing)
  + Building an ensemble of discord detectors: One at the building level, one at the site level, one at the building type level, etc.
    - Combining the individual results in a post-processing (Soft classification, ...).
    - Combining within the ALDI building level adaption:
      * KS test (site-wise) LOGIC OR discord in individual building MP time series. That way, the date becomes a discord for the building if at least one discord is detected.
  + Classifying “too good” motifs as discords
  + Introduction of a threshold value: Minimum number of buildings, within a site, to perform the KS test. E.g.: What happens if none but one single building has data for a specific period of time?
  + Try different periods used when calculating the MP values

**Future work/ ideas**

* Adaptation of ALDI to apply a supervised learning procedure with the discord labels (labels from the rank-1 team). The goal should be to determine a parameter (e.g. p-value or D-value) appropriately.
* Development of a discord detector (detached from the ALDI methodology) that applies a method from unsupervised or supervised learning to detect discords.
  + Compute clusters on the energy consumption data. Determining which clusters are discords. Assign new energy consumption data to one of the clusters and estimate whether it is a discord or not.

# 13th Weekly (May 31th) - draft

**What was done last week:**

* Upgrade of all ALDI adaptations, with the possibility of an hourly Discord determination.
  + This skips the existing step that filters out the "midnight MP values". MP values at all hours are used with this option (hourly processing).
  + The KS tests are calculated as follows:
    - KS(Event, Reference)
    - Event: All MP values within a site at a fixed but arbitrary time (hour) (e.g. 01.01.2016 (Mon) 01:00 am)
    - Reference: All MP values within a quarter that occur on the same day of the week and at the same time (e.g. MP values of all Mondays at 01:00 am)
* Tuning of the vanilla ALDI procedure (reduction of runtime from ~30 min to ~8 min).
* Running one structured experiment
  + vanilla ALDI + hourly processing option
  + Results:
    - The experiments with small p-values (0.001, 0.005, 0.01) provided the best discord classifications
    - Compared to the daily version from the vanilla ALDI approach, the hourly version is slightly worse (0.02 in ROC-AUC metric)
      * Note: In order to be able to compare the labels (true and predicted), the hourly labels of ground truth were combined with a logical OR.
* Implementation of a new plot function in ALDI + GMM + dyn weights. It is now possible to plot the ground truth labeled D-values in a histogram at the level of a site.
* Recalculation of the prediction error:
  + Previously, all metering types (including water, ...) were used in the calculation of the prediction and its error. However, ALDI is currently applied only to electricity data.
  + Recalculation of prediction error using only true electrical consumptions and electrical predictions.
  + Results: vanilla ALDI provides an improvement of the prediction result (compared to the predictions calculated with the manually labeled data (ground truth)). The ALDI adaptations also provide better prediction results, but perform worse compared to vanilla ALDI.

**Discussion and ToDos for the future:**

* Running multiple structured experiments: Testing all hourly ALDI variants.
* Visual inspection of the abnormalities in the plots of the D-values
* Applications of a cluster algorithm to differentiate the two classes (Discord and non-Discord).
  + Calculation of the class memberships based on:
    - the D-value
    - the p-value
    - both values (D and p values)
* Identify the time points where the ground truth predicts a discord but the D-values are very high (>0.5) (indicating a strong discord). Subsequent "correction" of the ground truth and calculation of the prediction on the adjusted ground truth.
  + If ALDI's intuition is correct (high D values are caused by a discord), then the prediction error must decrease. This is to be checked.
* Possible ALDI adaptations (carried over from previous weeks)
  + New feature to create plots visualizing the distribution of all labeled D-values at a database level (across all sites)
  + Combination of existing ALDI adaptations (site level) with a discord detector at building level (e.g. the IQR discord detector)
  + New ideas to take into account the different PSUs:
    - Aggregation of all data from a single PSU across all (or multiple) sites in the ASHRAE dataset. The background of this is to increase the amount of data that is used to calculate the GMM.
      * Note: It May be necessary to convert the timestamps
  + Combination of different discord detectors (soft (probability, depending on the possibility) or hard (logic or))
  + Find alternatives for the KS test (ongoing)
  + Building an ensemble of discord detectors: One at the building level, one at the site level, one at the building type level, etc.
    - Combining the individual results in a post-processing (Soft classification, ...).
    - Combining within the ALDI building level adaption:
      * KS test (site-wise) LOGIC OR discord in individual building MP time series. That way, the date becomes a discord for the building if at least one discord is detected.
  + Classifying “too good” motifs as discords
  + Introduction of a threshold value: Minimum number of buildings, within a site, to perform the KS test. E.g.: What happens if none but one single building has data for a specific period of time?
  + Try different periods used when calculating the MP values

**Future work/ ideas**

* Adaptation of ALDI to apply a supervised learning procedure with the discord labels (labels from the rank-1 team). The goal should be to determine a parameter (e.g. p-value or D-value) appropriately.
* Development of a discord detector (detached from the ALDI methodology) that applies a method from unsupervised or supervised learning to detect discords.
  + Compute clusters on the energy consumption data. Determining which clusters are discords. Assign new energy consumption data to one of the clusters and estimate whether it is a discord or not.

# 14th Weekly (June 7th) - draft

**What was done last week:**

* Upgrade of all ALDI adaptations, with the possibility of an hourly Discord determination.
  + This skips the existing step that filters out the "midnight MP values". MP values at all hours are used with this option (hourly processing).
  + The KS tests are calculated as follows:
    - KS(Event, Reference)
    - Event: All MP values within a site at a fixed but arbitrary time (hour) (e.g. 01.01.2016 (Mon) 01:00 am)
    - Reference: All MP values within a quarter that occur on the same day of the week and at the same time (e.g. MP values of all Mondays at 01:00 am)
* Tuning of the vanilla ALDI procedure (reduction of runtime from ~30 min to ~8 min).
* Running one structured experiment
  + vanilla ALDI + hourly processing option
  + Results:
    - The experiments with small p-values (0.001, 0.005, 0.01) provided the best discord classifications
    - Compared to the daily version from the vanilla ALDI approach, the hourly version is slightly worse (0.02 in ROC-AUC metric)
      * Note: In order to be able to compare the labels (true and predicted), the hourly labels of ground truth were combined with a logical OR.
* Implementation of a new plot function in ALDI + GMM + dyn weights. It is now possible to plot the ground truth labeled D-values in a histogram at the level of a site.
* Recalculation of the prediction error:
  + Previously, all metering types (including water, ...) were used in the calculation of the prediction and its error. However, ALDI is currently applied only to electricity data.
  + Recalculation of prediction error using only true electrical consumptions and electrical predictions.
  + Results: vanilla ALDI provides an improvement of the prediction result (compared to the predictions calculated with the manually labeled data (ground truth)). The ALDI adaptations also provide better prediction results, but perform worse compared to vanilla ALDI.

**Discussion and ToDos for the future:**

* Running multiple structured experiments: Testing all hourly ALDI variants.
* Visual inspection of the abnormalities in the plots of the D-values
* Applications of a cluster algorithm to differentiate the two classes (Discord and non-Discord).
  + Calculation of the class memberships based on:
    - the D-value
    - the p-value
    - both values (D and p values)
* Identify the time points where the ground truth predicts a discord but the D-values are very high (>0.5) (indicating a strong discord). Subsequent "correction" of the ground truth and calculation of the prediction on the adjusted ground truth.
  + If ALDI's intuition is correct (high D values are caused by a discord), then the prediction error must decrease. This is to be checked.
* Possible ALDI adaptations (carried over from previous weeks)
  + New feature to create plots visualizing the distribution of all labeled D-values at a database level (across all sites)
  + Combination of existing ALDI adaptations (site level) with a discord detector at building level (e.g. the IQR discord detector)
  + New ideas to take into account the different PSUs:
    - Aggregation of all data from a single PSU across all (or multiple) sites in the ASHRAE dataset. The background of this is to increase the amount of data that is used to calculate the GMM.
      * Note: It May be necessary to convert the timestamps
  + Combination of different discord detectors (soft (probability, depending on the possibility) or hard (logic or))
  + Find alternatives for the KS test (ongoing)
  + Building an ensemble of discord detectors: One at the building level, one at the site level, one at the building type level, etc.
    - Combining the individual results in a post-processing (Soft classification, ...).
    - Combining within the ALDI building level adaption:
      * KS test (site-wise) LOGIC OR discord in individual building MP time series. That way, the date becomes a discord for the building if at least one discord is detected.
  + Classifying “too good” motifs as discords
  + Introduction of a threshold value: Minimum number of buildings, within a site, to perform the KS test. E.g.: What happens if none but one single building has data for a specific period of time?
  + Try different periods used when calculating the MP values

**Future work/ ideas**

* Adaptation of ALDI to apply a supervised learning procedure with the discord labels (labels from the rank-1 team). The goal should be to determine a parameter (e.g. p-value or D-value) appropriately.
* Development of a discord detector (detached from the ALDI methodology) that applies a method from unsupervised or supervised learning to detect discords.
  + Compute clusters on the energy consumption data. Determining which clusters are discords. Assign new energy consumption data to one of the clusters and estimate whether it is a discord or not.

# 15th Weekly (June 21th)

**What was done last week:**

* Writing the Thesis
* Creating plots showing the D and p values together. Four plots were created per site:
  + 2D scatterplot with D and p values labelled (label for Discord & Non-Discord).
  + 2D scatterplot with D and p values labelled as Discord
  + 2D scatterplot with D and p values labelled as non-discord
  + 3D scatterplot with D-values, p-values and the labels {0,1}
* Creation of daily plots of energy consumption
  + Creation takes too much computation time.
  + Solution via a *creation script* that creates daily plots of the load profiles
    - Parameters:
      * Range of D-values
      * Range of p-values
      * Value of the true labels
      * Number of plots to be created
      * List of building IDs to be considered
      * List of sites to be considered

**Discussion and ToDos for the future:**

* Creation of the script to create the daily plots with different Query parameters.
* Implementation of a new idea: detection of missfunctional/untypical buildings:
  + Motivation: Which buildings behave permanently atypical
  + Classic run of ALDI (adaption)
    - Result: discord\_days (amount of days ALDI detected as discord days)
  + Iterate over the discord\_days:
    - Load the MP values for the corresponding discord day
    - Calculate a weight for each entry in the MP vector (measure of malfunctionality/untypicality)
      * weight = single mp value / sum over all mp values of the discord tag
  + Sum all weights per building.
  + The buildings with the highest MP values are the dysfunctional/untypical buildings
* Possible ALDI adaptations (carried over from previous weeks)
  + New feature to create plots visualizing the distribution of all labeled D-values at a database level (across all sites)
  + Combination of existing ALDI adaptations (site level) with a discord detector at building level (e.g. the IQR discord detector)
  + New ideas to take into account the different PSUs:
    - Aggregation of all data from a single PSU across all (or multiple) sites in the ASHRAE dataset. The background of this is to increase the amount of data that is used to calculate the GMM.
      * Note: It May be necessary to convert the timestamps
  + Combination of different discord detectors (soft (probability, depending on the possibility) or hard (logic or))
  + Find alternatives for the KS test (ongoing)
  + Building an ensemble of discord detectors: One at the building level, one at the site level, one at the building type level, etc.
    - Combining the individual results in a post-processing (Soft classification, ...).
    - Combining within the ALDI building level adaption:
      * KS test (site-wise) LOGIC OR discord in individual building MP time series. That way, the date becomes a discord for the building if at least one discord is detected.
  + Classifying “too good” motifs as discords
  + Introduction of a threshold value: Minimum number of buildings, within a site, to perform the KS test. E.g.: What happens if none but one single building has data for a specific period of time?
  + Try different periods used when calculating the MP values

**Future work/ ideas**

* Adaptation of ALDI to apply a supervised learning procedure with the discord labels (labels from the rank-1 team). The goal should be to determine a parameter (e.g. p-value or D-value) appropriately.
* Development of a discord detector (detached from the ALDI methodology) that applies a method from unsupervised or supervised learning to detect discords.
  + Compute clusters on the energy consumption data. Determining which clusters are discords. Assign new energy consumption data to one of the clusters and estimate whether it is a discord or not.