Manufacturing Variability; Effects and Characterization through Text-Mining

by

Alexander van Grootel

B.Eng., University of Edinburgh (2015)
M.Eng., Massachusetts Institute of Technology (2016)

Submitted to the Institute for Data, Systems, and Society
and the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degrees of

Master of Science in Technology and Policy

and

Master of Science in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2019

Signature redacted

Author ____

Institute for Data, Systems, and Society
Department of Electrical Engineering and Computer Science
May 10, 2019

Certified by ______________________ Signature redacted

Elsa Olivetti
Atlantic Richfield Associate Professor of Energy Studies, Materials Science and Engineering
Thesis Supervisor

Certified by ______________________ Signature redacted

Duane Boning
Clarence J. LeBel Professor, Electrical Engineering and Computer Science
Thesis Reader

Accepted by ______________________ Signature redacted

Noelle Selin
Director, Technology and Policy Program
Associate Professor, Institute for Data, Systems, and Society and
Earth, Atmospheric and Planetary Science

Accepted by ______________________ Signature redacted

Leslie A. Kolodziejski
Professor, Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Students

Manufacturing Variability; Effects and Characterization through Text-Mining

by

Alexander van Grootel

Submitted to the Institute for Data, Systems, and Society
and the Department of Electrical Engineering and Computer Science
on May 10, 2019, in partial fulfillment of the
requirements for the degrees of
Master of Science in Technology and Policy

and

Master of Science in Electrical Engineering and Computer Science

Abstract

Researchers and developers of new materials and processes often underestimate or neglect the effects of manufacturing variability and, as a result, make overly optimistic assumptions about their technologies. In this thesis, I explore the effects of manufacturing variability and find ways to characterize the manufacturing variability of emerging manufacturing processes. I develop a framework that connects manufacturing variability to environmental impact and economic costs through the concept of overdesign. I study examples using this framework and find that around 19% of concrete production is used solely to overcome issues of manufacturing variability, and that reducing the variability when producing fiber composite parts for a Boeing 787 reduces fuel consumption by millions of dollars and saves ktons of CO_2 from entering the atmosphere. I further explore the effects of manufacturing variability by considering its impacts on the commercialization process of new technologies. I consider Additive Manufacturing (AM), a promising technology, and argue that this technology has not reached commercial traction in great part due to our lack of understanding of the uncertainty associated with this process. I draw parallels to fiber composites, which faced similar issues in the 1980s before a collaborative effort, through the Advanced Composite Technology (ACT) and Advanced General Aviation Technology Experiments (AGATE) programs, was able to solve many of these challenges.

Finally, I consider the volumes of data available in published documents and analyze whether it is possible to extract this information using text mining techniques, and to use these data to characterize the manufacturing variability of upcoming technologies. Some important challenges obstruct our ability to extract all the important information from these documents, but important steps are made to remove some of these challenges and I demonstrate that useful information can be extracted.

Manufacturing engineers view processes as stochastic rather than deterministic. I ultimately argue for this view to also be adopted by environmentalists, materials researchers, and decision makers. I also further develop methods to extract and utilize manufacturing variation information.

Thesis Supervisor: Elsa Olivetti
Title: Atlantic Richfield Associate Professor of Energy Studies, Materials Science and Engineering

Thesis Supervisor: Duane Boning
Title: Clarence J. LeBel Professor, Electrical Engineering and Computer Science

Acknowledgments

I would like to express my deepest appreciation to Elsa Olivetti for all her guidance. You gave me the chance to follow my passions and interests and I will always be grateful for the opportunity you gave me. Thank you for your wisdom, insights, mentorship, patience and thoughtfulness. I have learned a lot from you.

I am extremely grateful to Duane Boning for sharing his time, experience and knowledge. You were able to cut to the core of an issue in minutes, even if it took me days of research before I could begin to describe it. Our conversations greatly helped me to narrow in on all the interesting problems.

Many thanks to Brian Wardle for all his expertise in everything composites, and for his collaborations and generosity. The entire idea of this thesis stems from our initial conversations.

I cannot begin to express my thanks to Edward Kim for his unbounded patience and inexhaustible source of thoughtful advice. I would not have been able to graduate without you. I hope we will work together many times in the future.

I am deeply indebted to Zach Jensen for always being there to brainstorm ideas and for always being willing to jump into the details with me. Your keen eye picked up many of my errors and saved me many days of work. I am so grateful that I had a friend to work on all these hard problems with.

I would like to thank Jiyoun Chang for all her wisdom and positive energy. Our discussions on variability brought so much clarity to me and my framework would not be nearly as good if it weren't for your discipline and thoroughness.

Alexander Denmark was invaluable in helping to shape the AM project. I am very grateful for your meticulousness and persistence.

Thank you to Eric Homer for all your advice on academia and life and for your help on all things metal.

I would like to thank Abbie Kaiser for her unstoppable support and unrelenting optimism. I could not have done any of this without you.

Thank you to the TPP community. I am honored to have learned with and from such a talented, kind, and thoughtful group of people. All of you have made me a better person. Special thank you to Barb DeLaBarre, Ed Ballo, and Frank Field. You are the hearts of TPP.

And last, but certainly not least, I am deeply grateful for my family who were always there when I needed them most. Without your love and support I could not have made it this far.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Manufacturing plays an important role in the world's economy, accounting for around 16% of global GDP [1], and is responsible for 20% of global CO_2 emissions [2]. These statistics underestimate how influential manufacturing really is. Manufacturing has direct impacts, as it consumes energy and materials to make products. Machinery, reagents, and energy are bought and used to turn raw materials into physical products. However, manufacturing also has far-reaching indirect impacts as all man-made objects we interact with were at one point manufactured. Furthermore, the design of objects is intrinsically linked with how they can be cheaply and reliably manufactured and as a result manufacturing is related to almost all designs of physical goods.

Manufacturing continues to play an important role in technological innovations. The stone age, bronze age and iron, age were characterized by our abilities to manipulate and shape new materials. The industrial revolution revolved around new approaches to mass manufacturing. Modern computation was enabled through impressive strides in ultra-precise fabrication techniques for producing computer chips. The future of renewable energy is in part reliant on the fabrication of cheap and efficient solar panels and batteries. Yet manufacturing and some of the challenges it faces may not be at the forefront of people's minds and can be easily forgotten. In this research, I find that a key and central idea in manufacturing, manufacturing variability, can often be neglected by those who propose new materials, processes, or policies and that this disconnect leads to over-optimistic predictions and delayed timelines. Manufacturing variability, including its impacts and the potential ways to characterize manufacturing variability, is the central topic of this thesis.

The idea of manufacturing variability starts from the fact that when two parts are manufactured they will never be exactly the same [3]. One of the parts will be slightly bigger, stiffer, stronger and/or will differ on any other property. As a result, these properties should be seen as random variables with distributions rather than as point estimates. Manufacturing variability is then a measure of how tight or spread out these distributions are.

High variability within a manufacturing process necessitates intervention. Ideally, the manufacturing variability can be reduced to the point where it can essentially be ignored, which is the premise of Six Sigma [4]–[6], a set of managerial and engineering tools that are used in organizations to systematically reduce manufacturing variability. However, if reducing variability to near zero is infeasible, as is typically the case, there are alternative strategies that must be adopted to deal with the remaining variability. These alternative strategies have environmental and economic implications. I explore this link between manufacturing variability and environmental/economic impacts in this work.

Furthermore, if manufacturers do not fully understand the variation in properties as a result of a manufacturing process, or do not understand how various processing parameters might influence this variation, then the commercialization process of such a technology is severely impeded. It therefore becomes important to characterize the manufacturing variability of emerging technologies. Unfortunately, a large amount of manufacturing data is required to understand these distributions. These data are hard to come by, especially for new manufacturing processes or materials. However, a wealth of data exists in published documents. Researchers investigate various manufacturing technologies and publish their results in academic journals. If the information contained in these documents can be extracted and put into a structured format, the data can be analyzed and we may be able to better understand the manufacturing variability. Recent developments in Natural Language Processing (NLP) allow us to extract information from documents at a large scale. However, while the field of NLP is developing rapidly, these techniques are not yet perfect, and although valuable information can be extracted using state-of-the-art and newly developed techniques, some important challenges remain.

This thesis answers the following questions:
1) What are the effects of manufacturing variability, and why is it important?
2) Can we gather data from published documents in order to characterize manufacturing variability, and what are challenges associated with that?

The rest of the thesis is split into three chapters plus a final chapter on conclusions. In Chapter 2, I develop a framework that can be used to think about manufacturing variability and the impacts it has on cost and ecological footprint.

Chapter 3 focuses on why it is difficult to get an accurate measure of variability, and how this translates into difficulties when commercializing a new technology. I also consider some historic (fiber composites) and modern (Additive Manufacturing (AM)) examples of how manufacturing variability can impede commercialization efforts, and what could be done to overcome this issue.

In Chapter 4, I continue with AM as a case study and discuss how we can text-mine manufacturing literature in order to characterize the manufacturing variability of this technology. I describe some problems associated with this approach, and also provide some insights and results from my text-mining efforts.

I provide a summary of the findings and conclusions in Chapter 5.

Chapter 2

Manufacturing Variability

Overview

This chapter looks into manufacturing variability, its causes, why it is important, and suggestions for how we should think about it. It begins with a brief overview of manufacturing variability as framed by the field of process control, the discipline which is tasked with managing variability in manufacturing processes. I then tie manufacturing variability to environmental impact and economic costs through the concept of overdesign, which lies at the heart of most certification frameworks. Some examples are given, with a particular emphasis on fiber composites which is a technology that has notoriously high manufacturing variability.

This work is largely based on two documents written by myself as the primary author and with substantial help from other authors [7], [8].

2.1 Properties as Random Variables

2.1.1 Overview of Process Control

No two fabricated parts are the same in manufacturing. Differences in the final properties of a manufactured part arise due to variations in the environment, materials or the machine itself [3]. For example, the temperature or humidity in a factory changes over the course of a day, which impacts the behavior of an injection molding machine. Or consider that the flow paths of the molten plastic will be slightly different each time plastic is injected into a cavity, which has some effect on the internal stresses, properties, and geometries of a part. Or potentially the materials used for the injection molding are not completely homogeneous because the feedstock itself has experienced its own set of variations in the environment, materials, or machines producing that feedstock. In this way, variations may propagate through the supply chain. The distribution of the properties of interest will depend on the process, the part being produced, the processing conditions and the standard operating procedures used by the machine operators. Because of these differences, one must consider properties as distributions rather than deterministic values.

The perspective of framing properties in a probabilistic way stems from work by Shewhart [9]. Shewhart grouped the sources of variation into two categories: common causes and assignable (or special) causes. Common causes refer to those causes that are expected and are inherent to the process. In contrast, assignable causes are more like disturbances and typically arise from some external sources. For example, when injection molding, hot liquid plastic is injected into a small cavity at an immense pressure. This chaotic fluid will always flow a little different every time this action is performed due to the nature of turbulent flow. As a result, the material and heat may spread a little differently in each run, resulting in some differences between each part that is made. Here the turbulent fluid flow is an example of a common cause, as it is inherent to the process.

As an example of an assignable cause, imagine that over the course of the day the injection molding machine becomes a few degrees hotter because the room itself is warmer. As a result, the plastic is injected at slightly higher temperatures, which causes the parts made in the morning to be slightly different compared to the parts made in the afternoon. This variability is not inherent in the process, and heating due to the environment is, therefore, an example of an assignable cause. Ideally, this temperature drift can be identified and corrected.

A process that has only common causes is said to be in control, while those that also have assignable causes is said to be out of control. Process control is the field that concerns itself with knowing when a process is out of control, and then "re-centering" the process after removing or adjusting for the assignable cause. Note that due to common causes, even if the process is in control there is some variation [10], [11].

2.1.2 Strategies for Dealing with Manufacturing Variability

If every part that is manufactured is slightly different, it raises the question of how manufacturers deal with this variability. For instance, if a hole is supposed to be 10 mm but ends up being 9 mm, then a bolt may have issues fitting in the hole at a later production step. Typically the design of a part will include a range of dimensions or properties that are acceptable, as defined by the Upper Specification Limit (USL) and Lower Specification Limit (LSL). The goal is to ensure that all manufactured parts will fall within these limits. Figure 2-1 illustrates this concept.

Note that this framework is binary; either the parts are within specification or they are not. This approach makes sense for situations such as certifying the safety of a building; either the part is safe or it is not. However, in many situations, the idea of a continuous loss function might be more realistic. In this framework, *any* deviation from the target leads to customer dissatisfaction and is therefore undesirable. An example of a continuous loss function might be the squared loss function. When adopting this kind of framework, the goal is to set up the process to minimize this loss. This approach was initially developed by Taguchi [12], and remains a popular framework in many process control problems. However, in this work, I concentrate on standards set by certification bodies. These are normally binary because the part either falls within the upper and lower bounds, or it does not.

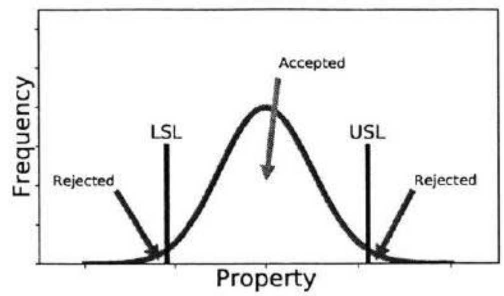

Figure 2-1: Diagram illustrating the Upper Specification Limit and Lower Specification Limit of a manufacturing process.

There are two main strategies that can be adopted to ensure that parts fall between the upper and lower specification limits: inspections and reducing the reject rate. Inspection requires the testing of all the parts. The ones that fall outside of the upper and lower specification limits are rejected or reworked. However, it is often impossible to inspect all desirable properties and there are significant costs associated both with inspection and with the possible rework or scrapping of out-of-spec parts [13].

The second group of strategies revolves around reducing the probability of having a rejected part to sufficiently low levels, such that it can be assumed that it is defect free. This idea is embodied by Six Sigma, a set of tools used for process improvement. Six Sigma's name comes from the idea that if we assume that the process is normally distributed, and that if the center of the process is six standard deviations away from the USL and six standard deviations away from the LSL (meaning that the USL and the LSL are twelve standard deviations apart), then the chance of error is 3.4 out of one million [11]. This might be sufficiently low odds that we can assume no parts will fall out of spec. A subset of parts is tested in order to verify that the distribution has not shifted and that it remains in control. If the distribution is as expected, then we continue to assume that only a negligible number of parts fall out of spec.

There are three ways to reduce the percentage of out of spec parts, as illustrated in Figure 2-2. The first is to increase the range of the specification limits by redesigning the part or redesigning the product in which the part will go. This approach makes a lot of sense in certain situations like in Liu et al. [14], where the researchers redesigned a subassembly that holds magnets for high precision applications. The redesign allows the magnets to be manufactured using traditional methods instead of high-precision methods, thereby saving significant costs. However, this approach is not always applicable. In some cases, redesign might not be possible or may be prohibitively expensive or complicated. This approach is mainly used when dealing with dimensions, rather than mechanical properties.

The second way to reduce out-of-spec rates is to reduce process variability. The first step would

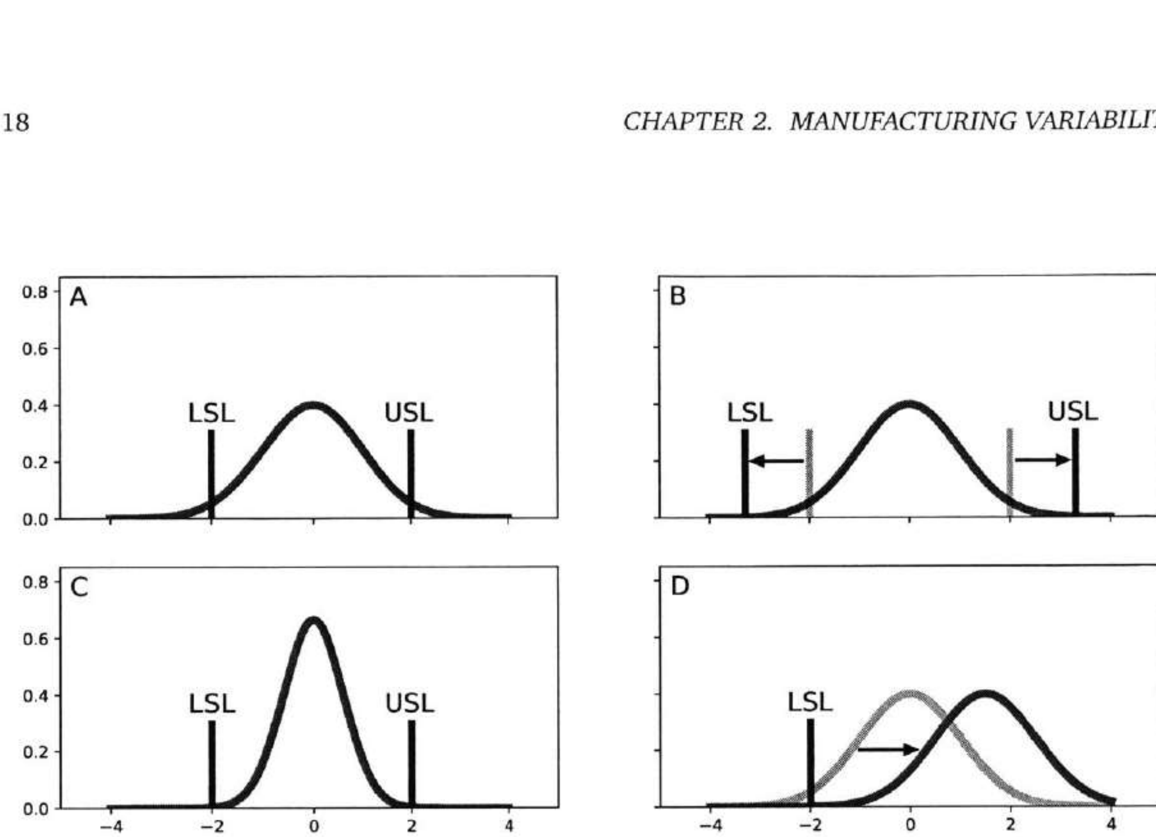

Figure 2-2: Different strategies to reduce the out-of-spec rate. A) the default scenario, B) shifting the USL and LSL, C) reducing the variance, D) shifting the distribution (overdesign).

be to remove all assignable causes, which should always be the aim. Reducing the common causes of variation requires changes in the process itself. This occurs either by upgrading important parts of the machines or by switching over to another process. This may therefore require more R&D spending than the other approaches. However, with concerted effort, this can be highly effective. For instance, the integrated chip industry was making features with sizes around ten micrometers in the 1970s and has improved this to around ten nanometers today [15]. These developments were only possible because the processes used by this industry became more and more precise and reproducible.

The third and final method to reduce out-of-spec rates is overdesign. This method can work well if there is either an upper or a lower specification limit, instead of both. An example may be the ultimate tensile strength of a beam. One might not care how high the ultimate tensile strength is, as long as it meets some lower limit. By shifting the mean to be sufficiently far away from the lower specification limit, it is possible to guarantee that only an acceptably small percentage of beams will fall below the limit.

It is possible to use multiple strategies. For instance, it might be possible to overdesign, while attempting to reduce the variability and perhaps also redesigning the part to reduce the lower specification limit.

2.1.3 Dominance of Overdesign

In many situations the strategy of overdesign is dominant. When dealing with structural parts such as beams or columns in a building, or a chassis in a car, the strength of the parts is critical. In these situations, it is not possible to reliably estimate properties like the ultimate tensile strength without destroying the part. It might be possible to test the strength of a part up to its design constraint and to approve it if it is able to withstand these forces, but this will likely deform and damage the part. It might also be possible to measure proxies, such as porosity measurement with X-ray in metal parts, which is typically correlated with tensile strength [16]. However, these proxies are imperfect and not always sufficient. The remaining strategy is therefore to reduce the out-of-spec rate to some negligible level. Because variability may be hard to reduce, the easiest approach here is to overdesign. It is probably for this reason that we find the idea of overdesign to be at the heart of most certification frameworks such as in [17] and fiber composites [18], [19]. The buffer for overdesign is typically decided by weighing the risk level that decision makers are willing to tolerate against the cost of mitigating this risk. The policy can either be set internally by a company or externally, through standards and certification frameworks. Overdesign in certification frameworks can take on various forms. In some cases, a safety factor is added (e.g., [18]), which follows the concept of overdesign directly and essentially sets a higher target for mechanical requirements to account for variation in mechanical properties. For example, one might say that instead of a steel bar needing to withstand 10 kN, it must withstand 13 kN (safety factor of 1.3) to account for variability in the strength. Assuming an average tensile strength of 400 MPa this turns a bar with 25 mm^2 of cross-sectional area into a bar of 32.5 mm^2.

In other cases, the preferred framework is that of design allowables [20]. Here the mechanical properties that are allowed to be assumed during design depend on the distribution of these properties. A material that has a very high spread in its properties will have lower allowable strength than a material that has a tighter distribution but the same average strength. For example, instead of assuming a tensile strength of 400 MPa for a steel bar, only a tensile strength of 300 MPa can be assumed. For our bar that has to withstand 10kN, this would mean that instead of 25 mm^2 of cross-sectional area, 32.5 mm^2 of area would be required. Ultimately these different approaches frame the same concept in different terms.

2.2 Manufacturing Variability, Environmental Impact, and Cost

2.2.1 Overdesign, Material Material Consumption and Ecological Footprint

For any design constraint that varies with the amount of material used, it is possible to overdesign by adding more of that material. This includes properties such as the stiffness, strength, and thermal capacity of a part among many others. In reality, the effect of overdesign will be subtler and more

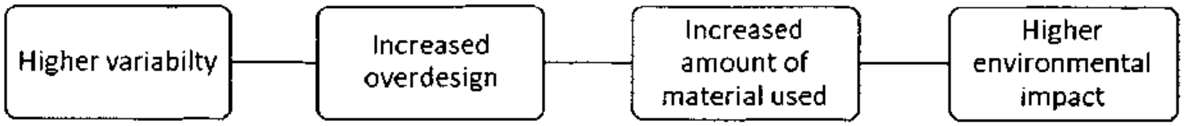

Figure 2-3: Schematic linking manufacturing variability and environmental impact.

complex than simply scaling the thickness of all parts at the end of a project's design. Variability will be taken into consideration from the onset of the project [20]. Significant testing will be conducted early on during the design process to account for variability and to establish the properties that can be assumed for the design. Alternatively, safety factors will be applied from the beginning of the project and parts will be designed to meet the requirements of more conservative design specifications. Ultimately, variability will lead to more material being used in the product.

Increased amounts of materials used in a part mean more embodied energy in that part. For many products, a substantial proportion of the lifetime energy is associated with the embodied energy in the materials [21], [22]. In addition, an increased weight of a final product due to more materials can also influence the environmental impact of the use phase of that product. For example, the fuel consumption of a vehicle heavily depends on the vehicle's weight.

The link between manufacturing variability and environmental impact can therefore be established under a few conditions. A simple schematic is shown in Figure 2-3. If manufacturing variability is high and the idea of safety factors is adopted, then overdesign will take place. Alternatively, the idea of design allowables is adopted and the effective strength that can be assumed by designers is adjusted based on the variability of the mechanical properties. In both cases, the result is overdesign which implies that more material will likely be used thereby leading to higher environmental impact.

2.2.2 Examples

Concrete

Concrete has a very high CO_2 emissions footprint [23], is a critical construction material, and the variance in the strength of concrete is notably high [24]. I will discuss variability in terms of the coefficient of variation (CV), which is defined as the standard deviation divided by the average of a distribution. A CV of 0.14 is rated to be 'fair' by industry standards [25].

In the compliance requirements of Section 26.12.3.1 in the standard developed by the American Concrete Institute (ACI) [17] the ACI requires the average of three consecutive tests to exceed the specified strength, with an allowable failure rate of one in a hundred tests. This can be translated into a safety factor as defined in Equation 2-1.

$$SafetyFactor = zscore\frac{CV}{\sqrt{n}} = 2.33\frac{0.14}{\sqrt{3}} = 18.8 \tag{2-1}$$

In other words, under these assumptions almost 19% additional concrete is required in order to overcome issues with manufacturing variability.

In 2016 the total CO_2 attributed to concrete was around 1.45 Gt CO_2 [23]. Assuming that all this concrete has a CV of 0.14 and that for all this concrete the above compliance requirements were followed, we can extrapolate our calculations to suggest that around 0.27 Gt of CO_2 is being put into the atmosphere every year to make up for the high variability in concrete. In reality, not all of the world's concrete will follow this standard, and not all concrete has a coefficient of variation of 14%. However, this figure presents itself as a reasonable estimate of the total opportunity that can be captured by reductions in variability for concrete.

Integrated Circuits

Manufacturers of Integrated Circuits (IC) are highly familiar with variability in manufacturing processes. There is clear motivation for this; most of the innovation in this industry is driven by the desire to create increasingly smaller transistors which requires increasingly more accurate and precise manufacturing operations. The trend has been powerful enough to shrink feature sizes from tens of micrometers in the 1970s, to tens of nanometers in today's state of the art [15].

With billions of transistors per chip, each individual transistor must have high probability of meeting specifications (high yield), otherwise too many chips may have too many defective transistors. To ensure extremely high yield for each transistor, each feature is made larger (overdesigned) relative to what current technologies may be able to produce on average.

IC manufacturers have to determine tolerable failure rates. This becomes a trade-off between the size of each transistor, and therefore speed and power draw, and the yield of each transistor, which manifests as the yield of the chips. Interestingly, the yield of chips has been falling as the transistors have shrunk. Transitioning from 350 nanometers to 90 nanometers in feature size, the yield went from around 90% to 50% for the highest performing products [26].

With current technologies around 14 nanometers, yield likely has dropped even further although the exact numbers remain difficult to obtain. These statistics give an interesting insight into the IC industry. The industry is willing to accept high levels of rejects just for the chance to build very small transistors, and this risk tolerance seems to have shifted over time. It also highlights a further dimension from a sustainability perspective. In this industry, there is a trade-off between yield and the amount of material required per part. With 350-nanometer technology Intel produced 31 million transistors on a chip with a die size of 242 mm^2 [27] while with 90-nanometer technology Intel produced 125 million transistors on 112 mm^2 [28]. This is an 8.7-fold decrease in area per transistor. Assuming that the kind of material used is comparable and the thickness remains the same, even if yield decreased from 90%

to 50% materials will still be saved. However, there might be a point where yield may be so low that it no longer leads to materials savings.

Nevertheless, process control in the IC industry should be seen as a success story. The field pays close attention to process capabilities, and has very sophisticated statistical tools and design strategies to continue progress [29]. This relentless push towards smaller transistors is a form of dematerialization that has been enabled by advances in process control.

2.3 Case Study: Fiber Composites in Aerospace

2.3.1 Motivation

Carbon Fiber Reinforced Polymers (CFRP) is an interesting case study to investigate the impacts of high manufacturing variability, and I go into depth here for several reasons. First, CFRP has amongst the highest weight-reducing potential relative to other commercially available materials [30] and is therefore of great interest to the aerospace industry. Because of the relatively high specific strength and stiffness of a CFRP, aircraft manufacturers have increased their use of CFRP and its growth is expected to continue over the next decade [31]. For example, the aviation industry recently launched aircraft made of CFRP at a significant portion of materials such as 50% and 52% in the Boeing 787 [32] and in the Airbus A350 [31], respectively.

Second, while the weight reduction capability of this material is incredibly promising, it comes at a high cost compared to other approaches, with some statistics putting it at 570 % higher than the cost of steel [33]. While these costs are dropping, the price of CFRP remains one of the main reasons why automotive and aerospace industries are not using the material more [34]. Potential reductions in the costs of this technology would be impactful.

Third, the manufacturing variability of CFRP is notoriously high [35] when compared to other competing materials such as aluminum, magnesium, or high-strength steel. The complex structure and manufacturing process for CFRP introduces many sources of variability [36]. Potter et al. identified at least 60 sources of variability and 130 sources of defects for the production of fiber composites [37].

Finally, CFRP is energy intensive to fabricate [38]. While research is ongoing, thermoset fiber composites still have no good end-of-life strategies [39]. Reducing the amount of CFRP required could therefore prevent substantial environmental degradation.

This case study explores the role of manufacturing variation of CFRP components within aircrafts. I analyze how the variation in manufacturing an aircraft part influences production cost and weight and also consider life-cycle energy and GHG emission implications. Through this case study, I suggest that variations in the product quality must be managed to achieve weight reduction, and to improve environmental performance.

2.3.2 Background of fiber composites

Interest in lightweight materials has been growing and the market size for lightweight materials exceeded 100 billion US dollars in 2016 [40]. While many industries can benefit from advanced lightweight materials, the largest demand is in the transportation sector, including the aviation and automotive industries. Here I focus on aviation. Volatile fuel prices and significant CO_2 emissions are driving the aviation industry to reduce vehicle weight [41] and enhance efficiency. The aviation industry is amongst the largest consumers of global transport fuel use and was responsible for 2.5% of global energy-related CO_2 emissions in 2013 [42] and 9% of global transportation greenhouse gas (GHG) emissions in 2010 [43]. Lightweight materials offer weight reduction and therefore fuel savings, but at higher cost than conventional structural materials. Thus, one objective of aircraft manufacturers is to achieve the highest structural strength per weight at minimum cost. While GHG emissions from the aviation sector are expected to roughly double in the next decade, the highest rate among transportation modes [43], the demand for fuel-efficient aircrafts is also strong [44]. The aviation industry has been utilizing lightweight materials and the share of lightweight materials is already almost 80% [33].

Although the decisions around designing a lightweight part for aerospace applications are made based on various constraints such as cost, manufacturability, etc., one hard limit is the strength requirement in structural components. A complicating factor is that there is uncertainty regarding both the strength that is required (i.e., what is the worst-case scenario that this plane will encounter?) as well as uncertainty regarding how strong a manufactured component is (i.e., how much force is required to break a specific part?). There are rules in place that adjust for both of these sources of uncertainties. For example, there is a standard safety factor of 1.5 that is applied to all aircraft designs. This safety factor accounts for inadvertent in-service loads that exceed the design limit, and does not account for variations in mechanical properties. Variations in mechanical properties are dealt with under different regulations as can be seen in the Code of Federal Regulations (C.F.R) 23.303 [45] and C.F.R 23.613 [46], and is discussed in a separate NASA document [47].

Setting uncertainty around strength design aside, I focus on the uncertainty of mechanical properties of fabricated parts. To overcome issues surrounding variations in mechanical properties, regulations like the Airworthiness standards from the Code of Federal Regulations will demand properties to be set on a statistical basis. For example, the properties that can be assumed must be set such that some majority (e.g., 95%) of all parts will fall above it [46]. In practice, this can be thought of as another safety factor.

If a part could be produced without any variability in material properties, then the factor of safety associated with the manufacturing could be one. However, as I argued, no manufactured product is exactly the same as any other, hence, neither is its performance. To ensure that products meet the required strength, the allowable strength must consider these variations, resulting in a safety factor higher than one. From this perspective, the safety factor is a measure of how much 'overdesign' is required to guarantee reliability. The US aerospace industry relies heavily on the Military Handbook 17 or MIL-HDBK-17[19] to specify how these values should be established.

2.3.3 Methodology

Overview

In order to investigate the effects of variability, I built a process based cost model (PBCM) [48] for a part made of CFRP for aerospace applications. The model coarsely changes the design of the part depending on the variability of the strength of the manufactured part. Strength is intended to be a generic representation of mechanical properties that scale with the weight of the part.

The manufacturing variability is set as an input to the model. This variability is measured on the strength as I define above, and because strength is an abstracted concept in our model, the variability is too. I frame manufacturing variability as variability on any property that scales roughly linearly with the amount of material, that comes as a result of the manufacturing process.

There are different ways to connect variability to overdesign. In this model, I adopt a standard as described in ASTM F3114-15 where the coefficient of variation is directly related to the test factor or the degree of overdesign required [18], as this provides a clean and intuitive mapping between variation and safety factor. This standard is replicated in Table 2-1. This safety factor determines how design (specifically the thickness) of the part should be adjusted to meet the required strength, which ultimately means that in our model the weight varies linearly with the degree of overdesign.

The PBCM models three process steps in manufacturing a CFRP part: prepregging, layup, and curing, which are each described below. The architecture of the whole model is shown in Figure 2-4 along with some key assumptions for input parameters. The outputs of the model are the cost of the part and the energy required to make a part.

Prepregging

A CFRP part consists of two components: carbon fibers, which act as the reinforcement element and provide the strength and stiffness of the part, and a resin matrix which provides the shape, some protection from handling, and forces the individual carbon fibers to engage together. In aerospace, prepregs (pre-impregnated) is the most popular approach to combine fiber and resin among various technologies. Here the fibers are arranged into aligned plies (unidirectional, or UD prepreg) or woven into mats, impregnated with resin, and are partially cured as sub-mm-scale sheets or plies. These plies

ASTM F3114-15	
Coefficient of Variation	**Overdesign Factor**
0.05	1
0.06	1.03
0.07	1.06
0.08	1.1
0.09	1.12
0.10	1.15
0.12	1.22
0.14	1.3
0.15	1.33
0.20	1.55

Table 2-1: ASTM F3114-15 standard which translates manufacturing variability into an overdesign factor [18].

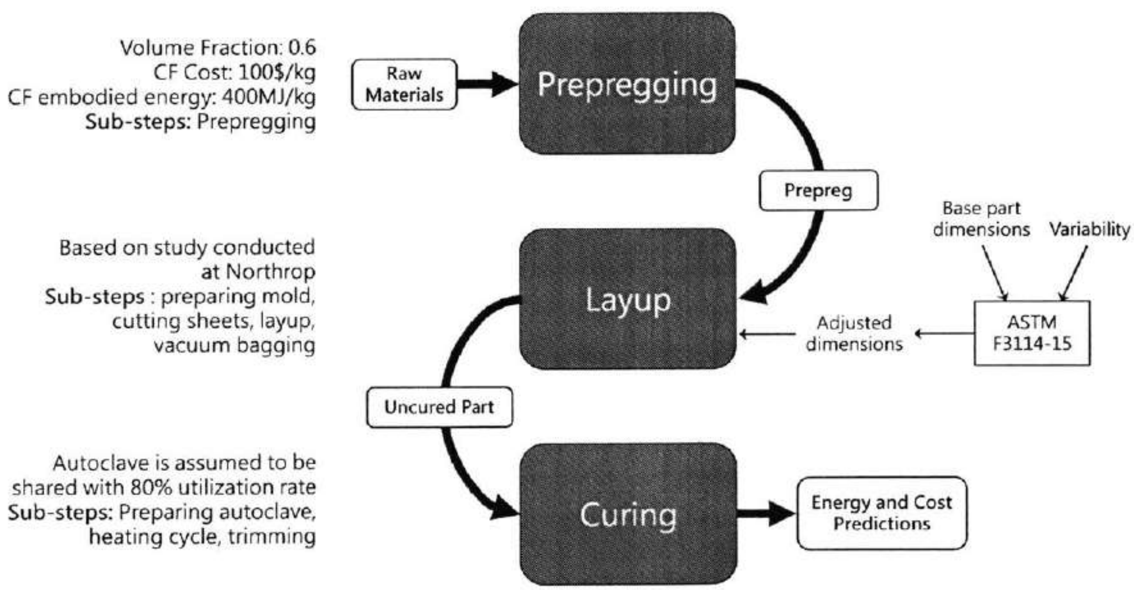

Figure 2-4: Process based cost model schematic.

are used during layup of the final composite laminate part.

In the model, prepregging is the first step and turns carbon fiber and resin into the raw material used in layup. I assume unidirectional prepregs, and very roughly base our model's prepreg on typical aerospace grade prepregs such as the Hexcel AS4/8552 UD CFRP. However, I allow some attributes of this material to deviate during analysis. Mainly, an important parameter in this step is the volume fraction, which is a measure of how much of the total volume of the prepreg is occupied by carbon fiber as opposed to resin. Higher volume fractions are typically preferred as long as the fibers are fully impregnated with resin.

I assume a volume fraction of 0.6 in our model unless stated otherwise. The other important assumptions are the price of carbon fiber (100 $/kg [49]) and the price of the epoxy resin (10 $/kg). I assume that the embodied energy of carbon fiber and epoxy resin is 400 MJ/kg and 132 MJ/kg [38], respectively. Prepreg processing is assumed to require 40 MJ/kg [38]. Finally, I assume that the machine used to make the prepreg costs $1,000,000, has a lifetime of 15 years and a throughput of 110 kg/hour [50].

Layup

There are multiple ways to process prepregs, but in many cases, manual layup remains popular [34]. In my model I make use of a study on manual layups conducted at Northrop [51] and funded by the USAF. While the study is dated, it is meticulous and the procedures of manual layups have not changed substantially. This study presents equations which estimate the labor-hours required for each of the various steps associated with manual layup. This includes steps like cleaning the mold, coating the mold with release agent, and applying the vacuum bag, amongst many others. Although there are some consumables used in this process, these are a small part of the total cost in comparison to the labor costs.

Curing

Curing of most aerospace CFRP is done in an autoclave, which has high capital costs. Here I make use of a Ph.D. thesis which collected information on costs and specifications of an autoclave [52]. The costs are adjusted for inflation, and a linear regression is conducted with the width, diameter, and volume as input parameters. This allows an estimate for an autoclave dependent on the part size. For a 10m x 3m part, this results in a cost of around $3 million. Curing cycles are estimated based on the recommended heating cycles as suggested by some material suppliers. During curing, I assume that the energy required scales with the size of the autoclave, using a benchmark of 27 kW for a 0.75m3 autoclave as per [53]. It is then assumed that the autoclave will run on 60% of maximum power during the warmup cycle and at 20% of maximum power during the hold phase of curing, similar to what is assumed by Witik et al. [53].

In my model, the ramp-up time in the autoclave is sufficiently slow (90 minutes), and the parts are sufficiently thin, that heat-flow through the part does not cause a bottleneck. This is verified using a 1-Dimensional lumped parameter model. In this model, I assume forced convection in the autoclave, with a convection heat transfer coefficient of 100 W/m2 [54]. The part is modeled as 10 nodes, with two additional nodes representing a 1 cm thick vacuum bag. The thermal conductivity of the vacuum bag and composite is assumed to be 0.4 W/m K and 0.8 W/m K [55], respectively. The autoclave is assumed to be stable at 180 °C throughout this process. Varying the thickness of our part up to 1 cm allowed the entire part to heat up to within 1 % of the target temperature within 15 minutes, which is well within the 90-minute ramp-up timeframe. In other words, for the range of thicknesses I consider, the cycle time is constant.

2.3.4 Monte Carlo Simulation

As expected, the outputs of the model vary dramatically depending on the input values. Almost all of the inputs have uncertainty associated with them or will be expected to change depending on the design or other circumstances. A single scenario therefore cannot generalize to give a robust representation of the cost or energy distribution. I use Monte Carlo simulations to address this limitation.

Table 2-2 below provides the range of the values used as input variables to the Monte Carlo simulation. All variables follow a uniform distribution, and the minimum and maximum that defines the uniform distribution are provided in the table. For specific and deterministic scenarios, I use the values provided in the default column. The coefficient of variation (CV) is defined as the standard deviation of a distribution divided by its mean. This is intended to be a normalized measure of manufacturing uncertainty in this study.

For some of these variables, there are strong arguments why they should not be independent. For example, the production volume of very large parts tends to be much lower than that of smaller parts, because every aircraft might only need a handful of these large parts but many small parts. In addition, the trim ratio of small parts will tend to be significantly higher than that of larger parts, due to the ratio of surface area and circumference. To capture these trends I also enforce correlations as defined by the correlation matrix in Table 2-3.

2.3.5 Modeling Fuel Consumption

The effects of lightweighting on the fuel consumption of cars has been studied in some detail [56], [57]. For aircrafts these models are more difficult to come by. While airline operators pay close attention to fuel consumption, and therefore have likely developed their own sophisticated internal models, they have not made their findings publicly available.

To estimate the fuel saving as a consequence of lightweighting, I use the study by Helms and Lambrecht, who assumed that for short-distance planes 100 kg leads to between 10 TJ to 30 TJ of fuel

Variable	Minimum	Maximum	Default
Part length (m)	0.5	25	5
Part width (m)	0.5	5	3
Part thickness (m)	0.001	0.01	0.005
Coefficient of variation	0.05	0.18	0.14
Production volume (units/year)	20	4,000	500
Volume fraction	0.5	0.65	0.6
Carbon fiber price ($/kg)	80	120	100
Resin price ($/kg)	5	15	10
Cost of direct labor ($/person-hr.)	45	70	50
Cost of indirect labor ($/person-hr.)	45	70	50
Prepreg yield (%)	85	99	95
Layup yield (%)	85	99	95
Curing yield (%)	85	99	95
Trim yield (%)	70	95	80

Table 2-2: List of the parameters used for the inputs of a Monte Carlo simulation. All distributions are assumed to be uniform over their range. Default parameters are as shown in the final column.

	Production Volume	Part Length	Part Width	Trim Ratio
Production Volume	1	-0.75	0	0
Part Length	-0.75	1	0	-0.54
Part Width	0	0	1	-0.54
Trim Ratio	0	-0.54	-0.54	1

Table 2-3: Correlation matrix of the Monte Carlo simulation.

savings over a commercial plane's 30-year life [58]. These authors base this estimate on correspondence with Luftsansa.

In order to corroborate the estimate by this study, I pulled in data from other studies. In an article in Composites World [59], some estimates are made on a Mitsubishi Regional Jet. Here it is proposed that if 3200 lb. of weight savings are observed, it would lead to fuel savings of 42,700 gallons. This works out to around 112 liters of fuel/kg saved/year. Over 30 years, and at 43MJ/kg of kerosene, this amounts to around 0.12 TJ/kg and therefore corroborates the estimates used by Helms and Lambrecht at between 0.1 and 0.3 TJ/kg. In this model, I set the fuel savings at 0.2 TJ/kg and provide the results together with the upper and lower bounds.

I model the total fuel consumption change in a Boeing 787-9 and assume that the plane's operating empty weight is 128,850 kg [60]. I then assume that 10% of the plane's operating empty weight is made out of composite and has opportunities to reduce variability. This is a conservative estimate, as around half of the plane's structural weight is made out of composite [61].

I further assume that airplane fuel costs 5 USD/gallon (1.32 USD/liter) [62]. To calculate net present value (NPV) calculations, a 30-year lifetime and 9% annual return rate are assumed. Finally, I assume there are 9.57 kg CO_2 per gallon of jet fuel [63].

2.3.6 Results: Production Cost and Energy

Typical costs to manufacture CFRP parts calculated by the model are around 230-280 USD/kg, which is roughly what one would expect for an aerospace part. One market report [49] puts the price of carbon fiber composite parts at 310 USD/kg for aerospace grade parts. Adding a markup of around 15%, which is reasonable for a Tier 2 aerospace company [64], and considering that some costs like inventory and transportation costs are not part of our model, my estimate falls in line with what is found in the industry.

The specific breakdown between material, capital, labor, and manufacturing energy costs will depend on the inputs to the model. In Figure 2-5, I show three different cases, whose inputs are defined in Table 2-4. Materials account for the majority of total costs in all three cases. This is due to the high cost of aerospace-grade carbon fiber. Capital and labor also account for a significant proportion of total costs. Energy costs are relatively small in comparison. The only difference between Case 1 and Case 2 is that Case 1 has a higher manufacturing variability. As a result, the required safety factor and hence the thickness of the part in Case 1 is higher, requiring more material to make, and higher labor costs to lay down the additional layers. Case 3 has modest changes across most of its parameters and illustrates that small changes in the part can lead to significant changes in the cost distribution. Most notably, in Case 3 almost 60% of the total cost is attributed to materials; about 15-20 percentile points higher than Cases 1 and 2.

	Case 1	Case 2	Case 3
Part Length (m)	5	5	9
Part Width (m)	5	5	2
Base Thickness (mm)	2	2	3
Coefficient of variation (%)	12	5	18
Effective thickness due to overdesign (mm)	2.4	2	4.6
Parts per year	500	500	100
Cost of CF ($/kg)	100	100	120

Table 2-4: Definitions of three different cases. All other inputs to the model are set to their default values.

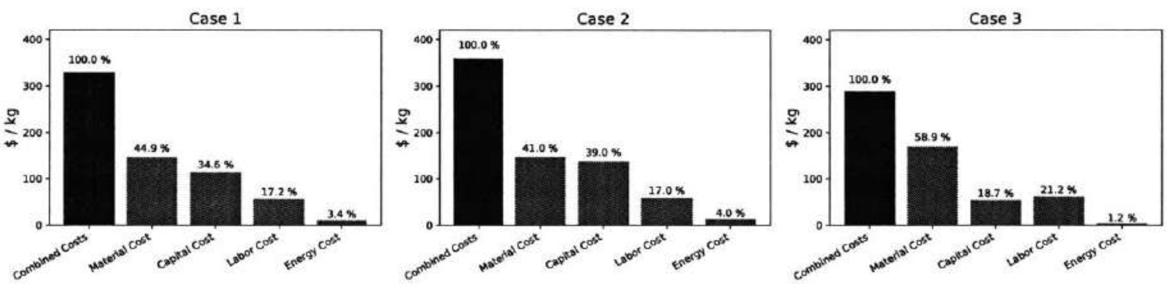

Figure 2-5: Cost distribution between material, energy, capital and labor for three different cases. For Case 1 the part weighs 122.06 kg and costs $40,028 for 327.94 $/kg. Case 2 weighs 100.05 kg and costs $35,859 for 358.41 $/kg. Case 3 weighs 111.66 kg and costs $32,192 for 288.30 $/kg.

Production Costs

The physical size of a part plays the largest role on production costs as it directly impacts all areas of the cost model. More material is required, a larger autoclave is needed, and layup requires more time. Part thickness relates to total cost through the materials required and the added man-hours required for the additional layups. In our model, the coefficient of variation impacts only the thickness of the part, and as a result, the manufacturing uncertainty has a significant indirect effect on material and labor requirements.

A more complete view of cost distributions can be generated by analyzing the results of the Monte Carlo simulation as is defined in Table 2-2. These results also suggest that materials tend to dominate the cost breakdown of carbon fiber parts across most scenarios. As can be seen in Figure 2-6 A, on average across all scenarios just over half of the total cost can be attributed to materials. The second largest contributor to cost is labor at around 26%. This is predominantly due to the manpower required during the layup process. Capital, driven predominantly by the high cost of autoclaves, comes in at around 22 % of total costs, followed by electricity costs at around 4%.

These analyses show that reducing the material requirements will have a strong impact on the final cost of the part, and that this assertion holds across a wide range of scenarios. Lowering the variability, and hence the amount of material that needs to be used, should, therefore, have a substantial

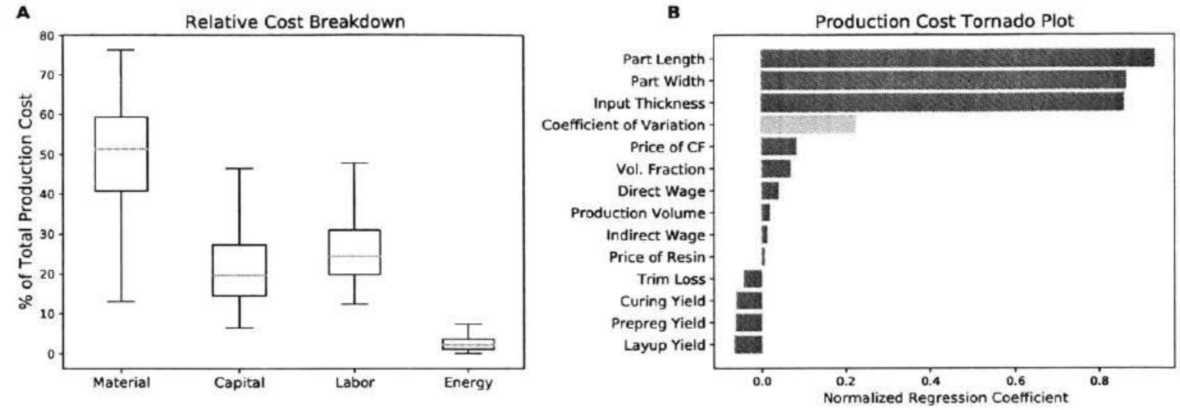

Figure 2-6: (A) Relative cost distribution of carbon fiber parts based on the Monte Carlo simulation. (B) Tornado plot. Input variables were normalized to have a mean of zero and a standard deviation of one. This plot shows the regression coefficient of the linear regression analysis on total part cost. The design of the part plays the most significant role in total cost and the coefficient of variation is the most important driver after this.

impact on the final cost.

We can determine the importance of manufacturing variability in our model more directly. In order to ascertain the relative importance of the different variables we first create a level playing field by taking the Monte-Carlo data and then normalizing each variable by first subtracting the mean and then dividing by its standard deviation. We then conduct a linear regression on all the normalized variables and report the regression coefficients in Figure 2-6 B. Because we normalize the variables, transformations like a change of units are accounted for. Instead, what becomes important is how the variable is used in our process as well as the assumed distributions for each of the variables. For instance, we could make the variable Direct Wage more important by allowing it to vary between 0 \$/hr. to 10,000 \$/hr. before normalizing. In other words, it is important that the distributions that we assume in Table 2-2 are reasonable representations.

We find that the variables related to design have the most influence on part cost by a wide margin. However, the design may be difficult to change and it is therefore worth considering the effects of other cost drivers. The next biggest influencer is the manufacturing variance as represented by the coefficient of variation variable. This suggests that there is a financial incentive to reduce the variability even when just looking at the production cost. These incentives are significant, and we find that reducing the variability from 14% to 9% reduces the production cost by 12.3% on average due to reduced labor and material requirements.

Production Energy

Figure 2-7 A illustrates the distribution of total production energy across its constituents. The production energy is assumed to be the combination of embodied energy from the resin and carbon fiber, the prepregging processing energy and the curing energy. As we can see from this figure, the embodied

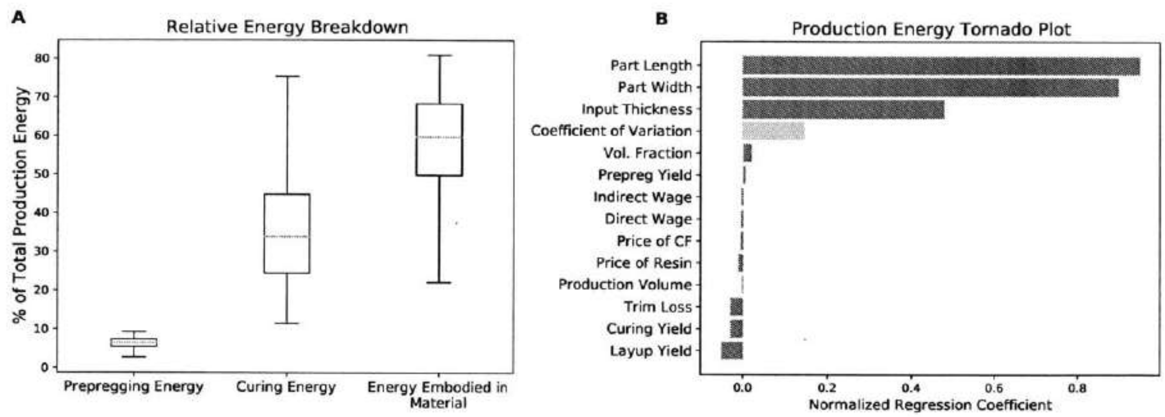

Figure 2-7: (A) Energy distribution between curing, embodied and prepregging. (B) Tornado plot of energy. The design of the part plays the most significant role in total cost and the coefficient of variation is the most important driver after this.

energy tends to dominate followed by the energy used during curing.

As changes in variance only manifest as changes of thickness, this also implies that the variance does not impact the curing cycle. However, the thickness does play a role in the embodied energy. We would therefore expect that the thickness of the part does not play as much as a role in the Tornado plot as the part length and width do. As we can see in Figure 2-7 B this is the case. In addition, we see once again that besides the part design, the coefficient of variation plays the most important role in determining the production energy. Similar to production costs savings, we find that reducing the coefficient of variation from 14% to 9% reduces the energy associated with materials and fabrication by 11.8% on average.

We therefore establish a link between the manufacturing variability and the production cost and production energy. However, we can go further by investigating the effects of manufacturing variability on the use-phase.

2.3.7 Results: Fuel Savings

Reducing the variability does not just impact the production costs, but for airlines it also reduces fuel consumption. Here I explore how the variability in a manufacturing process influences fuel consumption in the use-phase of aircrafts. Figure 2-8 shows that compared to the base case where the coefficient of variation is 14% (corresponding to a safety factor of 1.3), fuel saving can be realized or additional fuel consumption is required due to changes in variability. Reducing the variation below 14% leads to fuel savings and increasing the variation above 14% results in additional fuel consumption.

The opportunities are large, with net present values measured in millions of US dollars over a

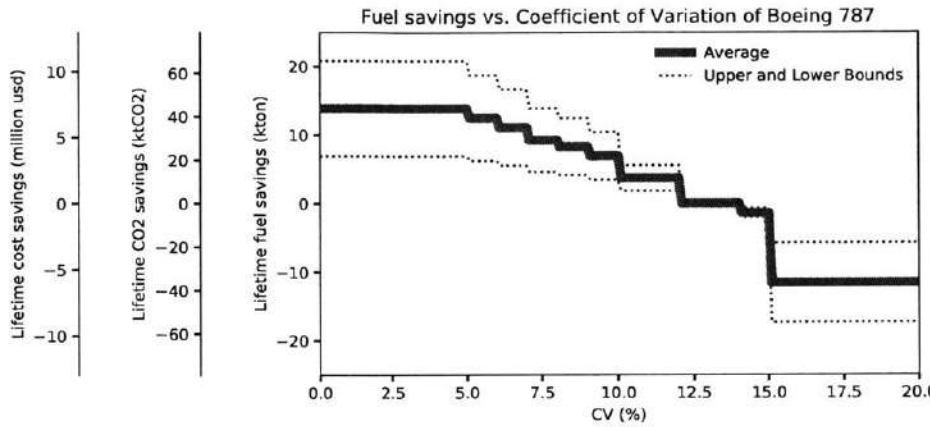

Figure 2-8: Plot of the impact of changing coefficient of variation on fuel savings for a Boeing 787. The jagged nature of this plot is due to the standard used for the model which is not continuous. See Table 2-1.

plane's lifetime and thousands of tons of potential fuel saved. Even at the lower bounds of the model, reducing the coefficient of variation from 14% to 10% saves three million USD in fuel savings per Boeing 787.

The model estimates that a kilogram saved in a plane results in an NPV of \$2,424 compared with the price of producing a carbon fiber part, which is around 310 \$/kg. Using the ASTM F3114-15 standard, reducing the coefficient of variation from 14% to 10% translates into safety factors of 1.3 to 1.15, respectively. For a kilogram part, this would imply weight savings of 0.12 kg, and the fuel savings would amount to an NPV of \$291. These calculations suggest that looking at the entire system it might be financially viable to spend significantly more money during production as long as it leads to reductions in manufacturing uncertainty.

It can therefore be argued that there is a financial incentive to reduce the coefficient of variation not just due to the lowered production cost but also due to the fuel savings, with a resulting reduction in environmental impact.

2.3.8 Case Study Discussion

Ultimately, the use-phase dominates the total energy use in our case study. Removing a kilogram of CFRP from an airplane's structure reduces the fuel consumption by around 0.2 TJ over the plane's lifetime, while the forgone embodied energy in the CFRP is closer to 400 MJ. However, while the use-phase may dominate energy requirements in this scenario, there are many other applications where this is not the case and where saving weight for the sake of saving embodied energy would be highly beneficial. The proposed framework is therefore not only applicable to vehicles, but also generalizes to other products.

In this study, we assume a direct relation between increased overdesign requirement and thickness of CFRP part design. While this assumption is reasonable for the purposes of our model, in industry this relation may be more complicated. In reality, there will be many different important properties to be considered simultaneously. There is work being done on engineering models that attempt to optimize the design of a part depending on strength and load constraints (e.g., [65]). Research in this area can provide a way to more realistically model the relationship between overdesign requirements and part weight. Integrating these design models with our framework can potentially generate a more realistic projection of how safety factors may relate to weight. However, we assert that our assumption and results presented here are reasonable to demonstrate the relationship between variability and overdesign.

Reducing manufacturing variability may require additional cost in order to better control the manufacturing process and may currently be constrained by technological limitations. However, the significant weight reduction potential as well as GHG emission reduction opportunity in a large commercial plane promises that there are both an environmental and a financial incentive for firms to reduce variability in manufacturing processes. These benefits seem to receive little attention in the academic literature. Therefore, we suggest that it is worthy to get a better understanding of the nature of variability in lightweight materials and their manufacturing processes. This calls for research efforts to identify processes and industries where manufacturing variability plays a significant role and provide guidelines in the earlier stage of technology's development.

2.4 Discussions

2.4.1 Reducing Variability

This framework suggests that the best outcome is to reduce manufacturing variability. If it were possible to reduce variability, then many opportunities to save materials would open up. For example, we find that around 19% of concrete is used to mitigate the effects of variance.

There are tools and methods that can be used to reduce variability. So far these methods mainly focus on business practices such as Six Sigma, and the results have been strong. For example, Gijo and Scaria were able to improve yields from 94.86% to 99.48% using process control methods [4]. Another study was able to reduce the standard deviation of a process from 2.17 to 1.69 [5].

There is also increasing interest in process-level solutions such as in-situ process monitoring in AM [66]. Here the process parameters are adjusted in-process based off of readings taken during the process, and might reasonably result in higher consistency. Furthermore, the trend of increasingly more precise fabrication techniques, like we see in integrated circuits, shows that there are ways to reduce variability at least in some circumstances.

While studies on variability for various materials and processes exist, there are few that attempt to understand and highlight the root causes of variability on a process level. Having a better understanding of underlying sources of variability, and potential ways to alleviate them may prove an effective way to reduce our material consumption.

2.4.2 Risk Neutrality

The framing outlined in this chapter implies that one strategy to reduce the magnitude of overdesign is to accept a higher risk of failure or a higher risk of being out of spec. For example, one might argue that instead of requiring less than 1% chance of failure, only 5% is required. With constant variance, this would lower the need for overdesign thereby saving material. Setting risk tolerances is a complicated task that is highly context dependent and I am not advocating that this is the correct strategy. However, in some scenarios, it may be insightful to get a better understanding of how the tolerable rate of failure is decided upon.

In the case of integrated circuits, it appears that manufacturers are accepting increasingly higher odds that there will be a chip-disabling defect and that the chip will need to be discarded, as this is the only way in which chips with very small transistors can be made at all. The scenario is very different when talking about setting risk tolerances for a structure. Here the result of failure might be catastrophic and should not be taken lightly.

2.4.3 Certification Processes

Throughout this work, a few interesting questions arise involving the way in which certification is done. In this chapter, I put forward a simple representation of the certification process. Distributions are established, safety factors are decided, designs are made, and parts are produced. In reality, this is likely to be much more complicated.

Certification and quality control are tightly related to the design process, which is non-linear. There are multiple departments within a company that will have an impact on the way some of these decisions are made. In addition, outside of the company there will be multiple stakeholders who are involved such as the material suppliers, the regulating bodies and the customers. Incentives between all these players may not be perfectly aligned. Companies might also decide to set internal safety factors which are more stringent than those imposed to them from a regulatory body. This may be because companies are risk averse on particular issues and would like to avoid quality problems. Alternatively, companies might prefer to set a more stringent safety factor early in the process to avoid needing to redesign their entire system when a different safety factor is applied due to a change in design or material later in the development process.

More research should be conducted to attempt to answer the questions around how companies actually interact with standards and certification procedures. These questions are especially interesting as additive manufacturing continues to commercialize and CFRP becomes more prominent in automotive applications.

2.4.4 Development of New Technologies

The link between variability and environmental impact warns us that the benefits of novel advanced materials can be undermined if the manufacturing variation is not properly accounted for. If a new technology promises 50% weight saving for the same strength compared to some conventional material, but also requires a 30% overdesign factor to make up for manufacturing variability, then the net benefit is closer to 15% instead of 50%.

The above example illustrates that decisions for a new process technology or material application should be made not on the basis of average values, but rather on effective values – i.e., values that are adjusted to account for manufacturing variability. Or in other words, we should view properties of new materials as distributions rather than point estimates.

Due to the influence of manufacturing variability, it is necessary to investigate the reproducibility of a new process or material early during development, rather than waiting to address these issues when on the production floor. It is admittedly not necessarily easy to do this. Measuring variability accurately requires a large number of samples as will be explored in Chapter 3. Furthermore, there may be a reasonable expectation that the manufacturing variability might be reduced when working with a bigger, stiffer and more precise machine than might be available during prototyping. Norms need to be developed to standardize the measurement and reporting of the distributions. Regardless of some of these issues, it appears that manufacturing variability plays an important role in material and energy consumption, and researchers that want to make a real difference should acknowledge and ideally address the questions surrounding manufacturing variability in the processes they study.

The results of this study also suggest that reducing variability can be an effective way to further light-weighting efforts. Instead of needing to develop and understand a completely new material system, significant weight savings can be achieved by reducing manufacturing variability in existing materials and processes. This suggests more research may be needed in this area. We need better understanding of sources of variability and methods to try and reduce this variability.

2.5 Conclusion

In this chapter, I established a link between manufacturing variability and environmental impact through overdesign. High manufacturing variability necessitates overdesign to ensure that the chance of a part's properties falling below the needed design level is sufficiently low. This overdesign requires more material, thereby, increasing not just the material consumption of the part, but also the weight. Weight has strong relations with fuel consumption, making this effect particularly important for lightweighting efforts in vehicles.

This connection between variability and environmental impacts implies that reducing variability could be an important part of light-weighting efforts. In conjunction with materials that are stronger, we could also develop technologies that reduce variability and thereby increase the effective strength.

Furthermore, the link between manufacturing variability and environmental impact should be taken into account by researchers proposing new processes or materials. Certification and quality assurance can play a significant role and greatly reduce the effective material properties that designers can assume in their designs. Researchers should recognize that deterministic values will never be realized and that a more realistic representation is that of a distribution and an associated risk-level that decision makers are willing to tolerate. As a result, the effective performance of a part, which is adjusted to account for the distribution's variance and the risk level, is more important than the average performance.

Manufacturing variability therefore plays an important role for materials and processes in that it establishes the effective performance of these technologies. In this way, initial predictions of new technologies may be overoptimistic. In the following chapter, I go a step further by also looking at how manufacturing variability, and our (lack of) knowledge of these distributions, impact the commercialization process of specific technologies.

Chapter 3

Variability and Commercialization

Overview

So far I have established that manufacturing variability has important implications for manufacturing processes. High manufacturing variability can imply high overdesign requirements, which in turn drives up cost and material usage, and reduces effective performance of a material. I took the examples of concrete, integrated circuits, and fiber composites and argued that in each of these cases the manufacturing variability increases cost and environmental impact.

In the considered cases so far the assumption is that we know key characteristics of the manufacturing variability. For example, we may know the form of the distribution and might even know how changing certain processing parameters shifts distributions one way or another. In this chapter, I consider examples where this is not the case. Characterizing manufacturing variability is difficult. Understanding how a process works and how distributions in properties change under different sets of processing conditions is even more difficult. Yet as I will discuss, manufacturers need a certain level of understanding before a new process can become commercialized. As a result, if we cannot characterize the manufacturing variability, the commercialization process becomes significantly impeded.

In this chapter, I first look at the issue of characterizing manufacturing variability, namely that it requires a large amount of data to accurately estimate. I then argue that not knowing the distribution of properties can impede or even halt the commercialization of new manufacturing technology. I explore this idea by looking at a historic example, Fiber Reinforced Polymers (FRPs), where issues around manufacturing variability were overcome through a concerted effort from industry and government. I then compare FRPs to a modern example that also struggles with manufacturing variability: AM. I close this section by arguing for a collaborative effort to better understand manufacturing variability in AM.

3.1 Estimating Manufacturing Variability

3.1.1 Confidence Intervals for Expected Value and Variance

Imagine that we have a way to fabricate a steel bar, and we want to understand the distribution of the tensile strength of this bar. As is typically done, we will assume that the tensile strength is normally distributed. The task is then to estimate the mean and variance of this distribution from a set of N samples. I denote the true mean as μ and the true variance as σ^2 and both are unknown. Our estimates will be denoted as $\hat{\mu}$ and $\hat{\sigma}^2$ for the mean and variance, respectively.

As is well known, obtaining a reliable estimate of the expected value requires relatively fewer samples than getting a comparably reliable estimate of the distribution's variance. In other words, as I will show, it takes a large amount of data to establish the variance of a distribution accurately.

We will look at the Confidence Interval (CI) of the expected value and the variance. The CI is an estimated range which is likely to contain the true parameter of a distribution, and is a standard way to measure a parameter. The level of confidence of the CI sets the probability that the interval contains the true value of the distribution's parameter. Equation 3-1 formalizes this relationship. Here Θ is the true parameter, the interval is defined between $[L,U]$, and $1 - \alpha$ is defined as the level of confidence.

$$Prob(L \leq \Theta \leq U) = 1 - \alpha \tag{3-1}$$

The idea is to estimate L and U for a given α, from a set of N samples that we draw from our true distribution. In other words, we manufacture N bars and measure their tensile strengths. I denote the tensile strength of bar i as x_i, where $\mathbf{X}$ is the vector of all $x_1 \ldots x_N$ and the mean over all N samples is $\bar{x}$.

Confidence Intervals of the Expected Value

Because the variance of the distribution is unknown, we first estimate the variance by using the sample variance as defined in Equation 3-2 [67].

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

The $100(1-\alpha)\%$ CI of the true mean μ is then as defined in Equation 3-3 [67]:

$$\bar{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{N}} \leq \mu \leq \bar{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{N}} \tag{3-3}$$

where $t_{\alpha/2,n-1}$ is the value of the student t-distribution with $N-1$ degrees of freedom such that $Prob(t \geq t_{\alpha/2,n-1}) = \alpha/2$ [68].

If the variance of our distribution was known, then we could use the Central Limit Theorem (CLT) to assume that $\bar{x}$ is normally distributed with variance $\frac{\sigma}{\sqrt{N}}$, but because we do not know σ we have to estimate it from our sample and have to use the student-t distribution to adjust for the added uncertainty.

Sample	**Value**
x_1	29.11
x_2	29.63
x_3	40.32
x_4	24.49
x_5	29.80
x_6	34.67
x_7	25.14
x_8	29.38
x_9	27.96
x_{10}	32.82

Table 3-1: 10 samples drawn from a normal distribution with mean 30 and standard deviation 4.2.

Now let us consider the CI of the variance.

Confidence Interval of the Variance

To get the CI of the distribution's variance, we instead have Equation 3-4 [68].

$$\frac{(N-1)s^2}{\chi^2_{\alpha/2,N-1}} \leq \sigma^2 \leq \frac{(N-1)s^2}{\chi^2_{1-\alpha/2,N-1}} \tag{3-4}$$

where $\chi^2_{\alpha/2,N-1}$ is chosen such that $Pr(\chi^2_{N-1} \geq \chi^2_{\alpha/2,N-1}) = \alpha/2$.

Comparing convergence of CI of mean and variance

Let us assume that on average the steel bars break at 30 kN and that the coefficient of variation is 0.14 and. This implies that $\mu = 30$ and that $\sigma = 30 \cdot 0.14 = 4.2$, or that $\sigma^2 = 17.6$. We set $N = 10$ and $\alpha = 0.05$. Using a simulation of a normal distribution with the given parameters, we get the 10 samples as shown in Table 3-1. The resulting 95% confidence interval for the mean is [27.05, 33.61]. This is roughly 30 +/- 3, meaning that we are within roughly +/- 10% of the true estimate with 95% confidence. The resulting 95% confidence interval for the variance is [10.62, 59.92]. This is roughly 17.6 -40%/+240%. The confidence interval is extremely large for the variance; we cannot establish it with high certainty.

In Expectation

We can repeat the above experiments in expectation for different levels of α and different sample sizes N. I denote the lower bound and upper bounds for the estimate of the mean as L_m and U_m respectively. Then noting that $L_m = L_m(\mathbf{X}; \alpha, N)$, the expected value of the lower bound over all draws is:

$$\mathop{\mathbb{E}}_{\mathbf{X}}[L_m(\mathbf{X};\alpha,N)] = \mathop{\mathbb{E}}_{\mathbf{X}}\left[\bar{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{N}}\right] = \mathop{\mathbb{E}}_{\mathbf{X}}[\bar{x}] - \frac{t_{\alpha/2,n-1}}{\sqrt{N}}\mathop{\mathbb{E}}_{\mathbf{X}}[s] \tag{3-5}$$

where only the $\bar{x}$ and s terms are dependent on $\mathbf{X}$ and therefore the expectation is only applied to these terms.

A complicating factor for Equation 3-3 is that s is a biased estimate of σ meaning that $\mathbb{E}[s] \neq \sigma$. However, a factor can be derived to correct for this unbiased estimate [69], [70]. At $N = 2$ this correction factors is around 0.797, meaning that in expectation s is 1/0.797 times larger than σ. This correcting factor quickly becomes negligible and goes to 0.98 or closer to 1 when $N \geq 15$. However all of this is accounted for in the t-statistic, and as a result we *can* use σ as an estimate for s.

Also note that $\mathop{\mathbb{E}}_{x}[\bar{x}] = \mu$. We therefore find that:

$$\mathop{\mathbb{E}}_{\mathbf{X}}[L_m(\mathbf{X};\alpha,N)] = \mu - t_{\alpha/2,n-1}\frac{\sigma}{\sqrt{N}} \tag{3-6}$$

And similarly for the upper bound:

$$\mathop{\mathbb{E}}_{\mathbf{X}}[U_m(\mathbf{X};\alpha,N)] = \mu + t_{\alpha/2,n-1}\frac{\sigma}{\sqrt{N}} \tag{3-7}$$

For the lower and upper bounds of the variance (L_v and U_v, respectively), we can go through a similar exercise. Noting that only s^2 is dependent on X, and that $\mathop{\mathbb{E}}_{\mathbf{X}} = \sigma^2$ we find that:

$$\mathop{\mathbb{E}}_{\mathbf{X}}[L_v(\mathbf{X};\alpha,N)] = \frac{(N-1)\sigma^2}{\chi^2_{\alpha/2,N-1}} \tag{3-8}$$

and

$$\mathop{\mathbb{E}}_{\mathbf{X}}[U_v(\mathbf{X};\alpha,N)] = \frac{(N-1)\sigma^2}{\chi^2_{1-\alpha/2,N-1}} \tag{3-9}$$

Now using Equations 3-6 through 3-9 we can plot the confidence intervals over a range of confidence levels, and a range of number of samples. The results are shown in Figure 3-1. In this figure, the confidence intervals are normalized by subtracting the true parameter and then dividing by the true parameter. For example, at $\alpha = 0.05$, and $N = 30$, the $E[U_v]$ is calculated to be 7.254. Subtracting the true value and dividing by the true value we get (7.254 - 4.2)/4.2 = 0.72 = 72%, which is the value that is plotted.

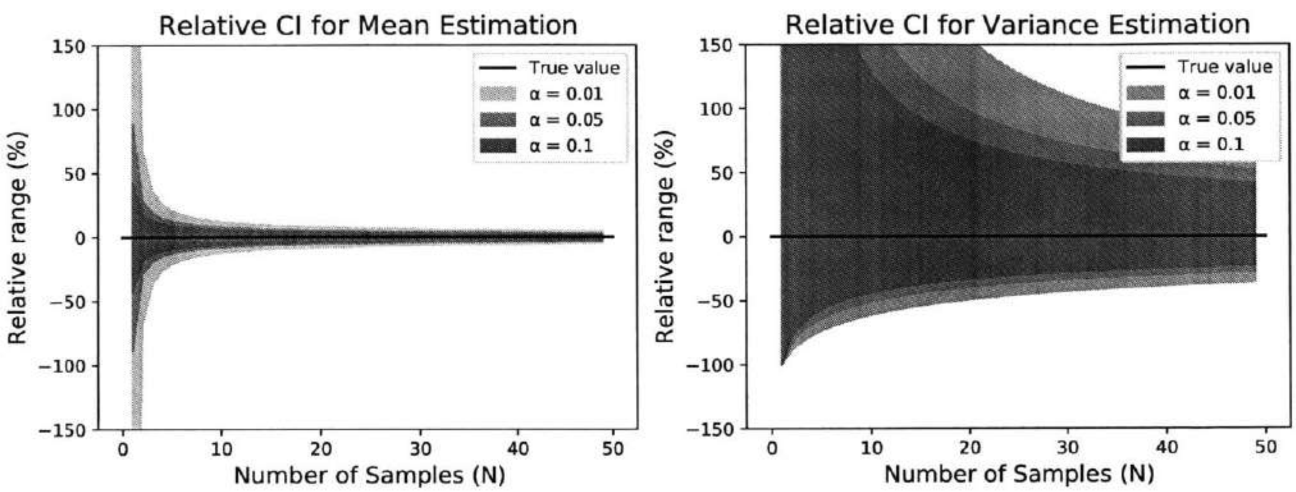

Figure 3-1: Expected Confidence Intervals for mean (left) and variance (right) when drawing from a normal distribution with mean = 30, and variance = 17.64. To allow comparison the confidence intervals were normalized by subtracting the true parameter and dividing by the true parameters (i.e., if the upper CI for mean was 33, then (33 - 30)/30 = 0.1 = 10% is plotted). As a result, the y-axis is measured in percent. The x-axis shows the number of samples. In both cases increasing the number of samples makes the relative distribution tighter, meaning that we get a more accurate estimate, but for the variance the convergence is much slower than for the mean.

The CI for both the mean and the variance estimations tend toward the true value, which in this figure is the horizontal line at 0. However, as is clear from the figure, estimates for the variance require many more samples to get a tight bound. Ultimately this is due to the way the chi-squared and t-squared distributions vary as N increases. Practically this shows that it is difficult to get an accurate measure of variability for a given confidence level.

Note that due to the linear scaling of variance in Equations 3-8 and 3-9, the plots of variance remain the same for every level of σ^2 that is set.

We therefore see that it requires a substantial number of data points to get an accurate bound on the variance. Hundreds of data points are required to get within $\pm 10\%$ of the true variance at a 95% confidence level, while only a dozen or so are required to estimate the mean at around this accuracy. From this point alone it can be imagined that trying to optimize the variability of a process might be difficult.

3.1.2 Difficulties in Characterizing Manufacturing Distributions and Commercialization Impediments

The previous section demonstrated that in order to be very accurate when measuring the variance of a distribution, dozens if not hundreds of samples are required. This suggests that it is hard to optimize the variability. If there are multiple parameters that can each be varied, and if it takes dozens of samples at each permutation to get an accurate sense of variability, then it would take a lot of data to find the processing conditions where the manufacturing variance is minimized. While this is true, it should be noted that reducing manufacturing variability remains a feasible goal, as has been demonstrated in various papers that do just this for various processes [71]–[73]. Nevertheless, due to the amount of data required, minimizing variability appears to require substantially more effort than optimizing an average.

Further complicating matters is the fact that multiple properties are typically important for a product. Beyond the ultimate tensile strength of a steel bar, we might also care about the stiffness, yield strength, hardness, corrosion resistance and toughness. This adds further dimensionality to a problem that has to be overcome.

For a new manufacturing process, the community might take some time to figure out basic mechanics like how the mean shifts when various processing conditions are changed. Studies on the variability might have a relatively lower priority. However, without an understanding of the variability of a process it is impossible to know what the rejection rate is, or how much overdesign is required to adhere to regulations. Furthermore, manufacturing process control requires one to know whether the process is in control or not, which once again requires an understanding of what the distribution should look like. Many standard and necessary procedures on the production floor are therefore impossible to implement without an understanding of these manufacturing distributions. All this uncertainty brings risks to manufacturers. Defective parts may lead the manufacturers to be held liable, or could tarnish the reputation of a manufacturer as a high quality producer. As a result, manufacturers are wary of processes they do not understand.

So if it takes a lot of data to characterize the manufacturing variability, how do new processes displace conventional processes? In the next sections, I explore how this transition is indeed a problem, but that there are some approaches that have worked in the past.

3.2 Fiber Composites - a History

3.2.1 Overview

Fiber Composites or Fiber Reinforced Polymers (FRP) are discussed in Chapter 1 as a case study of the impacts of manufacturing variability. This field still struggles with high manufacturing variability, but it has come a long way since it was first invented. Here I take a historic perspective on these composites and discuss how this technology was able to overcome obstacles around manufacturing variability and process control.

Time and cost are not the only barriers to implementing a new process on the production floor. The literature on FRP indicates that a low level of confidence in the reproducibility was a major reason why companies hesitated to adopt this new technology when it was first invented [74], [75]. The industry had built up substantial know-how about existing and conventional technologies, which makes shifting to a new technology comparatively risky. Composites were held back because of this confidence barrier until a NASA-led program in the 1990s was instigated. This program had the aim to create a shared database of material properties and important process parameters [76]. The program was deemed successful in giving manufacturers more confidence in composites technology, and is credited as one of the reasons why composites are relatively widespread in aerospace – a notoriously conservative and risk-averse industry.

3.2.2 Early Promises of Composites

Early advocates of fiber composites promised savings of 50% of a structure's weight compared to using an aluminum frame [76]. It is difficult to determine how much weight savings occurred in reality, but estimates place it closer to 20% of the structure's total weight [61], [77]. This remains true despite three decades of developments in the composites field.

Part of the shortcomings may be explained by the difficulties in manufacturing composites. The FRP production process proved to be more difficult to control than anticipated [76], and these challenges were not considered in initial optimistic predictions. As a result, additional safety factors must be applied when using composites [18]. These safety factors are in the range of 1.2 to 1.5. There is no indication that this was foreseen by early composite advocates, but it severely undermines the weight-saving potential.

It should be noted that FRPs are not considered a failure. A 20% reduction in weight is a significant boon for an industry where fuel accounts for a substantial proportion of total costs. Furthermore, there are other benefits like corrosion resistance, and part consolidation which continue to be attractive features for aerospace and other industries. The market for advanced composites grows at a steady pace - around 3-5% annually [78], [79] and the percentage of FRP in an aerospace structure continues to rise. However, the lesson to draw from these developments is that the predicted weight savings of 50% were unrealistic to achieve, at least within the first few decades. Perhaps when isolated from other factors it is possible to theoretically achieve 50% weight-savings for individual parts. However, once the technology reaches the factory floor the results are more sobering.

3.2.3 NASA ACT and AGATE Programs

In 1989 NASA started the Advanced Composite Technology (ACT) program. The aim of the program was to develop an understanding of composites to the point where they could become cost competitive with metal structures in aerospace applications [80]. The program involved a collaboration between industry, including Boeing, Lockheed-Martin, Northrop-Grumman, and McDonnell Douglas. In addition, research grants were given to universities to develop analytical models for resin flows and to do mechanical tests on new materials. The program finished in the year 2000 [81].

The program was started in order to provide further boosts to the field of fiber composites. The need for such a program was pointed out one or two years earlier in a report to NASA which claimed that the commercialization of advanced FRP was "disappointingly slow" [75]. While composites parts were already certified by the Federal Aviation Agency (FAA) and flown by various airlines, the steep investments required to develop, test, and certify parts impeded rapid progress [82]. Manufacturers were uncertain of the properties of the new materials in various loading conditions, which meant that for each new part substantial empirical and analytical work had to be done. This drove up production and design costs considerably.

The ACT program sought to alleviate barriers around process control through a number of activities. The program developed new analytical tools to predict the mechanical properties of a composite part, it developed best practices in design and production by fabricating and testing full-scale parts, and it collected the results of all experiments in a single database which it shared with manufacturers [80], [83], [84].

In order to create the database, it was recognized that standard testing procedures had to be developed [76], [85]. For these purposes, NASA decided to use the Military Handbook on composites: MIL-HDBK 17 [86]. A standardized input form was used to collect and store data in the shared database, and numerous quality checks were done to ensure that the results were reasonable. Further details on the operations of the database are given in a NASA report called "Database of Mechanical Properties of Textile Composites" [83].

The ACT program collected many data, but the impact was not fully felt until the Advanced General Aviation Transports Experiments (AGATE) program was founded in 1995. The AGATE program attempted to boost aviation innovation by shortening the time to develop small aircrafts. Along the way, AGATE developed a methodology for qualifying materials called the AGATE methodology [87]. Instead of requiring part manufacturers to individually qualify each material and part, material suppliers were required to do a large proportion of the certification. In this framework, the suppliers need to provide process parameters alongside any material that they supply. If part manufacturers could establish "equivalency" with previously tested parts, and the manufacturers adhered to the supplier's process parameters, then only a few additional tests would need to be conducted in order to gain certification. This should aid in bringing the field towards industry-wide specification, instead of relying on company specific specifications. Full details of the procedures can be found in a published report [88]. The AGATE methodology would not have been possible without the data and knowledge provided by the ACT.

It is difficult to determine how much of an effect these programs had on the commercialization potential of composites. However, the shared database seems to be looked at as an example of success [89]. By providing data on properties and testing procedures, and by clearly assigning responsibilities to parts of the supply chain, the resources required to develop and understand the properties of newly designed parts are significantly reduced. In this way, the programs served to reduce barriers to commercialization.

3.2.4 Lessons Learned from FRP

It appears that a few individual companies were able to get their FRP products certified and implemented in the 1980s. However, for the majority of companies and applications, the costs associated with certifying a new FRP product were simply too great. This meant that the entire industry was being held back due to issues with manufacturing variability. Only through concerted efforts did the entire industry grow to become what it is today. This observation implies that in some situations it could be worthwhile to share data and knowledge with competitors and suppliers as it moves the entire field forward. However, for an individual company, this might seem like a bad deal. Manufacturing data and knowledge are valuable assets, and simply sharing it with everyone seems like throwing away a competitive advantage. It is not immediately clear how the ACT was able to incentivize everyone to collaborate, but it proved to be effective.

Setting clear responsibilities for various stages of the certification process proved to be another effective strategy. Making material suppliers responsible for the majority of the certification process allowed for flexibility. If instead, the end products were to require full certification, it would take a lot of data to verify the quality of a new product. Now through the framework of equivalency, small companies looking to make FRP parts can use pre-approved materials.

There may therefore be some interesting lessons that can be drawn from fairly recent history. In the next section, I argue that some of these lessons could be applied to a contemporary example.

3.3 Commercialization Problem - AM

3.3.1 Similarities Between AM and FRP

Additive Manufacturing (AM) is being hailed as a technology which will disrupt many industries. By building up a part layer by layer, AM enables the fabrication of parts which were previously very difficult if not impossible to make. In theory, this removes a significant constraint when designing parts, and parts can, in turn, be designed to be much closer to their theoretical optimum instead of being held back by production limitations [90]. Furthermore, the flexibility of this production process allows mass-customization (e.g., see [91]). Advocates of AM boast of the technology's ability to save material and energy [92], to save cost and time [93], and lead to a new 'paradigm' in production [94]. However, despite its many benefits, and while AM has gained increasing interest since it was first invented in the 1980s, it has not yet gained widespread industrial use.

I will show that AM is currently in a similar position as FRPs were when the ACT began. AM has found some industrial applications and a few AM parts have been certified for use by either the FAA or the Federal Food and Drug Agency (FDA). However, there is still a low level of confidence by industry and not many functional parts are in use at this time. The uncertainty in the production process is impeding development efforts. In this section, I consider the status of AM in two industries: aerospace and medical devices. I will argue that the low level of confidence is a key barrier to widespread commercialization in both of these industries, and that a collaborative database approach, similar to what occurred with FRP, might serve to remove this barrier

3.3.2 AM Regulation by the FAA

In September 2015, the FAA and the Air Force Research Laboratory hosted a three-day workshop on AM in the aerospace industry. It was attended by NASA, the military, the National Institute of Standards and Technology, and by industry. The objective of the event was to discuss the status of AM in aerospace, with particular emphasis on qualification and certification processes. A summary report of the workshop was made available online [95].

A recurring theme of the event was the issue of process control. From the workshop's summary report: "The potential exists for a high degree of manufacturing variation because of the process, machines, suppliers, and input stock."[95]. In addition, participants agreed that the feedstock, specific AM process, and the resultant parts are "highly integrated and interdependent" and that this creates problems for the certification process. If all the parameters are indeed highly interdependent, it will be difficult to isolate different steps and to certify each step individually. Considering the entire process as a whole for certification makes for an inflexible system, as it does not allow modularity if changes need to be made. Standardization of testing and production parameters is ongoing, but is in early stages. The industry pointed out that these quality assurance issues are one of the main barriers to commercialization. Current nondestructive tests (NDT) fall short of requirements, and little is known about how the FAA will handle AM certification, which makes pursuing AM risky.

On a few separate occasions, the summary report mentions parallels between AM and composites, arguing that materials, processing conditions and the final part are highly related for both technologies. In other words, the material properties are not fully determined until the part has been fabricated. The report argues that this makes these technologies distinct from conventional processes like machining.

A complicating matter in aerospace is that the main failure mode is fatigue cracking [96]. Testing for fatigue is difficult as it takes a long time, and there are many inputs to the experiment (frequency, load, load profile) which means that if companies are looking to share and compare their data easily, they must use comparable parameters. This requires standardization. Furthermore, because it often takes thousands to millions of cycles to show signs of fatigue, it requires a lot of flight-hours to prove that a part is unlikely to fail via this mode.

The aerospace industry is therefore hitting a large barrier for implementing AM parts. Manufacturers need to be confident in their parts, but process control for AM is in early stages and tests for quality are not yet standardized. It may be possible to develop a bespoke testing procedure for a specific product and application, as demonstrated by the fact that AM parts have been certified and are currently flying [97]. However, this approach requires high investments, and is not scalable to the entire industry. A more standardized approach is required. This means that material qualification, production, and testing procedures must be defined.

3.3.3 AM Regulation by the FDA

Another promising prospect for AM is the medical device industry. The complex shapes and the ability to mass customize make AM an interesting technology in this area. AM has already found applications in medical devices. In an interview with MDDI, a medical device industrial magazine, Steven Pollack, Director of Office of Science and Engineering Laboratories at the FDA, said that at least 85 AM products have so far been approved by the FDA [98]. It should be noted that so far these devices have been approved through the 510(k) processes, meaning that they are "substantially equivalent" to devices which were already approved and manufactured through traditional means. However, there have not yet been AM parts which have successfully passed the more stringent Premarket Approval. Being limited to the 510(k) process is unsustainable in the longer term, because one benefit of AM is that it can make geometries which are otherwise almost impossible to make. Ideally, we should therefore be producing medical devices that are *inequivalent* to existing devices, due to the flexibility that AM provides. Certifying these kinds of devices will require Premarket Approval, which in turn requires proof of safety.

Due to the lack of material qualification, and production and testing procedures, it is difficult to adequately prove that parts are safely made and fit for medical applications. This industry therefore faces similar problems as aerospace, and it is therefore worthwhile to examine some of the literature on AM in medical devices.

The FDA held a workshop in 2014, where it invited members of industry and academia to discuss some of the challenges of AM in medical devices and to allow stakeholders to express their thoughts. The draft guidance was based on this workshop and came out 18 months later [99]. The purpose of the draft summary was for the FDA to share some initial thoughts and to give other stakeholders the opportunity to comment. This is therefore not a comprehensive regulatory document, but rather a way for the FDA to test how industry would react to some regulatory approaches. The document was updated about a year later [100], although the main messaging remained the same. One of the key takeaways of the workshop was that process control is one of the main critical barriers to more widespread AM medical devices.

The draft guidance mainly makes suggestions for parameters to consider when internally establishing an AM production procedure, but few specific settings or standards are proscribed. Some parameters or machine inputs the document suggests to keep in mind include the power of the energy delivery system (e.g., the laser), the build speed, build path, focal point or nozzle diameter, particle size and distribution, chemical composition of the material, location of the build and orientation of the part. This gives some insight into the kinds of parameters the FDA will be considering in the PMA process; however, there is little guidance for actually establishing the correct settings. The guidance notes that even if the same part were to be printed, using the same settings, on two printers of the same brand and model, there could still be significant variation between printers. The FDA therefore seems to imply that each printer would need to be individually validated, which may be a larger regulatory burden than anticipated by industry.

The position of the FDA is that AM does not significantly differ from other manufacturing processes and that the existing regulatory framework should be used to regulate AM devices. They state: "It is anticipated that AM devices will generally follow the same regulatory requirements as the classification and/or regulation to which a non-AM device of the same type is subject to" [99]. This indicates that the FDA does not intend to create a separate framework, but instead might make some adjustments to the current framework. However, the FDA also recognizes that there are aspects which makes AM different from conventional processes. For example, the FDA points out that there is a "relative lack of medical device history" for AM devices, which poses challenges for determining the optimal process settings or the most appropriate testing methods. This sentiment was shared by industry.

3.3.4 Environmental Impact of AM

Two environmental claims made about AM are that AM can save weight and save material. Weight savings lead to reduced environmental impact by lowering fuel consumption in vehicles. These weight savings are enabled by AM's ability to make complex parts, which removes many limitations that designers would typically be working with. In addition, part consolidation could make it possible to make much simpler parts which could also lead to weight savings. Some sources are claiming weight savings on the order of 40-60% [92], [101]. However, it is not yet clear that these savings will actually be realized. For example, it may turn out that due to the uncertainties in the manufacturing process a higher safety factor is required, just like was the case for FRP. This could negate a large amount of the weight-savings claimed by AM advocates.

Material savings could be expected with AM because of low buy-to-fly ratios. A buy-to-fly ratio is the weight ratio between the input material and the final part. A high ratio implies that there is significant scrap. In aerospace typical ratios are on the order of 10-20 [102], which suggests that the vast majority of the material ends up as scrap. AM brings that number much closer to one. However, this assumes that the un-used powder in the powder bed can be recycled. It has not yet been determined whether this is reasonable. Some experiments suggest that recycling up to 40 times results in little changes in properties [103]. Others suggest that after about four reuse times the powder becomes out of spec [104]. As the qualifying process of AM by itself is not yet fully understood, it will require many more robust studies before standardized procedures on powder reuse will emerge. It therefore may be possible that a buy-to-fly ratio of one will be unrealistic due to limitations imposed by process control requirements.

The discussions within the aerospace and the medical device industries appear to be running in parallel. Both industries are looking for certainty in the fabrication process, better nondestructive testing procedures, and clarity from their regulatory bodies. While the specifics do not have to be identical, it might make sense to develop our understanding of AM under a common framework. Both industries are essentially working on the same problem, and working together could prevent duplication of efforts. Although aerospace and medical devices are two typical industries cited in the AM literature, the same thinking could be applicable to other industries as well.

3.3.5 Similarities between AM and FRP

It appears that AM is going through a scenario that FRPs have already gone through. As a result, we might be able to learn from FRPs and apply some lessons to AM. AM made its first debut in the late 1980s [105]. Composites, on the other hand, are more developed and have been around since the 1950s and 1960s [106]. Composites may therefore provide a snapshot into the future for AM.

Some researchers have called for a top-down approach from the government to instill more structure in this industry [107]. There are individual cases where AM parts are being used as functional components [108]. However, it requires significant investment to get to this point for an individual company, and the process has to be restarted from the ground up for any new part or material. This is exactly the point at which the ACT program started for FRPs, and perhaps it makes sense to have a collaborative effort to understanding process variation and control in AM.

The AGATE methodology, where the majority of the regulatory burden is shifted to material suppliers, might too be applicable for AM. Perhaps it makes sense for the material suppliers to suggest operating conditions for specific materials, and for the companies that are doing the actual fabrication of AM parts to just have to establish equivalency. However, at this point, not enough appears to be known about the sources of variation in AM in order to back up this idea with robust evidence.

Finally, FRP might help us set more realistic expectations of the light-weighting potential for AM. Due to the effects of overdesign as described in Chapter 1, many lightweighting benefits initially promised by FRP advocates were nullified once the materials had to pass regulatory standards. For AM it appears that this effect is not considered as much as perhaps it should be.

3.4 Conclusions

In Chapter 2, I demonstrated that high manufacturing variability has cost and environmental implications. In this chapter, I argue that not knowing the manufacturing variability also has implications as it impedes the process from being commercialized at all. If the manufacturing variability for a given process is not understood it becomes impossible to guarantee a level of quality, and, as a result, the parts cannot be certified.

Manufacturing variability was shown to be a challenging statistic to estimate accurately because a high number of samples are required to get a tight confidence interval on the estimate. Uncertainty around the manufacturing distribution of FRP proved to be a difficult challenge for this technology in the 1980s. The ACT program was instrumental to remove this barrier and was able to get industry and regulatory bodies to collaborate and create a framework for certification that was not too cumbersome for individual companies. The new approach to certification, where the majority of the regulatory burden is placed with the material suppliers, together with all the knowledge that was shared as part of the ACT, seems to have been effective in boosting the commercialization path of FRPs.

In this chapter, I also drew parallels between FRPs and AM, and argue that many of the issues that FRP was facing regarding commercialization are also being faced by AM. Perhaps some of the approaches that worked for FRP will also work for AM.

More data are required to better understand the manufacturing variability of the AM process, which would, in turn, remove obstacles for further commercialization. Gathering this data is expensive, but many experiments have already been conducted on AM that might be able to help us understand the manufacturing distributions. All this information has not yet been combined and instead exists in various published documents. In the next chapter, I explore how we might be able to extract valuable manufacturing information from academic papers on AM using text-mining techniques. These techniques might be able to extract and structure data from the thousands of papers that researchers have written on AM. In doing so I attempt to characterize the manufacturing variability of this innovative technology.

Chapter 4

Text Mining for Manufacturing Data

Overview

In the previous chapters, I argued that it is important to be able to characterize the manufacturing variability of new process technology. First, this is because high manufacturing variability implies higher environmental and economic costs, and in order to reduce these costs, we would want to reduce manufacturing variability. This means we will need to know what the manufacturing variability is in the first place. Second, and probably more limiting, in order to manufacture a part with confidence manufacturers must know what the error rate is, which requires the distribution of the process to be known. In fact, in many situations, parts will not be approved for sale unless there is some kind of statistical guarantee that the part meets the specifications. However, I also argued that it requires a lot of data to accurately estimate the variability of a process and that companies are reluctant to share this information.

In this chapter, I look at a source of data that might be able to help characterize the manufacturing variability of Additive Manufacturing: published documents. The issue is that the data contained in these documents are typically unstructured; they appear as text, disparate tables and figures, and are not in a format that allow for easy analysis. However, through recent developments in machine learning and text mining, it becomes possible to extract and structure this information. There remain some challenges with this approach, which will be discussed, but even with these challenges valuable information can be extracted.

I first give an overview of the text mining pipeline used for this project. I then discuss two challenges for the pipeline. The first is that most text-mining techniques operate on a single sentence-level, but I need to be able to link information across multiple sentences and tables. I propose and implement some strategies to overcome this issue. Secondly, researchers are inconsistent with the information they provide in published articles. This makes it hard to have a complete set of process parameters and properties for analysis. Regardless of these issues, some information can be reliably extracted and analyzed and these results will be discussed. I finish this chapter by outlining areas of future work.

4.1 Text Mining Pipeline

4.1.1 AM and Conventional Manufacturing Papers

There are many researchers who do experiments and publish their results in academic papers. These papers therefore contain a wealth of information. For instance, one paper in the domain of AM [109] includes the composition of the material tested, processing parameters like the laser power, laser spot size, minimum offset height, laser wavelength, focal length, and scanning speed, as well mechanical properties like the yield strength, ultimate tensile strength and elongation for multiple samples. This is a representative example of these kinds of papers [110]–[114].

According to Engineering Village [115], around 3,000 papers on AM that include mechanical properties are available. When filtering to only papers from publishers that allow MIT to mass-download and text-mine their papers, the number is closer to 1,800 papers. Filtering this to the set of papers that deal with metals instead of manufacturing with polymers, we have around 800 papers remaining, but only 550 of these were actually about AM which is checked by searching the text for keywords such as known AM processes. Note that every paper tends to have multiple experiments, and, as a result, these 550 papers represent a few thousand datapoints.

Data from these 550 papers can be supplemented by looking at conventional processes. Often the same materials that are used in AM have been used in other fabrication methods like casting, hot forging or rolling. It might be useful to compare the results against each other and see what knowledge of conventional manufacturing carries over to AM. As a result, the pipeline is set up to also extract data from these conventional manufacturing papers. When considering papers on conventional processing of metals we find 10^4 papers using Engineering Village. Similar to the AM domain, these papers describe composition, processing conditions, and mechanical properties (e.g., [65], [116]–[118]).

4.1.2 Architecture of Pipeline

In recent years, publishers such as Elsevier and Springer have created interfaces to their databases that allows for mass downloads of their journals. We download these papers, preferably as HTMLs or XMLs, using the APIs (Application Programming Interfaces) written by these publishers. The papers are then processed with the text and tables being handled in parallel tracks. The extracted information is then saved as a JSON file and stored in the database. An overview of the process is shown in Figure 4-1.

This pipeline is based on previous work done in the Olivetti Group [119], [120], and further information on how the pipeline operates can be found there. The most important parts of the pipeline include the Token Classifier and Info-Extractor. The Token Classifier tags each word in the paper as a material, operation, condition, number, unit or some other label from our list of 27 labels. Info-

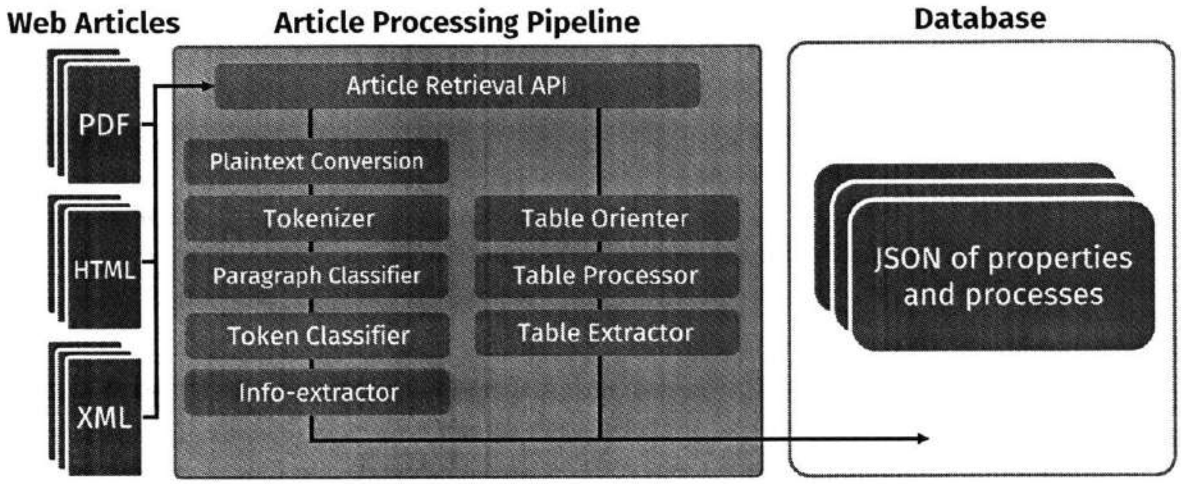

Figure 4-1: Schematic of Text Mining Pipeline.

Extractor then uses these tags alongside the grammar of the sentence to connect information together. For instance, a unit might be paired with a number, which might be paired with a condition. This process is described in the following section.

4.1.3 Token Classifier and Info-Extractor

The setup of our Token Classifier and Info-Extractor is representative of many text-mining pipelines across a wide range of domains such as nanotechnology [121], medicine [122], biology [123] and chemistry [124].

The first step is labeling the important words in a given sentence. This is a Named Entity Recognition (NER) problem [125]. These labels will differ depending on the domain the specific text-mining pipeline is designed for. In this case, there are a set of 27 labels such as "operation" or "material" A full list of labels can be found in the Appendix.

Next, every word in a sentence is represented as a vector through the use of word embeddings. There are multiple ways to do this, and we utilize two approaches in our model: FastText [126] and Word2Vec [127]. Both of these models were trained on around two million materials science and manufacturing papers [120]. These embedded words are concatenated before being fed into the bi-directional Gated Recurrent Unit (GRU) [128], which is a type of recurrent neural network architecture that is specifically designed for sequences of data such as text. The schematic of the Token Classifier is provided in Figure 4-2. Data were collected by Ph.D. students in Materials Science by manually annotating around 200 papers on materials science as described in [119].

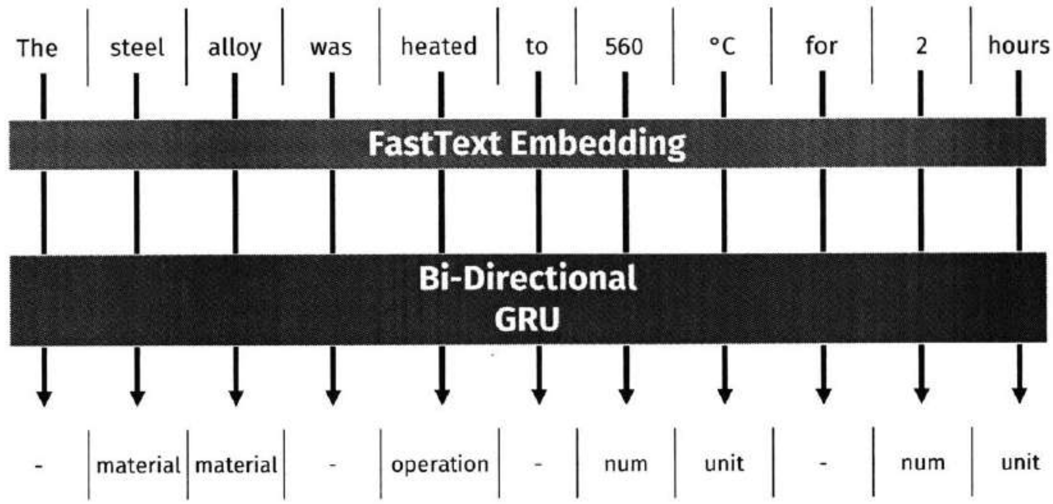

Figure 4-2: Architecture of the Token Classifier; a sequence-to-sequence prediction model. The embedding layer uses both FastText embeddings and Word2Vec embeddings. The bi-directional GRU is a recurrent neural network architecture typically used to process text.

The output of the Token Classifier is fed into the Info-Extractor algorithm. First, the sentence is parsed using Spacy [129], [130] to create a dependency tree parse as shown in Figure 4-3. Then the words in this sentence are linked to each other using the grammar tree. For instance, when we see that "2" is labeled as a "number", we then move up the grammar dependency tree from "2" to the nearest word that has been labeled as a "unit", which in this case is "hours". We then move further up the dependency tree until we come across a word labeled as "operation", which in this case is "heated". We can repeat the process for the temperature as well. Ultimately this allows us to know the temperature and time that the heating operation was conducted at. Once no more linking can be done in the sentence, the resulting relationships are captured in a structured dictionary which is saved as a JSON file in the database.

Figure 4-4 provides a sanity check on the extraction process. Here the temperature distribution is plotted when looking at various operations as extracted from papers on conventional manufacturing of steel alloys. Some operations do not make sense at all temperatures. For instance, by definition austenitizing requires the alloy to be heated above its critical temperature which for most steels is around 720 °C. As can be seen in Figure 4-4, all datapoints fall above this critical temperature, giving at least some indication that the extraction process is reasonably accurate.

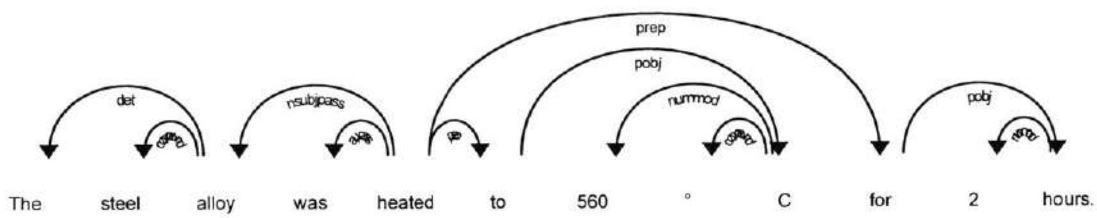

Figure 4-3: Dependency tree of an example sentence.

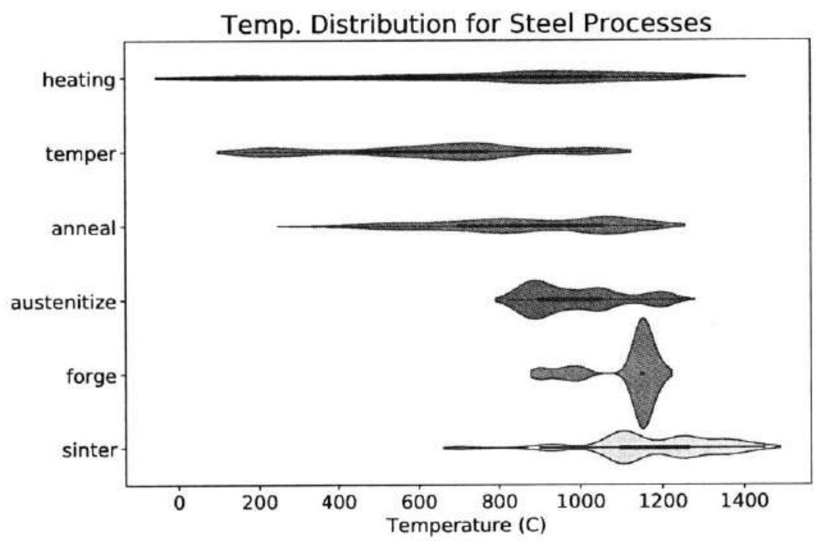

Figure 4-4: Distributions of temperatures for various operations as extracted from papers on conventional processes on steel alloys. This plot serves as a sanity check of the extraction process. Operations like "austenitizing" are generic and spread across a wide range of temperatures. However, some operations like "sintering" require the material to be heated beyond a certain critical threshold which for steel is around 720°C. The fact that we do not see any data below this number indicates that the extraction process is reasonable.

This kind of setup of a token classifier and info-extractor is quite representative of many text-mining pipelines, although they often go by other names. However, this approach has a significant flaw in that the information is only linked through grammar in an individual sentence. Because information can be spread all over the document, and because we need to be careful how all this information is linked, these current methods fall short. The following section discusses this challenge in more detail and proposes some solutions.

4.2 Challenge I: Intra-Document Information Extraction

4.2.1 The Problem of Intra-Document Extraction

Linking information beyond a single sentence remains a hard problem [131]. Within a sentence, we can link entities through grammar, but there is no obvious equivalent for cross-sentence relationships. This is forcing researchers to take other approaches, such as cross-sentence dependency based neural networks [132] or Graph LSTMs [133]. In some situations, it is possible to link information across sentences through rule based approaches. For example, in both [119], [124] entity relations are extracted across sentences by assuming that the sentences describe a process performed in chronological order. So far, all of these models are limited in their application to just text and fail to link information across an entire document including tables and figures.

In the domains that are being considered for this pipeline, there are often multiple experiments described in a single document. For instance, one paper might attempt to measure the effects of laser power on the ultimate tensile strength, and collect data with several different levels of laser power. The goal in this scenario is to extract information on the composition and processing conditions, and to link this to the mechanical properties while making sure that the information that is associated with one experiment is not accidentally linked to another experiment. The various pieces of information can exist far away from each other, such as in different parts of the main body of text, but also in captions, tables or figures. For instance, in Belkassa et al. [134] five experiments are performed, and the compositions of the materials are given in one table, the performance is given in another table, and the text contains processing information. This is a standard setup for papers in this kind of domain [111], [114], [135]. The way human readers are able to differentiate various experiments is through sample names, the name the authors of the document have given to each of the experiments. It appears then that for this class of problems, cleanly identifying the sample names is key to relating information across various parts of a document automatically.

As an example, take the sentence: "Two low-density steel specimens were prepared for experiments by cold-rolling and annealing at 800 °C (T800) and 900 °C (T900) for 2 min in an infrared heating furnace." [136]. Later in this document, the authors also write: "The yield strength of T800 is 718 MPa, which is much higher than the 561 MPa obtained for T900." Here the sample names are T800 and T900, and knowing these sample names would allow us to link 800 °C to a yield strength of 718 MPa by linking each to T800.

The first task is therefore to predict a list of sample names from a document. However, even for human, it is difficult to identify sample names when given a single sentence. Given the full document, humans are able to identify the samples with relative ease, indicating that information from other sentences aid in this classification problem. In this work, I find that developing a model that predicts sample names from individual sentences results in too many false positives being included. Just like for humans, more contextualized information is therefore needed. The issue is that parsing long-range dependencies remains challenging for current NLP methods. In this section, I propose an approach that can be used to extract the sample names from a paper. Using a unique dataset of 14,000 labeled papers, or over one million sentences, I train a two stage model that is able to take in context from across the entire document in order to predict the sample names contained in a paper.

4.2.2 Sample Names - Data Collection and Statistics

I build the dataset having access to a database of over two million academic articles on materials science [119], and a reliable table parser [120]. In around 0.5% of cases, papers will have multiple tables that each have a column "samples" or "specimens". This is a strong indicator that the entries in that column are the sample names of the paper. This can be further tested by checking to see if these sample names are contained in the text. Executing this procedure on our database results in a list of sample names for almost 14,000 papers. Some examples are provided in Table 4-1. Table 4-2 provides some summary statistics of this dataset. Manual verification of this dataset suggests that it is very clean, but not perfect.

This automated data extraction process allows for the generation of datasets that is significantly larger than typical annotated datasets used to train text-mining models normally found the literature. This enables the use of data-hungry algorithms such as deep learning methods.

DOI	Sample names
10.1016/j.jpowsour.2013.02.040	'F-MCs-800', 'F-MCs-900', 'F-MCs-600'
10.1016/j.actbio.2011.02.040	'Ti64-CB', 'Ti64-FB', 'Ti64-CB-7', 'Ti64-FB-7', 'Ti64-P'
10.1016/j.matchemphys.2011.03.034	'Sample 3', 'Sample 4', 'Sample 2', 'Sample 1'
10.1016/j.tsf.2007.08.086	'0.6 mTorr', '0.2 mTorr', '0.4 mTorr'
10.1016/j.jpowsour.2011.07.015	'MFC-5', 'MFC-3', 'MFC-1', 'MFC-0'
10.1016/j.jmbbm.2015.06.008	'PLG4B', 'PLGC', 'PLGA', 'PLGH', 'PLGIB'

Table 4-1: Examples of sample names as extracted using the automated system.

number of papers	13,827
number of sample names	59,892
number of unique sample names	43,152
avg sample names per paper (std)	4.3 (2.8)
% of sample names 1 token in length	88.0
% of sample names 2 tokens in length	8.6

Table 4-2: Summary statistics of dataset.

4.2.3 Base Model - tf-idf

Due to the uniqueness of sample names, a sample name appears infrequently in a corpus but frequently within a paper. This is perhaps a structure that can be exploited. Term frequency - inverse document frequency (tf-idf) score [137] was designed to give high values to just these kinds of situations.

The tf-idf score was initially designed to be a measure of how relevant a word is to describe a document, which is a useful metric for information retrieval. The tf-idf is a combination of two metrics. The first, tf or term frequency, is a ratio of the number of times a word appears in a document divided by the length of the document. Or formally, the tf score of term t in document d is:

$$\mathrm{tf}_{t,d} = \frac{f_{t,d}}{|d|} \tag{4-1}$$

where $f_{t,d}$ is the count of term t in document d and $|d|$ is the number of words in document d.

The second metric is the idf or inverse document frequency. This is a measure of how rare a term is across all the documents in a corpus. The inverse document frequency of term t is typically defined as:

$$\mathrm{idf}_t = \log \frac{N}{\mathrm{df}_t} \tag{4-2}$$

where N is the total number of documents in the corpus, and df_t is the number of documents that contain term t. Note that idf_t is high when a word shows up in only a few papers.

The tf-idf score of term t in document d is then the product of $t_{f,d}$ and idf_t :

$$\text{tf-idf}_{t,d} = \mathrm{tf}_{f,d} \cdot \mathrm{idf}_t \tag{4-3}$$

Because sample names are rare across documents, the idf is high. In addition, within a given

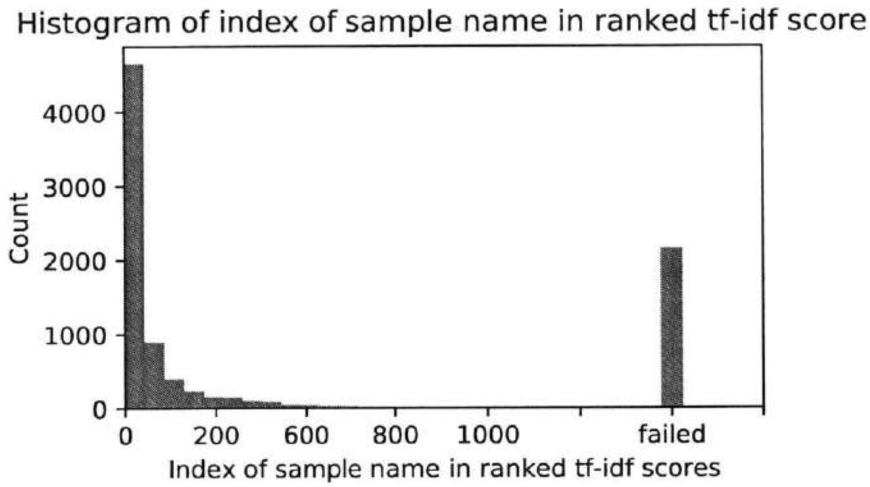

Figure 4-5: Position of sample names when ordered by tf-idf scores

document the tf is also high. We would therefore expect that for sample names, the tf-idf score should be relatively large.

I find that in many situations this is indeed the case. Many sample names show up in the top 50 highest tf-idf scores. However, there are also many sample names that only show up *after* the top 50 highest scoring tokens. Furthermore, many sample names do not show up at all because they are 2-grams or 3-grams and in this tf-idf setup we only consider 1-grams. A histogram is provided in Figure 4-5.

Expanding to 2-grams does help to capture a large percentage of the sample names, but the vocabulary grows immensely and the number of false positives increases substantially.

Regardless of some of these drawbacks, due to the simplicity and explainability of the algorithm as well as the structure of the problem, we argue that the tf-idf score is a reasonable baseline and will be used as such in our analysis.

4.2.4 Sample Name Classifier Model

Reasoning over every token in a document is computationally intractable if we need to draw in information from other parts of a document for each token. The strategy is therefore to come up with a list of candidate sample names, and to filter this list down by drawing in further information on each candidate of the list. The model is set up in two stages as shown in Figure 4-6.

Model 1
Model 1 takes every sentence in a document and makes a binary prediction for each token as to whether

it is a sample name or not. FastText [126] is used as the word embedding, which has been trained on our database of materials science papers [120]. The embeddings are used as inputs to a bi-directional LSTM, which is fed into a final time-distributed dense NN to make a prediction on each token. The predictions are then used to create a list of candidates. Adjacent tokens that are each positively predicted are assumed to be part of the same sample name and are combined for the rest of the pipeline. Model 1 is tuned to have high recall at the cost of lower precision.

Precision is defined as the number of relevant cases predicted (true positives) divided by the total predicted:

$$\text{precision} = \frac{\text{true positive}}{\text{true positive} + \text{false positive}} \tag{4-4}$$

Recall is defined as the number of total relevant cases predicted divided by the number of relevant cases that should be predicted:

$$\text{recall} = \frac{\text{true positive}}{\text{true positive} + \text{false negative}} \tag{4-5}$$

Good classifiers should have high recall and high precision, and in some cases the harmonic mean of mean and precision is used to represent a holistic score of classification power. This is also referred to as the F1 score:

$$\text{F1} = 2\frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{4-6}$$

With high recall and low precision, Model 1 aims to have anything that might be a sample name included in the candidate list, but as a result it also captures a large number of false positives.

Model 2:

The list of candidates predicted by Model 1 is filtered down by Model 2, which is able to bring in contextualized information from across the entire paper in order to determine whether a candidate is a sample name or not. For each candidate, ten sentences containing that candidate are randomly selected, embedded using FastText, and processed using an LSTM. The outputs of each of these LSTMs are added together to provide a order-invariant cumulative signal. In addition, we find that sample names within a paper tend to have a similar structure and that the Levenshtein [138], or edit-distance, between a candidate and all the other candidates can be an indicator of whether a candidate is a sample name or not, and we include this in our model. Sample names tend to have low edit-distances to all other sample names, but high edit-distances to everything else. Finally, the candidate itself is embedded. The signal from the LSTMs, the Levenshtein distance, and candidate embedding are concatenated and fed into a dense neural network before passing through a sigmoid to make a final binary prediction.

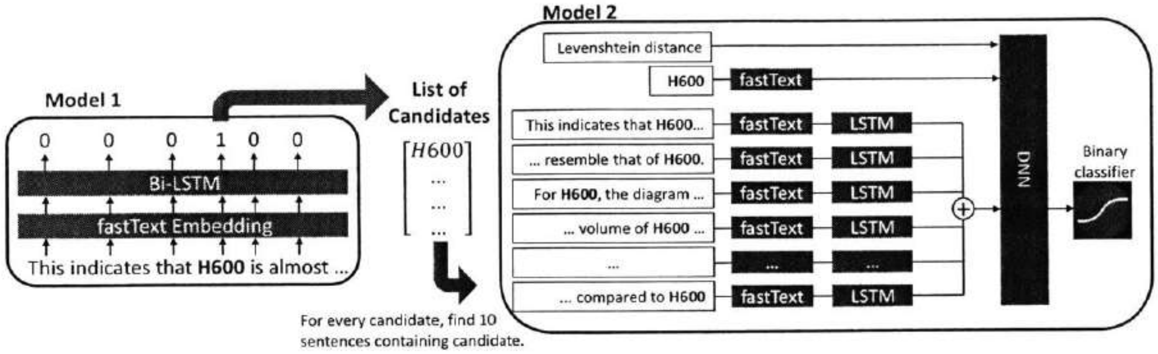

Figure 4-6: Schematic of architectures. The model is split into two; Model 1 iterates over every sentence in a document, and generates a list of candidate sample names. Model 2 filters the candidates and takes as input up to 10 randomly selected sentences that contain the candidate.

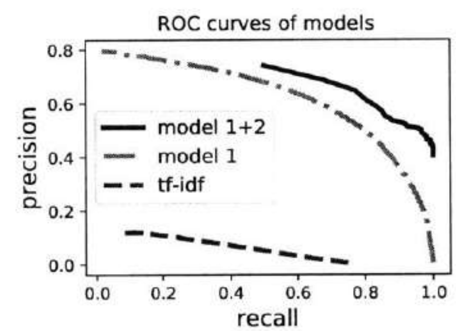

Figure 4-7: Precision-Recall plot of three different models.

4.2.5 Results: Sample Names Classifier

The best F1-score that can be achieved when only using the tf-idf score is only 0.144. This is increased to 0.646 when using Model 1. Combining Model 1 and 2 results in a top F1-score of 0.704, which is the best performing model. The corresponding recall and precision scores are given in Table 4-3, and the precision-recall curves are drawn in Figure 4-7.

4.2.6 Potential for Future Work

While this model is able to draw in information from various parts of a document, there is structure in this problem that remains unexploited. For instance, as can be observed in Table 4-1, sample names tend to be similar within a paper. While I tried to leverage this information by including the Levenshtein distance to all candidates, there may be better ways to exploit this information.

Model	Best F1	Recall	Precision
tf-idf	0.144	0.239	0.103
Model 1	0.646	0.689	0.609
Model 1+ 2	**0.704**	**0.772**	**0.648**

Table 4-3: Table of results of sample name classifiers.

As an example, we can take the list of candidates and for each candidate estimate the Levenshtein distance to all other candidates. The sample names will tend to cluster together as is shown in Figure 4-8 and Figure 4-9.

Furthermore, not just are the literal strings of sample names similar, but the usage of sample names are also very similar within a paper. For instance, taking the example used earlier [136], the text discusses the strains, phases, strengths and compositions of *both* "T800" and "T900". So we might believe that "T800" could be a sample name when looking at it in isolation. However, we could convince ourselves further by realizing that "T900" which is used in a very similar context as "T800" within this paper, also looks like a sample name and that perhaps both are sample names.

All this kind of structure suggests that instead of predicting one sample name at a time we might want to do something like a clustering analysis. Exploring ways to utilize this structure is a clear candidate for future work.

4.2.7 Using Sample Names to Link Data

Given the sample names, linking information across various parts of a document becomes feasible. The idea is to link every piece of information to a sample name, and if information cannot be linked to a sample name then assume the information is applied to all samples in that paper. This is distinguished as sample level data (which differs for different samples) and global level data (which is the same for all samples in a paper). For instance in [113] there are two samples (Ti_{64} and PH1). Some settings are different for the two samples, like the composition and scanning speeds, but both have the same maximum laser power and apparatus. In this case, the maximum laser power and apparatus are global level data, while composition and scanning speeds are sample level data.

Extracting information from tables and linking it to sample names is typically fairly straightforward, as in the majority of cases there is a column listing the sample name. In some cases, there is only a single row in a table (e.g [110], [141]), in which case the information in the table becomes global level data.

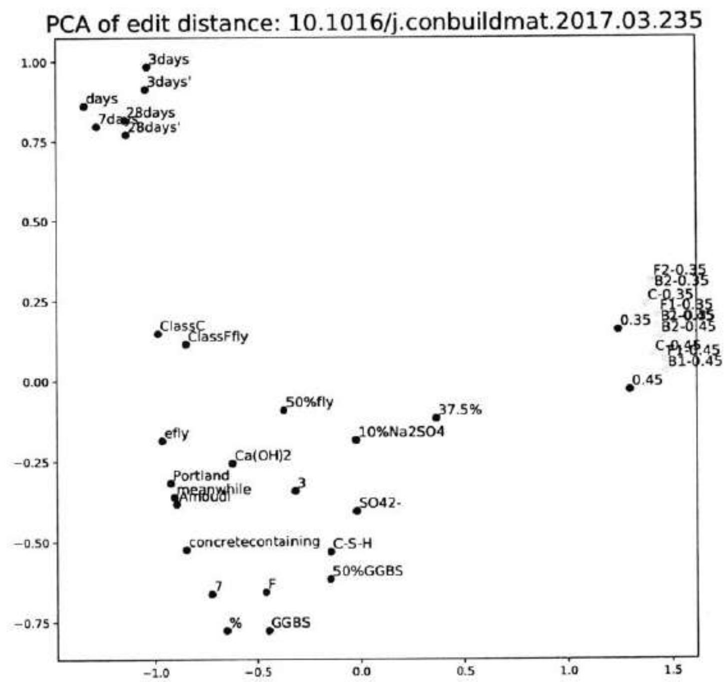

Figure 4-8: Edit distance of candidates projected onto two dimensions from Zhang et al. [139]. In green are the sample names, and in black are non-sample names. Note that the sample names form a fairly clean cluster on the right hand side.

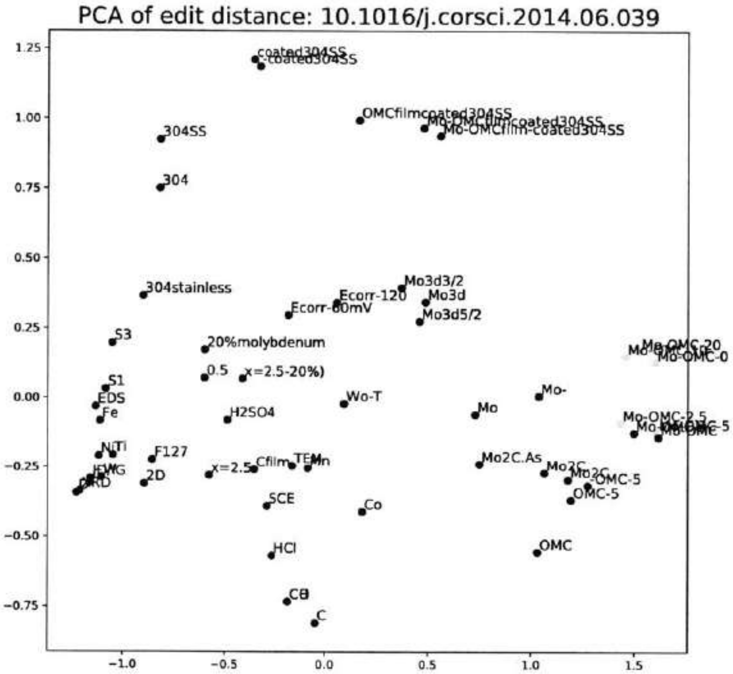

Figure 4-9: Edit distance of candidates projected onto two dimensions from Liu et al. [140]. In green are the sample names, and in black are non-sample names. The sample names seem to fall in two fairly clean clusters.

In some cases, extracting data from text is similarly straightforward. For instance in Rafi et al., [113] two separate sentences describe the heat treatments for the two samples : "Ti_{64} specimens were heat treated at 650 °C for 4 hours in an argon filled furnace for stress relief and cooled down in the furnace to room temperature. PH1 steel samples were heat treated to 482 °C for precipitation hardening prior to tensile testing." In this case, the information could be extracted using conventional means, and then attributed to the sample found in the sentence. However, frequently there are multiple sample names in one sentence, which requires a different approach. For example, in the same paper [113] we find: "For Ti64, the average particle size was 38 μm and for PH1 powders the particle size was 43 μm."

To extract data from this sentence we might need to use something more sophisticated than a per-sentence rule. A part of the dependency parse of this sentence is given in Figure 4-10. Note that we can connect Ti_{64} to a particle size of 38 μm through the word "was". The same thing can be done for the PH1 sample name. This potentially suggests that information can be linked using the grammar. However, this grammar dependency approach falls short in some cases, as can be seen in this quote

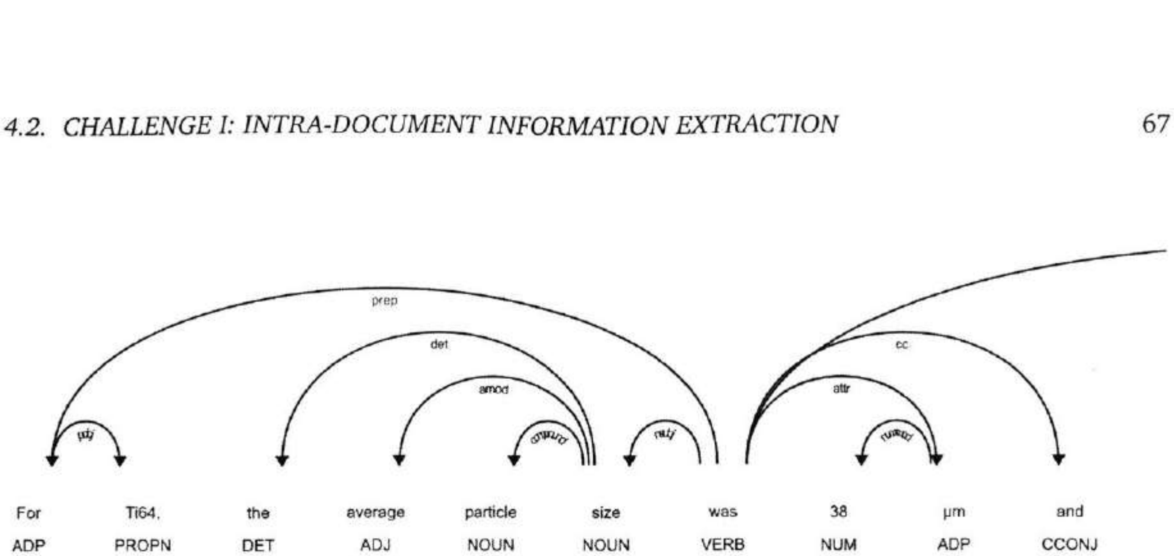

Figure 4-10: Part of the grammar dependency parse of a sentence from [113]

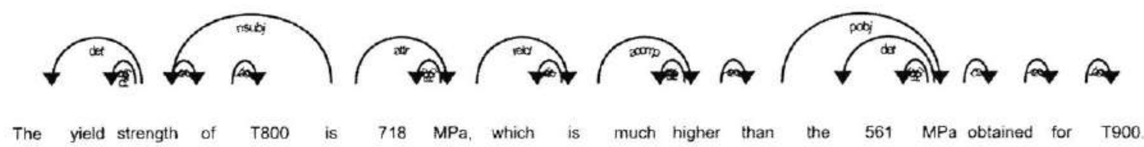

Figure 4-11: Dependency parse of a sentence from cite(lee)

from [136]: "The yield strength of T800 is 718 MPa, which is much higher than the 561 MPa obtained for T900." The corresponding dependency parse is given in Figure 4-11. In this case, it might be possible to link that the yield strength of T800 is 718 MPa, but while the dependency parse can be used to link 561 MPa to T900, it is hard to know that this refers to yield strength.

4.2.8 Conclusions on Intra-Document Extraction

Extracting usable data from academic papers requires the linking of information across multiple sentences and tables in a document. Current text-mining techniques focus on extracting data from a sentence at a time and are therefore not sufficient for the purpose of extracting manufacturing composition, processing conditions, and properties.

In this section, I proposed that sample names are the key for humans to link information found in a paper and argue that this is also the key for automated extraction of these kinds of papers. Sample names are hard to identify from single sentences, and more contextualized information is required to build a model with high classification capacity. The two stage model discussed in this section is able to outperform sentence-level approaches. However, the performance has room for improvement and future approaches might be able to better leverage further structure in the problem.

Given a full list of sample names it becomes feasible to link information across papers, but in some cases, it is still hard to link information found in text. Tables and simple sentence structures can be parsed accurately, but the more complicated sentences often found in academic papers still prove challenging, even with a perfect set of sample names for that paper.

4.3 Challenge II: Consistency

4.3.1 Lack of Documentation Standards

Besides the issue of whether it is possible to extract the relevant information from documents, there is a question of whether all the right information is contained in the document in the first place. It turns out that for many AM papers, the set of processing parameters described is incomplete. On some level, this is reasonable, as not all process parameters are necessarily relevant for every experiment. However, it makes replication of the experiment difficult, and also makes data analysis impossible without imputation of some of these values.

4.3.2 Documentation of Process Parameters

When first investigating the information contained in AM papers, we developed a schematic for data collection and manually collected information from 90 AM papers. This exercise helped to shape our thinking around the information contained in these documents, and how the information could be extracted. The template is available in Appendix C. Figure 4-12 shows the distribution of AM processes and materials used in these papers. Note that selective laser melting (SLM) and selective laser sintering (SLS) dominate the processes, and Ti6Al4V, a popular titanium alloy, dominates the dataset.

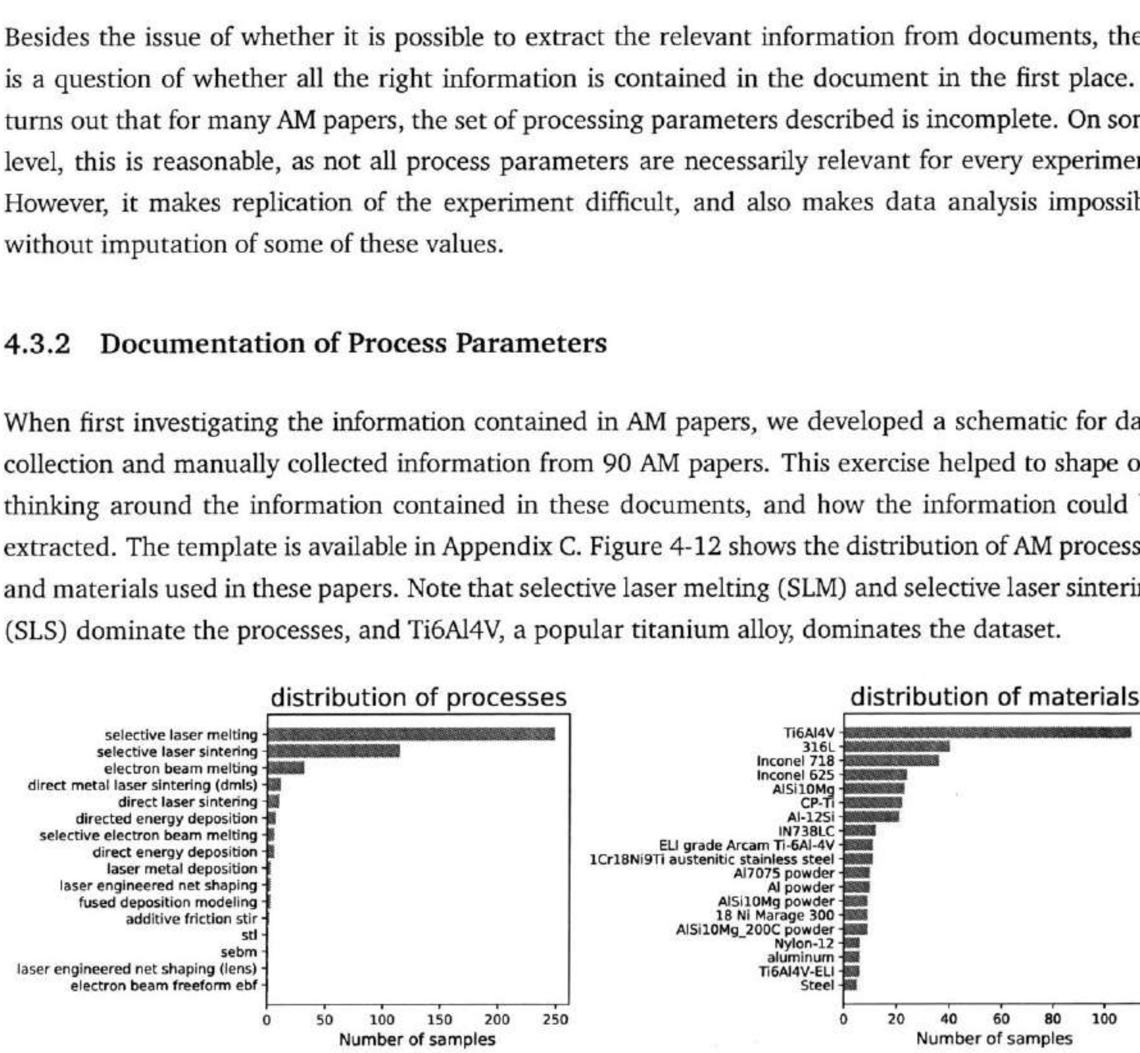

Figure 4-12: Distribution of the processes and materials used in the 90 papers that were manually extracted.

One observation that immediately stood out when collecting data is that researchers do not frequently publish the same set of information even for the same process. For instance, one paper on selective laser sintering (SLS) might include details on the laser power, scanning speed, and hatch

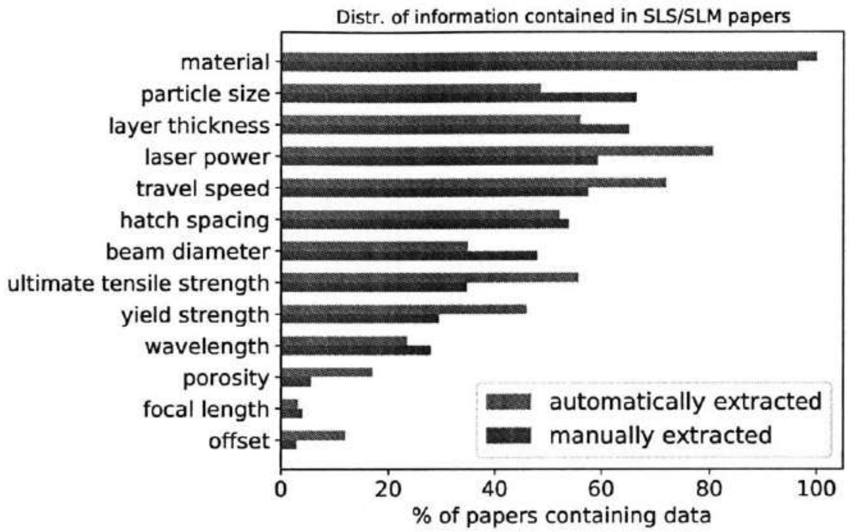

Figure 4-13: Bar plot showing what % of papers on selective laser sintering or selective laser melting include information on various parameters and properties.

spacing, while another using the same process talks about beam diameter and particle size. In reality, all of these parameters exist in both cases. Both researchers had to make a decision on what laser power to use, but not all researchers publish this information.

Figure 4-13 illustrates how significant this problem is. Here in blue is the percent of manually extracted papers on selective laser sintering or selective laser melting that contains the various pieces of information shown on the y-axis. In red is the same but on the full set of papers, automatically extracted using a dictionary approach (see Appendix B). The manually extracted data is very accurate, but the automated data is based on a larger number of papers. Both approaches tell the same story; papers do not tend to have all the import parameters in a process. For instance, I found that particle size, which is amongst the most discussed properties, is only mentioned in around 50%-70% of papers.

As a result of this inconsistency, there are only a few datapoints that have all the parameters that we might care about. A model that predicts the yield strength based on the material, particle size, layer thickness, and laser power only has a handful of datapoints with all those features. In order to conduct any statistical analysis, it is therefore necessary to do data imputation. There are various approaches to doing this. However, all approaches come at a cost of additional uncertainty and noise, especially considering that over 50% of the values would need to be imputed to make the above dataset complete.

4.3.3 Standardizing Documentation

As discussed in Chapter 3, a big part of the ACT program was standardizing tests and procedures so that data could more easily be shared between various parties. This included standardizing the way data were entered into a database [80], [81]. For academic papers, the primary objective is not to share data in order to better understand the manufacturing distributions, but rather to communicate findings of interesting theories and phenomena. From this perspective, it may be infeasible to expect researchers to standardize their data and enter it into databases. However, at the very least researchers should make this information available somewhere in their papers. As it currently stands, most of the research presented by these AM papers are not reproducible, because the set of processing parameters necessary to repeat the experiment are not published. Perhaps it is time to standardize the processing parameters published by AM researchers.

4.4 Compositions and Properties

4.4.1 Results Overview

Due to the limitations described in this chapter, it is challenging to extract information with the kind of granularity and accuracy that is required to fully characterize the manufacturing variability of AM. The greatest challenge is linking information across a document, and while some important steps have been taken in this work, there are remaining obstacles for this task. Regardless of these challenges, it is possible to extract some information, like material compositions and mechanical properties.

4.4.2 Tested Compositions

A powerful application of text-mining is mapping the landscape of an entire research field. For instance, using the pipeline it is possible to consider all the materials that have been studied, and determine if there are sets of materials that are under-researched. In order to create this landscape, I collect alloy compositions from all the conventional and AM papers in the database, and project these compositions onto a 2D grid using Principal Component Analysis (PCA). Each conventionally manufactured material is represented by a circle, and each AM material is represented by a cross. For the conventionally manufactured materials, I color each data point depending on what element has the highest proportion. The results are as shown in Figure 4-14.

Due to the nature of PCA, the axes do not necessarily have an intuitive interpretation. However, the idea is that materials that are similar also lie close together in Figure 4-14. Looking at the plot, four groups can be identified, where each corresponds to a different base element (iron, titanium, aluminum, and nickel). The apexes of the triangles tend to be pure elements. For instance, on the very left-hand-side of Figure 4-14 we can materials that are almost entirely pure iron. Moving towards the right brings us closer to high-alloyed steels.

It appears from this plot that AM researchers tend to focus on a select set of materials. There are clusters of AM papers around Ti6Al4V, Inconel 718, AlSi10Mg, and some steel compositions, but few papers exploring other materials. This plot allows the identification of materials that have not been published on for AM. Note, however, that this does not identify *why* these materials have not been investigated. For instance, perhaps researchers did attempt to manufacture these other materials but did not get publishable results.

4.4.3 Variability

Linking mechanical properties and compositions can often be done with high reliability. This is because both these data are given in tables, which can be linked using the predicted sample names, or because there is only a single material in the paper, thereby making data linking trivial. Furthermore, with a small dataset it is feasible to manually verify the data.

One question we can then ask is: how much do mechanical properties vary for a given material? To answer this question I selected some standard materials (316L SS, AlSi10Mg, Inconel 718, pure Fe, and Ti6Al4v), and found all materials that are similar to these materials. This is done by first taking the matrix of compositions and normalizing the values so that each element has a mean of zero and a range of one. I then draw a small hyper-sphere in this linear space that is centered around these materials. Similar materials are those that fall within the bounds of this sphere. The 0.2% yield strength data were then extracted from these papers and were manually checked and corrected where necessary.

The results are then split depending on whether the materials were manufactured with AM techniques or with conventional techniques, and are plotted as box plots as shown in Figure 4-15. Note that for both conventional and AM, the variability in properties is high, with ranges of a few hundred MPa within a material being fairly common. There are situations where the empirical variability of AM is larger than that of conventional manufacturing processes, but this is not always true.

The insight derived from this plot is that the range in mechanical properties can be quite large for a given material, which suggests that the processing conditions are extremely important in determining the behavior of a material.

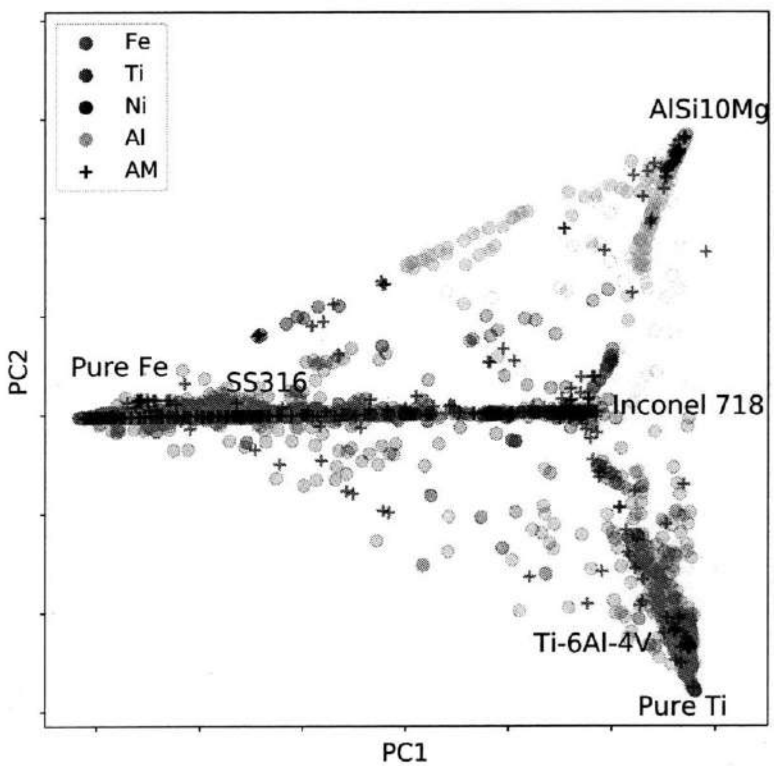

Figure 4-14: Composition of conventionally manufactured materials (circles) and AM materials (crosses), projected down onto two dimensions using Principal Component Analysis. Each point on the plot represents a material. The colors correspond to the most dominant element in that material. As can be seen, AM is not yet as diverse as conventional processes, and tends to clump together, especially in Aluminum and Titanium based material systems.

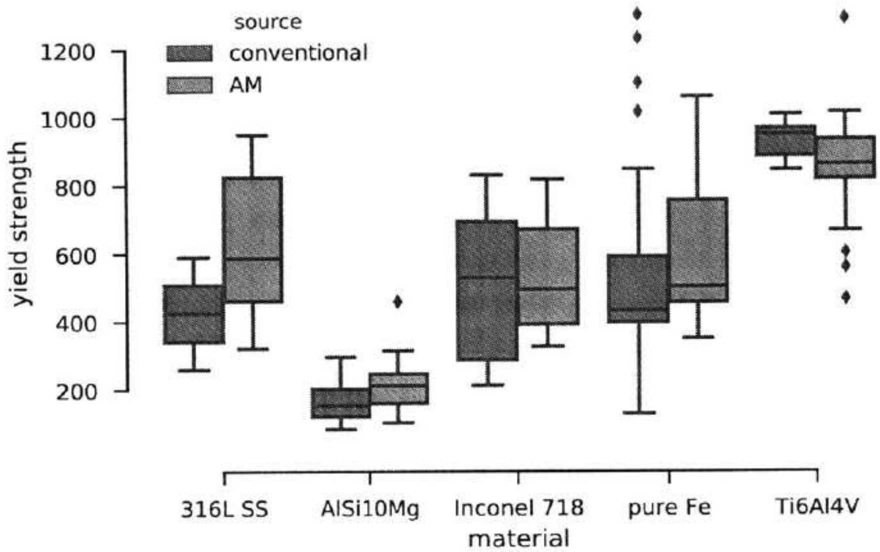

Figure 4-15: Boxplots of yield strengths for various materials manufactured using conventional processes or AM processes.

4.5 Conclusions

Significant valuable information exists as unstructured data in the form of published documents. The idea of extracting information from these documents is a powerful one, but current text-mining techniques are not yet in the position to reliably extract and link data found across a document. Without this linking, it becomes impossible to establish relationships between material composition, processing parameters, and mechanical properties, which is necessary in order to use these data to better understand manufacturing processes.

In this chapter, I demonstrated that in many manufacturing and materials science papers, multiple experiments are described and that it is not obvious how to keep these multiple experiments separated when doing information extraction. I argue for a strategy to identify the "sample names", which are the terms the researchers use to differentiate between various experiments in a single paper. Getting the sample names from a paper is nontrivial, as contextualized information is needed to accurately determine whether a word or sequence of words is a sample name or not. I develop a dataset of around 14,000 papers and create a two-staged model that is able to bring in contextualized information from the entire paper. This is an important step, but more work is required to more reliably extract the sample names from papers and to use these sample names to extract information.

This chapter also discusses the problem of missing data. Beyond the implications on data analysis, the lack of consistent information publishing makes the results of many of these papers unreproducible. For AM, which is a relatively well-established technology, it might make sense for researchers to agree to a standard list of processing parameters they will use when publishing their results.

Regardless of these limitations, some useful data can be extracted from papers which can be used to obtain a big-picture overview of what materials have or have not been published about in AM, or to compare the mechanical properties of AM versus conventional manufacturing for specific materials.

Chapter 5

Conclusions

Manufacturing variability is a central concept in manufacturing, but beyond select fields of study, manufacturing variability is often neglected by researchers. In this thesis, I have explored this topic by asking two questions: (1) what are the effects of manufacturing variability, and why is it important? And (2) can we gather data from published documents in order to characterize manufacturing variability, and what are some challenges associated with that?

In Chapter 2, I began to answer the first question by introducing the idea of manufacturing variability and arguing that there will always be some differences between manufactured parts. In order to deal with this uncertainty, most products have some degree of overdesign built in. As a result of this overdesign, products are rated to be less strong than they might be on average, which means that more material is required, driving up costs. Furthermore, this overdesign implies that parts tend to be heavier than if there were no overdesign. The additional materials lead to higher embodied energy, and also higher fuel consumption if the part is used in a vehicle. In this way, manufacturing variability can be linked to environmental and economic costs. Some examples illustrated this relation. For instance, roughly 19% of concrete was found to be produced solely to counter the effects of manufacturing variability. Or in the case of fiber composites, variability was found to play the largest role in production cost and energy after the design specifications. For a Boeing 787, under conservative assumptions, reducing the variability of fiber composite parts could save millions of dollars in fuel and ktons of CO_2 over the lifetime of a single aircraft.

In Chapter 3, I continued with the question of the effects of manufacturing variability by pointing out that manufacturers can be reluctant to adopt a new material or process if the distribution of properties are not fully understood. Underestimating the variance in a process might lead to too many out-of-spec parts, which could damage the reputation of a company, or worse, lead to catastrophic failure. Furthermore, certifying such a process becomes impossible, and as a result, the products will not legally be able to enter the market. It is therefore critical to be able to estimate the variance in a process, but as demonstrated, accurately estimating the variability of a distribution requires dozens if not hundreds of samples. Nevertheless, new processes are continuously adopted by industry. Fiber composites is an important example, where numerous issues around process control and manufacturing variability were overcome by an industry, academic, and government effort through the ACT and AGATE programs. Both programs continue to influence industry through the regulatory framework that was developed as a result of the work that was done under these programs. Additive Manufacturing (AM) finds itself in a similar situation as fiber composites in the late 1980s, and perhaps it is time to use the knowledge we have gained from the ACT and AGATE experience to develop a similar strategy for this technology.

In Chapter 4, I continued to focus on AM, but through the lens of the second thesis question: can we use published documents to characterize manufacturing variability? Thousands of experiments have been conducted on AM, but the results are mostly only available in published documents in unstructured formats like texts. Developments in text-mining make it feasible to extract some information at scale from these documents, but current techniques are insufficient to accurately extract all the important information. Specifically, most text-mining techniques focus on within-sentence extraction, while for this task the information needs to be linked across the entire document. This issue is reframed by looking at sample names in a paper, and a model is put forward to extract the sample names of a document. These sample names are then used to link information across a document. A clear area of future work is improving the sample names classifier further, and then using the sample names to extract more granular data from documents. The second issue in text-mining is that researchers do not standardize the process parameters they include in their papers. This makes it hard to build a complete dataset of alloy composition, processing parameters, and mechanical properties. It also raises questions around reproducibility. It seems that AM is mature enough as a technology to warrant a standardized way of describing this process, which would greatly aid the sharing of results. Ideally such a system would bypass any need for text-mining.

Ultimately this thesis argues that anyone developing new materials or manufacturing processes, or modeling the performance of materials, should view the properties of their products as distributions rather than point estimates. Processes are not deterministic. Without acknowledging this fact we will continue to over-promise and under-deliver the performance and timelines of emerging manufacturing techniques.

Bibliography

[1] World Bank, *Manufacturing, value added (% of GDP)*, 2016. [Online]. Available: https://data.worldbank.org/indicator/nv.ind.manf.zs.

[2] ——, *CO2 emissions from manufacturing industries and construction (% of total fuel combustion)*, 2014. [Online]. Available: https://data.worldbank.org/indicator/EN.CO2.MANF.ZS.

[3] D. E. Hardt, "Modeling and control of manufacturing processes: getting more involved", *ASME Journal of Dynamic Systems, Measurement, and Control*, vol. 115, no. 2B, pp. 291–300, 1993.

[4] E. V. Gijo and J. Scaria, "Process improvement through Six Sigma with Beta correction: A case study of manufacturing company", *International Journal of Advanced Manufacturing Technology*, vol. 71, no. 1-4, pp. 717–730, 2014, ISSN: 02683768. DOI: 10.1007/s00170-013-5483-y.

[5] V. Gupta, R. Jain, M. L. Meena, and G. S. Dangayach, "Six-sigma application in tire-manufacturing company: a case study", *Journal of Industrial Engineering International*, vol. 14, no. 3, pp. 511–520, 2018, ISSN: 2251712X. DOI: 10.1007/s40092-017-0234-6. [Online]. Available: https://link.springer.com/content/pdf/10.1007%5C%2Fs40092-017-0234-6.pdf.

[6] A. G. Prasad, S. Saravanan, E. V. Gijo, R. Dasari, R. Tatachar, and P. Suratkar, "Six Sigma-based approach to optimise the diffusion process of crystalline silicon solar cell manufacturing", *International Journal of Sustainable Energy*, vol. 35, no. 2, pp. 190–204, 2013. DOI: 10.1080/14786451.2013.861463.

[7] A. van Grootel, J. Chang, and E. Olivetti, "Economic and environmental cost of variability in manufacturing: the case of carbon fiber reinforced polymer composite in the aerospace industry", -, In progress.

[8] A. van Grootel, J. Chang, and E. Olivetti, "The Role of Manufacturing Variability on Environmental Impact", in *REWAS 2019*, G. Gaustad, C. Fleuriault, M. Gökelma, J. A. Howarter, R. Kirchain, K. Ma, C. Meskers, N. R. Neelameggham, E. Olivetti, A. C. Powell, F. Tesfaye, D. Verhulst, and M. Zhang, Eds., Cham: Springer International Publishing, 2019, pp. 19–32, ISBN: 978-3-030-10386-6.

[9] W. Shewhart, "Quality Control Charts", *Bells Labs Technical Journal*, vol. 5, no. 4, pp. 593–603, 1926.

[10] C. J. Spanos, "Statistical Process Control in Semiconductor Manufacturing", *Proceedings of the IEEE*, vol. 80, no. 6, pp. 819–830, 1992.

[11] D. Montgomery, "Quality Improvement in the Modern Business Environment", in *Introduction to Statistical Process Control*, 6th ed., 2009, ch. 1, pp. 3–42.

[12] G. May and C. J. Spanos, "Taguchi Method", in *Fundamentals of Semiconductor Manufacturing and Process Control*, 2006, ch. 7.4, pp. 262–268.

[13] K. Tang, "Economic design of product specifications for a complete inspection plan", *International Journal of Production Research*, vol. 26, no. 2, pp. 203–217, 1988, ISSN: 0020-7543. DOI:

10.1080/00207548808947854. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/00207548808947854.

[14] J. Liu, R. Dejus, A. Donnelly, C. Doose, A. Jain, and M. Jaski, "Field Quality From Tolerance Analyses In Eight-Piece Quadrupole Magnet", *Transactions on Applied Superconductivity*, vol. 28, no. 3, 2018.

[15] A. I. Kingon, J. P. Maria, and S. K. Streiffer, "Alternative dielectrics to silicon dioxide for memory and logic devices", *Nature*, vol. 406, no. 6799, pp. 1032–1038, 2000, ISSN: 00280836. DOI: 10.1038/35023243. [Online]. Available: https://www.nature.com/articles/35023243.pdf.

[16] J. A. Slotwinski, E. J. Garboczi, and K. M. Hebenstreit, "Porosity Measurements and Analysis for Metal Additive Manufacturing Process Control", *Journal of Research of the National Institute of Standards and Technology*, vol. 119, pp. 494–528, 2014.

[17] American Concrete Institute, *Building Code Requirements for Structural Concrete (ACI 318-14)*. 2014, vol. 11, p. 6858, ISBN: 9780870319303.

[18] ASTM International, *ASTM F3114-15*, 2015. DOI: 10.1520/F3114-15.2. [Online]. Available: https://compass.astm.org/download/F3114.8225.pdf.

[19] Department of Defense, *Composite materials handbook*, 2002. DOI: 10.1016/0378-3804(85)90127-5. [Online]. Available: https://www.library.ucdavis.edu/wp-content/uploads/2017/03/HDBK17-3F.pdf.

[20] US Department of Defense, "Polymer Matrix Composites", United States Department of Defense, Tech. Rep. January, 1997.

[21] F. Del Pero, M. Delogu, M. Pierini, and D. Bonaffini, "Life Cycle Assessment of a heavy metro train", *Journal of Cleaner Production*, vol. 87, no. 1, pp. 787–799, 2015, ISSN: 09596526. DOI: 10.1016/j.jclepro.2014.09.023. [Online]. Available: https://ac.els-cdn.com/S0959652614010658/1-s2.0-S0959652614010658-main.pdf?_tid=ed349182-2981-450c-aa59-62fa266307ca&acdnat=1539019747_97eb268401635f005aaf490a34e73f5e.

[22] I. Sartori and A. G. Hestnes, "Energy use in the life cycle of conventional and low-energy buildings: A review article", *Energy and Buildings*, vol. 39, no. 3, pp. 249–257, 2007, ISSN: 03787788. DOI: 10.1016/j.enbuild.2006.07.001. [Online]. Available: https://ac.els-cdn.com/S0378778806001873/1-s2.0-S0378778806001873-main.pdf?_tid=5ddd65b5-4b3d-4b42-b381-d56b29433faf&acdnat=1534519735_354ec4ff24b99b99c327b78c51674c1b.

[23] R. M. Andrew, "Global CO2 emissions from cement production", *Earth System Science Data*, pp. 1–52, 2018, ISSN: 1866-3508. DOI: http://dx.doi.org.kuleuven.ezproxy.kuleuven.be/10.5194/essd-10-195-2018. [Online]. Available: https://www.earth-syst-sci-data.net/10/195/2018/essd-10-195-2018.pdf.

[24] B. K. Obla, "Variation in Concrete Strength Due to Cement: Part III of Concrete Quality Series", *Concrete in focus*, no. November/December, pp. 8–12, 2010.

[25] J. E. Cook, J. Parnes, D. J. Akers, W. L. Barringer, J. L. Brown, and A. Graf, "Evaluation of Strength Test Results of Concrete", *Test*, pp. 1–20, 2011. [Online]. Available: http://www.aice.cl/es/archivos/wp-aice-old/uploads/2012/01/214r_02.pdf.

[26] S. Ozdemir, D. Sinha, G. Memik, J. Adams, and H. Zhou, "Yield-aware cache architectures", *Proceedings of the Annual International Symposium on Microarchitecture, MICRO*, pp. 15–25, 2006, ISSN: 10724451. DOI: 10.1109/MICRO.2006.52.

[27] M. Slater, "Intel Boosts Pentium Pro to 200 MHz", *Microprocessor Report*, vol. 9, no. 17, 1995. [Online]. Available: https://www.cl.cam.ac.uk/~pb22/test.pdf.

[28] D. Boggs, A. Baktha, J. Hawkins, D. Marr, J. A. Miller, P. Roussel, R. Signhal, B. Toll, and K. Venkatraman, "The Microarchitecture of the Intel® Pentium® 4 Processor on 90nm Technology", *Intel Technology Journal*, vol. 08, no. 1-18, pp. 119–130, 2004, ISSN: 1535864X. DOI: 10.1535/itj.0902. [Online]. Available: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Learning-Based+Computer+Vision+with+Intel?s+Open+Source+Computer+Vision+Library#0.

[29] K. Kuhn, C. Kenyon, A. Kornfeld, M. Liu, A. Maheshwari, W.-k. Shih, S. Sivakumar, G. Taylor, P. van der Doorn, and K. Zawadzki, "Managing Process Variation in Intel's 45nm CMOS Technology", *Intel Journal of Technology*, vol. 12, no. 45, pp. 77–85, 2008. DOI: 10.1535/itj.1201. [Online]. Available: http://developer.intel.com/technology/itj/index.htm.

[30] R. A. Witik, J. Payet, V. Michaud, C. Ludwig, and J. A. E. Månson, "Assessing the life cycle costs and environmental performance of lightweight materials in automobile applications", *Composites Part A: Applied Science and Manufacturing*, vol. 42, no. 11, pp. 1694–1709, 2011, ISSN: 1359835X. DOI: 10.1016/j.compositesa.2011.07.024. [Online]. Available: http://dx.doi.org/10.1016/j.compositesa.2011.07.024%20https://ac.els-cdn.com/S1359835X11002302/1-s2.0-S1359835X11002302-main.pdf?_tid=a77097b6-58eb-4c03-9d15-e732dc4ef852&acdnat=1539702218_d2d3e72d570445e81da99a0bedb3a51d.

[31] I. Yucel, "Reinforced strength: The industry will thrive as downstream manufacturing demand rises", *IBISWorld Industry Report*, Dec. 2016.

[32] J. Hale, *Boeing 787 from the Ground Up*, 2008. [Online]. Available: http://www.boeing.com/commercial/aeromagazine/articles/qtr_4_06/AERO_Q406_article4.pdf.

[33] R. Heuss, N. Muller, W. van Sintern, A. Starke, and A. Tschiesner, *Lightweight, heavy impact*, 2012.

[34] S. Rao, T. G. A. Simha, K. P. Rao, and G. V. V. Ravikumar, "Carbon Composites are Becoming Competitive and Cost Effective", *Infosys website*, pp. 2–3, 2018. [Online]. Available: https://www.infosys.com/engineering-services/white-papers/Documents/carbon-composites-cost-effective.pdf.

[35] G. W. Bright, J. I. Kennedy, F. Robinson, M. Evans, M. T. Whittaker, J. Sullivan, and Y. Gao, "Variability in the mechanical properties and processing conditions of a high strength low alloy steel", *Procedia Engineering*, vol. 10, pp. 106–111, 2011, ISSN: 18777058. DOI: 10.1016/j.

proeng.2011.04.020. [Online]. Available: http://dx.doi.org/10.1016/j.proeng.2011.04.020.

[36] T. S. Mesogitis, A. A. Skordos, and A. C. Long, "Uncertainty in the manufacturing of fibrous thermosetting composites: A review", *Composites Part A: Applied Science and Manufacturing*, vol. 57, pp. 67–75, 2014, ISSN: 1359835X. DOI: 10.1016/j.compositesa.2013.11.004. [Online]. Available: http://dx.doi.org/10.1016/j.compositesa.2013.11.004%20http://ac.els-cdn.com/S1359835X13003035/1-s2.0-S1359835X13003035-main.pdf?_tid=ff2e1f6c-b5b2-11e6-b844-00000aacb360&acdnat=1480369287_ad2fce0f200fd11575f160c070dc9781.

[37] K. D. Potter, "Understanding the origins of defects and variability in composites manufacture", *17th International Conference on Composite Materials*, pp. 27–31, 2009. [Online]. Available: http://iccm-central.org/Proceedings/ICCM17proceedings/Themes/Plenaries/P1.5%20Potter.pdf.

[38] M. F. Ashby, "Chapter 15 - Material profiles", *Materials and the Environment (Second Edition)*, pp. 459–595, 2013, ISSN: 13697021. DOI: http://dx.doi.org/10.1016/B978-0-12-385971-6.00015-4. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9780123859716000154.

[39] J. Rybicka, A. Tiwari, and G. A. Leeke, "Technology readiness level assessment of composites recycling technologies", *Journal of Cleaner Production*, vol. 112, pp. 1001–1012, 2016, ISSN: 09596526. DOI: 10.1016/j.jclepro.2015.08.104. arXiv: 0011032v1 [cs]. [Online]. Available: http://dx.doi.org/10.1016/j.jclepro.2015.08.104%20https://ac.els-cdn.com/S0959652615012056/1-s2.0-S0959652615012056-main.pdf?_tid=45d8bd5d-3bb4-4632-%2094de-049bbe68421f&acdnat=1541540691_2cf580cd28388b91ac78c42a5c8d4eaa.

[40] Grand View Research, "Lightweight Materials Market Size, Share & Trends Analysis Report", Tech. Rep., 2018, p. 96.

[41] E. Ozelkan, "Global Commercial Aircraft Manufacturing", *IBISWorld Industry Report*, Jul. 2018.

[42] International Energy Egency, *World Energy Outlook 2015*. Organization for Economic Co- operation and Development, Nov. 2015, ISBN: 9789264243651. DOI: 10.1787/weo-2015-en.

[43] C. Façanha, K. Blumberg, and J. Miller, "Global Transportation Energy and Climate Roadmap: The impact of transportation policies and their potential to reduce oil consumption and greenhouse gas emissions", *The ICCT Report*, p. 109, Nov. 2012. [Online]. Available: www.theicct.org.

[44] B. Sayler, "Global Airlines", *IBISWorld Industry Report*, Jun. 2017.

[45] Federal Aviation Agency, *Safety Factor*, 2010.

[46] ——, *Material strength properties and design values. AC No. 25.613-1*, 2010.

[47] C. Modlin and J. Zipay, *The 1.5 & 1.4 Ultimate Factors of Safety for Aircraft & Spacecraft - History, Definition and Applications*, 2014. [Online]. Available: https://ntrs.nasa.gov/search.jsp?R=20140011147.

[48] F. Field, R. Kirchain, and R. Roth, "Process cost modeling: Strategic engineering and economic evaluation of materials technologies", *The Journal of The Minerals, Metals & Materials Society*, vol. 59, no. 10, pp. 21–32, 2007, ISSN: 10474838. DOI: 10.1007/s11837-007-0126-0. [Online]. Available: https://link.springer.com/content/pdf/10.1007%5C%2Fs11837-007-0126-0.pdf.

[49] E. Witten, T. Kraus, and M. Kuhnel, "Composites Market Report 2016", Tech. Rep. September, 2016, pp. 1–46. [Online]. Available: http://www.eucia.eu/userfiles/files/20161128_market_report_2016_english.pdf.

[50] J. Sloan, "In-house prepregging: Cost/benefit calculus", *Composites World*, 2013. [Online]. Available: https://www.compositesworld.com/articles/(597).

[51] D. J. LeBlanc, J. Lorenzana, A. Kokawa, T. Bettner, and F. Timson, "Advanced Composite Cost Estimating Manual. Volume II", 1976. [Online]. Available: http://www.dtic.mil/dtic/tr/fulltext/u2/a041497.pdf.

[52] S. M. Haffner, "Cost Modeling and Design for Manufacturing Guidelines for Advanced Composite Fabrication", PhD thesis, Massachusetts Institute of Technology, 2002.

[53] R. A. Witik, F. Gaille, R. Teuscher, H. Ringwald, V. Michaud, and J. A. E. Månson, "Economic and environmental assessment of alternative production methods for composite aircraft components", *Journal of Cleaner Production*, vol. 29-30, pp. 91–102, 2012, ISSN: 09596526. DOI: 10.1016/j.jclepro.2012.02.028. [Online]. Available: http://dx.doi.org/10.1016/j.jclepro.2012.02.028%20https://ac.els-cdn.com/S0959652612001047/1-s2.0-S0959652612001047-main.pdf?_tid=c59bd7d2-28d0-4447-8823-1516f5ff4699&acdnat=1523631409_f450271dda80d04c8969c56d92a27705.

[54] J. H. Whitelaw, "Convective Heat Transfer", *Thermopedia*, 2011. DOI: 10.1615/AtoZ.c.convective\\heat\\transfer. [Online]. Available: http://thermopedia.com/content/660/.

[55] M. Srinivasan, P. Maettig, K. W. Glitza, B. Sanny, and A. Schumacher, "Out of Plane Thermal Conductivity of Carbon Fiber Reinforced Composite Filled with Diamond Powder", *Open Journal of Composit Materials*, vol. 6, no. 4, pp. 41–57, 2016. [Online]. Available: https://pdfs.semanticscholar.org/2303/e11ad0ae9ceab007b63cb6fbe2e7cfaff1ee.pdf.

[56] J. C. Kelly, J. L. Sullivan, A. Burnham, and A. Elgowainy, "Impacts of Vehicle Weight Reduction via Material Substitution on Life-Cycle Greenhouse Gas Emissions", *Environmental Science and Technology*, vol. 49, no. 20, pp. 12 535–12 542, 2015, ISSN: 15205851. DOI: 10.1021/acs.est.5b03192.

[57] A. C. Serrenho, J. B. Norman, and J. M. Allwood, "The impact of reducing car weight on global emissions: the future fleet in Great Britain", *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 375, no. 2095, p. 20 160 364, 2017, ISSN: 1364-503X. DOI: 10.1098/rsta.2016.0364. [Online]. Available: http://rsta.royalsocietypublishing.org/lookup/doi/10.1098/rsta.2016.0364.

[58] H. Helms and U. Lambrecht, "The Potential Contribution of Light-Weighting to Reduce Transport Energy Consumption.", *International Journal of Life Cycle Assessment*, vol. 12, no. 1, pp. 58–64, 2007. DOI: 10.1065/lca2006.07.258.

[59] C. Red, *Aviation Outlook : Fuel pricing ignites demand for composites in commercial transports*, 2008. [Online]. Available: http://www.compositesworld.com/articles/aviation-outlook-fuel-pricing-ignites-demand-for-composites-in-commercial-transports.

[60] B. C. Airplanes, *787 Airplane Characteristics for Airport Planning 787-10 - Information is Preliminary - Revision Record*, 2015.

[61] J. Hale, "Boeing 787 from the Ground Up", *Aeromagazine*, 2008. [Online]. Available: http://www.boeing.com/commercial/aeromagazine/articles/qtr_4_06/AERO_Q406_article4.pdf.

[62] Globalair, *Aviation Fuel - Current US Fuel Prices*, 2018.

[63] US Energy Information Administration, *Carbon Dioxide Emissions Coefficients*, 2016. [Online]. Available: https://www.eia.gov/environment/emissions/co2_vol_mass.php.

[64] Deloitte, "2017 Global aerospace and defense sector financial performance study", p. 48, 2017. [Online]. Available: https://www2.deloitte.com/content/dam/Deloitte/global/Documents/consumer-industrial-products/gx-cip-global-aerospace-defense-financial-performance-study.pdf.

[65] Z.-H. Zhang, H. Wang, S.-L. Li, X.-W. Cheng, F.-C. Wang, G. F. Korznikova, Z.-Y. Hu, and D. V. Gunderov, "The influence of defect structures on the mechanical properties of Ti-6Al-4V alloys deformed by high-pressure torsion at ambient temperature", *Materials Science and Engineering: A*, vol. 684, no. 12, pp. 1–13, 2016, ISSN: 09215093. DOI: 10.1016/j.msea.2016.12.033. [Online]. Available: http://dx.doi.org/10.1016/j.msea.2016.12.033.

[66] S. K. Everton, M. Hirsch, P. Stravroulakis, R. K. Leach, and A. T. Clare, "Review of in-situ process monitoring and in-situ metrology for metal additive manufacturing", *Materials and Design*, vol. 95, pp. 431–445, 2016, ISSN: 18734197. DOI: 10.1016/j.matdes.2016.01.099. [Online]. Available: http://dx.doi.org/10.1016/j.matdes.2016.01.099%20http://ac.els-cdn.com/S0264127516300995/1-s2.0-S0264127516300995-main.pdf?_tid=b597c516-4fa3-11e7-998c-00000aab0f02&acdnat=1497295200_afc7bd3c81932f24be86c56600ad1af7.

[67] D. Montgomery, "Inferences about process quality", in *Introduction to statistical quality control*, 6th, Wiley, 2009, ch. 4, pp. 103–176, ISBN: 9780470169926. DOI: 10.1002/1521-3773(20010316)40:6<9823::AID-ANIE9823>3.3.CO;2-C. [Online]. Available: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Introduction+to+Statistical+Quality+Control#0.

[68] G. May and S. Costas, "Statistical Fundamentals", in *Fundamentals of Semiconductor Manufacturing and Process Control*, 2006, ch. 4, pp. 122–147.

[69] B. Bolch, "More on Unbiased Estimation of the Standard Deviation", *The American Statistician*, vol. 22, no. 3, p. 27, 1968.

[70] R. Brugger, “A Note on Unbiased Estimation of the Standard Deviation”, *The American Statistician1*, vol. 23, no. 4, p. 32, 1969.

[71] T. Macak, J. Hron, and I. Ticha, “Reducing production variability using factorial optimisation: A case study from the food-packaging industry”, *Cogent Engineering*, vol. 32, pp. 1–8, 2018. DOI: 10.1080/23311916.2018.1455276. [Online]. Available: http://doi.org/10.1080/23311916.2018.1455276.

[72] J. Kwangkok, “Variability assessment and mitigation in advanced VLSI manufacturing through design-manufacturing co-optimization”, PhD thesis, UC San Diego, 2011.

[73] J. Antony, M. Hughes, and M. Kaye, “Reducing manufacturing process variability using experimental design technique: a case study”, *Integrated Manufacturing Systems*, vol. 10, pp. 162–170, 1999. DOI: 10.1108/09576069910264420.

[74] M. B. Dow, “The ACEE Program and Basic Composites Research at Langley Research Center (1975 to 1986)”, *NASA Reference Publications*, vol. 1177, 1987. [Online]. Available: https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19870020179.pdf.

[75] National Research Council, “Advanced Organic Composite Materials for Aircraft Structures: Future Program”, National Research Council, Washington DC, Tech. Rep., 1987, p. 99. [Online]. Available: https://books.google.com/books?id=r5orAAAAYAAJ&pgis=1.

[76] L. F. Vosteen and R. N. Hadcock, “Composite chronicles: A study of the lessons learned in the development, production, and service of composite structures”, NASA, Tech. Rep., 1994. [Online]. Available: http://ntrs.nasa.gov/search.jsp?R=19950010444.

[77] A. Quilter, “Composites in Aerospace Applications”, *Aviation Pros*, 2004. [Online]. Available: http://www.aviationpros.com/article/10386441/composites-in-aerospace.

[78] M. Gagliardi, “Ceramic Matrix Composites and Carbon Matrix Composites: Technologies and Global Markets.”, Tech. Rep., 2016.

[79] M. Schlechter, “The Global Market for Composites: Resins, Fillers, Reinforcements, Natural Fibers & Nanocomposite”, Tech. Rep., 2016.

[80] G. Davis, “Advanced Composites Technology Program”, NASA, Tech. Rep., 1995. [Online]. Available: https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19950022611.pdf.

[81] M. Rouse, D. Jegley, D. McGowan, H. Bush, and W. Waters, “Utilization of the Building-Block Approach in Structural Mechanics Research”, 2005, ISBN: 978-1-62410-065-9. DOI: 10.2514/6.2005-1874. [Online]. Available: http://arc.aiaa.org/doi/abs/10.2514/6.2005-1874.

[82] NASA, “First NASA Advanced Composites Technology Conference”, 1990, ISBN: 9781847359988. DOI: 10.2514/3.60379.

[83] J. Delbrey, “Database of Mechanical Properties of Textile Composites”, *NASA Contractor Report 4747*, p. 34, 1996. [Online]. Available: http://www.cs.odu.edu/%5C%7B~%5C%7Dmln/ltrs-pdfs/NASA-96-cr4747.pdf.

[84] NASA, “Second NASA Advanced Composites Technology Conference”, 1991. [Online]. Available: https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19940042252.pdf.

[85] C. C. Poe, H. B. Dexter, and I. S. Raju, "Review of the NASA Textile Composites Research", *Journal of Aircraft*, vol. 36, no. 5, pp. 876–884, 1999, ISSN: 0021-8669. DOI: 10.2514/2.2521. [Online]. Available: https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20040105589.pdf.

[86] Department of Defense, "Department of Defense Handbook Composite Materials Handbook Volume 1. Polymer Matrix Composites Guidelines for Characterization of Structural Materials", vol. 1, no. June, 1997.

[87] B. Berenberg, "AGATE Methodology Proves Its Worth", *Composites World*, 2003. [Online]. Available: http://www.compositesworld.com/articles/agate-methodology-proves-its-worth.

[88] Federal Aviation Agency, "Material Qualification and Equivalency for Polymer Matrix Composite Material Systems: Updated Procedure", no. September, 2003. [Online]. Available: http://www.niar.wichita.edu/agate/Documents/Materials/DOT-FAA-AR-03-19.pdf.

[89] D. Tenney, J. Davis Jr, R. Pipes, and N. Johnston, "NASA Composite Materials Development: Lessons Learned and Future Challenges", *NATO RTO Workshop on Support of Composite Systems*, pp. 1–58, 2009, ISSN: 12900958. DOI: 10.1016/S1290-0958(01)90057-7. [Online]. Available: http://ntrs.nasa.gov/search.jsp?R=20090037429.

[90] B. Vayre, F. Vignat, and F. Villeneuve, "Designing for additive manufacturing", *Procedia CIRP*, vol. 3, no. 1, pp. 632–637, 2012, ISSN: 22128271. DOI: 10.1016/j.procir.2012.07.108. [Online]. Available: http://dx.doi.org/10.1016/j.procir.2012.07.108%20https://ac.els-cdn.com/S2212827112002806/1-s2.0-S2212827112002806-main.pdf?_tid=72ece546-8b8a-4aff-bc0a-d0b9c37fcd91&acdnat=1528749342_4c88514caac6dfb2aaf831d2f8e079d5.

[91] J. H. P. Pallari, K. W. Dalgarno, and J. Woodburn, "Mass customization of foot orthoses for rheumatoid arthritis using selective laser sintering", *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 7, pp. 1750–1756, 2010, ISSN: 00189294. DOI: 10.1109/TBME.2010.2044178. arXiv: 9809069v1 [gr-qc].

[92] R. Huang, M. Riddle, D. Graziano, J. Warren, S. Das, S. Nimbalkar, J. Cresko, and E. Masanet, "Energy and emissions saving potential of additive manufacturing: the case of lightweight aircraft components", *Journal of Cleaner Production*, vol. 135, pp. 1559–1570, Nov. 2016, ISSN: 09596526. DOI: 10.1016/j.jclepro.2015.04.109. arXiv: 9809069v1 [gr-qc]. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0959652615004849.

[93] M. Thomas, G. J. Baxter, and I. Todd, "Normalised model-based processing diagrams for additive layer manufacture of engineering alloys", *Acta Materialia*, vol. 108, pp. 26–35, 2016, ISSN: 13596454. DOI: 10.1016/j.actamat.2016.02.025. [Online]. Available: http://dx.doi.org/10.1016/j.actamat.2016.02.025.

[94] M. K. Thompson, G. Moroni, T. Vaneker, G. Fadel, R. I. Campbell, I. Gibson, A. Bernard, J. Schulz, P. Graf, B. Ahuja, and F. Martina, "Design for Additive Manufacturing: Trends, opportunities, considerations, and constraints", *CIRP Annals - Manufacturing Technology*, vol. 65,

no. 2, pp. 737–760, 2016, ISSN: 17260604. DOI: 10.1016/j.cirp.2016.05.004. [Online]. Available: http://dx.doi.org/10.1016/j.cirp.2016.05.004%20http://www.gospi.fr/IMG/pdf/cirp-annals-dfam-1549-2016.pdf.

[95] Federal Aviation Agency, "Summary Report : Joint Federal Aviation Administration – Air Force Workshop on Qualification / Certification of Additively Manufactured Parts", Federal Aviation Administration, New Jersey, Tech. Rep., 2016, p. 227. [Online]. Available: http://www.tc.faa.gov/its/worldpac/techrpt/tc16-15.pdf.

[96] S. J. Findlay and N. D. Harrison, "Why aircraft fail", *Materials Today*, vol. 5, no. 11, pp. 18–25, 2002, ISSN: 13697021. DOI: 10.1016/S1369-7021(02)01138-0. [Online]. Available: http://dx.doi.org/10.1016/S1369-7021(02)01138-0%20http://ac.els-cdn.com/S1369702102011380/1-s2.0-S1369702102011380-main.pdf?_tid=fba16874-5d94-11e7-9a38-00000aab0f02&acdnat=1498828191_5d597b668501ddabff839449853d66fe.

[97] T. Kellner, "The FAA Cleared the First 3D Printed Part to Fly in a Commercial Jet Engine from GE", *GE News*, pp. 3–5, 2015.

[98] J. Hartford, *FDA's View on 3-D Printing Medical Devices*, 2015. [Online]. Available: http://www.mddionline.com/article/fdas-view-3-d-printing-medical-devices.

[99] Federal Food and Drug Agency, *Technical Considerations for Additive Manufactured Devices Draft Guidance for Industry and Food and Drug Administration Staff*, 2016.

[100] ——, *Technical Considerations for Additive Manufactured Medical Devices: Guidance for Industry and Food and Drug Administration Staff*, 2017. [Online]. Available: https://www.fda.gov/downloads/MedicalDevices/%20DeviceRegulationandGuidance/GuidanceDocuments/UCM499809.pdf?%20utm_campaign=Announcing%20Final%20Guidance%20and%20Webinar%20for%5C%3A%20Technical%20Considerations%20for%20AM&utm_medium=email&utm_source=Eloqua&elqTrackId=D34174C6F947F2.

[101] C. Emmelmann, P. Sander, J. Kranz, and E. Wycisk, "Laser additive manufacturing and bionics: Redefining lightweight design", *Physics Procedia*, vol. 12, no. 1, pp. 364–368, 2011, ISSN: 18753884. DOI: 10.1016/j.phpro.2011.03.046. [Online]. Available: http://dx.doi.org/10.1016/j.phpro.2011.03.046%20http://ac.els-cdn.com/S1875389211001258/1-s2.0-S1875389211001258-main.pdf?_tid=008c2bee-5dd1-11e7-8c7d-00000aab0f01&acdnat=1498853969_6eebc63291ca56dfbbab58330997f41d.

[102] Oak Ridge National Laborator, "Aerospace Workshop Planning Session Summary Report", Oak Ridge National Laboratory, Tech. Rep., 2010, p. 24. [Online]. Available: https://web.ornl.gov/sci/manufacturing/docs/Aerospace_Workshop_Report.pdf.

[103] S. M. Gaytan, L. E. Murr, F. Medina, E. Martinez, M. I. Lopez, and R. B. Wicker, "Advanced metal powder based manufacturing of complex components by electron beam melting", *Materials Technology*, vol. 24, no. 3, pp. 180–190, 2009, ISSN: 1066-7857.

[104] H. P. Tang, M. Qian, N. Liu, X. Z. Zhang, G. Y. Yang, and J. Wang, "Effect of Powder Reuse Times on Additive Manufacturing of Ti-6Al-4V by Selective Electron Beam Melting", *The Journal of*

The Minerals, Metals & Materials Society, vol. 67, no. 3, pp. 555–563, 2015, ISSN: 15431851. DOI: 10.1007/s11837-015-1300-4.

[105] T. Wohlers and T. Gornet, "History of additive manufacturing", Wohlers Report, Tech. Rep., 2014, pp. 1–34. [Online]. Available: http://wohlersassociates.com/history2014.pdf.

[106] D. Tenney, J. Davis Jr, R. Pipes, and N. Johnston, "NASA Composite Materials Development: Lessons Learned and Future Challenges", in *NATO RTO Workshop on Support of Composite Systems*, 2009, pp. 1–58.

[107] J. Bonnín Roca, P. Vaishnav, E. R. H. Fuchs, and M. G. Morgan, "Policy needed for additive manufacturing", *Nature Materials*, vol. 15, no. 8, pp. 815–818, 2016, ISSN: 1476-1122. DOI: 10.1038/nmat4658. [Online]. Available: http://www.nature.com/doifinder/10.1038/nmat4658.

[108] General Electric, *The FAA Cleared the First 3D Printed Part to Fly in a Commercial Jet Engine from GE*, 2015. [Online]. Available: http://www.gereports.com/post/116402870270/the-faa-cleared-the-first-3d-printed-part-to-fly/.

[109] B. E. Carroll, T. A. Palmer, and A. M. Beese, "Anisotropic tensile behavior of Ti-6Al-4V components fabricated with directed energy deposition additive manufacturing", *Acta Materialia*, vol. 87, pp. 309–320, 2015, ISSN: 13596454. DOI: 10.1016/j.actamat.2014.12.054. [Online]. Available: http://dx.doi.org/10.1016/j.actamat.2014.12.054%20http://ac.els-cdn.com/S135964541400980X/1-s2.0-S135964541400980X-main.pdf?_tid=61be871a-35cf-11e7-bed0-00000aacb35e&acdnat=1494455227_5fab555e5221948d242ae7827d9253c5.

[110] R. M. Hunt, K. J. Kramer, and B. El-dasher, "Selective laser sintering of MA956 oxide dispersion strengthened steel", *Journal of Nuclear Materials*, vol. 464, pp. 80–85, 2015, ISSN: 0022-3115. DOI: 10.1016/j.jnucmat.2015.04.011. [Online]. Available: http://dx.doi.org/10.1016/j.jnucmat.2015.04.011.

[111] D. Gu, H. Wang, F. Chang, D. Dai, P. Yuan, Y. C. Hagedorn, and W. Meiners, "Selective laser melting additive manufacturing of TiC/AlSi10Mg bulk-form nanocomposites with tailored microstructures and properties", *Physics Procedia*, vol. 56, pp. 108–116, 2014, ISSN: 18753892. DOI: 10.1016/j.phpro.2014.08.153. [Online]. Available: http://dx.doi.org/10.1016/j.phpro.2014.08.153%20http://ac.els-cdn.com/S1875389214002983/1-s2.0-S1875389214002983-main.pdf?_tid=9febd7ba-35d2-11e7-a57a-00000aacb35e&acdnat=1494456620_a975be82869f321782d5c729890b7e89.

[112] K. Kempen, L. Thijs, J. Van Humbeeck, and J.-P. Kruth, "Mechanical Properties of AlSi10Mg Produced by Selective Laser Melting", *Physics Procedia*, vol. 39, pp. 439–446, 2012, ISSN: 18753892. DOI: 10.1016/j.phpro.2012.10.059. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1875389212025862.

[113] H. K. Rafi, T. L. Starr, and B. E. Stucker, "A comparison of the tensile, fatigue, and fracture behavior of Ti-6Al-4V and 15-5 PH stainless steel parts made by selective laser melting", *International Journal of Advanced Manufacturing Technology*, vol. 69, no. 5-8, pp. 1299–1309,

2013, ISSN: 02683768. DOI: 10.1007/s00170-013-5106-7. [Online]. Available: %7Bhttp://download.springer.com/static/pdf/327/art%5C%253A10.1007%5C%252Fs00170-013-5106-7.pdf?originUrl=http%5C%3A%5C%2F%5C%2Flink.springer.com%5C%2Farticle%5C%2F10.1007%5C%2Fs00170-013-5106-7&token2=exp=1494515910~acl=%5C%2Fstatic%5C%2Fpdf%5C%2F327%5C%2Fart%5C%25253A10.1007%5C%25252Fs00170-013-510%7D.

[114] B. Baufeld, O. V. d. Biest, and R. Gault, "Additive manufacturing of Ti-6Al-4V components by shaped metal deposition: Microstructure and mechanical properties", *Materials and Design*, vol. 31, no. SUPPL. 1, S106–S111, 2010, ISSN: 02641275. DOI: 10.1016/j.matdes.2009.11.032.

[115] E. Village, *Quick Search*, 2019. [Online]. Available: https://www.engineeringvillage.com/.

[116] T. Wang, J. Zhao, C. Sheng, W. Wang, and L. Wang, "Multi-Layer Encoding Genetic Algorithm-Based Granular Fuzzy Inference for Blast Furnace Gas Scheduling", *IFAC-PapersOnLine*, vol. 49, no. 20, pp. 132–137, ISSN: 2405-8963. DOI: 10.1016/j.ifacol.2016.10.109. [Online]. Available: http://dx.doi.org/10.1016/j.ifacol.2016.10.109.

[117] G. Yuan, W. Hu, X. Wang, J. Kang, J. Zhao, H. Di, R. D. K. Misra, and G. Wang, "The relationship between microstructure , crystallographic orientation , and fracture behavior in a high strength ferrous alloy", vol. 695, pp. 526–539, 2017. DOI: 10.1016/j.jallcom.2016.11.141.

[118] R. D. K. Misra, H. Zheng, K. M. Wu, and L. P. Karjalainen, "Niobium-containing quenching and partitioning processed ultrahigh strength martensite – austenite dual phase steels", *Materials Science & Engineering A*, vol. 579, pp. 188–193, 2013. DOI: 10.1016/j.msea.2013.05.043.

[119] E. Kim, K. Huang, A. Saunders, A. McCallum, G. Ceder, and E. Olivetti, "Materials Synthesis Insights from Scientific Literature via Text Extraction and Machine Learning", *Chemistry of Materials*, vol. 29, no. 21, pp. 9436–9444, 2017, ISSN: 15205002. DOI: 10.1021/acs.chemmater.7b03500. [Online]. Available: https://pubs.acs.org/doi/pdf/10.1021/acs.chemmater.7b03500.

[120] O. Group, *The Synthesis Project*, 2019. [Online]. Available: https://www.synthesisproject.org/.

[121] D. E. Jones, S. Igo, J. Hurdle, and J. C. Facelli, "Automatic Extraction of Nanoparticle Properties Using Natural Language Processing: NanoSifter, an Application to Acquire Pamam Dendrimer Properties", *PLoS ONE*, vol. 9, no. 1, pp. 1–5, 2014, ISSN: 19326203. DOI: 10.1371/journal.pone.0083932. [Online]. Available: http://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0083932&type=printable.

[122] S. Pletscher-Frankild, A. Pallejà, K. Tsafou, J. X. Binder, and L. J. Jensen, "DISEASES: Text mining and data integration of disease-gene associations", *Methods*, vol. 74, pp. 83–89, 2015, ISSN: 10959130. DOI: 10.1016/j.ymeth.2014.11.020. [Online]. Available: http://dx.doi.org/10.1016/j.ymeth.2014.11.020%20https://ac.els-cdn.com/S1046202314003831/1-s2.0-S1046202314003831-main.pdf?_tid=3c49afb1-a88b-4e24-b267-9e8c89816791&acdnat=1551379717_fcf9425894a244d1fd841f6df5ef140f.

[123] A. Rzhetsky, I. Iossifov, T. Koike, M. Krauthammer, P. Kra, M. Morris, H. Yu, P. A. Duboué, W. Weng, W. J. Wilbur, V. Hatzivassiloglou, and C. Friedman, "GeneWays: A system for extracting, analyzing, visualizing, and integrating molecular pathway data", *Journal of Biomedical Informatics*, vol. 37, no. 1, pp. 43–53, 2004, ISSN: 15320464. DOI: 10.1016/j.jbi.2003.10.001. [Online]. Available: https://ac.els-cdn.com/S1532046403001126/1-s2.0-S1532046403001126-main.pdf?_tid=f981ac5e-0f1e-4993-bc88-9ac3f8fbbcdd&acdnat=1549994295_3568f9586c04a55ec23a515897b83230.

[124] L. Hawizy, D. M. Jessop, N. Adams, and P. Murray-Rust, "ChemicalTagger: A tool for semantic text-mining in chemistry", *Journal of Cheminformatics*, vol. 3, no. 1, pp. 1–13, 2011, ISSN: 17582946. DOI: 10.1186/1758-2946-3-17. [Online]. Available: https://jcheminf.springeropen.com/track/pdf/10.1186/1758-2946-3-17?site=jcheminf.springeropen.com.

[125] E. Strubell, P. Verga, D. Belanger, and A. McCallum, "Fast and Accurate Entity Recognition with Iterated Dilated Convolutions", 2017. [Online]. Available: http://arxiv.org/abs/1702.02098.

[126] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching Word Vectors with Subword Information", vol. 5, pp. 135–146, 2017, ISSN: 00385298. DOI: 10.1007/BF01959819. [Online]. Available: http://arxiv.org/abs/1607.04606.

[127] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space", pp. 1–12, 2013, ISSN: 15324435. DOI: 10.1162/153244303322533223. [Online]. Available: http://arxiv.org/abs/1301.3781.

[128] K. Cho, B. van Merrienboer, C. Gulcehre, F. Bougares, H. Schwenk, D. Bahdanau, and Y. Bengio, "Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation", in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, 2014, pp. 1724–1734.

[129] *An Improved Non-monotonic Transition System for Dependency Parsing*, September, Proceedings of the 2015 Conference on Empirical Methods in Natural Language Process, 2015, pp. 1373–1378.

[130] spaCy, *Industrial-Strength Natural Language Processing*. [Online]. Available: https://spacy.io/.

[131] D. Khashabi, S. Chaturvedi, M. Roth, S. Upadhyay, and D. Roth, "Looking Beyond the Surface : A Challenge Set for Reading Comprehension over Multiple Sentences", no. 2013, pp. 252–262, 2018.

[132] P. Gupta, S. Rajaram, H. Schütze, B. Andrassy, and T. Runkler, "Neural Relation Extraction Within and Across Sentence Boundaries", 2018. [Online]. Available: http://arxiv.org/abs/1810.05102.

[133] N. Peng, H. Poon, C. Quirk, K. Toutanova, and W.-t. Yih, "Cross-Sentence N-ary Relation Extraction with Graph LSTMs", vol. 5, pp. 101–115, 2017, ISSN: 2307-387X. DOI: 10.1111/j.1753-4887.2012.00493.x.Defining. [Online]. Available: http://arxiv.org/abs/1708.03743.

[134] K. Belkassa, F. Bessaha, K. Marouf-khelifa, I. Batonneau-gener, J.-d. Comparot, and A. Khelifa, "Physicochemical and adsorptive properties of a heat-treated and acid-leached Algerian halloysite", *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, vol. 421, pp. 26–33, 2013, ISSN: 0927-7757. DOI: 10.1016/j.colsurfa.2012.12.048. [Online]. Available: http://dx.doi.org/10.1016/j.colsurfa.2012.12.048.

[135] R. Liu, H. Tian, A. Yang, F. Zha, J. Ding, and Y. Chang, "Applied Surface Science Preparation of HZSM-5 membrane packed CuO – ZnO – Al 2 O 3 nanoparticles for catalysing carbon dioxide hydrogenation to dimethyl ether", *Applied Surface Science*, vol. 345, pp. 1–9, 2015, ISSN: 0169-4332. DOI: 10.1016/j.apsusc.2015.03.125. [Online]. Available: http://dx.doi.org/10.1016/j.apsusc.2015.03.125.

[136] K. Lee, S. J. Park, Y. S. Choi, S. J. Kim, T. H. Lee, K. H. Oh, and H. N. Han, "Dual-scale correlation of mechanical behavior in duplex low-density steel", *Scripta Materialia*, vol. 69, no. 8, pp. 618–621, 2013, ISSN: 13596462. DOI: 10.1016/j.scriptamat.2013.07.015. [Online]. Available: http://dx.doi.org/10.1016/j.scriptamat.2013.07.015%20https://ac.els-cdn.com/S1359646213003709/1-s2.0-S1359646213003709-main.pdf?_tid=ba06ca3a-3e10-4ed9-adc8-2812e81540a6&acdnat=1551384412_3ec14067c5d3742c6eb70b1a27a16182.

[137] C. Manning, P. Raghavan, and H. Schutze, "Scoring, term weighting, and the vector space model", in *Introduction to Information Retrieval*, 2008, ch. 6, pp. 100–122, ISBN: 9780521865715.

[138] V. Levenshtein, *Binary codes capable of correcting deletions, insertions, and reversals*, 1965. [Online]. Available: https://nymity.ch/sybilhunting/pdf/Levenshtein1966a.pdf.

[139] Z. Zhang, Q. Wang, H. Chen, and Y. Zhou, "Influence of the initial moist curing time on the sulfate attack resistance of concretes with different binders", *Construction and Building Materials*, vol. 144, pp. 541–551, 2017, ISSN: 0950-0618. DOI: 10.1016/j.conbuildmat.2017.03.235. [Online]. Available: http://dx.doi.org/10.1016/j.conbuildmat.2017.03.235.

[140] R. Liu, H. Tian, A. Yang, F. Zha, J. Ding, and Y. Chang, "Applied Surface Science Preparation of HZSM-5 membrane packed CuO – ZnO – Al 2 O 3 nanoparticles for catalysing carbon dioxide hydrogenation to dimethyl ether", *Applied Surface Science*, vol. 345, pp. 1–9, 2015, ISSN: 0169-4332. DOI: 10.1016/j.apsusc.2015.03.125. [Online]. Available: http://dx.doi.org/10.1016/j.apsusc.2015.03.125.

[141] M. A. Lodes, R. Guschlbauer, and C. Körner, "Process development for the manufacturing of 99.94% pure copper via selective electron beam melting", *Materials Letters*, vol. 143, pp. 298–301, 2015, ISSN: 18734979. DOI: 10.1016/j.matlet.2014.12.105. [Online]. Available: http://dx.doi.org/10.1016/j.matlet.2014.12.105%20https://ac.els-cdn.com/S0167577X1402271X/1-s2.0-S0167577X1402271X-main.pdf?_tid=09b0a4ce-69cc-4182-9da6-fb1d6e4db186&acdnat=1523380362_ac958e60dc22dc5a16d4d259c390f364.

Appendix A

Appendix: List of Labels for Token Classification

0: "null",

1: "nonrecipe-material",

2: "material",

3: "precursor",

4: "solvent",

5: "gas",

6: "target",

7: "unspecified-material",

8: "number",

9: "amount-unit",

10: "condition-unit",

11: "property-unit",

12: "apparatus-unit",

13: "condition-type",

14: "property-type",

15: "apparatus-property-type",

16: "meta" ,

17: "reference",

18: "brand",

19: "condition-misc",

20: "property-misc",

21: "amount-misc",

22: "apparatus-descriptor",

23: "material-descriptor",

24: "synthesis-apparatus",

25: "characterization-apparatus",

26: "operation"

Appendix B

Appendix: Dictionary Mapping of AM Terms

This dictionary was built while manually annotating 90 AM papers in order to map synonyms to the same word. For instance 'max laser power' and 'maximum output laser power' are really the same thing, but synonyms.

' ebm':'process'

' slm':'process'

' sls':'process'

'0.2 yield strength':'yield strength.1'

'0.2% offset yield strenght':'yield strength.1'

'0.2% offset yield strength':'yield strength.1'

'0.2% ys':'yield strength.1'

'2% yield strength':'yield strength.1'

'316l':'316l stainless steel'

'316l ss':'316l stainless steel'

'316l stainless steel':'316l stainless steel'

'3d printing':'additive manufacturing'

'? 0.2':'yield strength.1'

'?m':'micrometer'

'accelerating voltage':'accelerating voltage'

'additive manufacturing':'additive manufacturing'

'alsi10mg':'alsi10mg'

'alsi10mg powder':'alsi10mg'

'alsi10mg-200c powder':'alsi10mg'

'areal surface roughness':'surface roughness'

'areal surface roughness, sa':'surface roughness'

'average 0.2% yield strength':'yield strength.1'

'average carbon fibre length':'carbon fiber length'

'average diameter':'particle size'

'average particle size':'particle size'

'average powder size':'particle size'

'average uts':'ultimate tensile strength'

'bal':'balance'

'bal.':'balance'

'balance':'balance'

'baseplate temperature':'substrate temperature'

'beam current':'beam current'

'beam diameter':'beam diameter'

'beam spot size':'beam diameter'

'beam velocity':'travel speed'

'beam voltage':'beam voltage'

'beam wavelength':'wavelength'

'bed temperature':'substrate temperature'

'build direction':'build direction'

'build orientation':'build direction'

'build platform temperature':'substrate temperature'

'building plate size':'platform dimesion'

'building platform pre-heat temperature':'substrate temperature'

'building platform preheated temperature':'substrate temperature'

'carbon fiber length':'carbon fiber length'

'chamber pressure':'chamber pressure'

'contour spacing':'contour spacing'

'cube dimensions':'cube dimensions'

'cube height':'cube height'

'cubic specimen':'cube dimensions'

'cubic specimens':'cubic specimens'

'ded':'directed energy deposition'

'density':'density'

'density of carbon fibre material':'density of carbon fibre material'

'density of pa-12 pellets':'density of pa-12 pellets'

'diameter':'diameter'

'diameter fiber':'fiber diameter'

'dimension':'dimensions'

'dimensions':'dimensions'

'dimensions of specimens for flexural tests':'dimensions of specimens for flexural tests'

'direct melt electrospinning':'process'

'direct melt-electrospinning':'process'

'direct metal laser sintering':'process'

'direct metal tooling':'process'

'directed energy deposition':'directed energy deposition'

'e':'stiffness'

'e (modulus)':'stiffness'

'e-f':'stiffness'

'ebm':'electron beam melting'

'edge length':'edge length'

'el':'elongation to fracture'

'electron beam':'process'

'electron beam melting':'process'

’eli grade arcam ti-6al-4v’:’ti-6al-4v’

’elong’:’elongation to fracture’

’elongated coarse grains size’:’elongated coarse grains size’

’elongation’:’elongation to fracture’

’elongation at break’:’elongation to fracture’

’elongation at failure’:’elongation to fracture’

’elongation at fracture’:’elongation to fracture’

’elongation per astm e8’:’elongation per astm e8’

’elongation to failure’:’elongation to fracture’

’elongation to fracture’:’elongation to fracture’

’energy deposition’:’process’

’exposure time’:’exposure time’

’fatigue limit’:’fatigue limit’

’feed rate’:’powder feed rate’

’fiber diameter’:’fiber diameter’

’fill density of the device’:’fill density of the device’

’fill laser power’:’fill laser power’

’fill pattern’:’fill pattern’

’focal length’:’focal length’

’geometry’:’hatch angle’

’grain diameter’:’grain diameter’

’hardness’:’hardness’

’hardness ’:’hardness’

’hatch angle’:’hatch angle’

’hatch direction’:’hatch angle’

’hatch distance’:’hatch spacing’

’hatch space’:’hatch spacing’

'hatch spacing':'hatch spacing'

'hatch spacing overlap':'hatch spacing overlap'

'hatch-spacing':'hatch spacing'

'hatching distance':'hatch spacing'

'hatching spacing':'hatch spacing'

'height':'height'

'high temperature laser sinter':'process'

'horizontal overlap distance':'hatch spacing overlap'

'increment in z':'layer thickness'

'increment of z':'layer thickness'

'inkjet printing':'process'

'laser additive manufacturing':'process'

'laser beam':'laser beam'

'laser beam diameter':'beam diameter'

'laser beam rotation':'laser beam rotation'

'laser beam speed':'travel speed'

'laser beam spot size':'beam diameter'

'laser cladding':'process'

'laser diameter':'beam diameter'

'laser engineered net shaping':'process'

'laser focus offset':'laser focus offset'

'laser maximum power':'maximum laser power'

'laser melting deposition':'process'

'laser metal deposition':'process'

'laser powder deposition':'process'

'laser power':'laser power'

'laser power density':'power density'

'laser scan spacimg':'hatch spacing'

'laser scan speed':'travel speed'

'laser scanning speed':'travel speed'

'laser scanning velocity':'travel speed'

'laser solid forming':'process'

'laser spot diameter':'beam diameter'

'laser spot size':'beam diameter'

'laser spot size diameter':'beam diameter'

'laser stereolithography':'stereo lithography'

'laser travel speed':'travel speed'

'laser wavelength':'wavelength'

'layer height':'layer thickness'

'layer height (thickness)':'layer thickness'

'layer thickness':'layer thickness'

'layers of the slices':'layers of the slices'

'line energy':'line energy'

'line offset':'offset'

'linear heat input':'line energy'

'machine voltage':'machine voltage'

'max laser power':'maximum laser power'

'max power':'maximum laser power'

'maximium power':'maximum laser power'

'maximum elongation':'elongation to fracture'

'maximum laser output power':'maximum laser power'

'maximum laser power':'maximum laser power'

'maximum output':'maximum laser power'

'maximum output power':'maximum laser power' 'maximum power':'maximum laser power'

'maximum scanning speed':'travel speed'

'mean particle size':'particle size'

'measured elongation':'elongation to fracture'

'melt electrospinning writing':'process'

'microhardness':'hardness'

'micrometer':'micrometer'

'mixing duration':'mixing duration'

'modulus (young)':'stiffness'

'mwcnt diameter':'mwcnt diameter'

'mwcnt length':'carbon fiber length'

'mwcnt width':'mwcnt diameter'

'offset':'offset'

'offset height':'offset'

'offset spacing':'offset'

'optical fiber diameter':'fiber diameter'

'optical fiber width':'fiber diameter'

'output laser power':'laser power'

'particle size':'particle size'

'particle size distribution':'particle size'

'percent elongation':'elongation to fracture'

'plate size':'platform dimesion'

'platform dimension':'platform dimesion'

'platform dimesion':'platform dimesion'

'platform size':'platform dimesion'

'platform temperature':'substrate temperature'

'pore volume fraction':'porosity'

'porosity':'porosity'

'porosity levels':'porosity'

'powder average size':'particle size'

'powder bed fusion':'process'

'powder bed pre-heat temperature':'substrate temperature'

'powder bed temperature':'substrate temperature'

'powder diameter':'particle size'

'powder feed rate':'powder feed rate'

'powder feeding rate':'powder feed rate'

'powder grain size':'particle size'

'powder layer thickness':'layer thickness'

'powder mass flow':'powder mass flow'

'powder particle size':'particle size'

'powder size':'particle size'

'powder size range':'particle size'

'powder thickness':'layer thickness'

'power':'laser power'

'power density':'power density'

'pre-heating temperature':'substrate temperature'

'preheat':'substrate temperature'

'preheating temperature':'substrate temperature'

'rapid prototyping':'additive manufacturing'

'raster diameter':'raster diameter'

'raster orientation':'raster orientation'

'relative density':'density'

'remainder':'balance'

'roof and floor layers':'roof and floor layers' 'rotation':'laser beam rotation'

'rotation speed':'rotation speed'

'scan rate':'travel speed'

'scan rotation between successive layers':'scan rotation between successive layers'

'scan spacing':'hatch spacing'

'scan speed':'travel speed'

'scan vector length':'scan vector length'

'scanning rate':'travel speed'

'scanning speed':'travel speed'

'scanning velocity':'travel speed'

'sebm':'electron beam melting'

'selective electron beam melting':'electron beam melting'

'selective laser melting':'process'

'selective laser sintering':'process'

'setting power':'setting power'

'size':'size'

'slm':'selective laser melting'

'sls':'selective laser sintering'

'spacing':'hatch spacing'

'specific power':'specific power'

'speed':'travel speed'

'spherical pores average diameter':'spherical pores average diameter'

'spot diameter':'beam diameter'

'spot size':'beam diameter'

'ss316l':'316l stainless steel'

'start plate temperature':'substrate temperature'

'start-plate temperature':'substrate temperature' 'step size':'layer thickness'

'stereo lithography':'stereo lithography'

'stiffness':'stiffness'

'stl':'stereo lithography'

'strain to failure':'elongation to fracture'

'strain to fracture':'elongation to fracture'

'substrate dimensions':'platform dimesion'

'substrate temperature':'substrate temperature'

'surface roughess':'surface roughness'

'surface roughness':'surface roughness'

'tall':'tall'

'temperature':'substrate temperature'

'temperature of nozzle':'temperature of nozzle'

'tensile elongation':'elongation to fracture'

'tensile strain at break':'elongation to fracture'

'tensile strength':'tensile strength'

'tensile stress at break':'tensile stress at break'

'tensile stress at yield (offset 0.2

'thick':'thick'

'thickness':'thick'

'thickness of the layers':'layer thickness'

'thickness per layer':'layer thickness'

'three-dimensional printing':'process'

'ti-6al-4v':'ti-6al-4v'

'ti-6al-4v grade 23':'ti-6al-4v'

'ti6al4v':'ti-6al-4v'

'ti6al4v-eli':'ti-6al-4v'

'travel speed':'travel speed'

'ultimate compressive strength':'ultimate tensile strength'

'ultimate tensile strength':'ultimate tensile strength'

'ultrasonic additive manufacturing':'process'

'um':'micrometer'

'unnamed: 48':'unnamed: 48'

'uts':'ultimate tensile strength'

'vertical step size':'layer thickness'

'vicker hardness':'hardness'

'vickers hardness':'hardness'

'vickers micro hardness':'hardness'

'voltage':'beam voltage'

'wall layers':'wall layers'

'wavelength':'wavelength'

'wire arc additive manufacturing':'process'

'wire feed rate':'wire feed rate'

'working distance':'working distance'

'yield strenght':'yield strength'

'yield strength':'yield strength'

'yield strength.1':'yield strength.1'

'yield stress':'yield strength'

'young's modulus':'stiffness'

Appendix C

Appendix: JSON Template for Manual AM Extraction


```
---
metadata:
  doi:
  issn:
data:
  global_parameters:
    process:
    atmosphere:
      material:
      pressure:
        value:
        unit:
      flow_rate:
        value:
        unit:
    scanning_strategy:
    orientation:

    materials:
      - material_name:
        material_method:
        material_brand:
        composition:
          unit:
          Al:
          B:
          C:
          Co:
          Cr:
          Cu:
```

```
      Fe:
      H:
      M:
      Mg:
      Mn:
      Mo:
      N:
      Nb:
      Ni:
      O:
      P:
      Pb:
      S:
      Si:
      Sn:
      Ta:
      Ti:
      V:
      W:
      Zn:
      misc:
    blend_ratio:


apparatuses:
  - apparatus_type:
    brand:


heat_treatments:
  - treatment_name:
    treatment_step_number:
    type:
    time:
      value:
      unit:
      error:
    temperature:
```

```
      value:
      unit:
      error:
    rate:
      value:
      unit:
      error:
    pressure:
      value:
      unit:
      error:

  properties:
    - property_name:
      unit:
      value:
      error:

samples:
  - sample_id:
    process:
    atmosphere:
      material:
      pressure:
        value:
        unit:
      flow_rate:
        value:
        unit:
    scanning_strategy:
    orientation:

    materials:
      - material_name:
        material_method:
        material_brand:
        composition:
          unit:
```

```
      Al:
      B:
      C:
      Co:
      Cr:
      Cu:
      Fe:
      H:
      M:
      Mg:
      Mn:
      Mo:
      N:
      Nb:
      Ni:
      O:
      P:
      Pb:
      S:
      Si:
      Sn:
      Ta:
      Ti:
      V:
      W:
      Zn:
      misc:
    blend_ratio:


apparatuses:
  - apparatus_type:
    brand:

heat_treatments:
  - treatment_name:
    treatment_step_number:
    type:
```

```
    time:
      value:
      unit:
      error:
    temperature:
      value:
      unit:
      error:
    rate:
      value:
      unit:
      error:
    pressure:
      value:
      unit:
      error:

  properties:
    - property_name:
    unit:
    value:
    error:
...
```