

# Development of a SVM Prediction Model to Optimize the Energy Consumption of Industrial Installations by Detecting and Classifying Errors at an Early Stage

M. Stul and K. Stul

Department of Electrical Engineering, KU Leuven, Belgium

Email: {maarten.stul, koen.stul}@kuleuven.be

R. Leenders and L. Butaye

Departement of Energytechnology, Odisee, Belgium

Email: {rien.leenders, liesbet.butaye}@odisee.be

**Abstract**—Considerable energy savings in industrial environment are possible in an industrial environment by detecting installations not working at their optimum operating point. The present paper proposes a new generalized data driven FDD method capable of automatically detecting the abnormal energy demand of different types of installations or machines based on process data. The paper contains a comprehensive overview of the research, focusing on a trade-off between performance and computing time together with minimizing the human input. The proposed method contains an automated feature selection, a hyper-parameter optimization of the chosen SVM regression algorithm and a residual control algorithm. The method was tested in several industrial installations and two case studies are presented to demonstrate the performance of the proposed method, while underlining the significance of a decent number of relevant features.

**Index Terms**—energy prediction, SVM-regression, FFD, energy management

## I. INTRODUCTION

In Europe 25.06% of the total energy use can be assigned to the industrial sector [1]. A study shows a considerable energy reduction is possible by implementing and applying an Integrated Control System (ICS) and/or sub-metering the system [2]. A system, not located in its optimum operating point, causes overconsumption. Buildings (and systems) frequently absorb more energy than anticipated or desired [3]. As has been described in Mavromatid clearly demonstrate that there are three ways to deal with similar problems [4]. The first and most expensive approach will take action the moment a component of the system fails. A repeatedly applied second method used in industrial environment is based on a periodic maintenance, independent of the state of the installation. The third and

last approach makes use of a program to measure the performance of an installation and automatically demands maintenance the moment a significant loss of performance has been detected. During the last decades extensive research has been conducted into developing automatic Fault Detection and Diagnostics (FDD) algorithms. A comprehensive overview of relevant FDD methods, complemented with examples can be found in the papers [5]-[7]. In the literature specific results are presented for chillers [8], buildings [9], chemical processes [10] and centrifugal Pumps [11], [12]. The present paper's purpose is to propose a generalized method to automatically monitor the energy demand of an industrial system and to detect and diagnose faults at low severity levels. The method includes data preprocessing, automatic feature selection, parameter optimisation of the model and the development of a fault detection method. Since the proposed method applicable to different processes and the input of human expertise should be reduced to the bare minimum, a data driven approach is used. LIBSVM, an open source machine learning library for support vector machine regression (SVM regression) will be used as a framework for the regression model to predict energy consumption. The working principle of SVM regression (SVR) is extensively covered in [13]. Based on the difference between the actual and the predicted values (the residuals), the method can detect and diagnose deviant energy consumption of a system or machine.

## II. METHOD

Similar to other methods, using a data driven FDD algorithm, the proposed method contains both an offline and online section represented in Fig. 1. The offline section of the method is responsible for selecting the relevant features, optimizing the hyper- parameters for the SVR algorithm and defining the confidence interval of the prediction. The online section uses the results of

the offline section for the online validation of the energy consumption of a given installation.

The regression model of LIBSVM used within the scope of this paper is based on a number of relevant features as input, and a target (the energy consumption of the plant). In the offline training phase the labelled data, a set of features available at the time  $j$ :  $X_j$  corresponds with a target  $Y_j$ , allows the SVR algorithm to determine the relations between the various features and the target. The model has to be trained, based on error-free data. Error-free data does not contain any data measured during a fault situation or inefficient operation and is a representation of the working range of the installation.

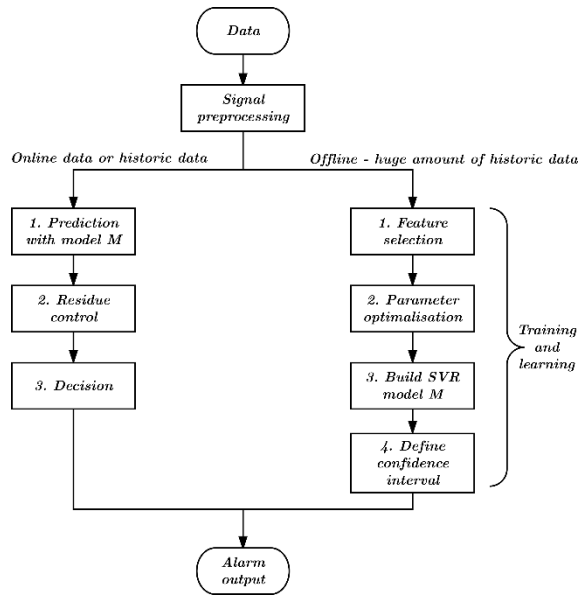


Figure 1. Proposed method

The model predicts the energy consumption at a given time. If the difference between the predicted energy consumption and the actual energy consumption exceeds a certain level, a non-optimal operation of the installation can be assumed.

#### A. Data Pre-Processing and Automatic Feature Selection

1) *Data Pre-processing*: The first stage of preprocessing checks the data for missing data points and non-realistic values. SVM algorithms assume that the variance of the features are in the same order and centered around zero. Stage two is responsible for the required scaling of the dataset, because a feature with a substantial variance will have an important influence on the model causing a negative impact on the performance. Normalization (1) scales the values of a given data set in the range  $[0,1]$  and will not be used for two reasons. First of all the feature is not centered around zero and more important outliers cause a majority of the data being located in a very small interval.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Standardization (2) on the other hand does not contain any absolute limits but uses the mean and the variation of the data set for scaling. This has the advantage that the dataset is centered around zero and the distribution of the relevant data (not the outliers) is approximately equal for any feature.

$$X' = \frac{X - \mu}{\sigma} \quad (2)$$

2) *Automatic feature selection*: As mentioned under [14], the performance of the SVR algorithm is sensitive to irrelevant or redundant features. There are multiple options to select the  $n_s$  significant features from a data set with  $n$  features. An obvious method is to investigate the performance of each possible combination of  $n$  features ( $(2^n-1)$  combinations in total). Because the number of features is not known in advance, this may lead to an extreme computation-intensive operation.

Automatic feature selection aims to retain the relevant features in a more efficient way and discard the remaining features. Feature selection can either be obtained by using the machine learning algorithm as evaluator (the wrapper approach) or by using a different evaluator (the filter approach). The computational time of the filter approach is generally smaller but results in a worse performance and has therefore been excluded from the method used.

According to [15], essentially four strategies can be applied for the wrapper approach to retrieve the  $n$  significant features in a more efficient way:

- Forward selection,
- Backward elimination,
- Stepwise regression,
- Least angle regression.

As forward selection approaches the performance of the brute-force method [14] and has a considerably smaller calculation time, it will be used as a simple but effective option.

Forward selection is a wrapped method starting with an empty subset of features. At each iteration of the algorithm a feature, not being part of the subset, is added to the subset of features. In the first iteration all  $n_s$  features are evaluated separately by training  $n_s$  SVR algorithms, each with one of the  $n_s$  features as input and the energy consumption as target. The hyper-parameters of each SVR are kept a constant (Cost= 1, kernel='RFB', gamma=1/number of features). The feature generating the most accurate prediction of the energy consumption using the SVR algorithm is added to the subset of relevant features. The accuracy of the prediction is determined by the standard deviation of the difference between the actual measured and the predicted energy consumption (residual).

$$s = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (3)$$

The second iteration compares the standard deviation of  $n-1$  SVR algorithms utilizing the feature in the

subset of features and one of the  $n - 1$  other features. The feature improving the standard deviation the most is added to the subset of features. This iterative process stops the moment the standard deviation of the residual diminishes on two successive iterations.

### B. SVR Model and Optimization

1) *Model*: The working principal of SVR will be kept minimal in this paper, as there are plenty of excellent papers on that topic [13].  $\epsilon$ -SVR is a regression technique using a training data set (4) where  $\mathbb{R}^d$  denotes the input feature space.

$$(x_1, y_1), \dots, (x_l, y_l) \subset \chi \cdot \mathbb{R}^d \quad (4)$$

The goal is to determine a function  $f(x)$  whereby the maximal deviation between the target  $y_1$  and the function  $f(x)$  amounts to  $\epsilon$  and the function  $f(x)$  itself is as flat as possible. Given that this form of hard margins is harmful for the generalization of a model (and results in overfitting), it is recommended to use soft margins. This is achieved by allowing a larger deviation than  $\epsilon$  but assigning a certain cost to this larger deviation. The cost is a trade-off between tolerance for deviations greater than  $\epsilon$  and the flatness of the function  $f(x)$ . The input features  $X$  with feature space  $\mathbb{R}^d$  will be mapped into a higher m-dimensional feature space  $\mathbb{R}^m$ . In this high-dimensional feature space the regression function  $f(x)$  will be constructed. Depending of the SVR kernel type

the function  $f(x)$  is linear (kernel: 'Linear') or non-linear (kernels: 'RBF', 'Polynomial', 'Sigmoid'). The non-linear kernels possesses additional parameters:

- $\gamma$  is a coefficient for RBF, Polynomial and Sigmoid kernels used to define the influence of a single training instance,
- Coef0 is an independent coefficient in the 'Polynomial' and 'Sigmoid' kernel functions,
- Degree is a coefficient for the Polynomial kernel and represents the degree of the polynomial kernel function.

2) *Hyper-parameter optimisation*: The  $\epsilon$ -SVR algorithm of LIBSVM contains a number of parameters such as:  $\epsilon$ , C, kernel type having specific setting parameters. For optimal performance of the model, these parameters should be set correctly. A grid search is a common solution to obtain the optimal value for each of the parameters. A grid search is a brute-force technique that optimizes the performance of the SVR algorithm by assigning a number of values ( $L^{(1)} \dots L^{(k)}$ ) to each of the P parameters (e.g. for  $\epsilon$ -SVR it would be  $\epsilon$ , cost C, kernel type and kernel parameters) and measures the performance of the SVR algorithm for every possible combination of the assigned values of the P parameters. Measuring the performance can be obtained by dividing the data set into a training and a test data set or by using cross-validation on the complete data set.

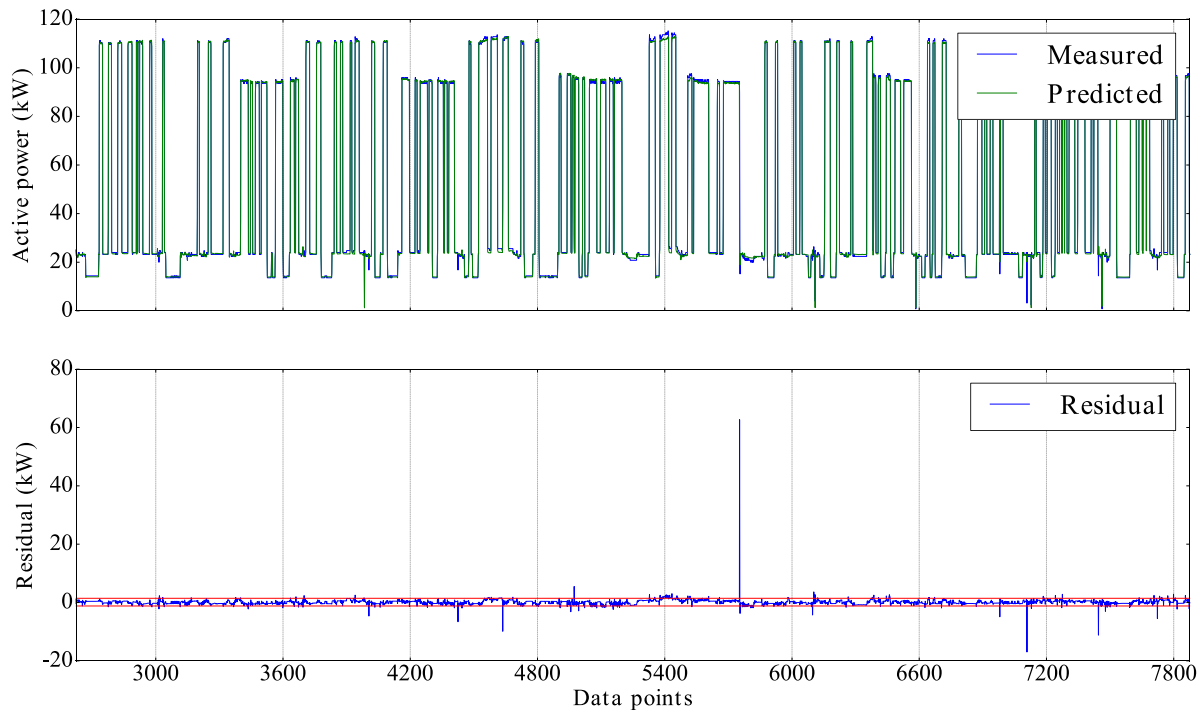


Figure 2. Measured and predicted energy consumption, case1

The grid search has to execute  $\prod_{p=1}^P |L^{(k)}|$  iterations, given  $P$  parameters and  $L^{(1)} \dots L^{(k)}$  values for a certain parameter. The amount of iterations goes up exponentially with the number of parameters and results in a huge computational time. Bergstra [16] indicates that the random search method is both faster and more

efficient than traditional grid search approach. Instead of using a fixed set of values for each parameter, random search picks random samples in a certain range for each parameter and the maximum number of iteration can be chosen instead of being dependent of  $\prod_{p=1}^P |L^{(k)}|$ . For the same or less amount of trials random search allows the

exploration of more values and has a positive impact on the performance of the algorithm when some parameters are more important.

### C. Residual Control

The predicted energy consumption, generated by the SVR algorithm is compared to the actual measured energy consumption. Since the predicted energy consumption of an installation does not correspond a 100% with the measured energy consumption, an uncertainty interval is added. As applied by [4] an uncertainty interval of 95% will be utilized. As long as the actual energy consumption is within the borders of the uncertainty interval, it can be assumed as normal. However, since the transients of a dynamic system can only be modelled with detailed physical models, see [17], it is possible that the actual energy consumption will be outside the uncertainty interval. To avoid a high false alarm rate, an alarm will be generated when 5 consecutive measurements or the measurements over a period of 30 minutes, whichever is longest, are outside the confidence bounds. This is considered by [4] as a “good compromise between false alarms and early detection”. The magnitude of the deviation between the actual and the predicted energy consumption of those measurements divided by the 95th percentile of the actual energy consumption of the training data set, will be used to distinguish the size of the error and the potential impact of the error on the system. If this deviation is less than 10% for the measurements, there is a “Very Low Probability of Failure”, a deviation between 10% and 20% corresponds to a “Low Probability of Failure”. When the difference between the actual and predicted energy consumption is between 20% and 30% a High Probability of failure can be assumed. If the difference exceeds 30%, a failure occurs.

## III. CASE STUDIES

In this section two case studies will be presented on which the proposed method is applied. The two selected case studies will show both the importance of sufficient relevant features for an efficient modelling and versatility of the method for predicting the energy consumption of an installation.

### A. Industrial Sterilization Installation

The first case study involves an industrial sterilization installation of a food company where 31 features are monitored and captured every minute. After scaling the data and removing erroneous measurements, 10 of the 31 features are selected by the automatic feature selection method. As shown in Fig. 3 the standard deviation reaches a minimum at the 10th iteration of the feature selection method. As mentioned earlier, the performance of the default model diminishes when too many irrelevant or redundant features are added to the subset of features. Fig. 4 shows the predicted and actual energy consumption of the installation using an optimized SVR algorithm by the random cross-validated search of the parameters.

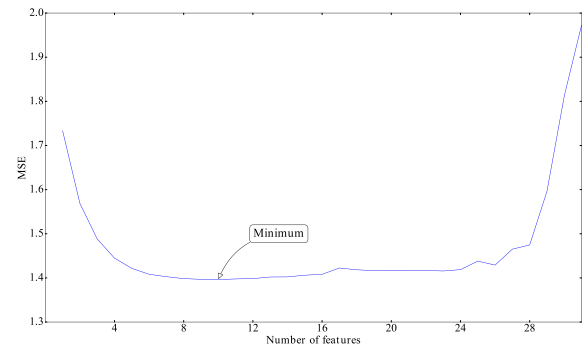


Figure 3. Automatic feature selection

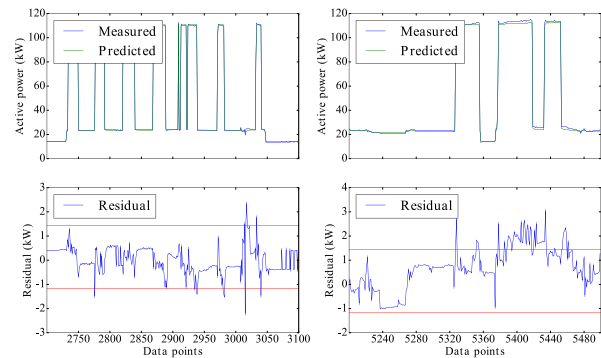


Figure 4. Normal and fault situation

Due to the dynamic behavior of the installation it is common the deviation between the actual and predicted energy consumption is temporally located outside the confidence bounds. This can be studied in Fig. 2 at measurement points 5753, 7183 and 7360. At those points no alarm is generated because five consecutive measurements need to exceed the confidence bounds. However the measurement values 5388 to 5433 in Fig. 4 show a fault situation which can be classified as a deviant energy consumption with a “Very Low Probability of Failure” (less than 10%). Because of the temporally nature of the alarm (total alarm time 45 min) and the indication “Very Low Probability of Failure” it was not possible to determine the cause of the temporally overconsumption.

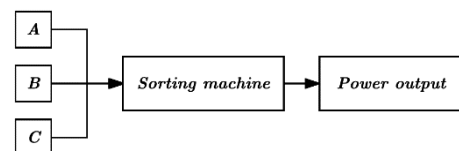


Figure 5. Shorting machine

### B. Paper Recycling Machine

The second case study is a sorting machine with three input parameters (A, B and C) and one output: the electric power consumption (Fig. 5). The automatic feature selection indicates all three input parameters appear to be correlated. The result of the training and validation can be seen in figure 6. There is no alarm detected in the dataset. It is important to notice the impact of the number of features on the performance of the proposed method and more specific on the width of the confidence bounds. In

the first case study, the width of the confidence bound equals 5% of the maximum energy demand, meanwhile in the second case study the confidence bound equals 33.7% of the paper recycling machine's maximum energy demand. It is important to have enough significant parameters to make a good prediction with an accurate model. The three available input parameters are insufficient to detect small errors in the installation. The number of false alarm can be reduced by the registration of more correlated process parameters

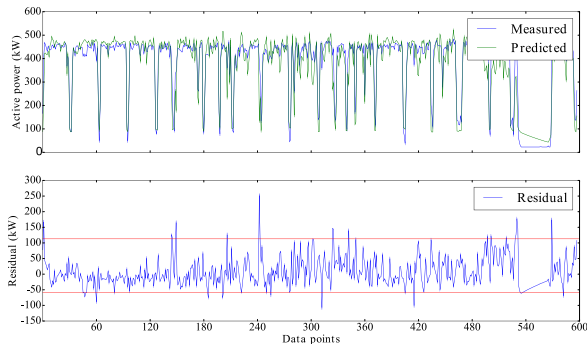


Figure 6. Measured and predicted energy consumption, case

#### IV. CONCLUSION

The development of a FDD method to detect and diagnose abnormal energy consumption was discussed. A trade-off was established between the performance of the method and the computational time. The goal is to develop a method able to implement into different types of installations with a bare minimum of human input. The proposed method contains an automated feature selection to remove redundant and irrelevant input features, a hyper-parameter optimization algorithm for the SVR model to obtain robust parameters and a residual control algorithm to detect and to classify errors based on confidence bounds. Case studies in industrial plants show the importance of the amount of relevant features to limit the confidence bounds and increase the accuracy of the method. An increase of accuracy has two main advantages. The FDD method will be able to detect faults in an early stage and will be capable to classify the faults more precisely. This general approach could be applied easily as one monitoring measure in the context of the ISO-50001 whenever material quality, ambient temperature and other specific thresholds are considered beside the process parameters of the system.

#### ACKNOWLEDGMENT

The authors would like to thank the IWT-TETRA fund and all the partners collaborating in the EnerProf project.

#### REFERENCES

- [1] A. Barbu, "Final energy consumption by sector and fuel," European Energy Agency, 2015.
- [2] Y. Chan, R. Kantamaneni, and M. Allington, "Study of energy efficiency and energy saving potential in industry and on possible policy mechanisms," Dec. 2015.
- [3] S. Treado and Y. Chen, "Saving building energy through advanced control strategie," *Energies*, vol. 6, no. 9, pp. 4769–4785, 2013.

- [4] G. Mavromatidis, S. Acha, and N. Shah, "Diagnostic tools of energy performance for supermarkets using Artificial Neural Network algorithms," *Energy and Buildings*, vol. 62, pp. 304–314, 2013.
- [5] Venkatasubramanian, R. Rengaswamy, and S. N. Kavuri, "A review of process fault detection and diagnosis. Part I: Quantitative model-based methods," *Computers and Chemical Engineering*, vol. 27, no. 3, pp. 293–311, 2003.
- [6] Venkatasubramanian, R. Rengaswamy, and S. Kavuri, "A review of process fault detection and diagnosis: Part II: Qualitative models and search strategies," *Computers & Chemical Engineering*, vol. 27, no. 3, pp. 313–326, July 13, 2009.
- [7] Venkatasubramanian *et al.*, "A review of process fault detection and diagnosis: Part III: Process history based methods," *Computers & Chemical Engineering*, vol. 27, no. 3, pp. 327–346, July 13, 2009.
- [8] Y. Zhao, S. Wang, and F. Xiao, "A statistical fault detection and diagnosis method for centrifugal chillers based on exponentially-weighted moving average control charts and support vector regression," *Applied Thermal Engineering*, vol. 51, no. 1-2, pp. 560–572, 2013.
- [9] Z. Fahim and X. Wang, "Improving the energy performance of a university building through fault detection and building systems diagnostics," Report ASHRAE Transactions, vol. 118, pp. 159–166, 2012.
- [10] S. Yin, *et al.*, "A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process," *Journal of Process Control*, vol. 22, no. 9, pp. 1567–1581, 2012.
- [11] P. Harihar and A. G. Parlos, "Fault diagnosis of centrifugal pumps using motor electrical signals," *Centrifugal Pumps*, 2012.
- [12] R. Zouari, S. Sieg-Zieba, and M. Sidahmed, "Fault detection system for centrifugal pumps using neural networks and neuro-fuzzy technique," Surveillance 5 Cetim Senlis, Oct. 11-13, 2004.
- [13] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, 2004.
- [14] M. Reif and F. Shafait, "Efficient feature size reduction via predictive forward selection," *Pattern Recognition*, vol. 47, no. 4, pp. 1664–1673, 2014.
- [15] O. Banfer, B. Hartmann, and O. Nelles, "Comparison of different subset selection algorithms for learning local model networks with higher degree polynomials," in *Proc. 11th International Conference on Control Automation Robotics & Vision. Institute of Electrical & Electronics Engineers*, 2010.
- [16] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *Journal of Machine Learning Research*, vol. 13, pp. 281–305, 2012.
- [17] S. Katipamula and M. Brambley, "Review article: Methods for fault detection, diagnostics, and prognostics for building systems—a review, Part I," *HVAC&R Res.* vol. 11, no. 1, pp. 3–25, 2005.



**Maarten Stul** (KU Leuven) graduated as an electrical engineer with the master's thesis: "Efficiency Determination of industrial electrical energy distribution based on a load profile." He works as a researcher at the University of Leuven. He collaborated on the project "Energy Management and Load Profile Monitoring". He is active in data mining applied to "intelligent energy control" algorithms.



**Rien Leenders** (Odisee) has over three year's experience in the field of energy monitoring. In the master thesis "Design and integration of an Energy Management System at Technology Campus Ghent" a total package was developed where all the energy flows are mapped. He is active in data-measurement, data-registration and electronic development of smart metering devices.



**Liesbet Butaye** (Odisee) graduated as electrical engineer and works as principal lecturer at the University College of Odisee, Technology Campus Ghent. She is active as lecturer in electrical machines and electricity.



**Koen Stul** (KU Leuven) graduated as electrical engineer and is active in the design and assessment of high and low voltage electrical grids. He is project leader of research in the context of energy management systems and is a lecturer at the catholic university of Leuven. E&A, KU Leuven, Gebroeders Desmetstraat 1, 9000 Gent, Belgium