A Data-Driven Approach to Improve Optical Fiber Manufacturing: Focus on Core Deposition

by

Maëlle J. Sardet

B.S., École des Ponts ParisTech (2020)

Submitted to the Department of Mechanical Engineering in partial fulfillment of the requirements for the degree of

Master of Engineering in Advanced Manufacturing and Design

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2023

Author . . . .

Department of Mechanical Engineering
January 20, 2023

Certified by . . . .

Dr. Brian W. Anthony
Principal Research Scientist
Thesis Supervisor

Accepted by . . . .

Nicolas Hadjiconstantinou
Graduate Officer

This page was intentionally left blank.

A Data-Driven Approach to Improve Optical Fiber Manufacturing: Focus on Core Deposition

by

Maëlle J. Sardet

Submitted to the Department of Mechanical Engineering
on January 20, 2023, in partial fulfillment of the
requirements for the degree of
Master of Engineering in Advanced Manufacturing and Design

Abstract

This thesis presents an in-depth investigation on characterization of optical fiber preform core manufacturing and the identification of underlying trends in measured production data. While walking through the different operations involved in the process, we explained the challenges associated with insuring refractive index profile precision and glass purity. Starting with unsupervised learning, process by process, we applied linear and non linear dimensionality reduction algorithms (PCA and t-sne) to features matrices created from time series data and have been able to connect data clusters with context information like machines or month of the year. Then considering the core fabrication process as a whole, we studied the propagation of trends in the data sets up to quality measurements using Dice's statistic to gauge similarities between samples sets. Finally, we developed some data-driven regression models in order to predict the refractive index measured at the end using data from all processes. As a result, Kernel algorithms performed the best and almost as well on raw statistics from all processes as on encoded information about machine sequences and dates. This supervised approach demonstrated some great potential for the development of prediction tools which could help design the optimized production line. An underlying objective is to support Sterlite Technologies Limited in using data-driven approach applied to process control for its plant in Waluj and Shendra starting by implementing good practices for variables measurement, logging and tracking.

Thesis Supervisor: Dr. Brian W. Anthony
Title: Principal Research Scientist

This page was intentionally left blank.

Acknowledgment

Words cannot describe my gratitude for all who have supported me directly and indirectly throughout my graduate career at MIT.

First and foremost, I would like to thank Dr. Brian Anthony, my thesis supervisor and also MEng program director, for his new inspiration for the program and opening the doors of the Device Realization Lab to my cohort. You have been a source of inspiration in my work and I am sincerely grateful for your guidance and encouragement on this thesis journey.

Furthermore, I am thankful for my labmate and friend Mohamed Ayman. You have been the greatest support under any circumstances and I am fortunate I had you by my side on this project.

Last but not least, I would like to thank my french engineering school École des Ponts ParisTech for giving the opportunity to study abroad while finalizing my master in France and accomplishing my dream of being part of MIT prestigious community. Of course, this wouldn't have been possible without the financial support from my parents who worked their all life in order to invest in my future. Thank you for that.

This page was intentionally left blank.

Contents

This page was intentionally left blank.

List of Figures

This page was intentionally left blank.

List of Tables

This page was intentionally left blank.

Chapter 1

Introduction

1.1 Motivation for Industry 4.0 in Optical Fiber Manufacturing

The increasing demand for scalable, dependable, and fast communication infrastructure is one of the leading market drivers of optical fiber manufacturing. Currently, fibre optics is the only technology that can effectively meet this need. Standard cables are ten times slower than fibre optic cables. Additionally, it carries more data than copper cables. In addition, fibre optics ensures high bandwidth connectivity for emerging technologies such as 5G and the internet of things (IoT). Despite the fact that 5G offers wireless connectivity, fibre optics are required to manage the massive backhaul traffic that 5G generates.

The fiber optics market is characterized by strong competition, with a few major worldwide competitors owning a significant market share. Some key players operating in the fiber optics market include Corning Incorporated (USA), Optical Cable Corporation (USA), Sterlite Technologies Limited (India), OFS Fitel (USA), LLC (USA), Prysmian Group (Italy), AFL (USA), Birla Furukawa Fiber Optics Limited (Japan) and Finolex Cables Limited (India). [5]

Most of those companies have been producing optical fiber for decades and manufacturing lines, representing a huge initial capital investment, haven't evolved much. One big limitation is how labor intensive the process still is. Indeed, fiber optic needs to remain affordable and versatile and to do so, modernization of manufacturing line at low cost is the key. Industry 4.0 technology helps manage and optimize all aspects of the manufacturing processes and supply chain. With a wiser usage of already acquired real-time data but also the installation of new smart sensors, there is strong opportunity for better monitoring and control which will ultimately boost process efficiency and reduce labor intensity.

1.1.1 Sterlite

Sterlite Technologies Limited (Formerly Sterlite Tech) is an Indian multinational technology company, headquartered in Mumbai. It is listed on Bombay Stock Exchange and National Stock Exchange of India. It has 636 patents [6] and is active in over 150 countries. The company is specialized in optical fiber and cables, hyper-scale network design, and deployment and network software. As presented on figure 1.1, Sterlite Technologies Limited has been manufacturing optical fiber for over 25 years.

STL has also partnered with other industry entities to design, build and manage such cloud-native software-defined network. It also has offices in China, US, SEA, Europe and MEA. It has facilities in India, Italy, China and Brazil and two software-development centres. To give a boost to Indian government make in India initiative, STL recently invested in 5G assembling an ecosystem of partners.

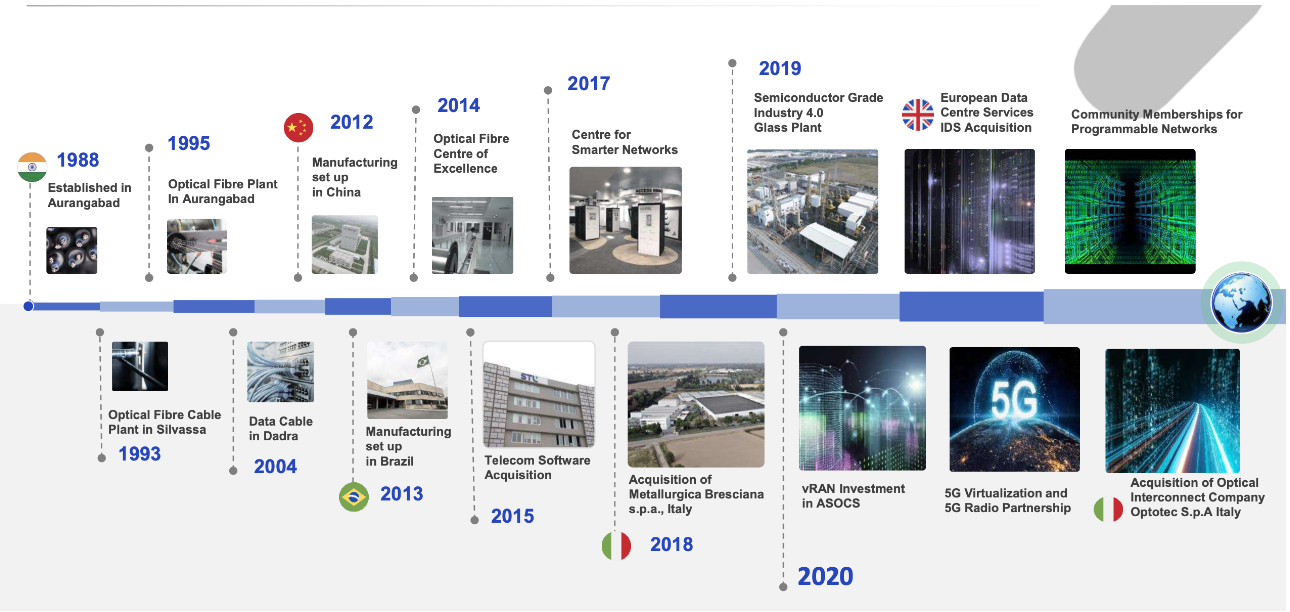

1988
Established in Aurangabad
1993
Optical Fibre Cable Plant in Silvassa
1995
Optical Fibre Plant In Aurangabad
2004
Data Cable in Dadra
2012
Manufacturing set up in China
2013
Manufacturing set up in Brazil
2014
Optical Fibre Centre of Excellence
2015
Telecom Software Acquisition
2017
Centre for Smarter Networks
2018
Acquisition of Metallurgica Bresciana s.p.a., Italy
2019
Semiconductor Grade Industry 4.0 Glass Plant
2020
vRAN Investment in ASOCS
European Data Centre Services IDS Acquisition
5G Virtualization and 5G Radio Partnership
Community Memberships for Programmable Networks
Acquisition of Optical Interconnect Company Optotec S.p.A Italy

Figure 1.1: Sterlite Technology Limited (STL) history as presented in their *Getting BSS on Cloud Strategy* slides deck. [1]

This project has been initiated by Dr. Badri Gomatam who is actually the Group Chief Technology Officer and concerned the manufacturing plants Waluj and Shendra based in India. Those plants are mainly dedicated to manufacturing single mode optical fiber transmitting 1490nm wavelength signal in accordance with internet communication standards. The project is currently managed by Ankush Bansal, Head of Product Management and directly involves Sudhakar Eddula and Vivek Singh, plants Engineers.

1.2 Process Overview

1.2.1 General Workflow

Sterlite manufactures single mode fibers with a step index which presents thinner core than multimode equivalents. Optical fiber manufacturing involves three distinct stages: core manufacturing, clad manufacturing and fiber draw. Those three layers insure the mechanical properties of the fiber, the clad being responsible for the mechanical strength but also for containing the light within the fiber due to its refractive index difference. Core and clad manufacturing are regrouped in the glass department since fiber draw is done in another location in the plant. As the name suggests, core refers to the core of the glass cylinder, id est the first layer deposited on a mandrel since clad refers to the envelope which is deposited afterward. Together they form a preform which will then be drawn to make the bare fiber ($125\mu m$ diameter). A fine layer of acrylic coating is also added bringing the final diameter of the fiber to $250\mu m$.

The goals of those three manufacturing processes are different in various aspects. For example, purity and acquisition is what matters the most in core manufacturing since clad manufacturing is all about volumes and scalability (how do you ensure it? How fast can you make it happen?). Geometrical control and thermal control are the most important in fiber draw. Geometrical control involves conserving the ratio between core and clad layers while drawing ($125/8.7 = 14.4$) so that the optical properties of the fiber stay within the limits. Thermal control involves ensuring a proper cooling of the fiber without adding any stresses to prevent it from breaking during the process as hundreds of kilometers of fiber need to be drawn without interruption. All in all, the three layers insure the mechanical properties of the fiber, the clad being responsible for the mechanical strength but also for containing the light within the fiber due to its refractive index difference. Ideally, 1kg of glass (including core, clad and coating) is equivalent to 37km of fiber drawn but actual yield is smaller.

The general workflow of the fiber manufacturing process is illustrate in the figure 1.2. This work focuses on the operations related to the fabrication of the preform core.

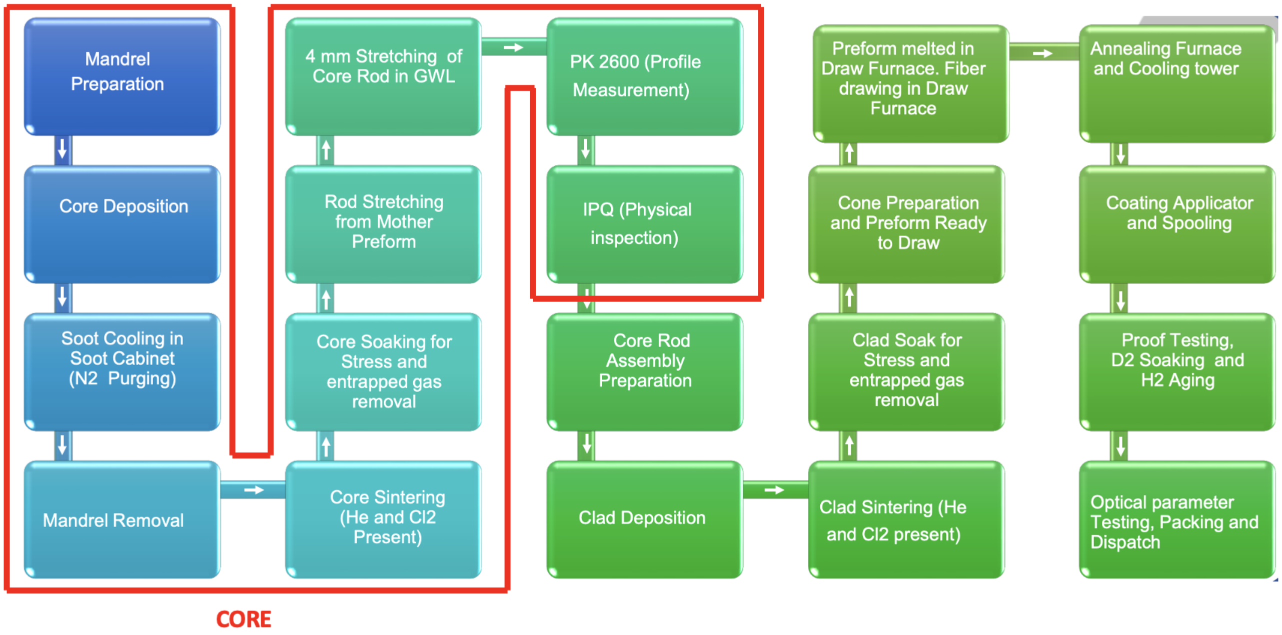
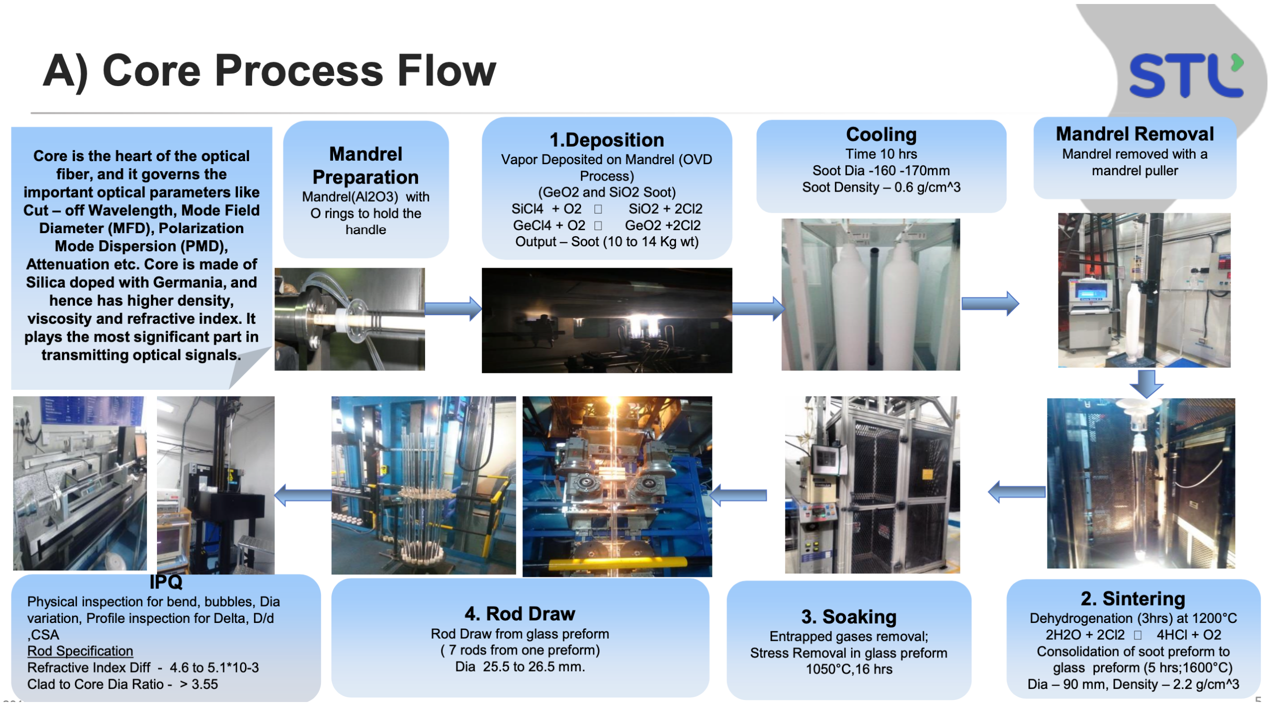

Mandrel Preparation
Core Deposition
Soot Cooling in Soot Cabinet (N2 Purging)
Mandrel Removal
Core Sintering (He and Cl2 Present)
Core Soaking for Stress and entrapped gas removal
Rod Stretching from Mother Preform
4 mm Stretching of Core Rod in GWL
PK 2600 (Profile Measurement)
IPQ (Physical inspection)
Core Rod Assembly Preparation
Clad Deposition
Clad Sintering (He and Cl2 present)
Clad Soak for Stress and entrapped gas removal
Cone Preparation and Preform Ready to Draw
Preform melted in Draw Furnace. Fiber drawing in Draw Furnace
Annealing Furnace and Cooling tower
Coating Applicator and Spooling
Proof Testing, D2 Soaking and H2 Aging
Optical parameter Testing, Packing and Dispatch
CORE

Figure 1.2: Optical fiber manufacturing general process workflow.[2]

1.2.2 Core Process

The core process workflow is presented in the following figure 1.3:



A) Core Process Flow

Core is the heart of the optical fiber, and it governs the important optical parameters like Cut – off Wavelength, Mode Field Diameter (MFD), Polarization Mode Dispersion (PMD), Attenuation etc. Core is made of Silica doped with Germania, and hence has higher density, viscosity and refractive index. It plays the most significant part in transmitting optical signals.

Mandrel Preparation
Mandrel(Al2O3) with O rings to hold the handle

1.Deposition
Vapor Deposited on Mandrel (OVD Process)
(GeO2 and SiO2 Soot)
SiCl4 + O2 □ SiO2 + 2Cl2
GeCl4 + O2 □ GeO2 +2Cl2
Output – Soot (10 to 14 Kg wt)

Cooling
Time 10 hrs
Soot Dia -160 -170mm
Soot Density – 0.6 g/cm^3

Mandrel Removal
Mandrel removed with a mandrel puller

2. Sintering
Dehydrogenation (3hrs) at 1200°C
2H2O + 2Cl2 □ 4HCl + O2
Consolidation of soot preform to glass preform (5 hrs;1600°C)
Dia – 90 mm, Density – 2.2 g/cm^3

3. Soaking
Entrapped gases removal;
Stress Removal in glass preform
1050°C,16 hrs

4. Rod Draw
Rod Draw from glass preform
(7 rods from one preform)
Dia 25.5 to 26.5 mm.

IPQ
Physical inspection for bend, bubbles, Dia variation, Profile inspection for Delta, D/d ,CSA
Rod Specification
Refractive Index Diff - 4.6 to 5.1*10-3
Clad to Core Dia Ratio - > 3.55

Figure 1.3: Core Process Workflow. [2]

Core Deposition The first step of the core process is the mandrel preparation. Typically, the mandrel is a reinforced rod made of ceramic on which material is deposited. It will typically have a diameter from 7mm to 8mm and be about 1.3 meters long. It is important that the mandrel is crystal clean and smooth and presents enough mechanical strength so that it can hold a weight of approximately 14kg to 15kg at temperature reaching up to 900°C. Once the mandrel is held in the chucks of the deposition machine at both ends, the Outside Vapor Deposition (OVD) can start. Although it is called vapor deposition, the terms refer to a reaction which is taking place during combustion of the input particles at about 800°C: those particles agglomerate and almost form a jet of vapor stream that is captured by the spinning mandrel. Basically, a mandrel mounted on a traverse will from left to right over a stationary burner to enable a slow formation of the layers of soot. For information, about half of the synthesize material (precursor chemical) goes into the exhaust and doesn't get captured. The process is controlled by stage according to the number of passes done by the traverse and but deposition won't stop until the weight specified in the recipe is reached.
Each core deposition machines has two spindles which means that two preforms are created simultaneously. Nevertheless, performance varies from on spindle to another.
Note: Sterlite chooses OVD over other conventional methods as Modified Chemical Vapor Deposition (MCVD) because it enables more freedom on the final diameter of the soot preform. Indeed, in the case of MCVD, chemicals are passed through the inside of a rotating glass tube made of pure synthetic SiO2 and deposition happens inside the tube from the sides to the center. Thus, preform diameter is limited by the size of the tube. [7]

Cooling and Mandrel Removal Once the deposition is done, the next step is cooling which happens in a separate cooling cabinet. Now cooled down, the soot preform goes for mandrel removal. This step is critical because any disturbance that could happen during the mandrel removal will disturb the soot and create uneven gaps in the core of the preform, which defect will propagate through the next stage of the process and in fine cause light path defects. This explain why the surface properties of the mandrel are so important. Actually, it is usually covered with a very thin layer of carbon obtained from burning Acetylene which acts as a lubricant. Also, at this stage, the soot is highly porous with a density of about $0.4 - 0.5 g/cm^3$.

Sintering Sintering is a two stages process (drying and then sintering) where the soot gets consolidated and the snow like preform turns into a crystal clear glass. Drying is done to remove any OH- from the deposited soot, it happens at 1050°C. Helium is used as a thermal medium to exchange the heat powdered material into the preform (from the outside to the inside) to a temperature below the melting point and enable the consolidation of the soot which diffuses and forms one solid piece. Dehydration also prevent from nuisance caused by moisture as change in refractive index and viscosity. Sintering happens in a quartz sinter tube at 1583°C. The centerline of soot formed by the mandrel removal is collapsed using vacuum torr (150 torr).

Soaking Soaking is used to purge the entrapped helium into the glass and relieve preform thermal stress. It is done at about 1100°C for around 15h to 16h. It is again very important not to have any bubbles of helium remaining that could pop up afterward in the draw process.

Rod Draw Rod Draw is similar to fiber draw and consists of stretching the preform which is about 1.2 meters long, 10 centimeters wide and weights about 13-14kg into very precise rods of around 25mm in diameter. The reason why the preform is cut is subsection is mostly a matter of bulk in the plant. About 7 core rods are sectioned from a mother preform. One constraint on the rod is that it needs to be perfectly straight. Unfortunately, the break energy (a notch is made and then the extremities are pulled until a break) often transfers back along the rods and can introduce some bends in the rods. Approximately, 35% of the rods will go to scrap at this stage of the process. It should be noted that, sometimes, the bend can be corrected using a flame but the rod cannot be touched with the flame more than three times otherwise it will damage it.

A 4mm piece from rod 1 and rod 5 are sampled and stretched in order to do profile measurements.

Quality Final step, the rods go to the Quality Department which task is to ensure their quality using various angles from various parameters. For example, rod diameter will be measured in order to estimate the core/clad ratio. As well, the refractive index profile across the rod section will be drawn and used to extract some profile indicator as the Center Line width and Center Line dip.

The timeline of each stage of the process are described in figure 1.4.

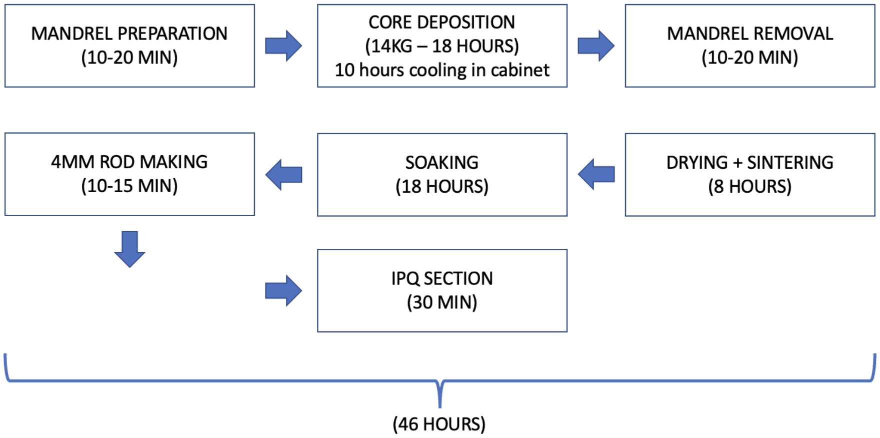


Figure 1.4: Time that takes at each stage of the core process. [3]

1.3 Project Scope

1.3.1 Previous Work

The collaboration between Device Realization Laboratory and Sterlite started in 2020. As data were the most consistently acquired during fiber drawing, it was decided to start with improving control of this stage of the process. As part of his master thesis, George C. Chen [8] worked on an in-depth investigation on modeling and simulation of one of Sterlite's drawing tower and its controllers. With measured production data during the fiber drawing process, a long short-term memory (LSTM) neural network has been created, implemented and trained to model the process dynamics and obtain the correlation between inputs (furnace power, preform speed, capstan speed and helium cooling temperature) and outputs (bare fiber diameter and fiber tension) of the fiber drawing process. The ultimate goal being to obtain an emulator of the tower and its controllers and use simulation results to train a smart controller which could adapt to shifts in the process happening throughout time. Indeed, once implemented in hardware, this new generation of controllers could replace laborious, iterative tuning process associated with conventional ones.

Because the current pre-processing pipeline restricts the training data by the upper and lower thresholds of the bare fiber diameter, George Chen's approach placed more emphasis on the steady-state region of the production data. As a consequence, the trained model learns and predicts the steady-state process dynamics with higher accuracy.

However, critical (and potentially nonlinear) dynamics happen each time the controllers is once again engaged after having been turned off because of a fiber break. Those breaks are the main reason for manufacturing efficiency loss during fiber drawing. If training the model on those regions where the bare fiber diameter starts with different initial conditions and is brought to steady-state could make it more robust to disturbances in simulated dynamics, question can be asked about the nature of those breaks and why they happen.

The reasons for those breaks are typically mechanical defects which origins are various: at a macroscopic level, little bubbles can appear during densification (liquid to solid) of the fiber. Handling scratches can also occur at the end of the process and moisture in the air can create crack propagation. At a microscopic level, we know for example that presence of un-sintered particles at the interface between clad and core could create a defect. Using Sterlite's rule of thumb, for one preform (from 20kg to 60kg), you don't want to have defects that are bigger than about half a micrometer because those defects propagate under tension leading to larger crack and inevitably breaks. One way to fight that is to exposed the preform to a laser beam and count the deflections. Using look up table, this number can give an estimation of the number of breaks per kilometers and decide if the preform should be scraped or send back to the draw. This time addressing root causes for fiber breaks, Sterlite is offering Device Realization Lab to investigate on the core manufacturing process. The project started in May 2022.

1.3.2 Thesis Overview

Problem Statement There are two important factors in core manufacturing: precision in how layers of material are deposited in order to achieve the kind of profile looked for and making sure that no impurity is introduced during deposition. Taking care of the first one prevents from optical and diameter defects since the second one prevents from mechanical defects, id est the breaks described in the previous section.

This thesis investigates the feasibility of a data-driven approach on the core deposition process in order to identify and improve efficiency losses.

Challenges Several challenges associated with the question this work tackles have been identified:

- **Delay:** In terms of control, we are talking about hours timescale even affecting feedback loops. For example, the bend introduced by a cut will show up many hours later.
- **Control checkpoints:** No irregularities detection is implemented from mandrel preparation to rod draw which makes the all core deposition process pretty opaque. The first physical inspection happens with the profile measurement after rod draw (ninth step of the process).
- **New to data-driven methods** Beside conventional feedback control, Sterlite is relatively new to data-driven approach and is willing to learn with Device Realization Lab.

Even though significant variables are being measured for control purposes, no consistent data logging and data tracking has been implemented yet. All data are pulled manually by an engineer on site each time using a method specific to the stage of the process. More over, IT network architecture differs from Waluj to Shendra plants.

- **Extreme operation condition** Currently, the quality of the acquisition is varying from one sensor to another to the extend that sometimes data can't be used. Also, machines being pretty old, bulky and required to be isolated from any kind of impurities and operating conditions being at very high temperature makes it hard to introduce new generation of sensors (specifically for sintering and soaking).

Approach Our work will mainly focus on the 35% rod going to scrap at the end of the core deposition process. Is this scrap the result only of cutting or can we identify reasons for this high scrap proportion upstream in the process ? (see subsection 1.2.2)
Our approach can be divided in two stages:

- **Stage 1:** Investigation using the base case system only

  1. Draw the systems base case: analysis of the current situation.
  2. Experiment using current available data. Test repeatability and parameter sensitivity with the current hardware.
  3. Run analysis on data from current hardware. Instrument characterization of trends, relationships. Start data driven models.

- **Stage 2:** Investigation using results from phase 1, introduction of more sensors and strengthening of data collection

  1. Update the super instrumented base case.
  2. Run experiments including parameters set beyond normal limits of range or rate.

3. Extend model to new data. Implement test and validation. Try data/physics combination models.

This page was intentionally left blank.

Chapter 2

System Base

2.1 Current Situation Analysis

2.1.1 IT Architecture Networks for Data Acquisition and Storage

The exact description of the data available to us for each process are detailed in the Annex section 6.1.

In this subsection is presented the example of deposition and sintering IT network architectures for data logging and storage. The storage is done is two steps: first time series runs are acquired by a PLC (Programmable Logic Controller), then, at the end of the run, those data are aggregated on a computer using an application from the family of the SAP software products (Systems Application and Products in data processing) called SAP-MII (Manufacturing Integration and Intelligence) which basically synchronizes manufacturing operations with both back-office business processes and standardized data. Both run data and aggregated data are stored on a cloud from where they can be pulled in csv or excel formats which is the type of files Sterlite is sharing with us.

Figures 2.1 and 2.2 illustrate the architecture just described which is specific to Shendra plant.

Architecture information regarding sintering, rod draw and IPQ haven't been shared for the plant in Shendra specifically but are similar to the one just presented.

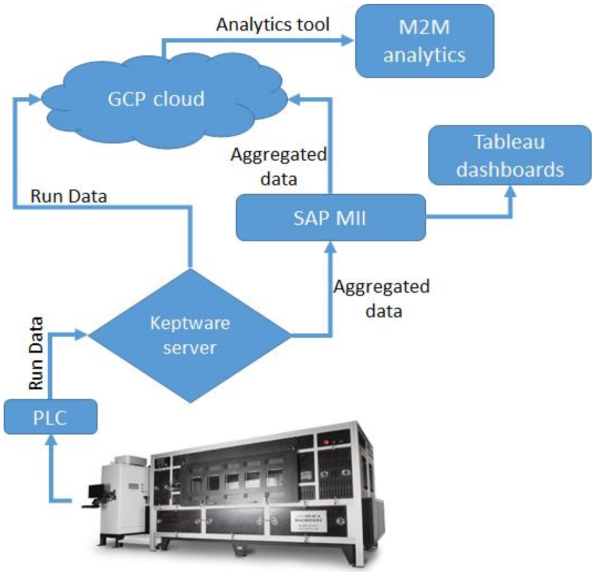


Figure 2.1: IT network architecture for core deposition process.[2]

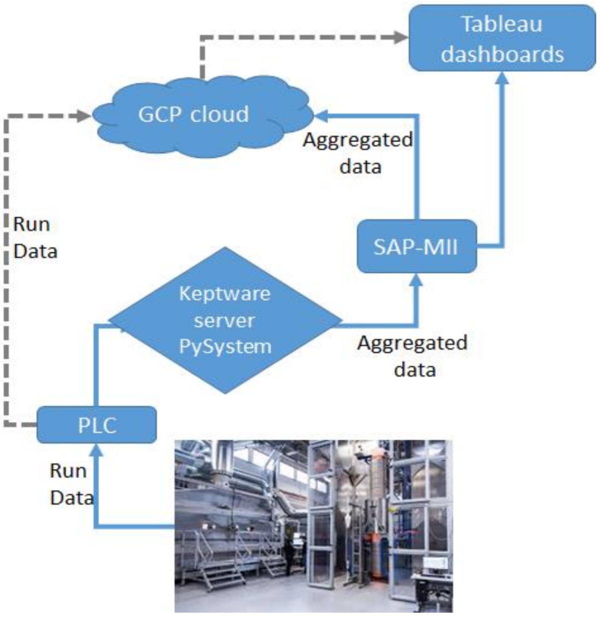


Figure 2.2: IT network architecture for sintering process.[2]

Here is an exhaustive list of variables captured during the two steps of acquisition.

For deposition operations, we have:

The PLC captures the data of variables during running operation till the end of operation i.e. till 18 hrs.

- ID: Preform ID, Recipe ID, Machine ID
- Time: Deposition Date, Elapsed Time, Deposition time
- Physical: weight deposited, Density and Soot Diameter
- Process stages: As per the recipe
- Chamber: Pressure and Temp of SIcl4 and Gecl4 Hotbox
- Pressure: Suction pressure
- Flows: H2, O2, SiCl4, GeCl4, C2H2, gases flows
- Position: Traverse movement

SAP MII data aggregates data at the end of core deposition operation.

- ID: Preform ID, Operator ID, Product type, *Si and Ge type*
- Time: Prod date, Elapsed Time, Deposition time
- Physical: Actual Dep wt, Actual Core wt, targeted core wt, Density, *Mandrel sequence and No*, Soot diameter
- Process parameters: Dep rate, Recipe, GeCl4 Passes
- Chamber: Pressure and Temp of Sicl4 and Gecl4 Hotbox
- Consumption: H2, O2, GeCl4, SiCl4, C2H2 total consumption
- *Any observation during or after operation*
- *Status: Final status of Preform*

For sintering operations, we have:

The PLC captures the data of variables during running operation till the end of operation.

- ID: Preform ID, Recipe ID, Machine ID
- Time: Elapsed Time
- Process stages: As per the recipe
- Pressure: Vacuum pressure
- Temperature: Main furnace, Zone-wise temperatures
- Flows: He, Cl2, N2 gases flows
- Position: Distance remaining

SAP MII aggregates data at the end of sintering operation.

- ID: Preform ID, Operator ID, Product type, core soak machine no
- Time: Production date, Elapsed Time, Deposition time
- Physical: Preform weight, Sintered preform diameter (max,min),Sinter length
- Process parameters: Column speed, vacum, drying time, speed in different zone, position in different furnaces
- Pressure: He, Cl2, Air, scrubber suction
- Consumption and flows: He, Cl2, N2
- Temperature: Main furnace, Zone wise temperatures
- *Any observation during or after operation (preform, centerline, cone, Bubbles, Top open capillary)*
- *Status: Final status of Preform*

It should be noted that those lists describe the theoretically overall data availability, but in reality, we haven't had access yet to the ones *in italic*.

2.1.2 Relevant Parameters Identified

Through our exchanges with Sterlite team, we have been able to do a brief summary of the most relevant variables per operation:

Core deposition

Input: gaz flow rate into the burners (H2, O2, SiCl4, GeCl4).

Process parameters: scrubber suction pressure, H2 pressure, O2 pressure, N2 pressure, SiCl4 line temperature, GeCl4 line temperature.

Machine parameters: mandrel rotation, number of burner (here only one).

Output: SiO2 and GeO2 soot.

Data measured: weight is measured with a load-cell mounted on the spindle. Mass flow controllers (MFC) acquire the different gaz flow rates cited above. Vaporizer pressure and temperature are measured. Two servo mortors give the traverse position.

Sintering

Input: He, CL2 and N2 flow rates.

Process parameters: drying furnace temperature, sintering furnace temperature, He and Cl2 pressures, vacuum to collapse central hole during starting and final stage of sintering and suction for chlorine and helium gaz.

Machines parameters: mandrel rotation, time in drying and sintering section (approximately 1hour changeover between both).

Output: glass preform.

Data measured: gaz/liquid flow rates through MFC, furnace temperature zone-wise, pressure into the sinter machine, traverse motors and rotational motors.

Soaking

Input: N2 flow rate.

Process parameters: soak furnace temperature, nitrogen pressure.

Machine parameters: preform rotation.

Output: bubble free glass preform.

Data measured: N2 flow rate and zone-wise temperatures.

Rod draw

Input: N2 and Argon consumption.

Process parameters: zone-wise temperature, rod drawing time, vacuum.

Machines parameters: preform speed feed.

Outputs: 6 to 7 glass rods.

Data measured: N2 and Argon flow rates (MFC), diameter measurement gauge, temperature, servo motors.

IPQ testing

Profile parameters: 1. Profile shape 2. Difference between core and clad refractive index (delta) 3. Core diameter over clad diameter ration (D/d).

Physical Parameters: 1. Length 2. Diameters 3. Bubble/ Un- Collapsed 4. Scratches 5. Rod Weight 6. Bend 7. Ovality

2.2 Previous Work Done by Sterlite

Sterlite has limited experience in data-driven approach applied to improving control and yield beside conventional controllers. Some investigations using data they currently have access to enabled to highlight some irregularities in the process but greater potential lies and should be exploited. Even so, it should be noted that Sterlite engineering team has worked in the past on developing some physics and chemistry models from various steps involved in the deposition process with the goal to improve preform characterisation based on the measurements they do. For example, some work on rod profile fitting using measurement points and parameters has been shared with us.

More relevant to this project, investigation were done a few years ago on uncontrolled weight gain in core deposition. By studying side by side traverse velocity and deposited weight, some discontinuities in weight acquisition by the load cell have been highlighted. Indeed, steps in the data can be observed. Using picks finder tools, the root cause has been identified: load cell measurement is made at the same moment the traverse is changing direction what necessarily disturbs the measurement. As a consequence, core deposition runs present weight control issues especially at the beginning of the run when the assembly weight is relatively low (signal is very noisy with negative excursions). Then initial offsets propagate and telemetry on end-points is lost. Additionally, some differences between the two spindles from the same machine have been observed and it is likely that, by improving weight control stage by stage (e.g. reducing weight fluctuation or using gentler and slower process set-points), both spindle will be able to perform equally well.
Those issues have not been addressed by the on-site team yet even though they are likely significantly impacting core radius uniformity. Therefore, knowing about those investigation has been very helpful as we will observe those irregularities throughout the first experiments. [9]

Comments: It was also pointed out that, in core deposition operation, the diameter acquisition is not consistent and is not taken into account for control.

2.3 Data Supply

2.3.1 Data Availability Overview

It took several try and fail iterations for the engineer on site to adjust the way the data were pulled and aggregated from each operation so that we receive consistent data sets. The overall process still have a lot of room for improvement. Until September 2022, only deposition data were available. Today, we have access to the following:

- Aggregated data per preform for deposition, sintering, soaking, rod draw and IPQ processes from (approximately) January 2021 to June 2022. Deposition, sintering and rod draw datasets are time series data for each preform run (each preform run being stored in an individual file) since soaking and IPQ datasets are directly aggregated results (only one file which each row corresponds to a preform/rod run).
- Some examples of rod profile measurements.
- One example of recipe file for deposition.
- Some examples of draw tower data and bobbin level data.

Comments Some context explanations and data dictionaries are still pending. We asked to get as least processed data as possible so to stay close to what on site acquisition looks like, therefor, the datasets present a mixed between process parameters (set points) and sensor outputs.

2.3.2 Ability to Track Preforms

In order to study irregularities propagation throughout process stages, it is necessary to be able to track a preform from deposition to IPQ measurement. As described in table 2.1, we reached a reasonable preforms overlap between data sets:

Number of preforms in both datasets	Deposition	Sintering	Soaking	Rod draw	IPQ
Deposition	2145	/	/	/	/
Sintering	1953 (91%)	3724	/	/	/
Soaking	1997 (93%)	3462 (93%)	3964	/	/
Rod draw	1914 (89%)	3335 (89%)	3446 (87%)	3677	/
IPQ	2090 (97%)	3652 (98%)	3864 (97%)	3632 (99%)	3816

Table 2.1: Two by two comparison between datasets: counting number of preforms present in both

Now when we look at the intersection between every process datasets, the number of preforms decreases to 1647 and after cleaning rows with missing values in the IPQ table, it shrinks to 964 (minus 40%). Therefor, improving the data logging during IPQ process could significantly increase the number of preforms that could be studied. This is particularly relevant when it comes to train models.

2.4 Partial Conclusion

The base case system analysis has shown great potential in data Sterlite is already acquiring without considering adding any sensors yet. However, work needs to be done in order to improve how those sets of data are logged, merged and conveyed to our team. Nevertheless, our weekly exchanges with the team on site have already lead to significant improvement in terms of data cleaning, organization and documentation. Good practices should be set up in order to maximize the success of this project.

This page was intentionally left blank.

Chapter 3

Experiments

For about half of the project, only core deposition data were available to us. Therefor, the first experiments aimed to familiarize ourselves with this specific operation.

3.1 Time Series Data from Deposition Operation

3.1.1 Zoom in Core Deposition Process

Doped Silica Deposition The aim of core deposition operation is to deposit GeO2 and SiO2 particles on the mandrel. Doping Silica with Germanium enables to modify refractive index of glass in order to obtain the desire cross sectional refractive index profile needed to guide the light through the fiber. For the recall, two preforms are deposited simultaneously inside one deposition machine which are called dual spindles machine. Three operations are involved in the process:

Step 1: The SiCl4 and GeCl4 vapours are formed in separate vaporizers (at respectively 75°C and 100°C) maintained at controlled temperature close to the saturating vapor pressure of the particles.

Step 2: Using O2 carrier gas, these vapours pass through a heated line and are finally injected in the center region of an annular H2/O2 burner. There is a gas flow rates controller to respect to the requirements.

Step 3: The temperature obtained from the reaction between H2 and O2 leads to the oxidation/hydrolysis of the Silica vapours, resulting in the formation of SiO2 particles and GeCl4 vapours resulting in the formation of GeO2 particles as described by the following reaction equations:

$$SiCl_4(g) + O_2(g) \longrightarrow SiO_2(s) + 2Cl_2(g)$$

$$SiCl_4(g) + 2H_2O(g) \longrightarrow SiO_2(s) + 4HCl(g)$$

$$GeCl_4(g) + 2H_2O(g) \longrightarrow GeO_2(s) + 4HCl(g)$$

For a normal process, the recipe starts with the Mandrel Flame Polish Mode. An aluminum mandrel of size of 67 mm with the length of almost 1.5 meter is used. Hydrogen auctions is provided during the flame polishing to avoid defects points on the mandrel. After that is done carbon coating through acetylene to maintain the smoothness of the mandrel post deposition process (5 to 6 passes). Then flame polishing is done again with hydrogen flow. The reason is that the mandrel got saturated in CO2 during the previous passes so for sticking purpose this operation is redone. Last but not least, deposition starts during the Centerline Mode with few passes of SiCl4 only again to enable the mandrel to be smoothly removed afterwards. One pass corresponds to one movement of the traverse in one direction. To give an idea, during Core Deposition Mode, 350 to 400 passes are equivalent to 1kg deposited. Now the mandrel is ready for core deposition.

Deposition is splitted into two sections/layers: one is core and the other is clad, but here the clad refers to the clad of the preform core and not the preform clad that will be deposited after in the Clad Deposition Process (see figure 1.2).

Each passes enable to split the core into layers so to slightly change the refractive index each

time and obtain the bell curve we see on the refractive index profile on figure 3.11.
Throughout the different passes, the suction may vary. So, out of the total quantity of SiO2 and GeO2 generated, maybe only 50-60% actually get deposited (the rest goes to scrap). So, because of these fluctuations, the last stage of core deposition Core J Mode is based on the weight of the preform and runs until the about 900g is reached. So it is a close loop on the weight considering the index profile calculated. Number of passes during this mode can vary from 50 to 100. All in all, out of 14kg, 1kg will be SiO2 and GeO2 for the core and the rest (13kg) will be SiO2 only for the clad.

Finally, Clad Mode is divided in 2 different sections to increase the deposition rate because there should be a gradual increment of the flow rate over this period of time.

The recipe sets up 17 stages which are the following:

Core Deposition Recipe Structure

- Mandrel Flame Polish Mode
- Carbon Coating Mode (Acetylene)
- Carbon Flame Polish Mode
- Centerline Mode (SiCl4 only)
- Core Deposition Mode (350-400 passes modes)
- Core A to I Modes (GeCl4 + SiCl4)
- Core J Mode (SiCl4 + GeCl4) controlled on Weight
- Clad Mode Oscillation 1 (SiCl4 only)
- Clad Mode Oscillation 2 (SiCl4 only)

Data are time series data acquired approximately every 5 seconds. Every pass is about less than 20 seconds so about three to four data points are acquired per pass.

Figure 3.1 describes the inside of the deposition machine and figure 3.2 the gaz flow before being injected into the burners head.

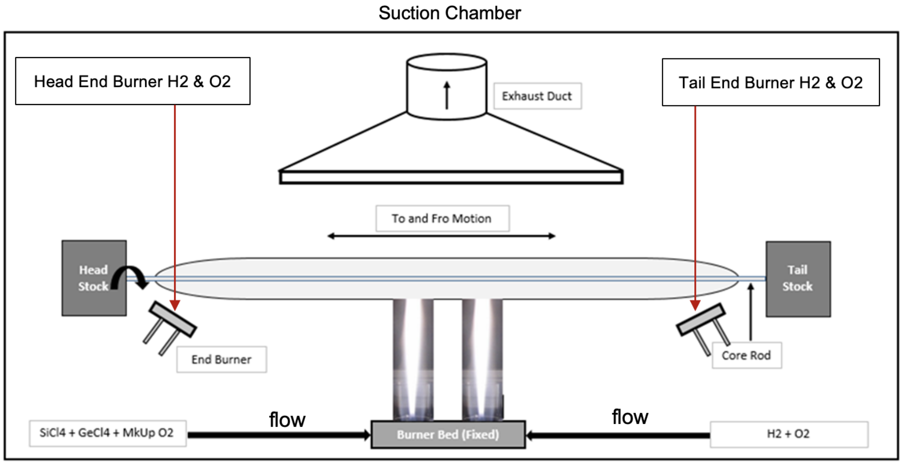


Figure 3.1: Schema of the inside of the deposition machine. [3]

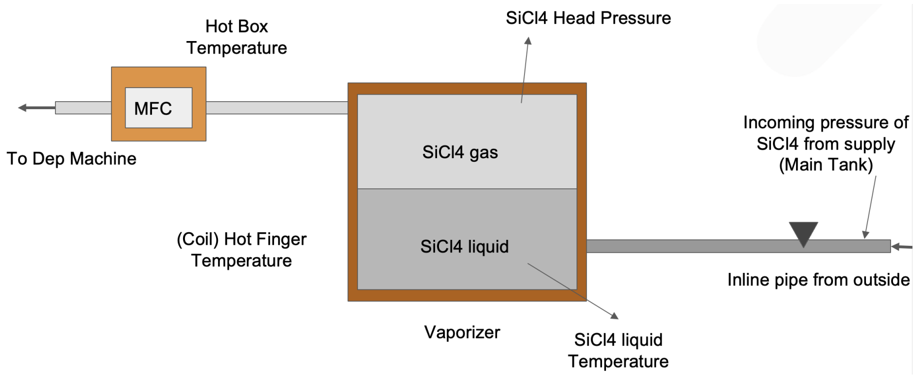


Figure 3.2: Schema of SiCl4 flow before being injected into the deposition machine. The schema is the same for GeCl4 flow. [3]

3.1.2 Some Plots

Time Series Throughout Stages Figures 3.3 and 3.4 present some time series run plots of the data communicated by Sterlite team. We can see the proportional importance of each stages one compared to each other.

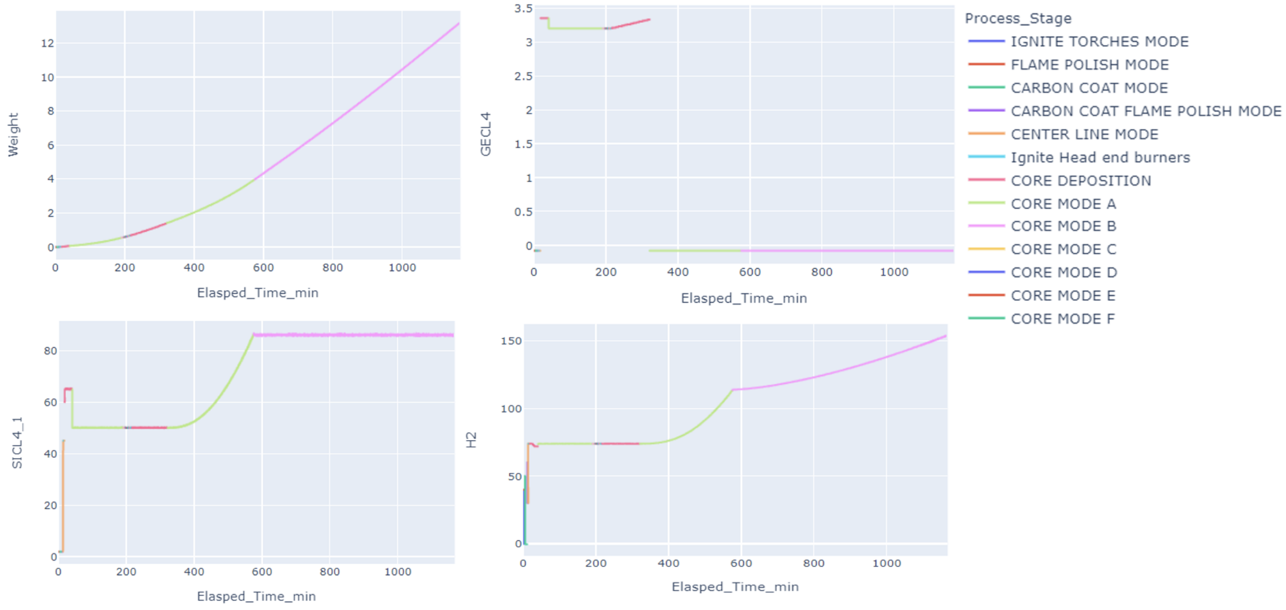


Figure 3.3: Weight, GeCl4 gaz flow (gpm), SiCl4 gaz flow (gpm) and H2 gaz flow (Lpm) plotted against Elapsed Time (min) showing the different stages of the deposition process.

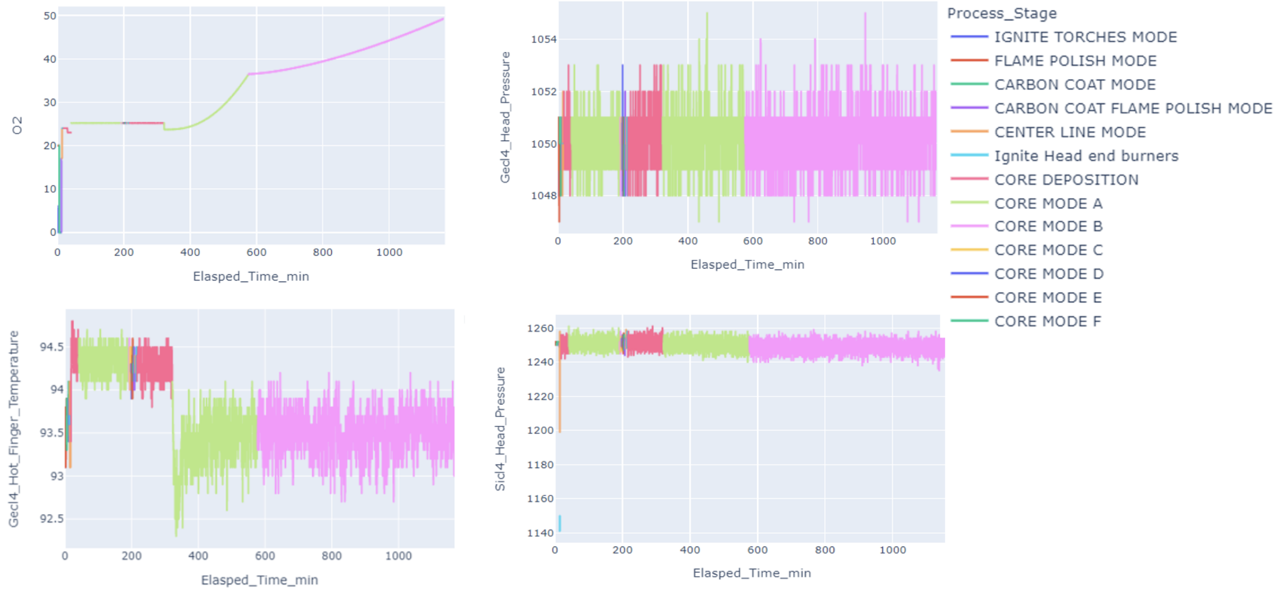


Figure 3.4: O2 gaz flow (Lpm), GeCl4 Vapouriser Head Pressure (Torr), GeCl4 Vaporiser Finger Temperature (°C) and SiCl4 Head Pressure (Torr) plotted against Elapsed Time (min) showing the different stages of the deposition process.

Correlation Matrices A way to study linear interdependence between measurements is to compute the Kendall's correlation coefficients between all pairs of variables in our data table. The function corrplot from MATLAB also returns tables for the correlation matrix R and matrix of p-values. Some examples are shown in figures 3.5, 3.6 and 3.7 when comparing all measurements with the same physical dimensionality one with each other.

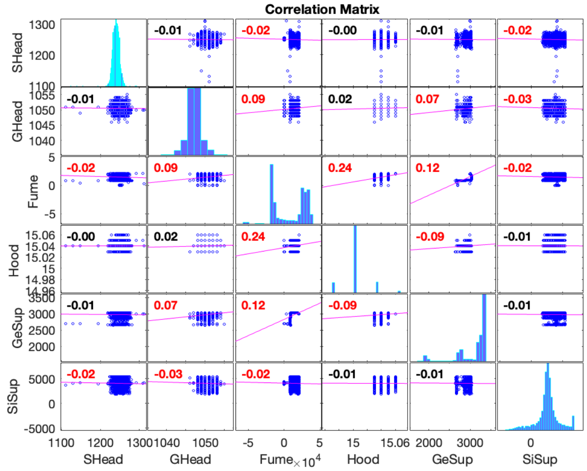


Figure 3.5: Kendall's correlation coefficients matrix between the time series variables SiCl4 Head Pressure, GeCl4 Head Pressure, Fume Pressure, Hood Pressure, SiCl4 Supply Pressure, GeCl4 Supply Pressure.

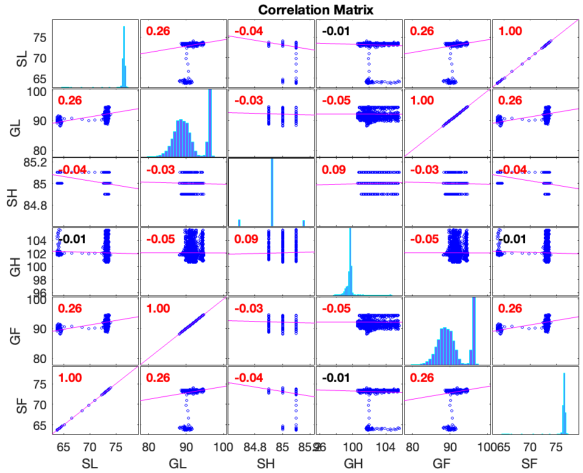


Figure 3.6: Kendall's correlation coefficients matrix between the time series variables SiCl4 Liquid Temperature, GeCl4 Liquid Temperature, SiCl4 Hotbox Temperature, GeCl4 Hotbox Temperature, GeCl4 Hot Finger Temperature and SiCl4 Hot Finger Temperature.

For example, on figure 3.6, we can observe that SiCl4 (resp. GeCl4) Hot Finger Temperature and SiCl4 (resp. GeCl4) Liquid Temperature are linearly related which makes sense considering the liquid SiCl4 is directly heated by the hot finger coil (see schema 3.2). On figure 3.5, we can see all pressure measurements are linearly independent one from each other.

Figure 3.7 focuses on every variables related to mass flow. We can first observe that O2 (resp. H2) flow rate at the tail and at the head of the end burner are behaving exactly the same which matches the fact that they are subject to the same controller. We can also validate that weight and diameter are linearly correlated (for recall, density of the preform is homogeneous and about $2.2kg/m^3$). Last but not least, we observe some strong correlation between the weight and the reactives of the oxidation/hydrolysis of SiCl4. We cannot see this correlation for GeCl4 since it is deposited only during specific staged of the process and in less quantity.

Of course, this type of analysis should be applied to all the other core operations.

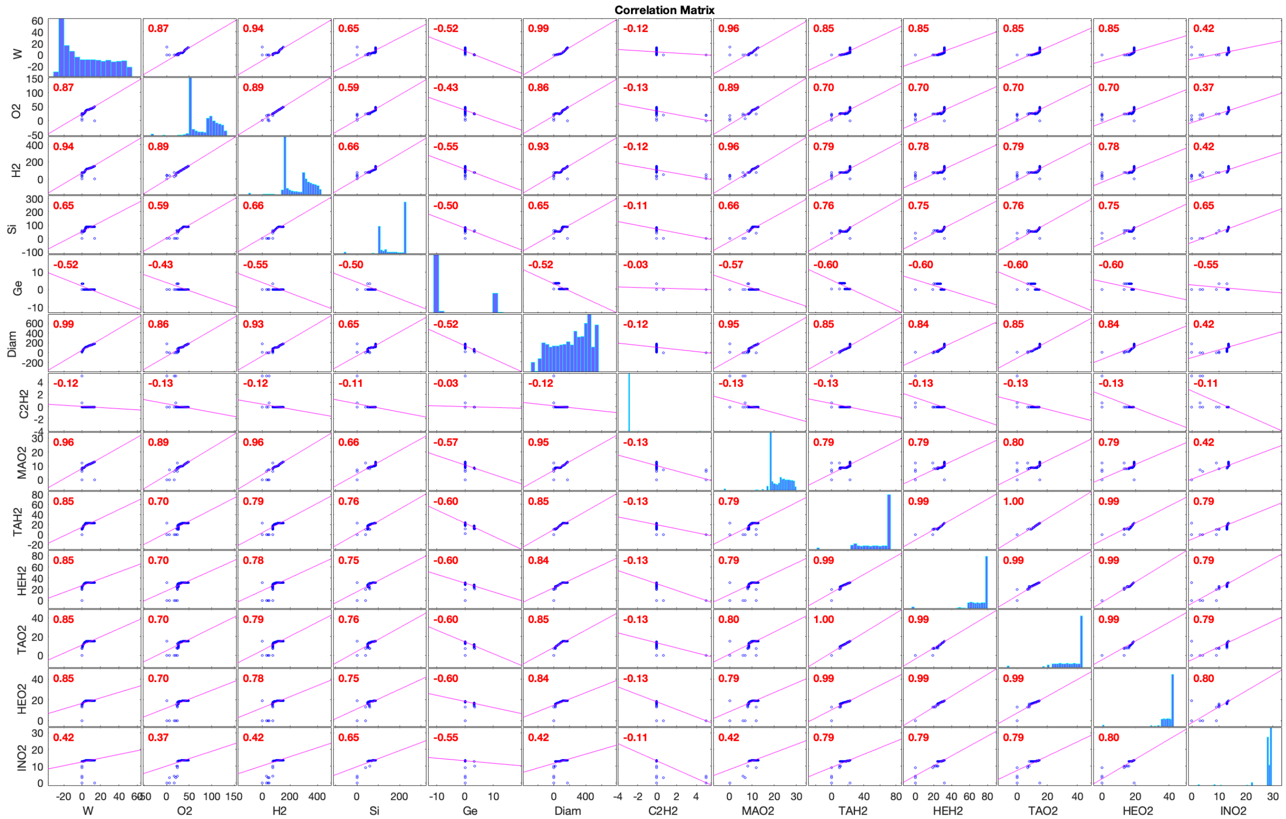


Figure 3.7: Kendall's correlation coefficients matrix between the time series variables Weight, O2 flow, H2 flow, SiCl4 flow, GeCl4 flow, Diameter, C2H2 flow, Make O2 flow, Tail End Burner H2 flow, Head End Burner H2 flow, Tail End Burner O2 flow, Head End Burner O2 flow and Inner O2 flow.

Traverse Position and Weight Acquisition Some work on traverse control and weight acquisition has also been done and has lead to the same conclusion as the study previously done by Sterlite (see section 2.2. As the deficiencies have already been identify but won't be attested for now, those results are not presented in this work and weight measurement will be used as it is.

3.1.3 Deposition Rate/Efficiency

Core Deposition efficiency calculation Sterlite references for deposition rate and efficiency calculation are described in the Annex section 6.2. Through our exchanges, we also learned that overall deposition efficiency of SiCl4 should about 50%-60%. Overall deposition efficiency results shown in figure 3.8 show slightly smaller values because they have been calculated for total GeCl4 and SiCl4 since GeCl4 is know to be a bit less effective.

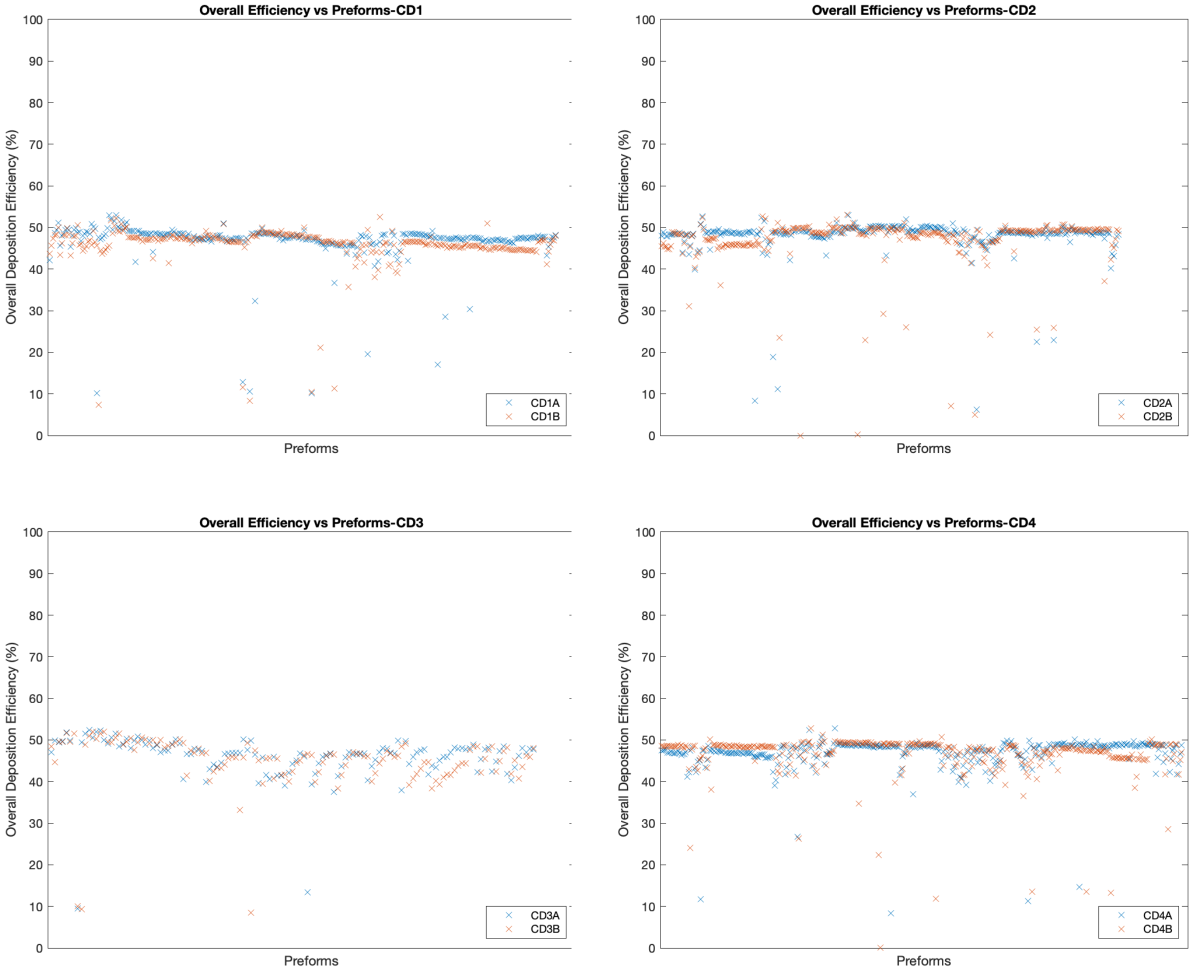


Figure 3.8: Overall SiCl4 and GeCl4 deposition efficiency associated to its preform (preforms are sorted chronologically) for deposition machines CD1, CD2, CD3 and CD4.

When looking at the evolution of the deposition efficiency throughout time (all data set are from Jan 2021 to June 2022), we can see in figure 3.8 that the process gets out of control during multiple periods. It also seems like machine 3 presents the worst results in terms of efficiency and regularity. It should also be noted that it happens that for certain periods of time, both spindles are able to perform almost equally (red and blue lines are confunded on the graph). Extending those periods so that they become majority is so a reachable goal for the future.

Now zooming on the periods of time where the process seems under control, we can observe in figure 3.9 the deposition efficiency/rate in function of the stage. Note: Efficiency is not plotted against diameter because diameter acquisition is known to be unreliable.

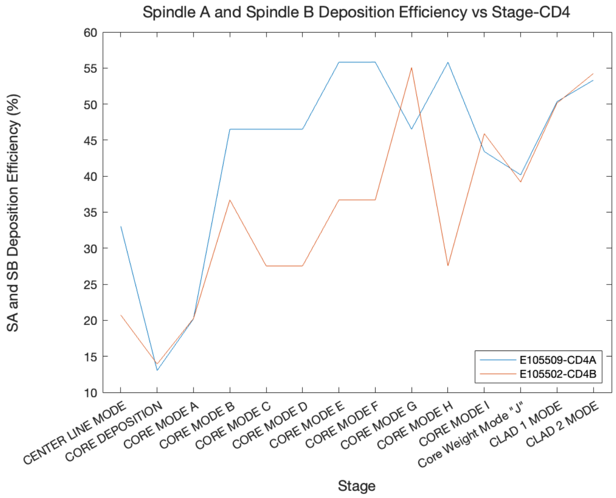


Figure 3.9: Deposition efficiency calculated per stage according to Sterlite guide 6.2.

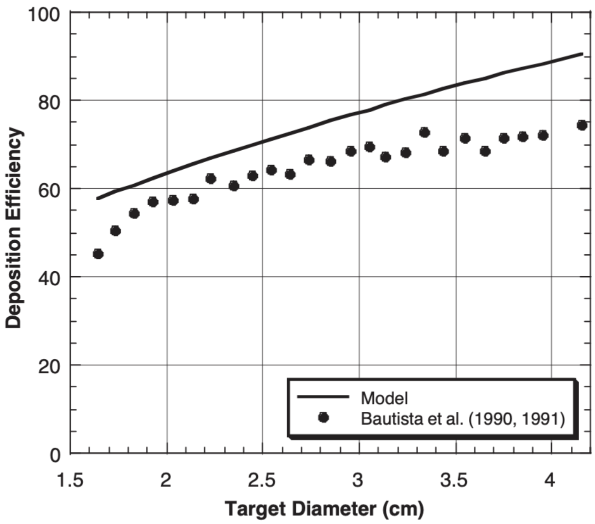


Figure 3.10: Comparison of predicted thermophoretic deposition efficiency as a function of target diameter with experiments of Bautista et al. (1990) and Bautista and Atkins (1991) [4].

Starting at Core Deposition, we can observe the same trend as expected in the literature when considering OVD process [4]: the deposition efficiency grows with the diameter of the preforms. Also, we can see that the big difference in deposition efficiency between the first and the second spindle happens during the core deposition phases (from Core Mode A to Core Mode I). As expected, we observed that Core Weight Mode 'J' re-establishes the balance between the two spindles. Indeed, after this mode, deposition efficiencies are equal.

Note: Figure 3.9 has been plotted for a preform which is in the stable zone identified in figure 3.8. When we plotted this graph for different preforms one next to each other, we can observe a large disparity between the plots which again suggests a lack of process control.

3.2 Experiments on Rod Profile data

3.2.1 Refractive Index Profile

The refractive index profile shows how the refractive index changes across the geometry of the fiber or preform. It is generally measured across the central axis of the end face, indicating the extent of core and cladding.

How is it measured? Preform Analyzer measures the deflection angle of a narrow laser beam when it refracts through a cross-section of the preform. As the beam scans across the preform, it encounters different index features in the preform causing its deflection angle to change.

Once the entire cross-section of the preform has been scanned by the beam, a mathematical transformation (the Abel transformation) is used to determine the index structure in the preform that created the measured deflection function. Sterlite also stretches a portion of the rod to be analysed in order to produce less deflection angle distortions. More explanation about rod profile measurement can be found in the documents Core Profiler and Methods [10] and STL-OF core manufacturing [2].

3.2.2 Spatial Characterisation of Rod Profile Variation

Using rod profile data, we attempted to define a metric to characterized the homogeneity of rods from the same mother preform. The idea is to compare profile curves point by point and for each point to compute the standard deviation between the rods. It gives the kind of profile we can observe on figure 3.12. For example, top right preform seems to present a greater homogeneity in rod profiles than top left preform.

There is no follow up on this method for now.

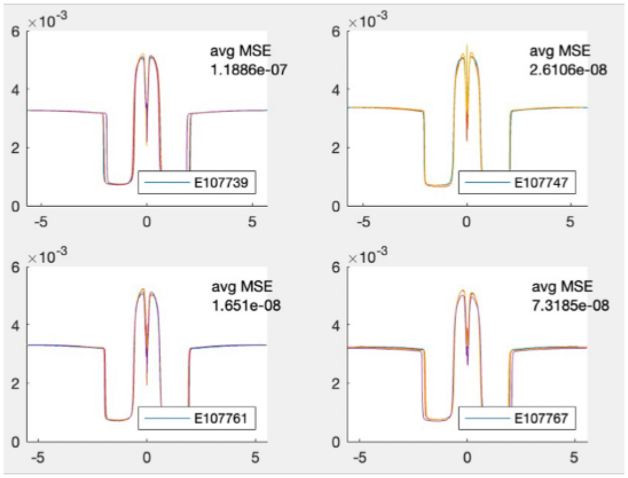


Figure 3.11: Rod profile (refractive index) plotted against position in the rod cross section. 4 rods from one preform are plotted each time.

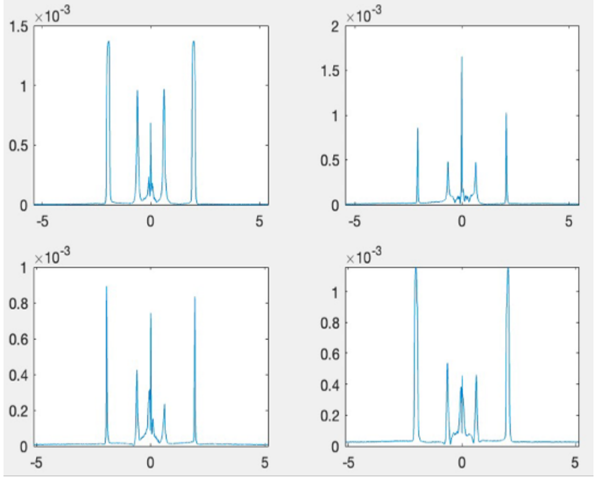


Figure 3.12: Standard deviation between rods refractive index in function of the position in the rod cross section.

3.3 Partial Conclusion

As explained in section 2.2, it seems like the control instabilities encountered by Sterlite are mainly related to inconsistent weight and diameter acquisition. In the case of weight for example, load cells measurement takes are disturbed by the extreme temperature inside the deposition machine (because the calibration is done at room temperature) and the movement of the traverse. Before even considering changing the sensors (which should not be taken lightly considered that the deposition is sealed and operating at very high temperature), one significant improvement could be increasing the data acquisition frequency. Indeed, more sample points could give better insight on the dynamics underlying during the acquisition. Also, if those dynamics can be fully characterized, there is a chance that they can be filtered out.

Last but not least, we know that some manual adjustments are done on site by directly modifying the recipe in order to balance performances over the machines since the 8 spindles should perform the same. Having access to the recipe for each preform could help understand those habits and their impact on the process variables measured.

This page was intentionally left blank.

Chapter 4

Analysis

The reception of new sets of data covering the entire process from deposition to IPQ motivated a new approach: to zoom out times series and study overall patterns and trends in the data. The goal is to demonstrate relationship between preform to IPQ measurements and even preform to fiber draw measurements.

4.1 Feature Engineering

4.1.1 Data Pre-processing

The first problematic was to translate times series data into feature vectors specific to each preform. To do so, the initial strategy was to consider that the more features the better. Thus, for each preform, we started by calculating the mean and variance of every variable measured per stage. So, one run of one variable measurements will be characterized by twice the number of stages in the process (mean and variance) and one preform run will be characterized by a feature vector of size being twice the number of stages time the number of variables acquired during the run. Then all those feature vectors are concatenated in a feature matrix in which each row is associated to one preform. One limit of this strategy is the large size the feature vectors can get which makes it hard to track back which feature is originally calculated using which stage/variable. This is especially limiting when doing Principal Component Analysis on the feature matrix. One way to fix this would be to document in a string array the name of each feature column.

Comments: If the process doesn't present any stages (rod draw for example), we average on the entire run. Sintering and IPQ data set are not time series so already consist of a sort of feature matrix once non numerical and process set up columns are removed. In general, some cleaning has been necessary in order to only consider varying measured variables relevant to the course of the operation. Last but not least, each column (corresponding to one feature over all preforms) is standardized so that, in the following steps, each of the variables contributes equally to the analysis.

4.1.2 Feature Wise Correlation Analysis

Linear dimensionality reduction Once feature matrices are obtained, we can learn about implicit trends in the data. One way to start is by applying Principal Component Analysis (PCA) which is a dimensionality reduction method that helps transform a large set of variable into a smaller one without losing information (variation).

How does PCA work? In short, the PCA algorithm first compute the covariance matrix of the data set (i.e. how the variables of the input data set are varying from the mean with respect to each other), then, the eigenvectors and eigenvalues of the covariance matrix are computed to identify the principal components. The principal components are new variables (axes) that are constructed as linear combinations or mixtures of the initial variables (axes). Those combinations are such that the new variables/axes (i.e., principal components) are uncorrelated and most of the information within the initial variables is 'captured by' the first components.

For example, a 10-dimensional data gives you 10 principal components, but PCA orders the new components/axes to put maximum information (variation) in the first component, then maximum remaining information (variation) in the second and so on. The algorithm also gives you the percentage of variation in the data set explained by each of the component so that you can chose only the components with the greater variation explanation in order to reduce dimensionality without losing much information. It should be noted that principal components may be less directly interpretable than the original individual variables but will have meaning relative to physical process.

Now that non significant axes are removed, the eigenvectors associated to the selected components form the new feature axes. Finally, we use the new feature axes to reorient the data from the original axes and are able to visualize variations in a lower dimensionality space.

As mentioned above, the larger the feature matrix is, the less interpretable the PCA analysis is going to be. For our case, it gave interesting insights on soaking operation and IPQ data. Anyway, first proceeding to a PCA dimensionality reduction before any other manipulation helped reducing significantly computation time.

Comments PCA confirmed that both mean and variance are relevant in capturing variation among the run measurements. Also, since PCA uses the global covariance matrix to reduce data, you can get that matrix and apply it to a new set of data with the same result. That's helpful when you need to try to reduce your feature list and reuse matrix created from train data.

Results The strategy for principal components selection was to chose eigenvectors explaining more than 0.5% of the data set variations. The results are describe in the table 4.1. We will use those dimensionality reduced space for the rest of our analysis in order to decrease computing time.

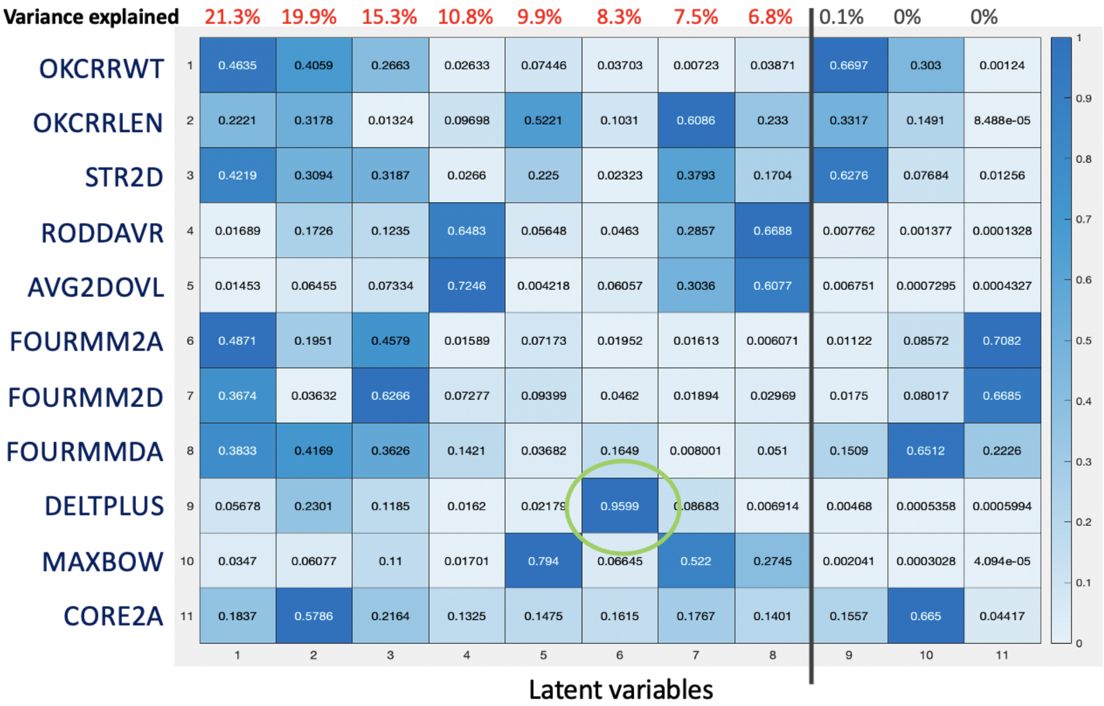
The soaking data set should be considered as an exception as only two of its columns (Soaking Time and N2) are independent. The other columns are either proportional or constant.

Therefor, there is no need to apply PCA for dimensionality reduction. Also, plotting soaking variation on the two independent axis hasn't given any satisfying results.

The IPQ data set presenting only 11 variables, it has been interesting to visualize the eigenvector coefficients associated to each initial variable. On figure 4.1, you can see for example that DELTPLUS (refractive index, cf annex dictionary 6.1) alone is explaining about 8% of the overall variations observed in the data set. This highlights the refractive index as a strong IPQ feature.

	Before PCA	After PCA
Deposition	720	30
Sintering	289	33
Soaking	/	/
Rod draw	76	40
IPQ	11	8

Table 4.1: Data reduction after PCA for deposition, sintering, rod draw adn IPQ.



Variance explained		21.3%	19.9%	15.3%	10.8%	9.9%	8.3%	7.5%	6.8%	0.1%	0%	0%
OKCRRWT	1	0.4635	0.4059	0.2663	0.02633	0.07446	0.03703	0.00723	0.03871	0.6697	0.303	0.00124
OKCRRLEN	2	0.2221	0.3178	0.01324	0.09698	0.5221	0.1031	0.6086	0.233	0.3317	0.1491	8.488e-05
STR2D	3	0.4219	0.3094	0.3187	0.0266	0.225	0.02323	0.3793	0.1704	0.6276	0.07684	0.01256
RODDAVR	4	0.01689	0.1726	0.1235	0.6483	0.05648	0.0463	0.2857	0.6688	0.007762	0.001377	0.0001328
AVG2DOVL	5	0.01453	0.06455	0.07334	0.7246	0.004218	0.06057	0.3036	0.6077	0.006751	0.0007295	0.0004327
FOURMM2A	6	0.4871	0.1951	0.4579	0.01589	0.07173	0.01952	0.01613	0.006071	0.01122	0.08572	0.7082
FOURMM2D	7	0.3674	0.03632	0.6266	0.07277	0.09399	0.0462	0.01894	0.02969	0.0175	0.08017	0.6685
FOURMMDA	8	0.3833	0.4169	0.3626	0.1421	0.03682	0.1649	0.008001	0.051	0.1509	0.6512	0.2226
DELTPLUS	9	0.05678	0.2301	0.1185	0.0162	0.02179	0.9599	0.08683	0.006914	0.00468	0.0005358	0.0005994
MAXBOW	10	0.0347	0.06077	0.11	0.01701	0.794	0.06645	0.522	0.2745	0.002041	0.0003028	4.094e-05
CORE2A	11	0.1837	0.5786	0.2164	0.1325	0.1475	0.1615	0.1767	0.1401	0.1557	0.665	0.04417
		1	2	3	4	5	6	7	8	9	10	11

Figure 4.1: PCA results on IPQ data set.

4.2 Consistent Labeling Mapping

4.2.1 Non-Linear Dimensionality Reduction

PCA being limited to linear dimensionality reduction (which is likely not the best fit for the data we have access to), it can be interesting to use an algorithm that deals with linearly non-separable data.

t-distributed stochastic neighborhood embedding (t-SNE) t-distributed stochastic neighborhood (t-SNE) was developed by Laurens van der Maaten and Geoffrey Hinton in 2008. This probabilistic algorithm is mostly used to understand high-dimensional data and project it into low-dimensional space (like 2D or 3D). Specifically, it models each high-dimensional object by a two- or three-dimensional point in such a way that similar objects are modeled by nearby points and dissimilar objects are modeled by distant points with high probability. t-SNE is iterative so, unlike PCA, it cannot apply on another dataset.

How does t-SNE work? The t-SNE algorithm comprises two main stages. First, t-SNE constructs a probability distribution over pairs of high-dimensional objects in such a way that similar objects are assigned a higher probability while dissimilar points are assigned a lower probability. Second, t-SNE defines a similar probability distribution over the points in the low-dimensional map, and it minimizes the Kullback–Leibler divergence (KL divergence) between the two distributions with respect to the locations of the points in the map. While the original algorithm uses the Euclidean distance between objects as the base of its similarity metric, this can be changed as appropriate.

t-SNE has been used for visualization in a wide range of applications, including genomics, computer security research, natural language processing, music analysis, cancer research, bioinformatics and biomedical signal processing.

While t-SNE plots often seem to display clusters, the visual clusters can be influenced strongly by the chosen parameterization (e.g. perplexity, number of steps, learning rate, distance) and therefore a good understanding of the parameters for t-SNE is necessary[1]. Such "clusters" can be shown to even appear in non-clustered data, and thus may be false findings. Interactive exploration is thus necessary to choose parameters and validate results [11]. If it has been demonstrated that t-SNE is often able to recover well-separated clusters, it should be used for visualisation only and rather not for clustering afterwards.

Comments t-SNE might be less useful when wanting to perform dimensionality reduction for ML training (cannot be reapplied in the same way). It's not deterministic and iterative so each time it runs, it could produce a different result.

[1] See MATLAB documentation

4.2.2 Examples of Visualization

First attempts on deposition data per variable The idea behind visualizing data sets in a 2-dimensional space is to test if we can visually identify some clusters/trends. In addition to that, we want to colorize each point (representing a preform feature vector reduced in a 2D space) according to some labels we know and see if this underlying knowledge can explain the clusters observed or not. For example, figures 4.2 and 4.3 show the machine 1 statistics associated with Head Endburner Hydrogen Flow for each preform reduced in a 2D space. We can see that both chronology and spindles can explain the clusters observed.

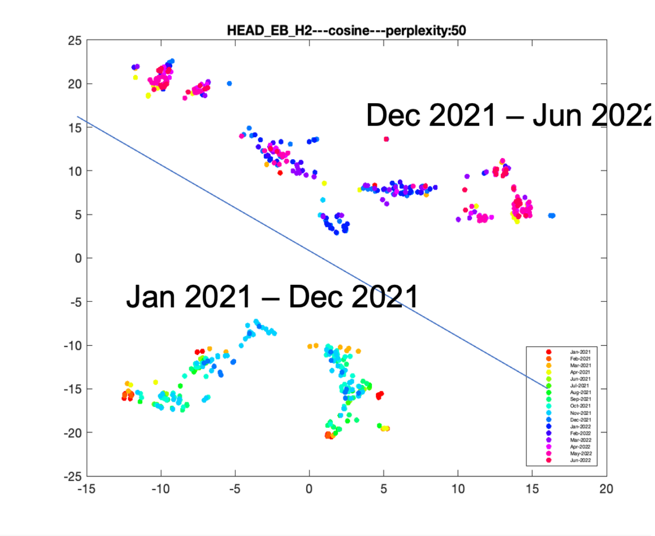


Figure 4.2: t-sne applied to Head Endburner Hydrogen Flow measurements (distance: cosine, perplexity:50) with operation dates labeling.

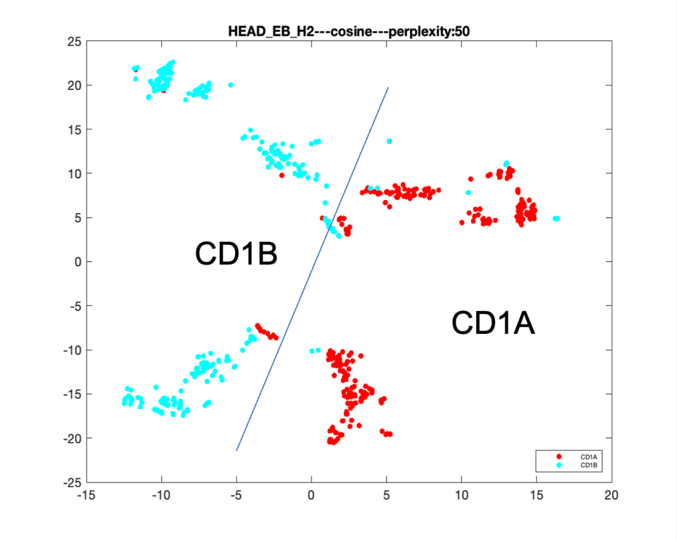
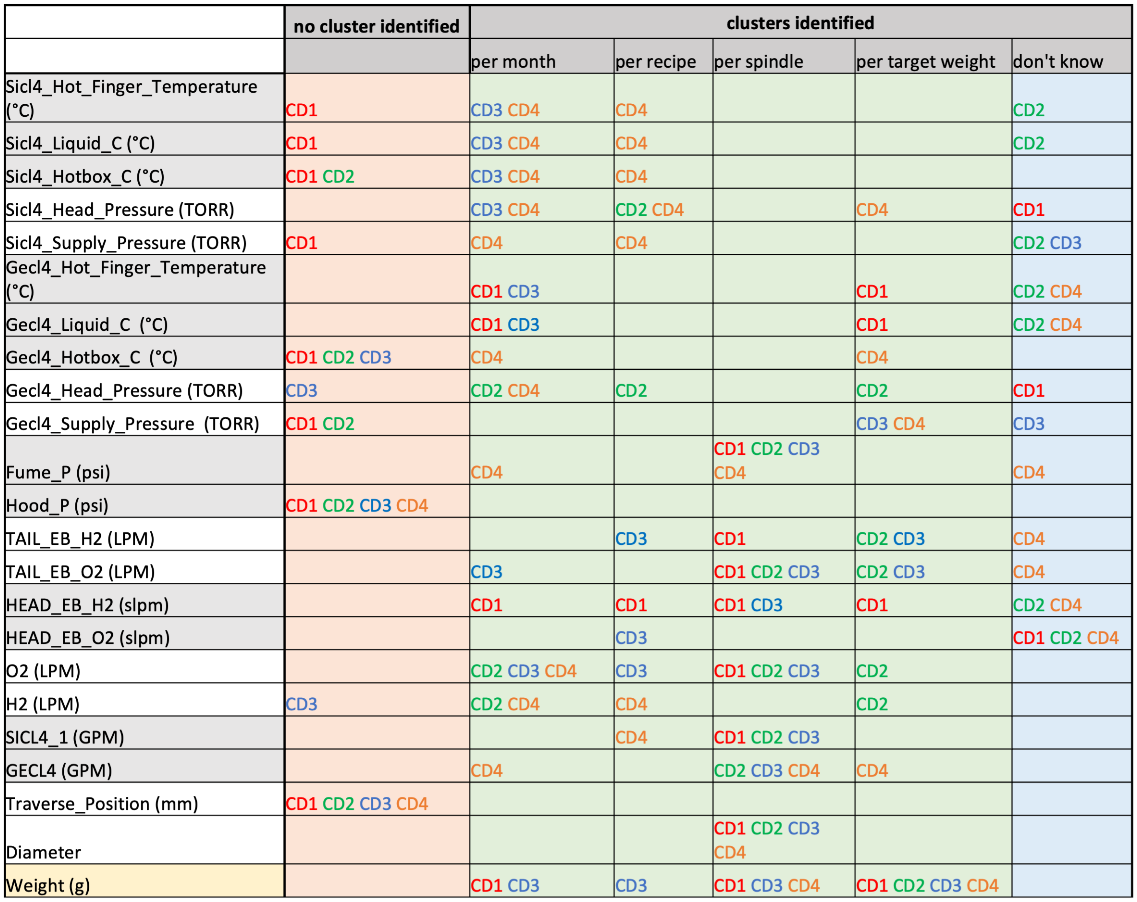


Figure 4.3: t-sne applied to Head Endburner Hydrogen Flow measurements (distance: cosine, perplexity:50) with spindle number labeling.

We iterated the methods on every variables for the four deposition machines and summarized the observation results in figure 4.4. Each time the name of a machine is written in a cell, it means that, for the corresponding row variable, clusters have been observed to be correlated to the column corresponding label (if clusters have been identified). We can observe that, overall, pressure and temperature variable from machine 1 are controlled more uniformly as, most of the time, no cluster have been identified in the variable data.



	no cluster identified	clusters identified				
		per month	per recipe	per spindle	per target weight	don't know
Sicl4_Hot_Finger_Temperature (°C)	CD1	CD3 CD4	CD4			CD2
Sicl4_Liquid_C (°C)	CD1	CD3 CD4	CD4			CD2
Sicl4_Hotbox_C (°C)	CD1 CD2	CD3 CD4	CD4			
Sicl4_Head_Pressure (TORR)		CD3 CD4	CD2 CD4		CD4	CD1
Sicl4_Supply_Pressure (TORR)	CD1	CD4	CD4			CD2 CD3
Gecl4_Hot_Finger_Temperature (°C)		CD1 CD3			CD1	CD2 CD4
Gecl4_Liquid_C (°C)		CD1 CD3			CD1	CD2 CD4
Gecl4_Hotbox_C (°C)	CD1 CD2 CD3	CD4			CD4	
Gecl4_Head_Pressure (TORR)	CD3	CD2 CD4	CD2		CD2	CD1
Gecl4_Supply_Pressure (TORR)	CD1 CD2				CD3 CD4	CD3
Fume_P (psi)		CD4		CD1 CD2 CD3 CD4		CD4
Hood_P (psi)	CD1 CD2 CD3 CD4					
TAIL_EB_H2 (LPM)			CD3	CD1	CD2 CD3	CD4
TAIL_EB_O2 (LPM)		CD3		CD1 CD2 CD3	CD2 CD3	CD4
HEAD_EB_H2 (slpm)		CD1	CD1	CD1 CD3	CD1	CD2 CD4
HEAD_EB_O2 (slpm)			CD3			CD1 CD2 CD4
O2 (LPM)		CD2 CD3 CD4	CD3	CD1 CD2 CD3	CD2	
H2 (LPM)	CD3	CD2 CD4	CD4		CD2	
SICL4_1 (GPM)			CD4	CD1 CD2 CD3		
GECL4 (GPM)		CD4		CD2 CD3 CD4	CD4	
Traverse_Position (mm)	CD1 CD2 CD3 CD4					
Diameter				CD1 CD2 CD3 CD4		
Weight (g)		CD1 CD3	CD3	CD1 CD3 CD4	CD1 CD2 CD3 CD4	

Figure 4.4: Summary of trends observed in deposition data (per variable) which match known labels (like time, machines, recipes and so).

Extending the method to each process Since the ultimate goal is to compare clusters process by process, we apply the same methodology to every process feature matrix. In what we call a process feature matrix, the information of the variable is lost as all variables feature matrix have been concatenated side by side. Note: soaking data set presenting very few variation i.e. information, it has been left on the side for now.

When looking at process data from deposition in figure 4.5 and 4.6, we can see both strong correlation with time of the year and with the machine, even the spindle use. Indeed, it appears that the cluster in the center corresponds almost exclusively to preforms manufactured between January 2021 and December 2021 since the five clusters around it would correspond to the period from January 2022 to June 2022. We can also observe a strong correlation between clusters and spindles especially for machines 1 and 4 where the separation is clear.

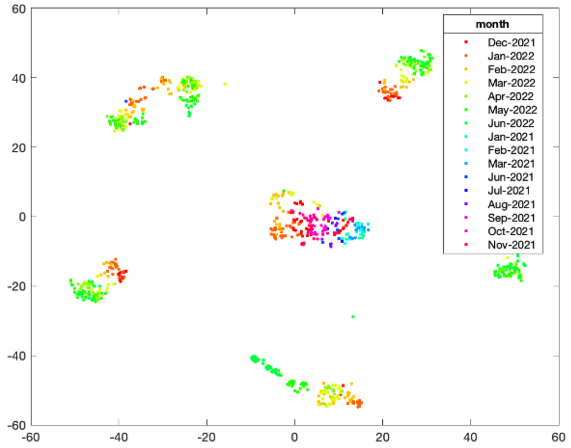


Figure 4.5: t-sne visualization of month labeling for deposition data set.

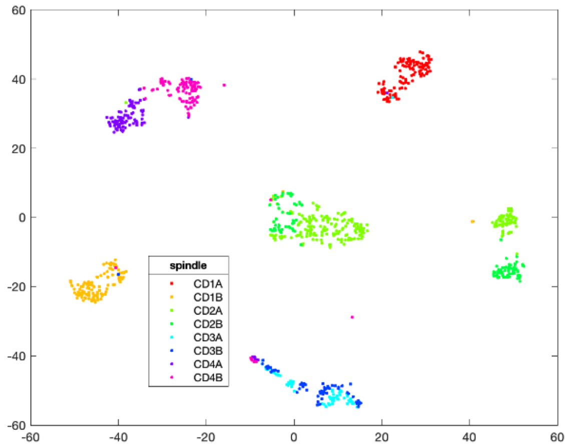


Figure 4.6: t-sne visualization of spindle labeling for deposition data set.

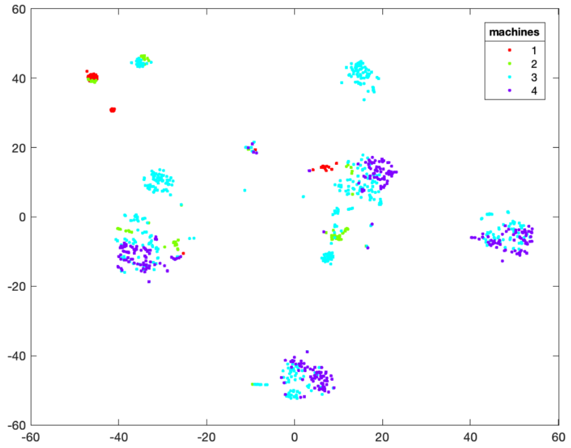


Figure 4.7: t-sne visualization of machines labeling for sintering data set.

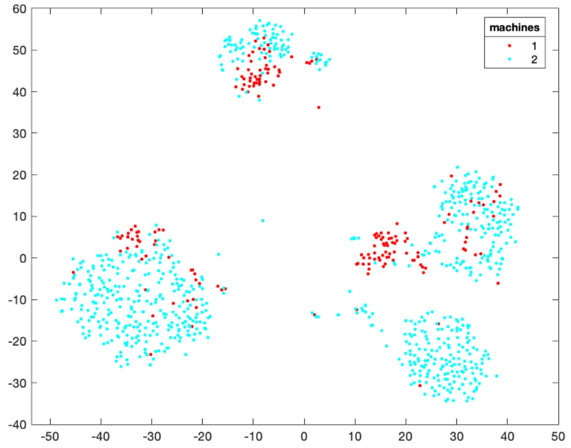


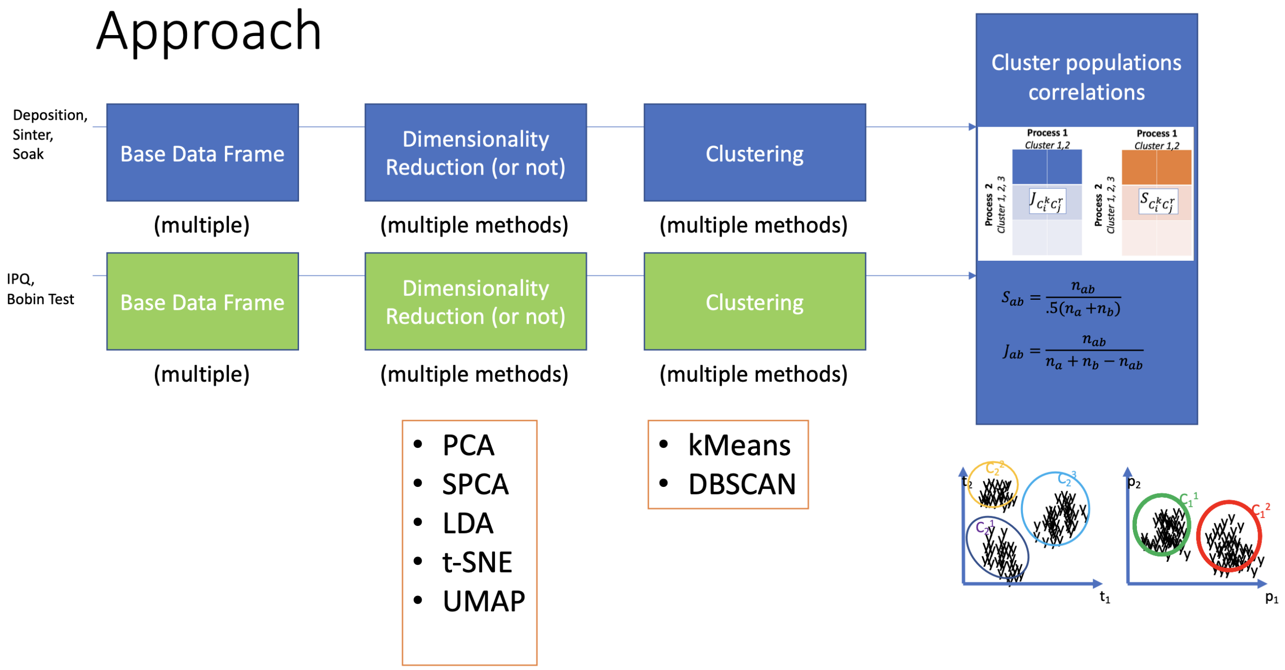
Figure 4.8: t-sne visualization of machines labeling for rod draw data set.

Comments Other methods have been experimented for dimensionality reduction like SPCA, LDA and UMAP but results have not been successful enough to be presented in this work.

4.3 Clusters Counting for Detecting Changes

4.3.1 Approach: to Study Clusters Overlap Between Processes

The figure 4.9 bellow shows the approach as it has been presented to Sterlite team. As we just explained the two first steps (base data frame and dimensionality reduction) in the previous sections, this section focuses on the strategy to actually cluster the data and study the overlap between clusters from one process to another.



Approach

Deposition, Sinter, Soak → Base Data Frame (multiple) → Dimensionality Reduction (or not) (multiple methods) → Clustering (multiple methods) → Cluster populations correlations

IPQ, Bobin Test → Base Data Frame (multiple) → Dimensionality Reduction (or not) (multiple methods) → Clustering (multiple methods) → Cluster populations correlations

- PCA
- SPCA
- LDA
- t-SNE
- UMAP

- kMeans
- DBSCAN

$$S_{ab} = \frac{n_{ab}}{.5(n_a + n_b)}$$

$$J_{ab} = \frac{n_{ab}}{n_a + n_b - n_{ab}}$$

Figure 4.9: Approach presented to Sterlite team for demonstrating relation between Preform-to-Test data.

Clustering It is a good thing to know that there are underlying trends and clusters in the data set and to be able to partially explain them with intuitive labeling but for the sake of the analysis, we need to be able to partition the data according to those clusters using clustering algorithms. We started with the most common one: k-means clustering.

k-means clustering is a method of vector quantization, originally from signal processing, that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean (cluster centers or cluster centroid), serving as a prototype of the cluster. The input taken by the algorithm is the number of clusters k that should be identified. Choosing k is one of the most challenging task of this clustering algorithm. One way to do it could be to use the intuition from the t-sne observations but more conventionally, we prefer to use conventional clusters evaluation methods as the Elbow Method and the Silhouette Method. The two methods are equivalent and should be used in combination the one with the other. However, the Elbow Method is not directly implementable in MATLAB so, in a first approach, we used the Silhouette Method only.

Silhouette Method for Clusters Evaluation This technique is based on the calculation of the silhouette coefficient which measures how similar a data point is within a cluster (cohesion) compared to other clusters (separation). $S(i)$ being the silhouette coefficient pf the data point i, it can be expressed in function of $a(i)$ the average distance between i and all the other data points in the cluster to which i belongs and $b(i)$ the average distance between i and all clusters to which i does not belong the following way: $S(i) = \frac{b(i)-a(i)}{max(a(i),b(i))}$. The coefficient is between [-1,1]. A score of 1 denotes the best meaning that the data point i is very compact within the cluster to which it belongs and far away from the other clusters. The worst value is -1. Values near 0 denote overlapping clusters.

To find the optimal number of clusters, we first select a range of values of k and then the overall average silhouette coefficient for each value of k. Thus, the optimal k is the one maximizing the average silhouette coefficient. The figure 4.10 shows the evaluation for deposition, sintering, rod draw and IPQ data sets. We chose cosine for the distance as it was the one giving the best results in term of data partitioning when using t-sne visualization (i.e. clearest separation of clusters the one from the others).

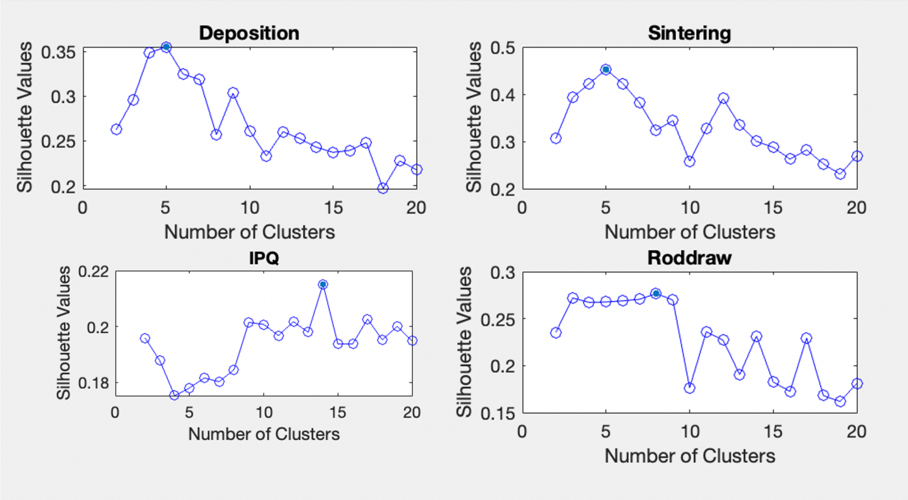


Figure 4.10: Clusters evaluation using Silhouette method for deposition, sintering, rod draw and IPQ data (distance cosine)

Now that we have k for each data set, we can directly apply the k-means algorithm which will produce an array that assigns each preform to a cluster.

Metrics to study overlaps between clusters The Dice's coefficient and the Jaccard coefficient are statistic used to gauge the similarity of two samples. In our case, we will use them to estimate the overlap between clusters from process 1 (e.g. deposition) and clusters from process 2 (e.g. sintering) and so on following the process chain until IPQ measurements. Given two sets of data X and Y, the Dice's coefficient equals twice the number of elements common to both sets divided by the sum of the number of elements in each set: $DSC = \frac{2|X \cap Y|}{|X|+|Y|}$ where |X| and |Y| are the cardinalities of the two sets (i.e. the number of preforms in each cluster). Whereas the Jaccard coefficient is defined as the size of the intersection divided by the size of the union of the sample sets. Both are equivalent with a close difference: Dice does not satisfy the triangle inequality, hence $1-DSC$ can not properly be used as a metric, while $1-J$ can. However, Dice's coefficient is used more often because it gives bigger coefficient than Jaccard's. For the same reason, we chose to present results in the form of a Dice's coefficient matrix.

4.3.2 Results

First thing we can observe especially on figure 4.12 is that, although the silhouette method indicated a number of clusters that matches the ones visualized with t-sne, k-means identified some out layers which haven't been considered as so by the cosine metrics. This created some single points clusters (clusters 3, 5 and 6) which are inconvenient for our approach. All in all, we don't observe any strong overlap between clusters from deposition and from sintering.

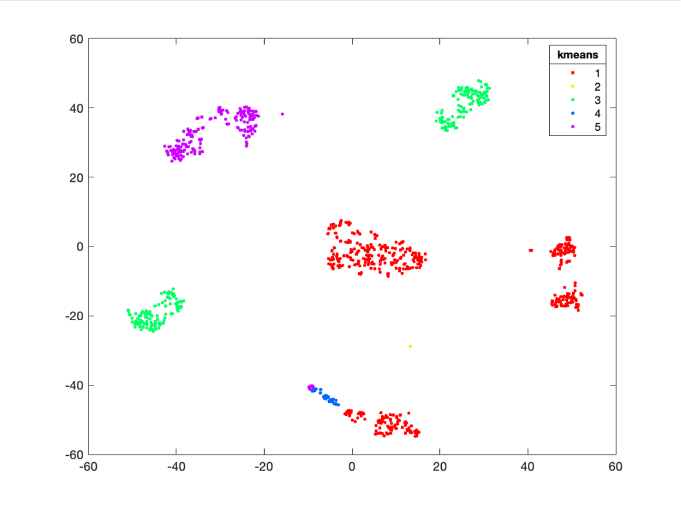


Figure 4.11: t-sne visualization of k-means clustering for deposition data set.

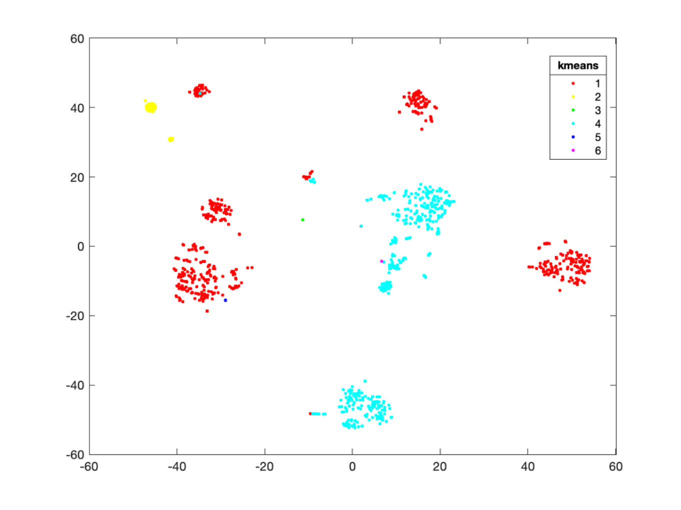
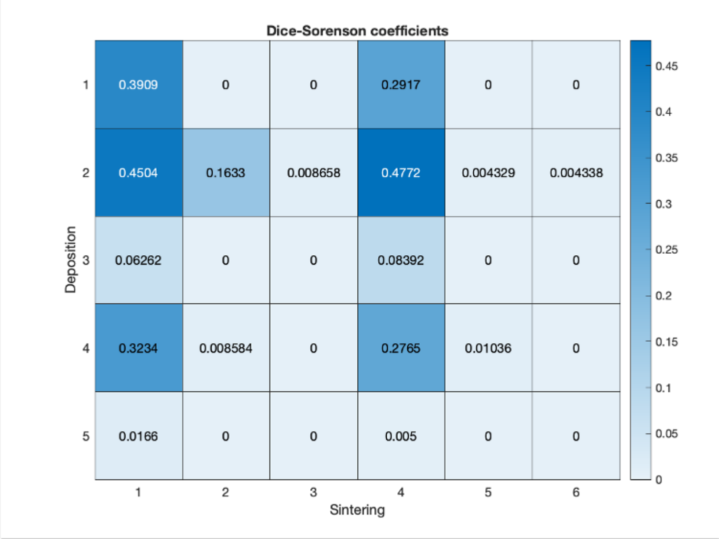


Figure 4.12: t-sne visualization of k-means clustering for sintering data set.



	1	2	3	4	5	6
1	0.3909	0	0	0.2917	0	0
2	0.4504	0.1633	0.008658	0.4772	0.004329	0.004338
3	0.06262	0	0	0.08392	0	0
4	0.3234	0.008584	0	0.2765	0.01036	0
5	0.0166	0	0	0.005	0	0

Figure 4.13: Dice's coefficient matrix showing the overlap between clusters from deposition and sintering data sets.

Comparing each process clusters to IPQ clusters (figures 4.14, 4.16 and 4.15) also gave inconclusive results as none of the Dice's coefficients observed have greater value than 0.4.

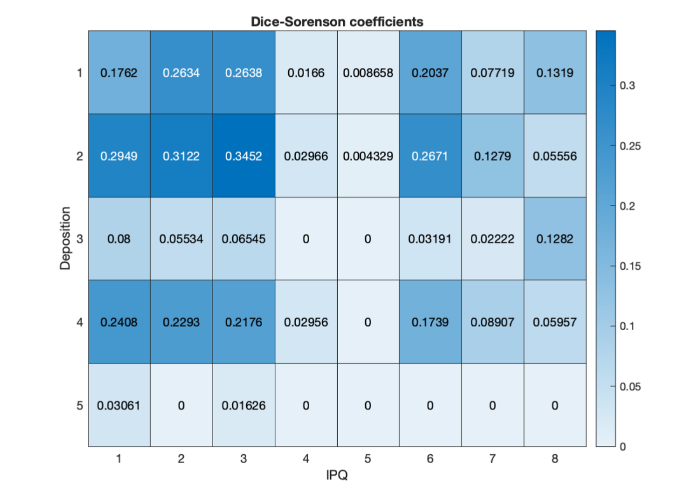


Figure 4.14: Dice's coefficient matrix showing the overlap between clusters from deposition and IPQ data sets.

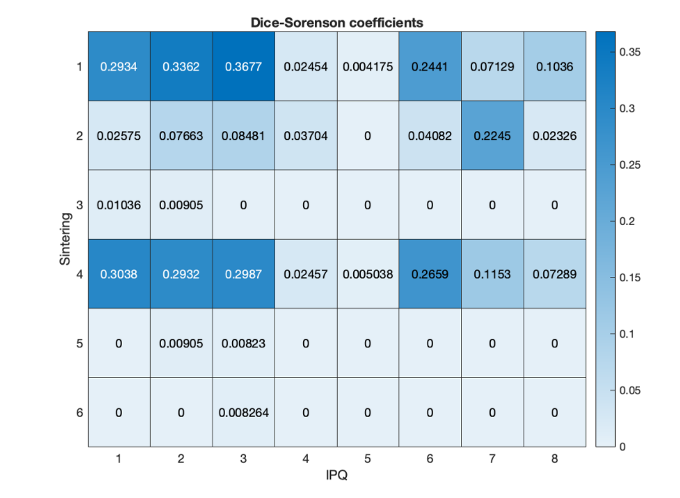


Figure 4.15: Dice's coefficient matrix showing the overlap between clusters from sintering and IPQ data sets.

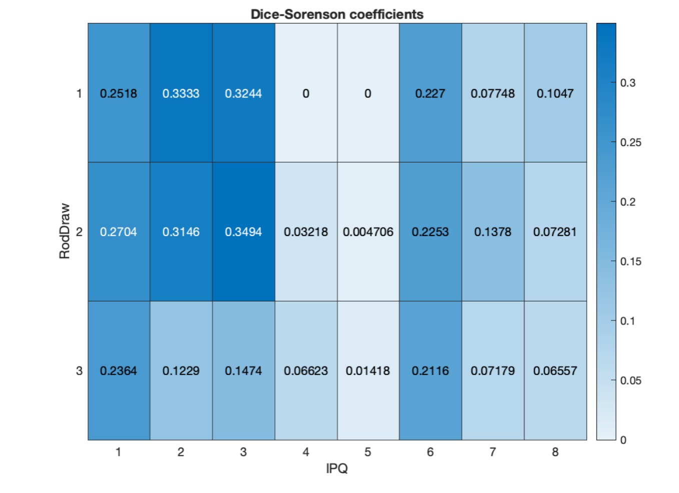


Figure 4.16: Dice's coefficient matrix showing the overlap between clusters from rod draw and IPQ data sets.

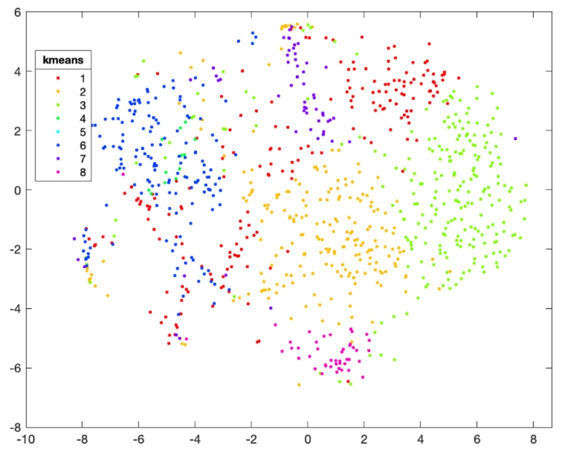


Figure 4.17: t-sne visualization of k-means clustering for IPQ data set.

One last tentative was to concatenate all stats from all processes together and compare those clusters (figure 4.18) to the ones from IPQ (figure 4.19). This time, we made sure to adjust the number of clusters suggested by the silhouette method in order to limit single point clusters. This way, we have been able to obtain slightly greater Dice's coefficients up to 0.46.

Figure 4.21 shows the results of considering not only the clusters from all processes data set but also their intersection when computing the Dice's coefficient. This highlights again no particular trend in clusters overlap.

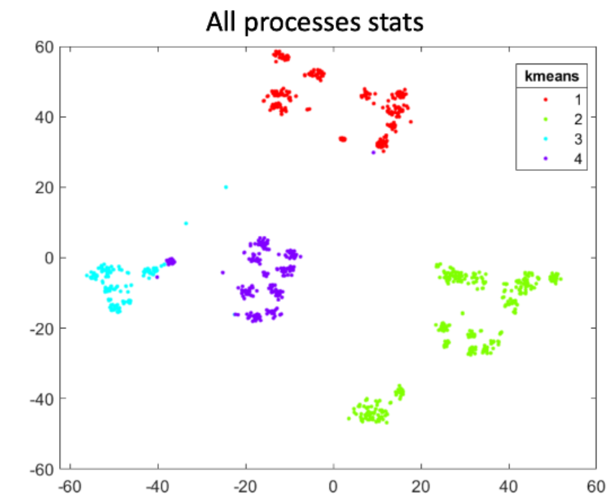


Figure 4.18: t-sne visualization of k-means clustering for IPQ data set.

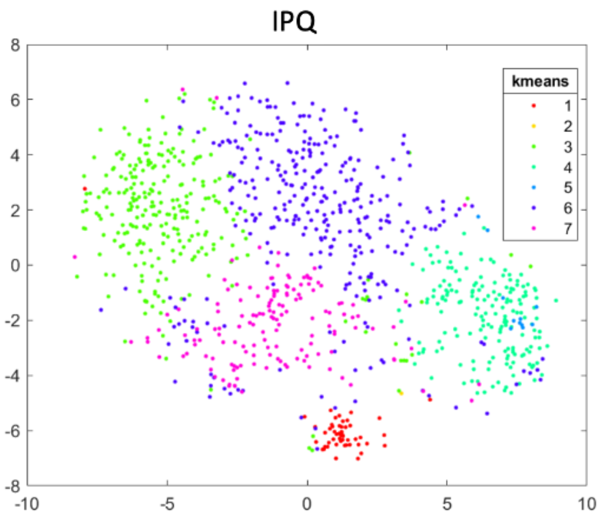
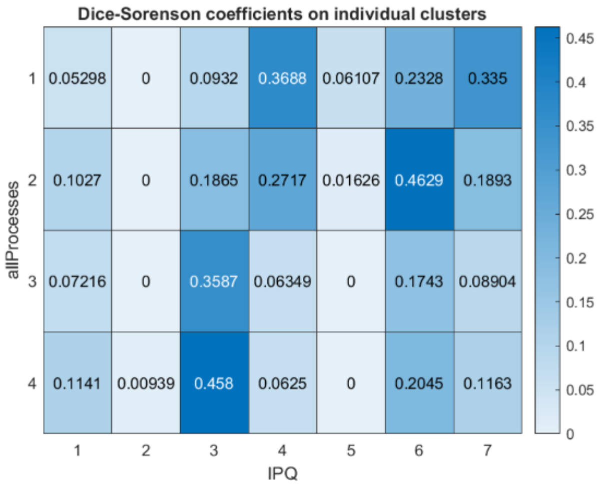
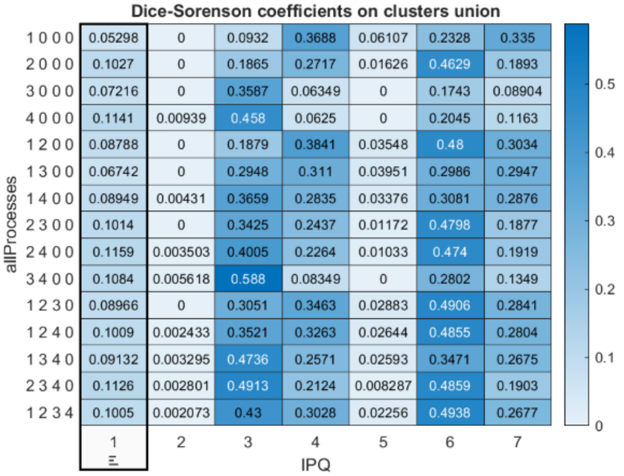


Figure 4.19: t-sne visualization of k-means clustering for IPQ data set.



	1	2	3	4	5	6	7
1	0.05298	0	0.0932	0.3688	0.06107	0.2328	0.335
2	0.1027	0	0.1865	0.2717	0.01626	0.4629	0.1893
3	0.07216	0	0.3587	0.06349	0	0.1743	0.08904
4	0.1141	0.00939	0.458	0.0625	0	0.2045	0.1163

Figure 4.20: Dice's coefficient matrix showing the overlap between clusters from rod draw and IPQ data sets.



	1	2	3	4	5	6	7
1 0 0 0	0.05298	0	0.0932	0.3688	0.06107	0.2328	0.335
2 0 0 0	0.1027	0	0.1865	0.2717	0.01626	0.4629	0.1893
3 0 0 0	0.07216	0	0.3587	0.06349	0	0.1743	0.08904
4 0 0 0	0.1141	0.00939	0.458	0.0625	0	0.2045	0.1163
1 2 0 0	0.08788	0	0.1879	0.3841	0.03548	0.48	0.3034
1 3 0 0	0.06742	0	0.2948	0.311	0.03951	0.2986	0.2947
1 4 0 0	0.08949	0.00431	0.3659	0.2835	0.03376	0.3081	0.2876
2 3 0 0	0.1014	0	0.3425	0.2437	0.01172	0.4798	0.1877
2 4 0 0	0.1159	0.003503	0.4005	0.2264	0.01033	0.474	0.1919
3 4 0 0	0.1084	0.005618	0.588	0.08349	0	0.2802	0.1349
1 2 3 0	0.08966	0	0.3051	0.3463	0.02883	0.4906	0.2841
1 2 4 0	0.1009	0.002433	0.3521	0.3263	0.02644	0.4855	0.2804
1 3 4 0	0.09132	0.003295	0.4736	0.2571	0.02593	0.3471	0.2675
2 3 4 0	0.1126	0.002801	0.4913	0.2124	0.008287	0.4859	0.1903
1 2 3 4	0.1005	0.002073	0.43	0.3028	0.02256	0.4938	0.2677

Figure 4.21: Dice's coefficient matrix showing the overlap between clusters and their intersections from all processes concatenated data set and IPQ data sets.

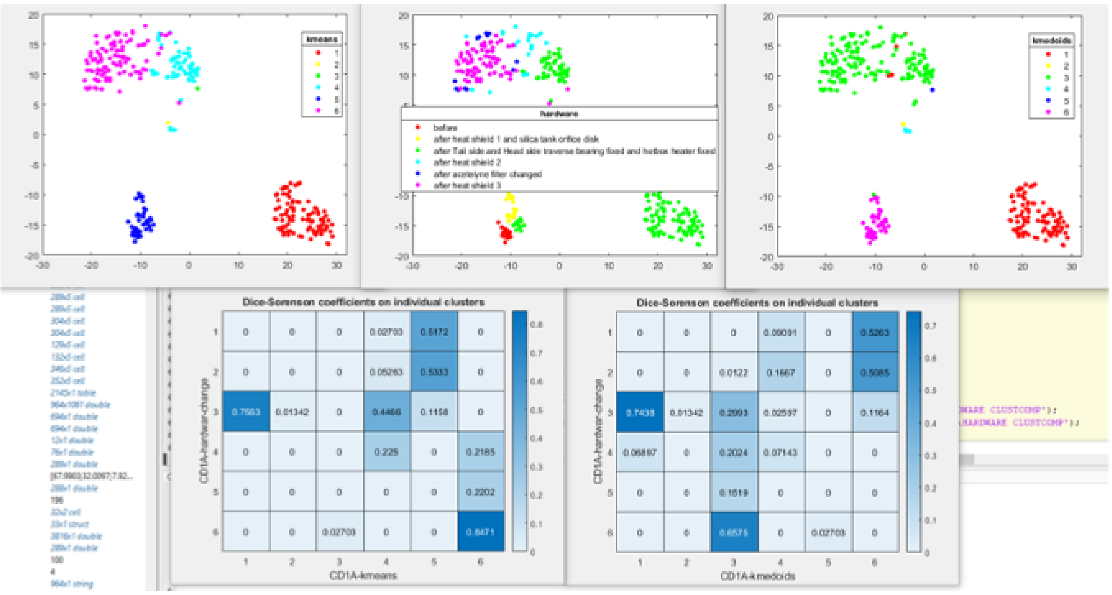
Comments Results from this approach are mixed as we would have expected less mix match between the clusters and more strong coefficients standing out in the Dice's coefficient matrix. Dice's coefficient may also not be the best suited metric for this type of exercise. Indeed, coefficient order of magnitude is dependant from the relative sizes of the two clusters, so it is hard to compare between each other coefficients on the same row or column.

However, this method was initially applied to the entire set of preforms available (i.e. not the preforms only present in every operation data sets) and has helped to identify and track disappearance of group of preforms throughout the processes.

4.3.3 Approach: to Study Clustering Algorithms Known Information Capturing

The reception of hardware change history for deposition machines has created new opportunities for the cluster overlap comparison methodology. In fact, Dice's coefficient matrices can be used for studying how well clustering algorithms can capture know information about the data.

In figure 4.22, we studied the performance differences between k-mean algorithm and k-medoids algorithm. To do so, we compared clusters corresponding to different hardware periods (those have been assigned by hand) to the clusters found by the algorithm. An ideal case would be a matrix with only one coefficient equal to 1 in each row and each column.



Dice-Sorenson coefficients on individual clusters

CD1A-hardware-change \ CD1A-kmeans	1	2	3	4	5	6
1	0	0	0	0.02703	0.5172	0
2	0	0	0	0.05263	0.5333	0
3	0.7563	0.01342	0	0.4466	0.1158	0
4	0	0	0	0.225	0	0.2185
5	0	0	0	0	0	0.2202
6	0	0	0.02703	0	0	0.8471

Dice-Sorenson coefficients on individual clusters

CD1A-hardware-change \ CD1A-kmedoids	1	2	3	4	5	6
1	0	0	0	0.09091	0	0.5263
2	0	0	0.0122	0.1667	0	0.5085
3	0.7438	0.01342	0.2993	0.02597	0	0.1164
4	0.06897	0	0.2024	0.07143	0	0
5	0	0	0.1519	0	0	0
6	0	0	0.6575	0	0.02703	0

Figure 4.22: Example of studying how coherent clusters identified by k-means/k-medoids algorithms are with know deposition machines hardware changes.

4.4 Regression

Now that we have a better understanding about the underlying trends in the data sets and what seem to be the sources of influence, we can start developing predictive models using those existing data to identify and model relationships between input controls and material and output properties.

4.4.1 Input Data Preparation

First, we want to create matching size input and output data sets. Indeed, all processes information are logged per preform since IPQ measurements are logged per rod. To do so, we choose to consider each rod measurement as an output observation for the associated mother preform. So we will replicate the mother preform row in the input the number of times a measurement on one of its rod has been logged. The approach is illustrated in figure 4.23.

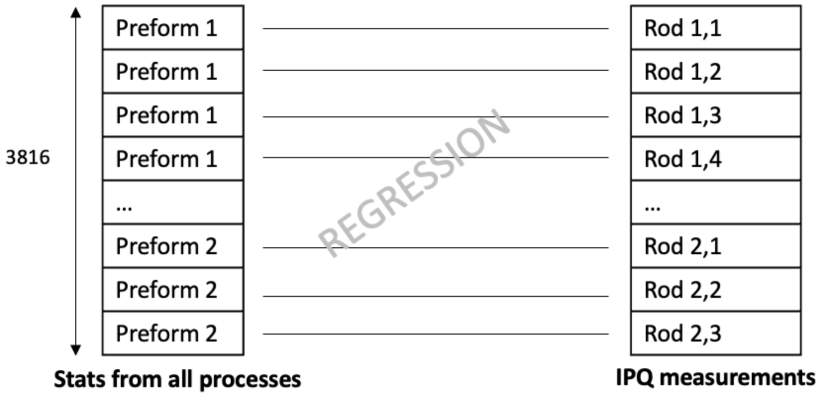


Figure 4.23: Input data preparation for IPQ measurements prediction using regression.

Input We will consider several data sets as inputs in order to compare regression performances when fed with different types of data. Those are the following:

- All statistics from all processes concatenated in a big matrix (1061 variables). Raw data are used but also PCA reduced data.
- Sequence of machines the preform has been going through and deposition dates encoded in hot-ones vectors.
- Sintering and deposition recipes encoded in hot-ones vectors.
- Rod number (measurements being always made in the same order) encoded in hot-ones vectors.

Output At first we have considered predicting all IPQ measurements at one time (11 variables), then only the first principal component (using PCA) and finally each variable individually. Only predicting DELTPLUS (refractive index) alone gave significantly good results so that is the direction we chose to take.

Performance Metric To compare the different models, we will compute the coefficient of correlation between test output and predicted output generated using test input and trained network.

4.4.2 Regression Experiments

We built and trained different families of algorithm (neural networks, regression tree and random forest, kernel, lasso regressor, support vector machines) on several combinations of the inputs and outputs cited above, the best results are summarized in the following tables 4.2, 4.3, 4.4 and 4.5.

"PC" refers to the fact that only the principal components have been selected (i.e. explaining more than 1% of the variations).

Neural Network

Input	Output	Output test vs predicted correlation coefficient
all_processes_stat_PC Machines Dates	DELTPLUS	0.64
all_processes_stats Machines Dates Recipes	DELTPLUS	0.61
all_processes_stats	DELTPLUS	0.60
all_processes_stat_PC	DELTPLUS	0.60
all_processes_stats_PC Machines Dates Recipes	DELTPLUS	0.58
all_processes_stats Machines Dates	DELTPLUS	0.52
all_processes_stat_PC	all_IPQ_PC	no convergence when training
all_processes_stats	all_IPQ	no convergence when training

Table 4.2: Regression testing results when training a 6 layers Neural Network (FullyConnectedLayer - 500 - 300 - 150 - 100 hidden units - FullyConnectedLayer).

Regression Tree

Input	Output	Output test vs predicted correlation coefficient
all_processes_stats Machines Dates Recipes	DELTPLUS	0.75
all_processes_stats	DELTPLUS	0.75
Machines	DELTPLUS	0.73
all_processes_stats Machines Dates	DELTPLUS	0.72
all_processes_stats_PC	DELTPLUS	0.71
all_processes_stats_PC Machines Dates Recipes	DELTPLUS	0.70
all_processes_stat_PC Machines Dates Recipes	all_IPQ_PC	0.11
all_processes_stats	all_IPQ_PC	0.01

Table 4.3: Regression testing results when training a Regression Tree (MATLAB automatic hyperparameters optimization).

Random Forest

Input	Output	Output test vs predicted correlation coefficient
all_processes_stats	DELTPLUS	0.73
all_processes_stats Machines Dates	DELTPLUS	0.73
all_processes_stats_PC	DELTPLUS	0.73
all_processes_stats Machines Dates Recipes	DELTPLUS	0.73
all_processes_stats_PC	all_IPQ	0.147

Table 4.4: Regression testing results training a Random Forest (80 trees).

Kernel

Input	Output	Output test vs predicted correlation coefficient
all_processes_stats	DELTPLUS	0.75
all_processes_stats Machines Dates	DELTPLUS	0.75
all_processes_stats Machines Dates Recipes	DELTPLUS	0.75
Machines Dates Recipes	DELTPLUS	0.73
Machines Dates	all_IPQ	0.73
Machines Dates Rod Numbers	DELTPLUS	0.73

Table 4.5: Regression testing results training a Kernel (automatic hyperparameters optimization).

Comments Rod number have been introduced after that this first serie of experiment has been done, therefor this data set doesn't appear in the table. However, different combination with encoded machines sequence and dates but also raw data were tried for Kernel training and gave slightly worst results for predicting DELTPLUS than when training without this additional information about the rods. However, influence on training of rods number as an input has not been tried yet for predicting other variables than DELTPLUS.

Results We obtain best prediction results (in terms of correlation coefficient and robustness) when applying a Kernel algorithm on raw statistics data from all processes to predict the refractive index with a correlation coefficient of 0.75. Kernel algorithm is all the more interesting as it is the only algorithm which has been able to predict all IPQ measurement at once with a good correlation coefficient of 0.73.

In addition to this, the main takeaways are the following:

- For almost every algorithm used, we obtain similar results when predicting the refractive index DELTPLUS using the raw data directly or only the encoded information about the machines sequence the preform has seen and the fabrication date. This opens the way for input ablation strategy. Indeed, we could train the algorithm only on certain sequences to see if it increase the correlation coefficient and in fine develop some tool to suggest the best machines sequence for a desired refractive index.

- Adding information about the recipes doesn't seem like to improve DELTPLUS prediction. It is still to be tested for other IPQ variables.
- DELTPLUS is the only variable which we are able to "predict" for now (expect with Kernel). Some other strategies should be considered for the rest of the variables.

Nevertheless, it is great news to already see some signal in the data even though data measurement, logging and tracking have room for improvement.

4.5 Partial Conclusion

Several strategies have been implemented in order to analyse existing data:

- Dimensionality reduction with linear and non linear algorithms
- Unsupervised learning using clustering algorithms
- Supervised learning using regression algorithms to predict IPQ data

This way, we have been able to identified strong trends (especially for deposition) in the data sets with data behaviours mainly explained by the information of the machine and the date. This statement has been demonstrated using each of the three strategies listed above.
Last but not least, regression gave the most promising results by attesting to a signal in the data mainly in machines sequence and date information which opens the door to the design of a large variety of predictive tools.

Note: another approach to regression for predicting IPQ measurement could be to do classification using IPQ identified clusters as a label.

This page was intentionally left blank.

Chapter 5

Conclusion

5.1 Conclusion

This thesis demonstrated the feasibility of a data-driven approach on core deposition process in order to identify and improve efficiency losses. Indeed, we pointed out the potential of the current data sets available with the example of an in-depth investigation into core deposition operations. Unsupervised learning methods, zooming out time series data and create features statistics associated with each process, led to the most promising results: we learned about underlying trends in the process which are mainly time (month) dependencies and machine sequence dependencies. We also defined some methods to track trends throughout the process following the path the preform has taken. All in all, those analyses have set the pavement for quality measurement prediction using intelligently chosen input data sets and kernel regression algorithm. We are definitely able to identify some signal in the data beside we are still in the very early phase which looks good for the future of this project.

5.2 Room for Improvement

Considering our results, we are about to reach the end of the stage 1 described in the thesis approach. However, even before considering introducing new sensors in the process, several area of improvement have been identified. The overall priority is in fact to strengthen the data acquisition pipeline which can be presented in three categories:

Data Measurements

- Desynchronize measurements from disturbance sources (e.g. weight measurement vs traverse movement).
- Increase measurement frequency (= decrease sampling time) especially for nosy variables (it would help having a better understanding of the underlying dynamics and possibly help filter noise out).
- Develop new protocols for calibration based on the data.
- Acquire better sensors (diameter, rod profile).

Data Logging

- Make sure that sampling time is constant.
- Codify everything that falls under comments in data set (so it can be used as categorical labeling).
- Log more information about where the measurement has been made (especially for rods) and the overall context (e.g. hardware changes).

Data Tracking

- Dedicate one person to verify that all data acquired are coming through up to the cloud until the pipeline is robust enough.
- Create consistent protocols for pulling request. Same for merging data sets.
- Document more (e.g. operation recipes).

The scope of this project could also be extended by collaborating with Sterlite R&D teams who worked on physics modeling of the process (COMSOL, Fluent/CFD) as well as by having access hot stage microscope images from quality measurement (open the way to vision machine learning).

5.3 Future Work

While waiting for more data (or cleaner data set), the following tasks can be attested:

- Study each process in-depth as it has been presented for deposition operations (see chapter 3). Map data with new type of labelling like hardware changes, operators comments or information from IPQ. A way to keep track of the statistics origin (e.g. mean of which variable during which stage) when creating the features vectors could be very helpful.
- For regression, develop ablation strategy for training set: split dataset according to hardware changes, narrow down to only certain machines combinations.
- Try out reverse approach using supervised learning: train algorithm to predict from which machines the raw data have been acquired.

With more data, we could:

- Develop models for nanoscale structure (i.e. porosity) using hot stage microscope (because for now we can know only by destructive methods used one the preform is done), the idea would be to characterize the right temperature and then control it (deposition process).
- Identify, for sintering process, sensors which could be put inside the quartz tube to get temperature profile during sintering (it is more important than better temperature acquisition during deposition).

This page was intentionally left blank.

Bibliography

[1] Sterlite Technologies Limited. Getting bss on cloud strategy. https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwi-5vTe4sr8AhVUrYkEHa5KAWYQFnoECBsQAQ&url=https%3A%2%2Fwww.stl.tech%2Fwebinars%2Fadmin%2Fstorage%2FGetting%2520BSS%2520on%2520Cloud%2520Strategy%2520Right_Webinar_compressed.pdf&usg=AOvVaw3SjRR1CbjpRCX_f8QBItSs. Type: slides deck.

[2] Stl-of core manufacturing. https://www.dropbox.com/home/_MIT_mengm_2022_fiber_sterlite/From%20Sterlite?preview=STL-OF+core+manufacturing.pdf. Type: slides deck.

[3] Stl-of core manufacturing. https://www.dropbox.com/s/sjt61uzda46ch7h/Core%20Dep%20Data.pptx?dl=0. Type: slides deck.

[4] Jitendra Balakrishnan Pushkar Tandon. Predicting heat and mass transfer to a growing, rotating preform during soot deposition in the outside vapor deposition process. *Chemical Engineering Science*, 60:5118–5128, April 2005.

[5] Grand View Research. Fiber optics market size, share and trends analysis report by type(single mode, multimode), by application (telecom, military and aerospace, medical), region, and segment forecasts, 2023 - 2030. https://www.grandviewresearch.com/industry-analysis/fiber-optics-market. Report ID: GVR-1-68038-860-2.

[6] Patents assigned to sterlite technologies limited. https://patents.justia.com/assignee/sterlite-technologies-limited. Actualized January 2023.

[7] Pushkar Tandon. Fundamental understanding of processes involved in optical fiber manufacturing using outside vapor deposition method. *International Journal of Applied Ceramic Technology*, 2(6):504–513, 2005.

[8] George C. Chen. A data-driven approach to system dynamics modeling and control design. *Massachusetts Institute of Technology*, May 2022.

[9] Badri Gomatam. Traverse velocity and deposited weight analysis: Uncontrolled weight gain in core preform. https://www.dropbox.com/s/wwtdzonclaj939c/Traverse%20Velocity%20%26%20Deposited%20Weight%20Analysis.pdf?dl=0. Type: slides deck.

[10] Badri Gomatam. Core profiler and methods. https://www.dropbox.com/s/na7k1o1y6n2rpno/For%20MIT-STL%20Core%20Program%20-%20Core%20Profiler%20and%20Methods.pptx?dl=0. Type: slides deck.

[11] Fernanda Viégas Martin Wattenberg and Ian Johnson. How to use t-sne effectively. https://distill.pub/2016/misread-tsne/. Retrieved: 4 December 2017.

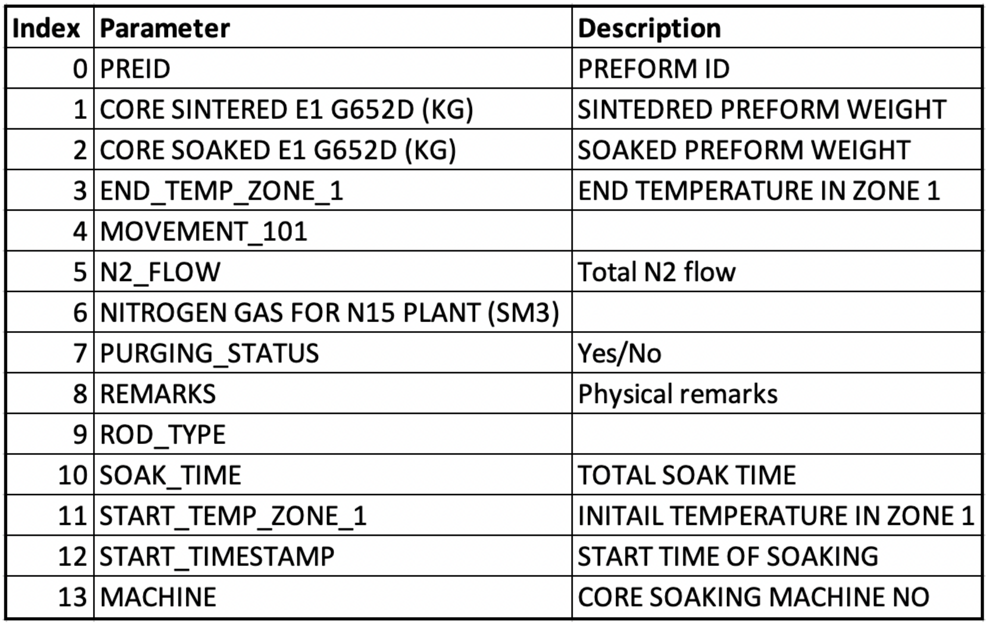
This page was intentionally left blank.

Chapter 6

Annexes

6.1 Data Dictionaries

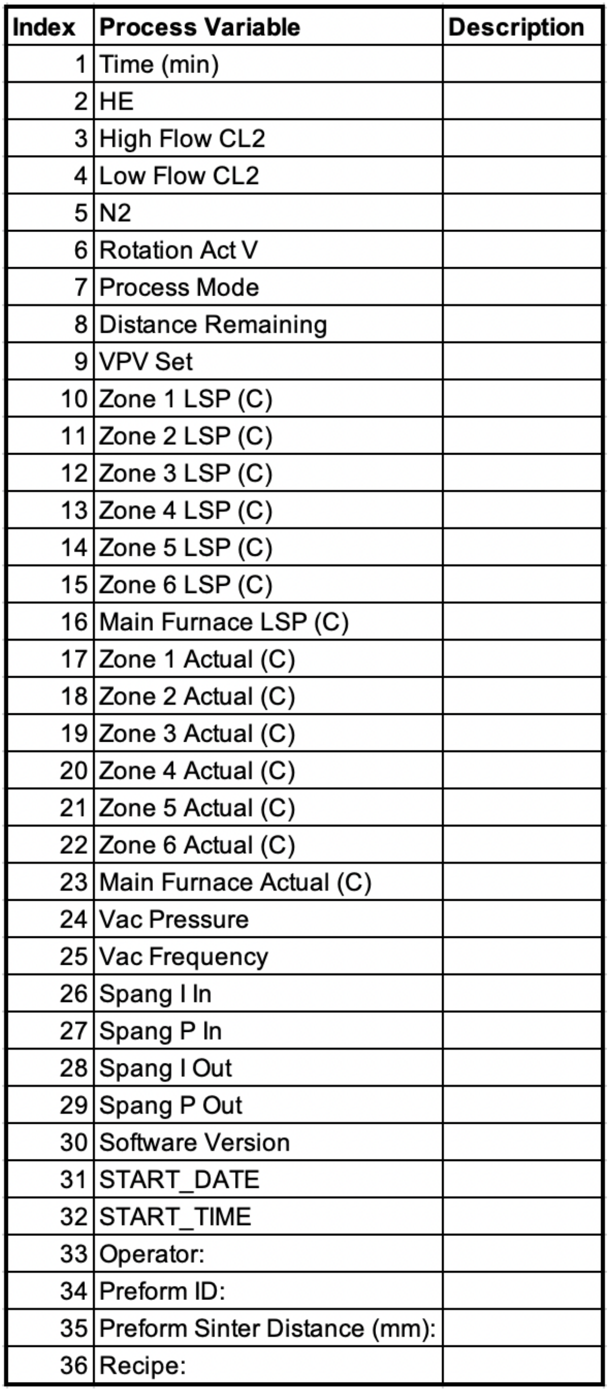
6.1.1 Soaking



Index	Parameter	Description
0	PREID	PREFORM ID
1	CORE SINTERED E1 G652D (KG)	SINTEDRED PREFORM WEIGHT
2	CORE SOAKED E1 G652D (KG)	SOAKED PREFORM WEIGHT
3	END_TEMP_ZONE_1	END TEMPERATURE IN ZONE 1
4	MOVEMENT_101	
5	N2_FLOW	Total N2 flow
6	NITROGEN GAS FOR N15 PLANT (SM3)	
7	PURGING_STATUS	Yes/No
8	REMARKS	Physical remarks
9	ROD_TYPE	
10	SOAK_TIME	TOTAL SOAK TIME
11	START_TEMP_ZONE_1	INITAIL TEMPERATURE IN ZONE 1
12	START_TIMESTAMP	START TIME OF SOAKING
13	MACHINE	CORE SOAKING MACHINE NO

Figure 6.1: Dictionary for soaking operation data set.

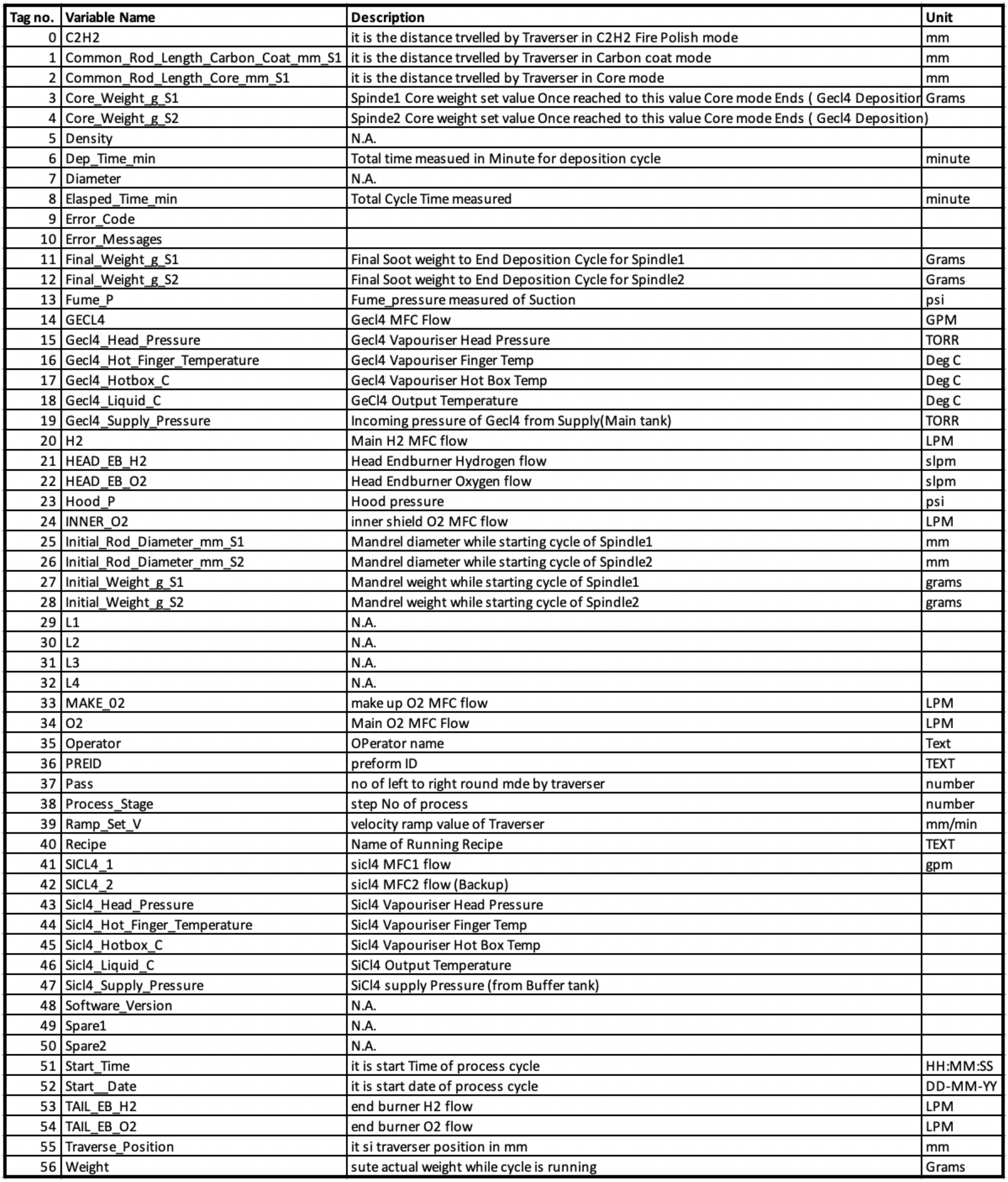
6.1.2 Sintering



Index	Process Variable	Description
1	Time (min)	
2	HE	
3	High Flow CL2	
4	Low Flow CL2	
5	N2	
6	Rotation Act V	
7	Process Mode	
8	Distance Remaining	
9	VPV Set	
10	Zone 1 LSP (C)	
11	Zone 2 LSP (C)	
12	Zone 3 LSP (C)	
13	Zone 4 LSP (C)	
14	Zone 5 LSP (C)	
15	Zone 6 LSP (C)	
16	Main Furnace LSP (C)	
17	Zone 1 Actual (C)	
18	Zone 2 Actual (C)	
19	Zone 3 Actual (C)	
20	Zone 4 Actual (C)	
21	Zone 5 Actual (C)	
22	Zone 6 Actual (C)	
23	Main Furnace Actual (C)	
24	Vac Pressure	
25	Vac Frequency	
26	Spang I In	
27	Spang P In	
28	Spang I Out	
29	Spang P Out	
30	Software Version	
31	START_DATE	
32	START_TIME	
33	Operator:	
34	Preform ID:	
35	Preform Sinter Distance (mm):	
36	Recipe:	

Figure 6.2: Dictionary for sintering operation data set.

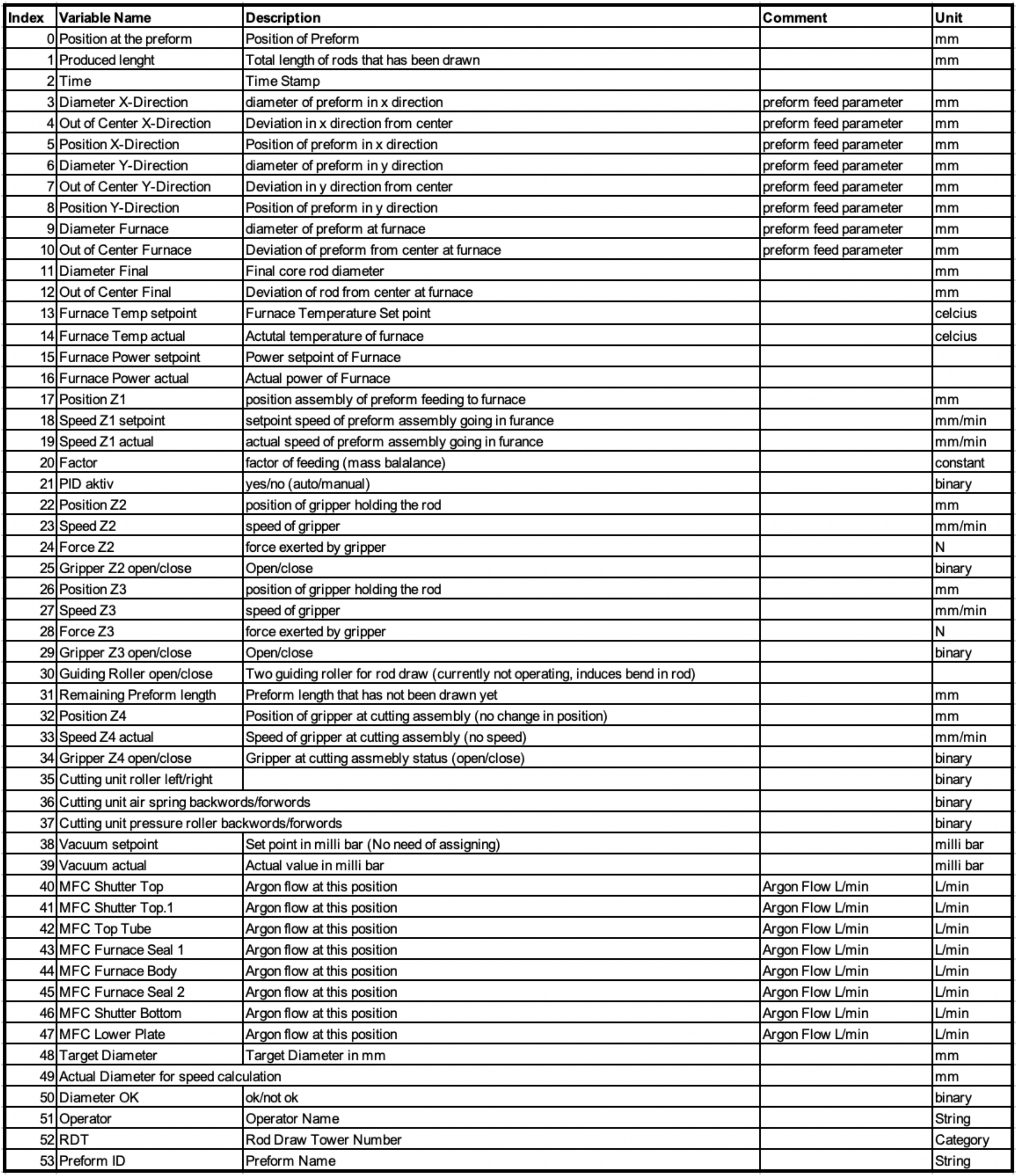
6.1.3 Deposition



Tag no.	Variable Name	Description	Unit
0	C2H2	it is the distance trvelled by Traverser in C2H2 Fire Polish mode	mm
1	Common_Rod_Length_Carbon_Coat_mm_S1	it is the distance trvelled by Traverser in Carbon coat mode	mm
2	Common_Rod_Length_Core_mm_S1	it is the distance trvelled by Traverser in Core mode	mm
3	Core_Weight_g_S1	Spinde1 Core weight set value Once reached to this value Core mode Ends (Gecl4 Depositior	Grams
4	Core_Weight_g_S2	Spinde2 Core weight set value Once reached to this value Core mode Ends (Gecl4 Deposition)	
5	Density	N.A.	
6	Dep_Time_min	Total time measued in Minute for deposition cycle	minute
7	Diameter	N.A.	
8	Elasped_Time_min	Total Cycle Time measured	minute
9	Error_Code		
10	Error_Messages		
11	Final_Weight_g_S1	Final Soot weight to End Deposition Cycle for Spindle1	Grams
12	Final_Weight_g_S2	Final Soot weight to End Deposition Cycle for Spindle2	Grams
13	Fume_P	Fume_pressure measured of Suction	psi
14	GECL4	Gecl4 MFC Flow	GPM
15	Gecl4_Head_Pressure	Gecl4 Vapouriser Head Pressure	TORR
16	Gecl4_Hot_Finger_Temperature	Gecl4 Vapouriser Finger Temp	Deg C
17	Gecl4_Hotbox_C	Gecl4 Vapouriser Hot Box Temp	Deg C
18	Gecl4_Liquid_C	GeCl4 Output Temperature	Deg C
19	Gecl4_Supply_Pressure	Incoming pressure of Gecl4 from Supply(Main tank)	TORR
20	H2	Main H2 MFC flow	LPM
21	HEAD_EB_H2	Head Endburner Hydrogen flow	slpm
22	HEAD_EB_O2	Head Endburner Oxygen flow	slpm
23	Hood_P	Hood pressure	psi
24	INNER_O2	inner shield O2 MFC flow	LPM
25	Initial_Rod_Diameter_mm_S1	Mandrel diameter while starting cycle of Spindle1	mm
26	Initial_Rod_Diameter_mm_S2	Mandrel diameter while starting cycle of Spindle2	mm
27	Initial_Weight_g_S1	Mandrel weight while starting cycle of Spindle1	grams
28	Initial_Weight_g_S2	Mandrel weight while starting cycle of Spindle2	grams
29	L1	N.A.	
30	L2	N.A.	
31	L3	N.A.	
32	L4	N.A.	
33	MAKE_02	make up O2 MFC flow	LPM
34	O2	Main O2 MFC Flow	LPM
35	Operator	OPerator name	Text
36	PREID	preform ID	TEXT
37	Pass	no of left to right round mde by traverser	number
38	Process_Stage	step No of process	number
39	Ramp_Set_V	velocity ramp value of Traverser	mm/min
40	Recipe	Name of Running Recipe	TEXT
41	SICL4_1	sicl4 MFC1 flow	gpm
42	SICL4_2	sicl4 MFC2 flow (Backup)	
43	Sicl4_Head_Pressure	Sicl4 Vapouriser Head Pressure	
44	Sicl4_Hot_Finger_Temperature	Sicl4 Vapouriser Finger Temp	
45	Sicl4_Hotbox_C	Sicl4 Vapouriser Hot Box Temp	
46	Sicl4_Liquid_C	SiCl4 Output Temperature	
47	Sicl4_Supply_Pressure	SiCl4 supply Pressure (from Buffer tank)	
48	Software_Version	N.A.	
49	Spare1	N.A.	
50	Spare2	N.A.	
51	Start_Time	it is start Time of process cycle	HH:MM:SS
52	Start__Date	it is start date of process cycle	DD-MM-YY
53	TAIL_EB_H2	end burner H2 flow	LPM
54	TAIL_EB_O2	end burner O2 flow	LPM
55	Traverse_Position	it si traverser position in mm	mm
56	Weight	sute actual weight while cycle is running	Grams

Figure 6.3: Dictionary for deposition operation data set.

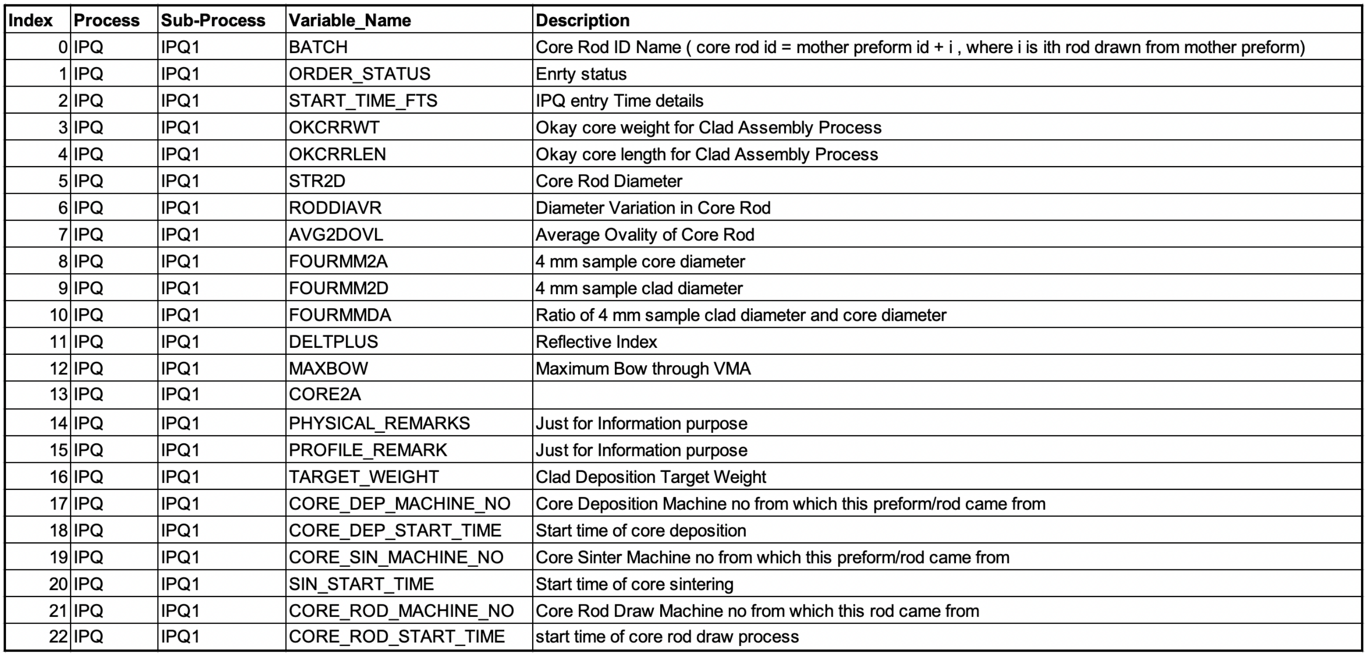
6.1.4 Rod Draw



Index	Variable Name	Description	Comment	Unit
0	Position at the preform	Position of Preform		mm
1	Produced lenght	Total length of rods that has been drawn		mm
2	Time	Time Stamp		
3	Diameter X-Direction	diameter of preform in x direction	preform feed parameter	mm
4	Out of Center X-Direction	Deviation in x direction from center	preform feed parameter	mm
5	Position X-Direction	Position of preform in x direction	preform feed parameter	mm
6	Diameter Y-Direction	diameter of preform in y direction	preform feed parameter	mm
7	Out of Center Y-Direction	Deviation in y direction from center	preform feed parameter	mm
8	Position Y-Direction	Position of preform in y direction	preform feed parameter	mm
9	Diameter Furnace	diameter of preform at furnace	preform feed parameter	mm
10	Out of Center Furnace	Deviation of preform from center at furnace	preform feed parameter	mm
11	Diameter Final	Final core rod diameter		mm
12	Out of Center Final	Deviation of rod from center at furnace		mm
13	Furnace Temp setpoint	Furnace Temperature Set point		celcius
14	Furnace Temp actual	Actutal temperature of furnace		celcius
15	Furnace Power setpoint	Power setpoint of Furnace		
16	Furnace Power actual	Actual power of Furnace		
17	Position Z1	position assembly of preform feeding to furnace		mm
18	Speed Z1 setpoint	setpoint speed of preform assembly going in furance		mm/min
19	Speed Z1 actual	actual speed of preform assembly going in furance		mm/min
20	Factor	factor of feeding (mass balalance)		constant
21	PID aktiv	yes/no (auto/manual)		binary
22	Position Z2	position of gripper holding the rod		mm
23	Speed Z2	speed of gripper		mm/min
24	Force Z2	force exerted by gripper		N
25	Gripper Z2 open/close	Open/close		binary
26	Position Z3	position of gripper holding the rod		mm
27	Speed Z3	speed of gripper		mm/min
28	Force Z3	force exerted by gripper		N
29	Gripper Z3 open/close	Open/close		binary
30	Guiding Roller open/close	Two guiding roller for rod draw (currently not operating, induces bend in rod)		
31	Remaining Preform length	Preform length that has not been drawn yet		mm
32	Position Z4	Position of gripper at cutting assembly (no change in position)		mm
33	Speed Z4 actual	Speed of gripper at cutting assembly (no speed)		mm/min
34	Gripper Z4 open/close	Gripper at cutting assmebly status (open/close)		binary
35	Cutting unit roller left/right			binary
36	Cutting unit air spring backwords/forwords			binary
37	Cutting unit pressure roller backwords/forwords			binary
38	Vacuum setpoint	Set point in milli bar (No need of assigning)		milli bar
39	Vacuum actual	Actual value in milli bar		milli bar
40	MFC Shutter Top	Argon flow at this position	Argon Flow L/min	L/min
41	MFC Shutter Top.1	Argon flow at this position	Argon Flow L/min	L/min
42	MFC Top Tube	Argon flow at this position	Argon Flow L/min	L/min
43	MFC Furnace Seal 1	Argon flow at this position	Argon Flow L/min	L/min
44	MFC Furnace Body	Argon flow at this position	Argon Flow L/min	L/min
45	MFC Furnace Seal 2	Argon flow at this position	Argon Flow L/min	L/min
46	MFC Shutter Bottom	Argon flow at this position	Argon Flow L/min	L/min
47	MFC Lower Plate	Argon flow at this position	Argon Flow L/min	L/min
48	Target Diameter	Target Diameter in mm		mm
49	Actual Diameter for speed calculation			mm
50	Diameter OK	ok/not ok		binary
51	Operator	Operator Name		String
52	RDT	Rod Draw Tower Number		Category
53	Preform ID	Preform Name		String

Figure 6.4: Dictionary for rod draw operation data set.

6.1.5 IPQ



Index	Process	Sub-Process	Variable_Name	Description
0	IPQ	IPQ1	BATCH	Core Rod ID Name (core rod id = mother preform id + i , where i is ith rod drawn from mother preform)
1	IPQ	IPQ1	ORDER_STATUS	Enrty status
2	IPQ	IPQ1	START_TIME_FTS	IPQ entry Time details
3	IPQ	IPQ1	OKCRRWT	Okay core weight for Clad Assembly Process
4	IPQ	IPQ1	OKCRRLEN	Okay core length for Clad Assembly Process
5	IPQ	IPQ1	STR2D	Core Rod Diameter
6	IPQ	IPQ1	RODDIAVR	Diameter Variation in Core Rod
7	IPQ	IPQ1	AVG2DOVL	Average Ovality of Core Rod
8	IPQ	IPQ1	FOURMM2A	4 mm sample core diameter
9	IPQ	IPQ1	FOURMM2D	4 mm sample clad diameter
10	IPQ	IPQ1	FOURMMDA	Ratio of 4 mm sample clad diameter and core diameter
11	IPQ	IPQ1	DELTPLUS	Reflective Index
12	IPQ	IPQ1	MAXBOW	Maximum Bow through VMA
13	IPQ	IPQ1	CORE2A	
14	IPQ	IPQ1	PHYSICAL_REMARKS	Just for Information purpose
15	IPQ	IPQ1	PROFILE_REMARK	Just for Information purpose
16	IPQ	IPQ1	TARGET_WEIGHT	Clad Deposition Target Weight
17	IPQ	IPQ1	CORE_DEP_MACHINE_NO	Core Deposition Machine no from which this preform/rod came from
18	IPQ	IPQ1	CORE_DEP_START_TIME	Start time of core deposition
19	IPQ	IPQ1	CORE_SIN_MACHINE_NO	Core Sinter Machine no from which this preform/rod came from
20	IPQ	IPQ1	SIN_START_TIME	Start time of core sintering
21	IPQ	IPQ1	CORE_ROD_MACHINE_NO	Core Rod Draw Machine no from which this rod came from
22	IPQ	IPQ1	CORE_ROD_START_TIME	start time of core rod draw process

Figure 6.5: Dictionary for IPQ measurements data set.

6.2 Deposition Efficiency Calculation

Calculation of Dep Rate and Dep Efficiency of the Core Preform Deposition Process at STL Waluj:

1. Raw runtime Data from the Maintenance Department is collected in the form of .xlsx or .csv format.

2. Data sets/Columns that needs to considered for calculating Deposition Rate:

   (a) DEP TIME (in mins)

   (b) Process Stage (name of the stages)

   (c) S1/S2 weight (in Kg)

3. **Deposition Rate (in grams per minute or gpm):**

   Deposition Rate of the soot preform is the rate of increase in the weight of the preform per unit time. It is calculated in two different ways: Stage-wise and Overall Deposition Rate.

(A) Calculation of Stage-wise Deposition rate:

(a) Select the desired Process Stage by filtering out the rest.

(b) Note down the initial and final Dep time and S1/S2 weight.

(c) Depo Mode wise Dep Rate is calculated as follows:

$$DepoModeDepRate = \frac{(S1Weight_{finalpass} - S1Weight_{initialpass}) * 1000}{DepTime_{finalpass} - DepTime_{initialpass}}$$

Multiplying by 1000 is done to convert the Kg values to gm values.

(d) Cumulative Deposition Rate based on the Depo Mode Dep Rate:

$$CummDepRate = \frac{\sum_{CenterlineMode}^{Clad2Mode} S1Weight_{ineachpass}}{\sum DepTime_{ineachpass}}$$

(B) Calculating Overall Deposition Rate:

Overall Deposition Rate of the Preform is calculated as follows:

$$OverallDepRate = \frac{S1Weight_{endofCladMode2}}{DepTime_{endofCladMode2}}$$

Cumulative and Overall deposition rates are very close to each other with an error pc. of 0.16%. For accurate calculations, stage-wise Dep rate may be considered.

4. Deposition Efficiency (in %)

Deposition efficiency is the percentage of soot that is deposited on the rotating mandrel as compared to the stoichiometric quantity of soot generated from the burner. It is also calculated in 2 different ways: Stage-wise and overall deposition efficiency.

(A) Calculation of stage-wise Deposition Efficiency:

Data sets/Columns that needs to be considered for calculating:

- DEP TIME (in mins)
- Process Stage (name of the stages)
- S1/S2 weight (in Kg)
- S1/S2 SiCl4 1 (in grams/min)
- S1/S2 GeCl4 (in grams/min)

(a) Select the desired Process Stage by filtering out the rest.

(b) Note down the initial and final Dep time and S1/S2 weight.

(c) Note down the average S1/S2 SiCl4 weight over that stage (=X).

(d) Note down the average S1/S2 GeCl4 weight over that stage (=Y).

(e) Depo Mode wise Dep Efficiency is calculated as follows:

$$D = \frac{(S1Weigth_{finalpass} - S1Weight_{initialpass}) * 1000 * 100}{(X * \frac{60}{170} + Y * \frac{104}{214}) * (DepTime_{finalpass} - DepTime_{initialpass})}$$

Multiplying by 1000 is done to convert the Kg values to gm values and by 100 is done to convert into percentage values.

(f) Cumulative Deposition Efficiency based on the Depo Mode wise Dep Eff:

$$CD = \frac{S1Weight_{endofCladMode2} * 100 * 1000}{\sum_{CenterlineMode}^{Clad2Mode}(AvgGeCl4flowrate * \frac{104}{214} + AvgSiCl4flowrate * \frac{60}{170}) * DepTime}$$

(B) Calculating overall Deposition Efficiency:

$$OD = \frac{S1Weight_{endofCladMode2} * 100}{(AvgGeCl4in1run * \frac{104}{214} + AvgSiCl4in1run * \frac{60}{170}) * DepTime_{endofCladMode2}}$$

NOTE: Here we have not multiplied the numerator by 1000 as the final weight is already given in the gram unit.
Cumulative and Overall deposition efficiency are very close to each other with an error pc. of 1.84%. For accurate calculations, stage-wise Dep Efficiency may be considered.

Glossary:

- The entire deposition process of the preform is divided into 14 steps where called deposition modes or deposition stages. In each of the stages, the flow rate of the Germanium and Silica varies. This causes a change in the Deposition rate and deposition efficiency.
- Dep Rate = Deposition Rate (in grams/minute)
- Dep Eff = Deposition Efficiency (%)
- The factor (60/170) is multiplied in the deposition efficiency term to incorporate the SiCl4 to SiO2 molecular weight conversion factor.
- The factor (104/214) is multiplied in the deposition efficiency term to incorporate the GeCl4 to GeO2 molecular weight conversion factor.