



universität
wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Particle Swarm Optimisation of Chemical Passivation of
Steel Substrate with Zirconium Oxide Layers through
Sol-Gel Process“

verfasst von / submitted by
Johann Dorn, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Master of Science (MSc)

Wien, 2023 / Vienna 2023

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

A 066 862

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Masterstudium Chemie UG2002

Betreut von / Supervisor:

Dr. Philipp Marquetand, Privatdoz.

Abstract

Passivated steel can act as fit substrate for thin-film photovoltaics. A thin zirconium oxide layer was applied via doctor blading onto a steel foil substrate with the goal of get a homogeneous and insulating layer. Layers were qualitatively characterized with SEM and XRD and quantitatively charaterized via current-voltage curves. The process variables were optimized by particle swarm optimization (PSO) algorithm in combination with multivariate adaptive regression splines (MARS). A strong correlation between the calcination temperature and the conductance of ceramics has been revealed. The MARS model performed well compared to linear regression, kernel ridge regression and support vector regression.

Preface

I thank Leticia, Sebastian and Philipp for introducing me to the intricate world of theoretical chemistry, dynamics and machine learning. I thank Theo, Adhi, Neha and Philipp for giving me fruitful advice during my master's project. I thank Quyhn, Maria, Jana, Vivien and Katharina who kept me company in the lab and inspired me. I thank Elif who got the practical part rolling of this project. I thank Karin, Johanna, Dennis and Gala who supported me with finishing this thesis. I thank my parents for my upbringing, teaching me to stay curious and to never give up. And finally I thank this project which taught me how import is to plan (experiments), chose fitting methods, reproducibility and how important proper research and constant reading of literature is.

Contents

1	Introduction	8
2	Theoretical Background	9
2.1	Photovoltaics	9
2.1.1	History of Photovoltaics	9
2.1.2	Photovoltaics Basics	9
2.1.3	CIGS	10
2.2	Materials and Their Analysis	11
2.2.1	Zirconium oxide	11
2.2.2	Sol-Gel and Doctor Blading	11
2.2.3	Sputtering	11
2.2.4	Scanning electron Microscopy	13
2.2.5	UV/Vis/NIR Spectroscopy	13
2.2.6	X-Ray Diffraction	14
2.3	Machine Learning and Statistics	14
2.3.1	Artificial Intelligence and Machine Learning	14
2.3.2	Machine Learning Methods	15
2.3.3	Population Based Optimization Algorithms	17
2.3.4	EMMA	17
2.3.5	Design of Experiment	18
2.3.6	Analysis of Variance	19
3	Experimental	20
3.1	Substrate Preparation	20
3.2	Solutions	20
3.2.1	Aquatic solution	20
3.2.2	Buthanolic solution	21
3.3	Doctor blading	22
3.4	Calcination	22
3.5	Characterization	23
3.6	Pre-optimization	24
4	Computational Details	25
4.1	Data Processing	25
4.2	Sample Selection	25
4.3	Fitting via Machine Learning	26
5	Results and Discussion	27
5.1	Material Scientific and Pre-optimization	27
5.1.1	XRD	28
5.1.2	Spectrometry	28
5.1.3	Current Voltage Curves	29
5.1.4	Preoptimization	29
5.2	EMMA	30
5.3	Post-EMMA	34

5.3.1	ANOVA	34
5.3.2	Linear Regression	34
5.3.3	Grid Search, KRR and SVM	35
5.4	Hindsight is easier than foresight	36
5.4.1	Research, exploration and design phase	36
5.4.2	Narrowing, pre-optimization and constraint phase	37
5.4.3	Optimization, modelling and data generation phase	37
6	Summary and Outlook	38
A	Pre-EMMA Experiments	46
B	EMMA Experiments	47

Acronyms

- H₂O** deionized water. 20, 21
- H₂SO₄** sulfuric acid. 11, 20, 21, 27
- N₂** nitrogen. 20
- NaOH** sodium hydroxide. 20, 21, 27
- Zr(PrO)₄** zirconium(IV)propoxide. 20–22, 27
- ZrO₂** zirconium dioxide. 11, 13, 22–24, 27–30
- 1F** one-fold concentrated solution. 21, 27
- 2F** two-fold concentrated solution. 21, 22
- 3F** three-fold concentrated solution. 21
- 4F** four-fold concentrated solution. 21, 27
- 5F** five-fold concentrated solution. 21
- AcOH** glacial acetic acid. 21, 22, 24, 27
- AI** artificial intelligence. 14, 15
- ANN** artificial neural networks. 15
- ANOVA** Analysis of Variance. 19, 34
- BF** basis function. 33, 34
- BuOH** Butan-1-ol. 20–22, 27
- CIGS** CuIn_xGa_{1-x}Se₂. 9, 10
- DB** doctor blading. 11, 22–24, 27, 29
- DOE** design of experiment. 18
- EM** electro magnetic. 13
- EMMA** evolutionary model-based multiresponse approach. 17, 25–29, 33–38
- EPV** events per predictor variable. 33, 37
- FEG** field emission gun. 13
- FT** Fourier transform. 14, 23
- FTO** fluorine doped tin oxide. 20, 27, 28

GA genetic algorithm. 17

GCV generalized cross validation. 18

Hacac acetylacetone. 20–22

HCl hydrochloric acid. 20, 21, 27

I-V current-voltage. 23, 25, 29

IPO Propan-2-ol. 20–22, 24, 27

IR infrafred. 13

ITO indium doped tin oxide. 20

KRR kernel ridge regression. 15, 26, 35

MAE mean absolute error. 15, 26, 35

MARS multivariate adaptive regression splines. 17, 33, 36

ML machine learning. 14, 15, 17

MSE mean squared error. 15, 18, 25, 26, 33, 35

NIR near-infrafred. 13, 23, 28

PB Plackett-Burman. 18

PSO particle swarm optimization. 17, 21, 24, 25

PV photovoltaic. 9–12

PVD physical vapor deposition. 11

RBF radial basis function. 26

RF regression function. 17, 30, 33–35

SDS sodium dodecyl sulfate. 20, 21, 27

SEM scanning electron microscopy. 13, 20, 23, 28

SVM support vector machine. 16, 17, 26

UV ultraviolet. 13, 23, 28

Vis visible. 13, 23, 28

vs. versus. 21

XRD X-ray diffraction. 14, 23, 28

Chapter 1

Introduction

Photovoltaics (PV) is one viable option when becoming carbon neutral. Furthermore, it uses the sun's energy directly in contrast to other energy sources (e.g. wind, water or even carbon based) and therefore it is fit to be used in energy harvesting projects like futuristic Dyson spheres [1] which harness the whole power output of the sun. One sort of PV are CIGS (copper indium gallium sulfide) cells [2]. Due to their large absorption coefficient, less material is needed and they can be made thinner and flexible. In order to make a module, multiple cells are operated in series. The cells must be applied to a non-conducting surface. Glass is a good non-conducting substrate, but very rigid and brittle. An alternative is steel, which is ductile, inexpensive and highly available, but conducting. An insulating layer must therefore be applied to the steel substrate before any CIGS cells can be synthesised on top. Polymere would be a obvious choice if not for its low melting point which makes high prodction temperatures impractical. A non-toxic material which is suitable for the insulation is ZrO_2 . An economic and scalable method is doctor blading via a sol-gel process. Sol-gel processes often produce porous layers. In this work a dense, insulating and homogeneous layer is pursued. Machine learning can help to uncover complex non-linear relations, such as the dependence of the thickness and resistance of the resulting layer on production factors. The minimization of the conductance is performed with a particle swarm optimization algorithm.

The remainder of this work is organized with chapter 2 giving some background information on topics of PV, material science and computational science which were used in this work. In chapter 3 and 4 the experimental and computational procedures are described. Section 5 is split into three sections: material specific results are presented in section 5.1, results regarding particle swarm optimization in section 5.2 and further analysis in section 5.3. Finally, chapter 6 summarizes and discusses the outlook and next steps.

Chapter 2

Theoretical Background

This chapter can be broken down into three sections. The first section tries to shine light on the evolution of PV and give some background information on $\text{CuIn}_x\text{Ga}_{1-x}\text{Se}_2$ (CIGS). The middle section reads about material-scientific methods which were used during the practical part of this work. The last and third part focuses on the algorithmic, statistical and analytical methods used to optimize and predict material properties.

2.1 Photovoltaics

The world wide energy consumption has more than doubled between 1970 and 2015 [3] and according to recent studies both fossil [4] and uranium sources [5] will be exhausted within the next 100 years. Even though this time period is not exact and highly dependent on detection methods of resources, this number is rather small and brings us in *zugzwang* to develop sustainable energy sources. One viable option is photovoltaic (PV).

2.1.1 History of Photovoltaics

The photoelectric effect was first described in 1839 by french scientist Alexandre Edmond Becquerel [6], the father of Henri Becquerel (the person after whom the unit is named). Another relevant piece in the PV jigsaw was brought to light with the discovery of photo conductivity of selenium by British engineer Willoughby Smith [7]. In 1876 William Adams and Richard Day [8] showed that the energy of light can be directly converted into electrical energy by a bar of selenium with attached platinum electrodes. And finally, in 1905 Einstein described the physical background of the photoelectric effect with his light quantum theory [9]. In the 1950s the first solar cells (with efficiencies under 10 percent) were used in niche applications such as space flight. Eventually, the interest in photovoltaic and other alternative energy sources rose - fuelled by the oil crisis in 1973 - and the development of photovoltaic for the consumer market was boosted. This development lead to a drop in average price for PV module from \$ 100 per watt in 1975 to \$ 4 per watt in 2007 [10].

2.1.2 Photovoltaics Basics

The process of conversion of photons into voltaic energy can be broken down into two essential steps (both in silicon based PV or in plants): the creation of an electron-hole pair and then its separation by the structure of the material [11]. This means that PV cells are basically diodes, which have a low resistance in one direction and a high resistance in the other direction. This is often achieved through a p-n junction. The n-type side has excess electrons and p-type side has excess electron holes. If the n- and p-type are of the same basis material, they are called homojunctions (e.g. silicon); if not, heterojunctions e.g. CdS/CIGS and CdS/CdTe [12]. The first marketable PV were crystalline silicon photovoltaic modules which still have the biggest market share in the PV segment (including polycrystalline and monocrystalline silicon). In 2022 over 88% of all sold PV were made of silicon (including amorphous silicon) [12]. A comprehensive overview on different PV technologies can be found in [11].

2.1.3 CIGS

CIGS ($\text{CuIn}_x\text{Ga}_{1-x}\text{Se}_2$) is of the chalcopyrite (CuFeS_2) group (tetragonal crystal system). CuInSe_2 and CuGaSe_2 were first synthesised in 1953 by Harry Hahn et al. [13]. The potential use of CuInSe_2 as PV material in combination with CdS was first mentioned in 1974 [14]. Cells with efficiencies of over 10% were achieved already by 1975 [15]. Today, CIGS modules (multiple PV cells in series) reach efficiencies of up to 16% [16], monocrystalline silicon PV modules reach efficiencies of 14-20% and polycrystalline silicon PV modules 12-17% [17].

Just like CdTe , GaAs and amorphous silicon, CIGS has much higher absorption coefficients (lower penetration depth) of visible light than crystalline silicon (see table 2.1). This is due to a direct band gap rather than an indirect band gap (like crystalline silicon). These thin film PVs not only use less material, but also can be used in flexible applications.

Material	Type	Band Gap [eV]	λ [nm]	Abs. coef. α [cm^{-1}]	Penetration Depth [μm]
c-Si	indirect	1.12	600	4000	2.5
c-Si	indirect	1.12	1000	64	150
c-Si	indirect	1.12	1100	3.5	290
a-Si	direct	1.7	600	40 000	0.25
CdTe	direct	1.45	600	37 000	0.3
GaAs	direct	1.42	600	40 000	0.2

Table 2.1: Photonic properties of several established PV materials (data from [18])

Empirical Formula	Name	Band Gap [eV]	Abbreviation
CuInSe_2	copper indium di selenide	1.04	CISe
CuInS_2	copper indium di sulfide	1.5	CIS
CuGaSe_2	copper gallium di selenide	1.7	CIGSe
CuGaS_2	copper gallium di sulfide	1.55	CIGS

Table 2.2: Band gaps of different chalcopyrites (data from [18])

Amazingly, the band gap of CIGS can be varied between 1 eV and 1.7 eV by varying the indium gallium ratio. This is a result of the large difference of band gaps of CuInSe_2 and CuGaSe_2 (see table 2.2). In figure 2.1 the schematic layer sequence of a standard CIGS thin film cell is shown. Typically a 1 μm thick molybdenum layer is deposited on soda lime glass. The sodium in the glass diffuses through the molybdenum layer and increases efficiency and reliability by directing the growth of CIGS in 112 direction [20]. A CIGS layer of 1-2 μm thickness is applied on top via co-evaporation from elemental sources controlled by a mass spectrometer. [20] The p-type doping of CIGS is achieved by adding more than stoichiometric copper to the mix. The heterojunction is then completed by deposition of CdS (typically 50 nm thick). CdS was earlier used as front contact, but now only acts as n-type wide-gap window and buffer. ZnO has a band gap of 3.2 eV and is therefore transparent for visible light. The ZnO window layer (usually of thickness 50-70 nm is highly conductive (especially the Al doped layer) and acts as front contact.

Instead of glass as substrate, steel and plastics (e.g. polyimide [16]) can be used to create flexible CIGS modules. They both come with their own inconveniences. Plastics generally have very low melting points which restrict preparation temperatures. Steel, on the other hand, is temperature resistant enough, but is an electric conductor. The conductance of steel can be masked by a insulator layer (e.g. ZrO_2 , Al_2O_3).

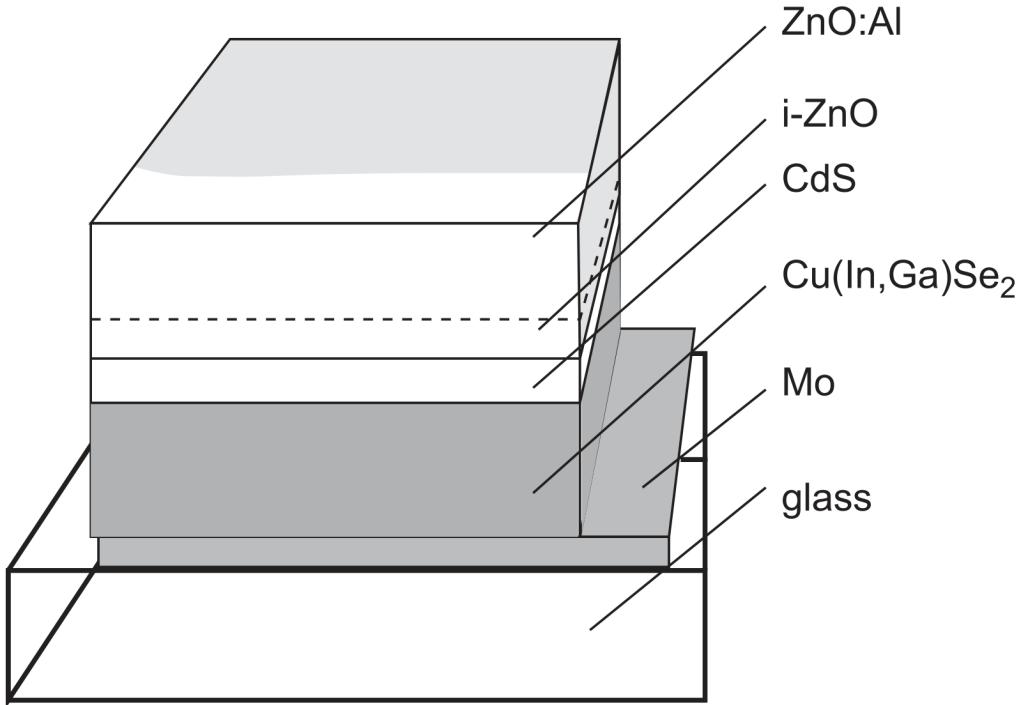


Figure 2.1: Schematic layer sequence of a standard ZnO/CdS/Cu(In,Ga)Se₂ thin-film solar cell. [19]

2.2 Materials and Their Analysis

2.2.1 Zirconium oxide

Zirconium oxide zirconium dioxide (ZrO_2) is a ceramic with a band gap of 5-7 eV [21] and relative permittivity of circa 20 at room temperature [22]. This makes it attractive as an insulator for semiconductor and PV industry. It is monoclinic below 1050 °C, tetragonal between 1170 °C and 2370 °C, and cubic above 2370 °C [23]. The cubic phase can be stabilized down to room temperature by the addition of magnesia (MgO), calcia (CaO) or yttria (Y_2O_3) which avoids mechanical failing due to shrinkage when cooling and undergoing phase transition [23]. It is very resistant to acids (except HF and hot sulfuric acid (H_2SO_4)) and alkalis [23].

2.2.2 Sol-Gel and Doctor Blading

Doctor blading (DB) coating (also known as tape casting) is widely used in the textile, paper, photographic film, printing and ceramic industries. The roll-to-roll compatible process can fabricate highly uniform flat films over large areas [24]. A blade is moved over a substrate spreading a slurry at a fixed distance with a fixed velocity. In roll-to-roll processes the substrate moves instead of the blade.

2.2.3 Sputtering

Sputtering is the processes of highly energetic ions hitting a surface and atoms or molecules being expelled. This is called a physical vapor deposition (PVD) technique in contrast to chemical vapor deposition. PVD can be divided into activation by thermal energy and activation by energetic particle bombardment. Sputtering is of the latter kind, which is advantageous if substrates can't

withstand high temperatures. Sputtering evolved from being a curious experiment in the middle of the 20th century to having various applications in research and engineering. Use cases vary from thin films depositions for PV, for electrical circuits or for storage media such as CDs and DVDs over sputter cleaning and etching to analysis. Advantages of sputter-deposited thin films include good adhesion to the substrate and good step coverage [25].

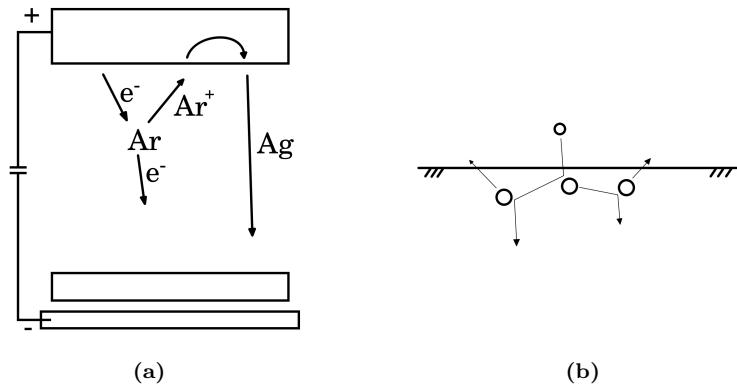


Figure 2.2: Schematic representation of the creation of sputter particles during sputter deposition. (a) creation of ions in the chamber: the top rectangle depicts the source/cathode/target, the bottom rectangle the anode/substrate holder and the middle rectangle the substrate to be sputtered. (b) example series of collision processes leading to sputtering of atoms: horizontal line represents the target surface, the small circle a high energetic ion and the large circles target particles.

A high voltage is applied to two parallel electrodes with low pressure gas in between. The target acts as a feedstock anode and the substrate (holder) as cathode (see figure 2.2a). The anode, then, emits electrons which collide with gas particles (mostly argon because of its inert properties and potential to transfer more kinetic than lighter noble gases). Some gas particles may get ionized by the collision and the gas cations are accelerated to the anode. If a cation has enough energy, it will bump off one or more atoms or molecules from the surface. This happens by a cascade of momentum transfers, which can reach the surface again (see figure 2.2b). If a surface particle obtains momentum pointing away from the bulk and its kinetic energy is higher than the binding energy, the particle is sputtered. This neutral sputtered particle travels – unaffected by the electrical field nearly perpendicular to the surface – towards the substrate and condenses with other particles to form a layer. The pressure in the chamber should be small, such that the sputtered particle has a long mean free path, but on the other hand a minimum pressure is needed to keep the plasma “alive”. Usual pressures buzz around 1 Pa (10^{-2} mbar) or lower [25].

A magnetron can be placed behind the cathode (target) in order to trap the plasma close to the source. This lowers the pressure needed to keep the plasma “alive”. Furthermore, magnetrons prevent high energy electrons from reaching the substrate and undoing the deposited layer. This also increases the probability of an electron colliding with an argon atom and ionizing it, increasing the yield. When oxygen or nitrogen are added to the gas this is called reactive sputtering. Sputtered atoms will react with the gas and result in oxide or nitride layers, respectively. The stoichiometry of the resulting layer can be regulated by gas ratios, but too much reactive gas can lead to target poisoning; meaning that the target is being passivated by an insulating layers which can lead to defects in the growing film [26]. The limitation of only being able to use conducting materials as targets can be circumvented by using a radio frequency electrical field. This prevents charge from building up on the target. Although radio frequency sputtering is more versatile, directional current sputtering is more common because of its simpler system and economical reasons.

2.2.4 Scanning electron Microscopy

The history of scanning electron microscopy (SEM) can be traced back to 1843 when Scottish clockmaker Alexander Bain filed a patent for dissecting an image by scanning. The first SEM was built in 1933, the first images from scanning electron beams were published in 1935 and the first SEM was marketed in 1965. A more detailed history of SEM can be read in an open-to-read paper by McMullan from 1995 [27]. SEM is a microscopical technique which allows visualization of surfaces with features in the nano meter regime. While optical microscopes use visible light and optical lenses, SEM uses accelerated electron beams and electrostatic and electromagnetic lenses. This allows the generation of much more detailed images due to the shorter wavelengths of electrons compared to visible light [28]. The electron beam produces X-rays, elastically backscattered (primary) electrons, inelastic (secondary) electrons and Auger electrons in the examined material. Secondary electrons carry information to conclude morphology and topology of the sample while X-rays can be used to identify the elements. Electrons originate from either a field emission gun (FEG), where a strong electrical field rips electrons from the bulk, or from thermionic guns where a filament (tungsten W or LaB₆ (brighter and longer lasting but more expensive)) is heated until electrons are emitted. Electrons are then accelerated by a voltage of 2 kV to 40 kV and bundled into narrow beams by lenses [29]. A high mean free path is needed for electrons to travel from the source to the sample (and to the detector). Thus, a very low pressure is needed inside the microscope. In this work SEM was used as a preliminary way of checking the quality of ZrO₂ layer.

2.2.5 UV/Vis/NIR Spectroscopy

Ultraviolet (UV)/visible (Vis)/near-infrared (NIR) spectroscopy is a molecular spectroscopic method using interactions of UV, Vis and NIR light (wavelengths of 10⁻⁷ to 10⁻⁶m) with molecules [30]. In general, light is described as periodic electro magnetic (EM) wave which - in vacuum - moves with the speed of light ($c = 299\,792\,458\text{ m/s}$) [31]. In other materials light moves with the velocity $v = \frac{c}{n(\lambda)}$, where $n(\lambda)$ is the refractive index which is dependent on the wavelength λ . The relation of the energy of a photon E , its frequency ν , wavelength λ and wave number $\bar{\nu}$ are as follows:

$$\begin{aligned} E &= h \cdot \nu \\ \nu &= c/\lambda \\ \bar{\nu} &= 1/\lambda, \end{aligned}$$

where h is Planck's constant. In practice the spectrum of EM waves is sectioned into different ranges; from high to low energy: X-ray, UV, Vis, (near, middle and far) infrared (IR), microwaves and radio waves. X-rays interact with core electrons, UVVis with valence electrons, IR with molecular vibrations, microwaves with molecular rotations and radio with electron and nuclear spins. Two prerequisites need to be fulfilled in order for a UVVis photon to be absorbed by a molecule or atom. Firstly, the energy must be the same as the difference between two electronic states, a lower populated and a higher empty orbital. The electron will then be excited into a energetically higher orbital. Secondly, the spin state of both states has to be the same otherwise the transition is "forbidden" and the absorption is extremely unlikely.

IR photons interact with molecular vibrations, which is only a part of the degrees of freedom of a molecule. A molecule has $F = 3N$ (N number of atoms) degrees of freedom, including translational $F_T = 3$ and rotational $F_R = 3$ (2 for linear molecules) movements. The number of vibrations can therefore be calculated as

$$\begin{aligned} F_V &= 3N - F_T - F_R = 3N - 6 \\ F_V &= 3N - F_T - F_R = 3N - 5 \text{ (for linear molecules).} \end{aligned}$$

Vibrations are classified in valence vibrations (change of bond length) and deformation (change of bond angle) vibrations [32]. Only vibrations that change the dipole moment of the molecule are IR active. The absorption of NIR electromagnetic waves involves excitations of non-fundamental vibrations, overtones and combination modes which lead to intricate patterns called the fingerprint region. These transitions are mostly "forbidden" quantum mechanically and allow analysis of bulk matter without extensive preparation due to the deep penetration depth of NIR light [33].

Fourier Transform Spectrometer

In the classic two-beam-spectrometer the light emitted from the light source is split, sent through the sample and the reference and one of the two beams is alternately sent through a monochromator to a detector (often a thermopile). In the Fourier transform (FT) spectrometer the beam is sent through the sample, split and reflected from a static and from a moving mirror, recombined and detected by a photo multiplier (a device which transforms photons into electrical signals). How the two beams will interfere upon recombination depends on the optical path difference (also called retardation) of the two light beams [30]. In a FT spectrometer the reference has to be measured before the sample.

2.2.6 X-Ray Diffraction

X-ray diffraction (XRD) is used to study the crystalline structure of materials. Since X-rays wavelengths (0.2 to 10 nm) are comparable to the interatomic spacing of crystalline solids, the beams get reflected and contain information about the structure [28]. Each crystalline material has a discreet atomic structure, which upon irradiation with X-rays causes constructive and destructive interference according to Bragg's law and generates unique diffraction patterns. XRD diffraction plots of crystalline materials feature distinct peaks, whereas amorphous materials exhibit a broad curve with a maximum extending over several degrees (2θ).

2.3 Machine Learning and Statistics

Some might argue that machine learning (ML) is just overglorified statistics, but that's not the case. Although ML uses several statistical methods, their goal and frame conditions are different. ML tries to predict unseen data points accurately and statistics is a subfield of maths which tries to get insight into a given data. For example, in a statistical model, it is desirable to reduce the number of inputs. This allows a statistician to better study how a change in input variables can directly affect a output variable [34]. More precisely, mathematical statistics (stochastic) [35] is the subject of finding mathematical models to describe the data whereas (classical) statistics is the domain of representing the data. Nonetheless, statistics and maths is at the very basis of ML.

2.3.1 Artificial Intelligence and Machine Learning

Artificial intelligence (AI) is a trans-disciplinary field with roots in logic, statistics, cognitive psychology, decision theory, neuroscience, linguistics, cybernetics, and computer engineering [36]. The history of AI goes back to the middle of the 20th century. Researchers from the emerging field came together at the Dartmouth conference and the term "AI" was coined [37]. AI's history is beautifully depicted by McCorducks' 1982 book "Machines Who Think" [38, 39], which focuses on the great minds behind the advances. Pioneers like Alan Turing thought a lot about how to define, test and implement AI [36]. One example how to measure AI is to let it play chess against a human [40] (in 1997 a chess computer called Deep Blue won against the World Chess Champion Garry Kasparov for the first time [41]). Another test *ingenious* by Alan Turing is the imitation game [42], nowadays known as Turing test: an interrogator communicates with two unknown entities **A** and **B** (a woman and a man) and must find out who is who. **A** will try to make to interrogator misjudge, whereas **B** is on the interrogators side. The question is if **A** is replaced with a computer how the ratio of outcomes would deviate from the original ratio. At the moment it's hard to imagine a computer getting a higher ratio than a human but when reading AI written articles [43], it's easy to see this test being passed in the near future; especially with ChatGPT's convincingly human-like answers [44].

But fear not, that doesn't mean that computers are more intelligent than humans or sentient [45, 46] even though some claim [47] and it certainly does not mean that research is over. AI is still a young field, which is strongly growing and is gaining ubiquitous status. It is slowly creeping into every aspect of modern human life just like electricity around one hundred years ago. Realms in which AI is gaining traction are: playing board games (and beating humans) [40, 41, 48], image recognition (very popular for medical diagnosis) [49–52], chemistry [53–55], cyber security [56],

facial recognition (to prevent theft of toilet paper [57]), financial sector (as robo-advisors) [58], natural language processing (NLP) [59–61] (which can also create code and pictures through scalable vector graphics (SVG)) and even creative tasks like creating non existing faces [62], create graphic artwork (DALL-E 2) [63] or making video games [64]. It is nearly hard to find a field where AI isn't used in some way. This steady incorporation of AI leads to the so called AI effect [38, 65]: certain fields get incorporated into AI research and practice, such that after some time of general use it is no more considered AI (e.g. spam filter or web searches). Google CEO Sundar Pichai even goes as far and said: "AI is one of the most important things humanity is working on. It is more profound than [...] electricity or fire" [66].

2.3.2 Machine Learning Methods

ML is at the base of most AIs. It is an umbrella term for programs with instructions to learn from data, i.e. gain knowledge, categorize, predict and make decisions based on data. There is a platitude of different machine learning methods: they can be divided into supervised (training set is labeled) and unsupervised (exploratory). An orthogonal division can be made by regression (continuous data) versus classification (discrete, categorical data). Independent from these 2×2 categories there are multiple ways to let machines learn from data. Artificial neural networks (ANN) (one of the most popular architectures for big data [67]) are loosely modeled after the brain [68]. Artificial neurons (also called nodes), which are arranged in layers, are connected to each of the neurons of previous and next layers and the weights (parameters of intensity), with which the data is routed from one neuron to another, are optimised during training. Convolutional-layer ANN excel in picture recognition [69] and are useful in quantum mechanics too [70]. Other common methods include linear regression, kernel ridge regression and support vector regression.

Linear regression is one of the simplest methods to predict data. It persuades by its computational simplicity and easy interpretation.

$$\mathbf{y} = \mathbf{X}\mathbf{k} + c \quad (2.1)$$

Where $\mathbf{y} \in \mathbb{R}^n$ (n is the number of data points) is a vector of dependent variables which shall be predicted and $\mathbf{X} \in \mathbb{R}^{n \times d}$ (d is the the number of independent variables) is a matrix of independent variables. The parameters $\mathbf{k} \in \mathbb{R}^d$ and $c \in \mathbb{R}$ are chosen by minimizing an objective function (loss function in ML chargon). A typical function to minimize are the L_1 and L_2 error, i.e. mean absolute error (MAE) and mean squared error (MSE), respectively.

$$L_1 = \sum_i |y_i - \hat{y}_i| \quad (2.2)$$

$$L_2 = \sum_i (y_i - \hat{y}_i)^2 \quad (2.3)$$

Ridge Regression is like linear regression with an extra term which penalizes steep regression functions. The extra term reduces overfitting and is scaled by a correction parameter α . The larger α is, the larger is the regularization and the flatter is the regression functions. When α tends to infinity, we get an intercept-only model. When α is zero, the regularized loss function becomes the L_2 loss function.

$$L_{RR} = L_2 + \alpha \sum |k_i| = \sum (y_i - \hat{y}_i)^2 + \alpha \sum |k_i| \quad (2.4)$$

Kernel ridge regression (KRR) combines ridge regression with the kernel method. A kernel transforms data in such a way that a linear hyperplane (a point in one dimension, a line in two dimensions, a plane in three dimensions) can fit data in regression problems or separate the data in classification problems without actually doing the transformation for every data point, which lowers computational costs. A kernel is some kind of similarity measure which fulfills the requirements of nonnegativity, symmetry and linearity [71]. The following equations show definitions for linear (eq. 2.5), polynomial (eq. 2.6), sigmoidal (eq. 2.7) and radial basis functional kernel (eq. 2.8), with γ as fixed hyperparameter and c_0 as parameter to optimize. A requirement for using kernels is having a dot product in the loss function. This can be accomplished by expressing \mathbf{k} in terms of \mathbf{X} : $\mathbf{k} = \mathbf{X}^\top \mathbf{r}$ [72].

$$k_{lin}(\mathbf{x}, \mathbf{y}) = \mathbf{x}^\top \mathbf{y} \quad (2.5)$$

$$k_{pol}(\mathbf{x}, \mathbf{y}) = (\gamma \mathbf{x}^\top \mathbf{y} + c_0)^d \quad (2.6)$$

$$k_{sig}(\mathbf{x}, \mathbf{y}) = \tanh(\gamma \mathbf{x}^\top \mathbf{y} + c_0) \quad (2.7)$$

$$k_{rbf}(\mathbf{x}, \mathbf{y}) = \exp(-\gamma \|\mathbf{x} - \mathbf{y}\|^2) \quad (2.8)$$

Support vector machine (SVM) is a versatile machine learning algorithm first mentioned in 1992 [73]. The SVM was initially developed by Vladimir Vapnik for the binary classification of separable data, then improved to handle non-separable data and eventually adapted to solve regression problems. The concepts of SVM will be discussed in the same chronological order. Classification works by spanning a hyperplane between two linearly separable categories in a way such that the closest points from each category have the largest distance to the hyperplane. The distance from the points to the hyperplane is called margin τ . The points with the shortest distance to the hyperplane are called support vectors and are used to define the hyperplane. A SVM avoids overfitting by only using a subset of the data - the support vectors - to fit the model. The goal is to find the decision boundary which correctly classifies all samples with the biggest margin. The decision boundary can be expressed as an hyperplane

$$\hat{\mathbf{y}} = h(\mathbf{X}) = \mathbf{X} \cdot \mathbf{w} + b \quad (2.9)$$

with $\mathbf{y} \in \{+1, -1\}^n$, $\hat{\mathbf{y}} \in \mathbb{R}^n$, $\mathbf{X} \in \mathbb{R}^{n \times d}$ and $\mathbf{w} \in \mathbb{R}^d$. The constraint of the positive and negative support vectors (\mathbf{x}^+ and \mathbf{x}^- , respectively) satisfying

$$\mathbf{x}^+ \cdot \mathbf{w} + b = 1 \quad (2.10)$$

$$\mathbf{x}^- \cdot \mathbf{w} + b = -1 \quad (2.11)$$

can be generalized to

$$y_i(\mathbf{x}_i \cdot \mathbf{w} + b) \geq 1 \quad (2.12)$$

where y_i being the labels of the training data. The width of the margin can be inferred by projecting the vector spanning between two support vectors on opposite sides of the decision boundary onto the unit vector perpendicular to the hyperplane. If we now take a vector from a positive support vector \mathbf{x}^+ to a negative support vector \mathbf{x}^- and project it onto the unit vector of \mathbf{w} (which is perpendicular to the hyperplane), we get the width of the margin.

$$\begin{aligned} 2\tau &= (\mathbf{x}^+ - \mathbf{x}^-) \cdot \frac{\mathbf{w}}{\|\mathbf{w}\|} = (\mathbf{x}^+ \cdot \mathbf{w} - \mathbf{x}^- \cdot \mathbf{w}) \cdot \frac{1}{\|\mathbf{w}\|} \\ &= ((1 - b) - (-1 - b)) \cdot \frac{1}{\|\mathbf{w}\|} = \frac{2}{\|\mathbf{w}\|} \end{aligned} \quad (2.13)$$

Thus maximizing the margin is equivalent with minimizing $\|\mathbf{w}\|$ and minimizing $\frac{1}{2}\|\mathbf{w}\|^2$ (mathematical convenience for further steps). By incorporation the constraint (eq. 2.12) via the Lagrangian multiplier method we mathemagically arrive at the loss function which should be maximized:

$$\mathcal{L} = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_i^n \alpha_i[y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1] \quad (2.14)$$

which can be rewritten by some mathematical wizardry (setting the partial derivatives of the Lagrangian function $\frac{\partial \mathcal{L}}{\partial \mathbf{w}}$ and $\frac{\partial \mathcal{L}}{\partial b}$ to zero and inserting into eq. (2.14)) [74, 75]:

$$\mathcal{L} = \sum_i^n \alpha_i - \frac{1}{2} \sum_i^n \sum_j^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j \quad (2.15)$$

Soft margin SVM – in contrast to hard margin SVM – is used if the data is non-separable due to outliers [76]. Such data can be handled by introducing a penalization term for wrongly categorized samples into the loss function.

$$L_{SM} = \frac{1}{2}\|\mathbf{w}\|^2 + \sum_i^n \max(0, 1 - y_i(\mathbf{w} \cdot \mathbf{x}_i + b)) \quad (2.16)$$

If the prediction $\mathbf{w}\mathbf{x}_i + b$ and the true category y_i don't agree, they have opposite signs and their product will be a negative number. The subtraction of a negative number will result in a positive penalization. If the sample is correctly predicted, the product will result in a positive number, the subtraction will lead to a negative number and the maximum will be 0, if the sample is outside of the margin. The function we want to optimize (eq. 2.15) and its soft margin equivalent (eq. 2.16) can both be expressed as inner products. This allows us to separate data not only by hyperplanes but also by complex decision boundaries due to the kernel trick. Again, a kernel $K(\mathbf{x}_i, \mathbf{x}_j) = T(\mathbf{x}_i)^\top \cdot T(\mathbf{x}_j)$ allows to calculate the innerproduct of two vectors in a transformed space without the need of transforming each vector, which turns out to be computationally much cheaper.

SVMs can also be used for regression (support vector regression). The decision function becomes the regression function and the margin includes all data points instead of none. In soft margin support vector regression a few outliers are allowed. Non-linearity can be likewise introduced via the kernel trick.

2.3.3 Population Based Optimization Algorithms

There are also lesser known ML algorithms like evolutionary algorithms (e.g. genetic algorithm (GA) and particle swarm optimization (PSO)). Evolutionary algorithms take advantage of the ability to cope with local optima by evolving several candidate solutions simultaneously [77]. One particular feature of evolutionary algorithms is that they start with a small data set and periodically request new data in order to solve the problem iteratively.

A GA is a search algorithm that uses principles of natural selection and genetics to optimize a search space. A GA starts with a population of randomly generated solutions, or chromosomes, and then proceeds to breed them together to create new solutions. The new solutions are then tested for fitness, and the best solutions are selected to create the next generation of chromosomes. This process is repeated until a satisfactory solution is found.¹

A PSO also uses a starting population of particles where each experiment (particle) is represented by its independent and dependent variables. It was originally inspired by the behavior of bird flocks and fish schools [77, 78]. Each particle has an associated position and velocity. Every movement across the search space is, additionally to a stochastic term, influenced by its particle velocity and position as well as its and the swarm's best visited position.

2.3.4 EMMA

Evolutionary model-based multiresponse approach (EMMA) is an implementation of PSO in the R programming language. Each time step the dependent variables for all possible input variable combinations are predicted with the help of multivariate adaptive regression splines (MARS). The MARS regression function (RF) is then used to choose the next position for each particle. MARS is a regression method introduced by Friedman's 1991 paper "Multivariate Adaptive Regression Splines" [79]. The CRAN package which implements MARS is called `earth` due to MARS being trademarked [80]. MARS is an extension of multivariate linear regression. Friedman presented MARS as an alternative to piecewise polynomials (splines) and local averaging methods (e.g. kernel functions), whose number of parameters quickly exceeds the practical for moderate dimensional problems ($n > 5$). The main advancements of MARS is the ability to fit (possibly complex) interactions and non-linearities without losing interpretability. The RF is developed in a forward/backward stepwise recursive manner [81]. During the forward recursion terms are added in pairs until a certain number of terms is reached. Each pair consists of two hinge functions (also called rectifier functions) multiplied with a term already part of the RF (including the constant term). The maximum degree of interaction is 2 for EMMA, but can be varied. This means that only two basis functions can be multiplied (excluding the constant term) to form a subsequent term. Pairwise added hinge functions are of the simple form $h(x - c)$ and $h(c - x)$ where $h(e) = \max(0, e)$ and with x being an independent variable, c being a constant and e being any expression. Then, the backwards algorithm regularizes the function by removing individual terms.

¹This paragraph was written by GPT [60] given the input "Introduction to genetic algorithms: "

The metric which decides if a term should be removed is called generalized cross validation (GCV) and was introduced by Wahba and Cravenin 1969 [82].

$$GCV(M) = \frac{1}{N} \sum_{i=1}^N \frac{(y_i - \hat{f}(x_i))^2}{\left(1 - \frac{C(M)}{N}\right)^2} \quad (2.17)$$

with M being the number of terms, N the number of data points, a correction term

$$C(M) = (d + 1)M + 1 \quad (2.18)$$

and penalty is $d = 3$ for interactions larger than 1 and $d = 2$ otherwise. The variable of the hinge functions and its knot location at the forward step and which terms to delete at the backwards step are selected by minimizing GCV. The coefficients for each term are then chosen via regular MSE minimization [81].

2.3.5 Design of Experiment

"The real purpose of experiment design is to maximize the information content of the data within the limits imposed by the given constraints." - Grahem C. Goodwin [83]

In two cases a deliberate design of experiment (DOE) is especially beneficial. (1) If the query of a new data point is very expensive, it is favourable to actively chose the query (e.g. drilling for oil or quantum chemical calculations). (2) If the query space is so vast, that randomly querying might explore domains which might lead to uninteresting or even misleading information. At the beginning of any experiment its constraints must be determined. Constraints for a given experiment include range of input and output variables as well as total time available and total number of samples/experiments that can be taken. [83]

A naive approach to an experiment design is the full factorial DOE. Each possible combination of discrete values is tested. While this is the most informative design it will more often than not be infeasible due the curse of dimensionality [75]. The 2-level factorial design provides an alternative with 2^d experiments (where d is the number of independent variables). Drawbacks of 2-level factorial DOEs include no data about the inside of the search space and infeasibility for high dimensional problems. In a full factorial or 2-level factorial design most experiments are redundant and most resources will be spent exploring high-order interaction effects [84], which are often minimal to non-existent. In order to overcome these obstacles a certain number of experiments can be chosen randomly from the search space. When a subset is chosen from the factorial design, it is called a fractional factorial design. The Plackett-Burman (PB) [85–87] design is a special case of 2-level fractional factorial design, where the number of needed experiments n is $n < d + 4$ (more precisely $n = (\lfloor d \div 4 \rfloor + 1) \cdot 4$, where $\lfloor x \rfloor$ denotes the floor function on x). The PB design ensures that each combination of levels for any pair of factors appears the same number of times. A drawback of 2-level factorial (incl. PB) and random fractional designs is that the sample set is likely not evenly distributed across the search space [88]. The Hammersley design [88, 89] is based on the Hammersley sequence and produces space filling data points. The Latin hypercube DOE [88, 89] is a type of orthogonal DOE, which has the feature that each level for each variable will be tested only once. A Latin hypercube DOE can be also created such that data points distribute more uniformly over the search space. Latin hypercube DOEs are mainly used in computer simulations which are purely deterministic and therefore are very precise.

2.3.6 Analysis of Variance

Analysis of Variance (ANOVA) is a statistical test for estimating the influence of multiple categorical independent variables on a dependent numerical variable. At the heart of ANOVA lies the F-test. The F-test was *ingeniered* by Ronald Fisher [90], an important figure in modern statistics. The F-test uses a ratio of variances to determine if a null-hypothesis (observed difference is due to chance alone) is true. The variance in the numerator measures the "between-group variability" and the denominator measures the "within-group variability". This ratio is unaffected by units, scaling errors and constant bias. Additional assumptions of ANOVA include: groups and levels should be independent, residual error should follow normal distribution and variance within groups should be equal (homoscedasticity).

Chapter 3

Experimental

In this section the used chemicals and substrates, experimental procedures and any used equipment are described. The experiments can be split into three sections: In the first section the base recipe and the base process were sought. In the pre-optimization section the boundaries for the optimization were investigated. And during the third and final step the experiments for the optimization were performed.

3.1 Substrate Preparation

Five different substrates were used throughout this work: microscope glass slides ($2.5\text{ cm} \times 7.5\text{ cm}$), squared glass plates ($2.5\text{ cm} \times 2.5\text{ cm}$), indium doped tin oxide (ITO) glass plates ($2.5\text{ cm} \times 2.5\text{ cm}$), fluorine doped tin oxide (FTO) glass plates ($5\text{ cm} \times 5\text{ cm}$) and steel foil ($10\text{ cm} \times 10\text{ cm}$) provided by Sunplugged GmbH (<http://sunplugged.at/>). Glass was used because of its availability and to measure transmission and reflectance spectra. FTO and ITO were used in order to have a conducting substrate, which is needed to measure the conductivity of the applied layer and for SEM micrographs. The glass slides and FTO were scored with a diamond scribe and broken with a running plier into pieces of dimensions $2.5\text{ cm} \times 2.5\text{ cm}$. The steel foil was cut with a foil cutter into squares of the same dimensions. All substrates were cleaned in three steps before usage:

1. 15 min in 50 mL deionized water (H_2O) and 1 mL of Hellmanex III alkaline concentrate in a sonic bath
2. 15 min in H_2O in a sonic bath
3. 15 min in Propan-2-ol (IPO) in a sonic bath

After the last cleaning step, the samples were blown dry with dry nitrogen (N_2) gas and kept in a clean plastic container until the doctor blading step.

3.2 Solutions

All recipes for solutions can be divided into two categories: the first recipe - adopted from Anwar et al. [21] - was based on zirconium(IV)propoxide ($\text{Zr}(\text{PrO})_4$) in acetylacetone (Hacac) and H_2O (aqueous solution). The second recipe - adopted from Hu et al. [91] - was based on $\text{Zr}(\text{PrO})_4$ in Butan-1-ol (BuOH) (buthanolic solution).

3.2.1 Aquatic solution

$\text{Zr}(\text{PrO})_4$ was added to Hacac while stirring and in a separate vessel H_2O (including any optional additives such as sodium dodecyl sulfate (SDS), hydrochloric acid (HCl), H_2SO_4 or sodium hydroxide (NaOH)) was added to IPO and both were stirred for one hour. These additives were added to influence pH and surface tension with the goal to influence the resulting layer. The H_2O -IPO mixture was added to the other solution and stirred over night. The exact volumes can be taken



(a) Aquatic solution



(b) Buthanolic solution

Figure 3.1: Aquatic and buthanolic solution with magnetic stirring bars in beaker glass sealed with Parafilm

from table 3.1. Unfortunately, this solution failed to produce anything near homogeneous layers. Thus, an alternative solution was found.

Table 3.1: Compositions of different aquatic solutions

recipe	1	2	3	4	5	6	7
Zr(PrO) ₄ [mL]	8	8	8	8	8	8	8
Hacac [mL]	8	8	8	8	8	8	8
IPO [mL]	2	2	2	2	2	2	2
H ₂ O [mL]	2.6	2.6	2.5	2	2	2	2
SDS [mg]	-	5.9	-	-	-	-	-
HCl [mL]	-	-	-	-	0.5	-	-
H ₂ SO ₄ [mL]	-	-	-	-	-	0.5	-
NaOH [mL]	-	-	-	-	-	-	0.5

3.2.2 Buthanolic solution

Five different concentrations of the buthanolic solutions were prepared. The one-fold concentrated solution (1F) was closest to the recipe proposed by Hu et al. [91]. The other four solutions (two-fold concentrated solution (2F), three-fold concentrated solution (3F), four-fold concentrated solution (4F) and five-fold concentrated solution (5F)) were similar with higher concentrations of Zr(PrO)₄ (see table 3.2) with the aim of producing thicker layers. Not all chemicals mentioned by Hu et al. [91] were available, so chemically similar starting materials had to be chosen from the available inventory. Different solvents (Butane-1,2-diol, BuOH and Propan-1-ol) and chelating agents (Hacac and citric acid) were tested and later in the process - just before the PSO started - the stabilization compound (glacial acetic acid (AcOH)) was changed to IPO. The most promising combination of solvent, chelating agent and stabilization reagent was BuOH, Hacac and IPO, respectively, which was used for the final optimization. The main differences between the used recipe and the recipe mentioned by [91] were precursor (transition metal complex Zr(PrO)₄ versus (vs.) post-transition metal complex aluminium isopropoxide), solvent (BuOH vs. 2-ethoxyethanol), reaction temperature (room temperature vs. 105 °C), stirring time (15/15/30 min vs. 30/30/120 min), application process (doctor blading vs. spin coating), the heat treatment between layers (200 °C vs. 200 °C and then 400 °C) and the calcination temperature (400 °C vs. 500 °C).

The solvent (BuOH) was put into a beaker glass (or similar, preferably with an air-tight cap) with a magnetic stirring bar and Zr(PrO)₄ was added while stirring. After stirring 15 min one mole equivalent chelating agent (Hacac) was added and stirred for another 15 min. Finally, the

Table 3.2: Composition of different buthanolic solutions

recipe	1F	2F	3F	4F	5F
BuOH [mL]	4.95	4.9	4.85	4.8	4.75
Zr(PrO) ₄ [mL]	0.05	0.1	0.15	0.2	0.25
Hacac [mL]	0.0125	0.025	0.0375	0.05	0.0625
IPO/AcOH [mL]	2	2	2	2	2

stabilization solvent [91] (IPO or AcOH) was added to the mixture and stirred for additional 30 min. In order to make a 2F solution, the volume of Zr(PrO)₄ and Hacac was doubled and the volume of BuOH was decreased by the increase of volume of Zr(PrO)₄.

3.3 Doctor blading

All glass substrates were doctor bladed manually with a smooth stainless steel wire bar coater. On two opposing edges adhesive tape was applied to create a valley in between. After the layer was applied and dried the tape was removed and the substrate treated with heat. Lower DB velocities resulted in less homogeneous layers. Thus, the DB velocity was not altered in the manual DB process. One and two layers of adhesive tape were used to alter the depth of the valley.

All steel substrates were doctor bladed with an Erichsen Coatmaster 510 film applicator with a heatable plate. The rest was doctor bladed on a glass plate, initially fixed with double sided adhesive tape and partially heated with a heat gun. The blade height was varied from 0 mm to 0.35 mm in 0.05 mm steps and did not substantially alter the results. For the rest of the experiments a blade height of 0.2 mm was used. Eventually, a heatable vacuum plate was used. It had equally spaced circular 2.5 mm diameter patches of porous metal where underpressure could be applied to keep the substrate in place (see figure 3.2). Most of them were covered with tape to increase the suction intensity at the remaining ones. After setting the temperature of the heating plate to 200 °C, the temperature of the vacuum plate (room temperature, 40 °C, 50 °C, 60 °C, 70 °C or 80 °C) and the DB velocity (10 mm/s, 12 mm/s, 14 mm/s, 16 mm/s, 18 mm/s or 20 mm/s), the blade was put into its initial position, the sample was placed on the vacuum plate and the vacuum was switched on. During pre-optimization samples, lower DB velocities (0.1 mm/s, 0.5 mm/s, 1 mm/s, 2 mm/s and 5 mm/s) were tested. 100 µL of solution were applied with a 10 - 1000 µL pipette and the blade moved over the sample distributing the liquid evenly. After evaporation of the solution, the vacuum was turned off, the 'blade pusher' was put into initial position, the blade was removed and excess solution was removed from the plate with a wipe. The doctor bladed substrate was transferred to the 200 °C heating plate and rested on there for 5 min. This process of applying a ZrO₂ layer was repeated as many times as needed.

3.4 Calcination

A LabTech EH45C heating plate and a Naberterm LB410 muffle furnace were used to calcinate the doctor bladed samples. The heating plate could hold temperature for a certain amount of time, but heated with a fixed rate of circa 10 °C/min. In order to achieve a lower overall heating rate several temperature ramps and plateaus were alternated (see table 3.3a). This heating procedure was called HP1. The HP1 procedure was optimized for the available hardware by a colleague working on the project prior to the author.

The NT1 heating program was used to mimic the HP1 heating procedure from the heating plate in the Naberterm muffle furnace. NT2 is an simplification of NT1 and programs NT3 - NT6 are the same as NT2 with altered heating rate and NT5 additionally used a calcination temperature T_{Cal} of 500 °C. NT2-NT6 had 2 variables (heating rate and one calcination temperature) in contrast to NT1 which had 4 (three different heating rates and calcination temperature). All heating programs were held at the calcination temperature for one hour.



(a)

(b)

Figure 3.2: (a) Temperature regulator (1) on the left, vacuum pump (2) in the background and Erichsen Coatmaster 510 film applicator (3) with heatable vacuum plate. (b) Close up of the DB blade in position with blade height adjusted to 0.2 µm. The majority of the suction areas is sealed with tape to increase the underpressure at the remaining ones.

Table 3.3: Heating programs

(a) Heating program HP1 used with the Labtech EH45C heating plate

T [°C]	80	100	150	160	170	180	190	200	250	300	350	400
t [min]	10	10	5	5	5	5	5	10	10	10	10	60

(b) Heating programs NT1 - NT6 used with the Naberterm LB410 muffle furnace

Name	80-150°C [°C/min]	150-200°C [°C/min]	200°C-T _{Cal} [°C/min]	T _{Cal} [°C]	t _{Cal} [min]
NT1	2		2	400	60
NT2	2	2	2	400	60
NT3	3	3	3	400	60
NT4	4	4	4	400	60
NT5	4	4	4	500	60
NT6	1	1	1	400	60

3.5 Characterization

All SEM micrographs were taken with a Zeiss Supra 40. FTUV/Vis/NIR transmittance and reflectance spectra (at 20° incident angle) were recorded with a Bruker Vertex 70 spectrometer with a quartz beam splitter, 0.5 mm aperture and Gallium-Phosphide detector for ultra violet light (303 nm–588 nm) and a Silicon detector for visual and NIR light (500 nm–1.2 µm). For transmittance the light entered the sample from the side with the layer. The UV and Vis/NIR spectra were merged in Opus software. XRD spectra were obtained with a Thermo Scientific ARL Equinox 100 X-Ray Diffractometer. All XRD spectra were taken at 5° incident angle and compared to the internal database.

The current-voltage (I-V) curves were measured with Agilent 4156C Precision Semiconductor Parameter Analyzer from -0.5 V to 0.5 V with steps of 10 mV. Prior to I-V measurements the samples were sputtered with aluminium through a mask to produce a multitude of equidistant contacts with a Leybold UNIVEX450C Sputter System. Directional current sputtering was used with Argon as inert gas at 0.005 mbar and with a power of 40 W for 700 s. Between sputtering and I-V measurements direct contacts to the steel substrate were created by removing a patch of the ZrO₂ layer with sand paper on two opposing edges of a sample and then by applying silver paste. See figure 3.3 for a sketch of the connectivity.

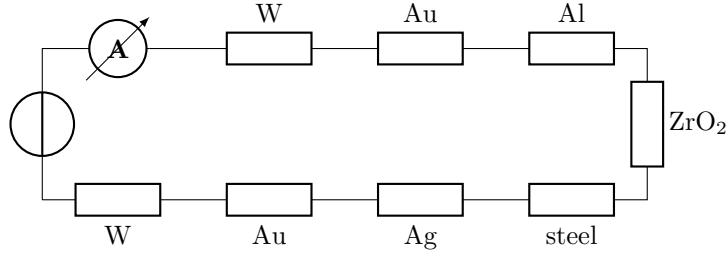


Figure 3.3: Sketch of circuit, from the source clockwise: varibale voltage source, built-in amperometer, gold plated tungsten probe, sputtered aluminium contact, ZrO_2 layer to be measured, steel substrate, dried silver paste contact, gold plated tungsten probe.

3.6 Pre-optimization

The main part of all samples (around 90%) were pre-optimization samples, from which the bulk were screening experiments optimized via biotic reinforcement learning (i.e. my brain) in order to obtain reasonable constraints for computerized pre-optimization. Composition of the aquatic solution, heating program, blade distance (distance between blade and substrate) and layer count were varied but no homogeneous layers resulted. Thus, the buthanolic recipe was introduced, which gave rise to homogeneous films. After the blade moved over the substrate the remaining liquid needed about a minute to evaporate. The then current procedure resulted in clear and continuous films, but was plagued from inhomogeneities due to drying stains. The drying stains could be circumvented by using a heat gun, but the results were not reproducible. The glass plate of the film applicator was therefore exchanged with a metal plate which allowed to hold the sample in place through underpressure and simultaneously heat it to a certain temperature. The bounds of the process variables were then explored in a preliminary study using the Plackett-Burman [92] design implemented in the python3 library pyDOE. With (1,5), (4,10), (400,500), (120,480), (0.1,5) and (20,80) as nominal and extreme values for relative concentration of Zr c_{zr} , number of layers λ , calcination temperature $T_{\text{cal}}[\text{°C}]$, heating rate $v_{\text{cal}}[\text{°C/h}]$, DB velocity $v_{\text{DB}}[\text{mm/s}]$ and DB temperature $T_{\text{DB}}[\text{°C}]$, respectively. After testing the first samples, the lower limit for DB velocities was altered from 0.1 to 1 (see table 3.4a).

Table 3.4: Pre-optimization experiments

(a) Plackett-Burman style experiments						(b) Hand picked experiments; *varied stabilization agent					
c_{zr}	λ	v_{DB}	T_{DB}	v_{cal}	T_{cal}	c_{zr}	λ	v_{DB}	T_{DB}	v_{cal}	T_{cal}
1	10	5	20	120	400	2	8	0.5	40	360	470
1	4	0.1	20	120	500	2	6	2	40	360	430
5	10	0.1	20	120	500	1	4	12	70	120	500
5	4	5	20	480	400	1	9	18	80	240	400
1	5	5	80	120	500	4*	6	14	60	240	500
1	10	1	80	480	400						
5	5	1	80	120	400						
5	10	5	80	480	500						

Some additional hand picked experiments (see table 3.4b) were introduced to further narrow down the limits of the optimization as every reduction in variables or levels meant a faster convergence. Shortly before the PSO the stabilization agent was changed from AcOH to IPO because of much better solution stability. The - up to this moment - best process was chosen to be tested with 1 mL AcOH, 0.5 mL AcOH plus 0.5 mL IPO and 1 mL IPO. The sample produced with IPO as stabilization agent showed comparable passivation results and better stability. Therefore IPO was used as stabilization in further experiments. After the boundaries of the solution were fixed the process variables to produce an insulating layer was examined. See table 4.1 for the input space spanned during the optimization.

Chapter 4

Computational Details

4.1 Data Processing

Multiple I-V curves were measured for every sample. The main difficulty faced in processing the data was to present the measurements obtained from a sample in one representative number. For every I-V curve (i.e. for every aluminium dot contact) the gradient g at $V = 0$ (conductance) is calculated by taking 5 points before and 5 points after the origin with at one point distance from zero (which boils down to the data points from ± 0.02 V to ± 0.07 V), calculating averages \bar{V} and \bar{I} of each point cloud and taking the ratio of differences:

$$g = \frac{\bar{I}_+ - \bar{I}_-}{\bar{V}_+ - \bar{V}_-} \quad (4.1)$$

The average of each conductance would be an easy and naive choice but difficult to represent a sample correctly since the possible values of conductance span across several magnitudes. So the average of base 10 logarithm of conductances is the next nearby *ansatz*. In order to make the metric closer MSE (and to penalize deviation from ideal more than linearly) the average of the squared difference to the ideal non-conduction case ($g = 10^{-13}$) was chosen instead of the average of logarithms and shall be called conductivity hereafter for lack of a better term.

$$\gamma = \sum_i^N \frac{(-\log_{10}(g_i) - 13)^2}{N} \quad (4.2)$$

Another measure is the density of shorted species ρ (pin hole density), calculated in following way:

$$s_i = \begin{cases} 1 & \text{if } -\log_{10}(g_i) < 5 \\ 0 & \text{if } -\log_{10}(g_i) \geq 5 \end{cases} \quad (4.3)$$

$$\rho = \sum_i^N \frac{s_i}{N} \quad (4.4)$$

4.2 Sample Selection

An evolutionary approach was chosen, namely a multi-objective PSO algorithm called EMMA [77, 78, 93, 94]. Initially, the input parameters (independent variables), their boundaries and number of equidistant levels for each parameter are declared (see table 4.1). Next, the output variables (dependant variables), their weights in the objective function (the function which should be optimized) are specified and if they should be minimized or maximized is noted.

The *optimizands* where chosen to be γ , ρ , λ and v_{cal} , where γ , ρ and λ are to be minimized and v_{cal} to be maximized. The next step is to generate an initial population (ensemble of experiments), which is chosen randomly from the population space. The experiments are executed, measured and

c_{zr} [22 mmol/L]	λ	T_{DB} [°C]	v_{DB} [mm/s]	T_{cal} [°C]	v_{cal} [°C/h]
2	4	40	10	300	120
3	6	50	12	400	360
4	8	60	14	500	600
5	10	70	16		840
	12	80	18		1080
			20		

Table 4.1: Discrete levels of each input parameter

evaluated according to section 3 and the responses conductivity γ (see eq. 4.2), pin hole density ρ_s (see eq. 4.4), λ (numbers of layers) and v_{cal} (heating rate of calcination process in °C/min) are supplied to the program. The program, then, uses this data to estimate a response for each output variable for the entire population space (and to choose a fraction of the initial population which is allowed to propagate). The current population - each of the particles independently - are given a random velocity vector associated to them which is directed to the best predicted sample. The population for the next time step is generated and the experiments are again executed, measured and evaluated. Each time step thus includes experiment identification, solution preparation, coating deposition and layer analysis.

4.3 Fitting via Machine Learning

Python 3.10.9 and sci-kit learn 1.1.2 [95] were used to implement a linear fit model, KRR and SVM with polynomial, radial basis function (RBF) and sigmoid kernels. A grid search over a set of hyperparameters was executed. The hyperparameters which were examined in KRR were kernel, γ_{ml} and α . The kernel took the form of RBF, sigmoid or polynomial with degree 1–3. The parameters α and γ_{ml} (subscript to distinguish from *optimizand*) took the values 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1, 2, 5 and 10. Note that γ_{ml} is used differently in each kernel. The kernels for SVM were the same as for KRR. Further hyper parameters were C , ε and γ_{ml} , where C and ε took the values 0.05–1 with step size 0.05 and was the same as for KRR plus **scale**. Where **scale** equals to $\frac{1}{d \cdot \sigma(\mathbf{X})}$ with d being the number of features and $\sigma(\mathbf{X})$ the variance of the input variables. The input variables (c_{zr} , λ , v_{DB} , T_{DB} , v_{cal} , T_{cal}) of the EMMA data set were scaled before each calculation. For each calculation the 5-fold cross validation score was calculated. The score equals to the coefficient of determination R^2 which defined the following for non linear regression functions:

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)}{\sum(y_i - \bar{y}_i)} \quad (4.5)$$

For each method (KRR and SVM) and each *optimizand* (γ and ρ) the parameters with the highest score were noted. Finally, the different data sets were predicted for each model and MAE and MSE were calculated. These values were also calculated for EMMA and the linear model in order to compare the different methods.

Chapter 5

Results and Discussion

In this section the results will be presented, explained and discussed. Starting with physical measurements and material properties. Followed by the presentation of the results obtained by optimization and exploration with the EMMA method. Finally, the results of alternative optimization methods were let loose on the data and compared with the EMMA method.

5.1 Material Scientific and Pre-optimization

As already indicated in the experimental section the aquatic recipe adopted from Anwar et al. [21] failed to produce a homogeneous layer. The solution was opaque from the beginning and the resulting layers were patchy and laced with cracks (see figure 5.1a). Altering the ratios of ingredients, adjusting pH (by adding NaOH, H₂SO₄ or HCl) or adding SDS as surfactant didn't improve the resulting ZrO₂ layers. Even after the consecutive application of two layers, there was no improvement and the results reminded rather of a sparse crust than of a film. Figure 5.1a shows a typical detail view of ZrO₂ (dark) on top of FTO (finely polycrystalline). Large cracks, non coated areas and inhomogeneity were shared characteristics among all samples created with the aquatic recipe.

The sol-gel recipe optimized by Hu et al. [91] for Al₂O₃ worked splendidly for our use case even after (not so) minor changes (see section 3.2.2). Figure 5.1b shows a top view of 10 layers of the buthanolic solution on top of a steel substrate. The large irregularity on the bottom left could depict a pin hole, a (tiny) hole in the layer that reaches to the substrate. These irregularities are rather the exception. Alternatively, the small, hardly visible slits spread across the surface could depict such pin holes. These - I think - are rather only one layer deep irregularities since they are so prevalent and hardly any pin holes were electrically observed on similar samples.

As soon as the solution showed initial cloudiness, it was declared as unstable and no more used for DB. The visually asserted stability of the solution could be increased a lot by replacing the stabilizing agent (AcOH) with IPO. A 1F and a 4F solution with AcOH - sealed with Parafilm - was stable for approximately 24 h and 2 h, respectively. Whereas a 4F solution with IPO as stabilizing agent stayed stable for circa 96 h. An increase of the Zr(PrO)₄ concentration accelerates the aging process of the solution. The aging process was accelerated by being in contact with "fresh" air in contrast to being sealed. It therefore dependent on the water or oxygen in the air corresponding to an oxidation or hydrolization process. Only the introduction of IPO into the recipe allowed higher concentrated solutions to be practicable as they could be stored until the next time needed, instead of producing a separate solution for each sample. It was discovered that after a 4F AcOH solution has aged beyond cloudiness the addition 50% of the volume in IPO can even reverse the aging effect for a respectable amount of time. This effect is not only due to dilution as additional solvent (BuOH) did not re-stabilize the solution, even after 5 fold dilution.

Figure 5.1c shows the cross section of a layer of ZrO₂ on FTO glass at a deliberate fracture. The large crystalline structure which makes up most of the upper part of the image is FTO. The boundary to glass is visible at the very top. On the lower edge of FTO a circa 100nm thick homogeneous layer of ZrO₂ (produced with 5 layers of 1F buthanolic AcOH recipe) can be

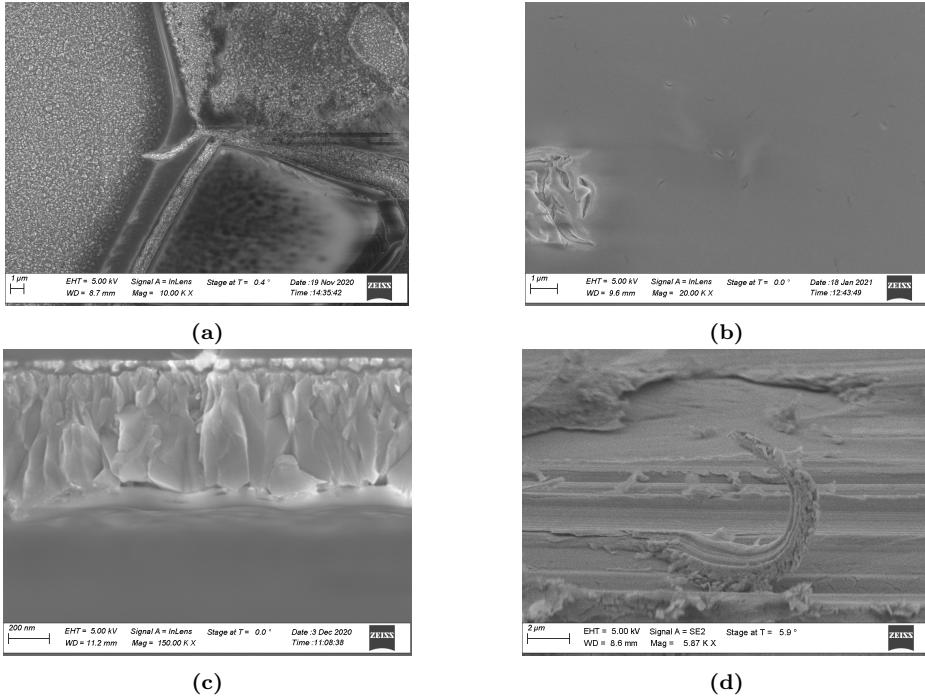


Figure 5.1: SEM images of of ZrO_2 : (a) ZrO_2 produces by aquatic solution on FTO (b) 10 layers of ZrO_2 on steel substrate (c) cross section of crack of 5 layers of ZrO_2 on FTO (d) side view of scratched ZrO_2 on steel substrate

observed. Measuring the cross section on steel was not as straight forward because the ductility of steel prevents fractures of the substrate. Resorting to scratching the surface with an diamond glass cutter allowed us to get an insight into the thickness of the layer on steel substrate. The result can be seen in figure 5.1d. 5 layers of double concentrated solution were applied to this sample. Assuming that the raised structure resembling a tentacle is ZrO_2 , the thickness of the layer can be estimated to be in the magnitude of several 100 nm.

5.1.1 XRD

XRD was used in order to confirm that the produced layer is indeed ZrO_2 . Figure 5.2 shows XRD spectra of the substrate (steel) and the substrate with ten layers of double concentrated solution (EMMA experiment number 6113, see Appendix B). In addition, two idealized spectra from the Crystallography Open Database of cubic ZrO_2 (COD ID 1521753 [96]) and of $\alpha\text{-Fe}$ (COD ID 1100108) are depicted. The four largest peaks of cubic ZrO_2 at $\theta = 30, 51, 60, 35$ (highest to lowest intensity) clearly stand out against the noise from the experimental spectrum. They confirm that the produced substance is ZrO_2 . Raman spectra are needed to further clarify the phase, since cubic and tetragonal phases are difficult to distinguish from XRD alone [97].

5.1.2 Spectrometry

In figure 5.3a transmission spectra of UV/Vis/NIR light can be seen. Each sample had different numbers of layers. The incident angle was 0 degree for each sample. The more layers, the more of the light is absorbed (at 600-1100 nm) by the ZrO_2 layers. The thicker the layer is, the more wavy the graph, which can be attributed to interference [98]. This trend can also be observed in figure 5.3b. The weak interference patterns at low wavelengths indicate that the thickness of the film is in the order of magnitude of the wavelengths [99] – agreeing with SEM measurements. Figure 5.3b shows reflectance spectra of UV/Vis/NIR at an incident angle of 45 degrees. Thicker layers reflect more light between approximately 800 nm and 1000 nm. Again, this effect is most likely not observed across the whole spectrum because of interferences.

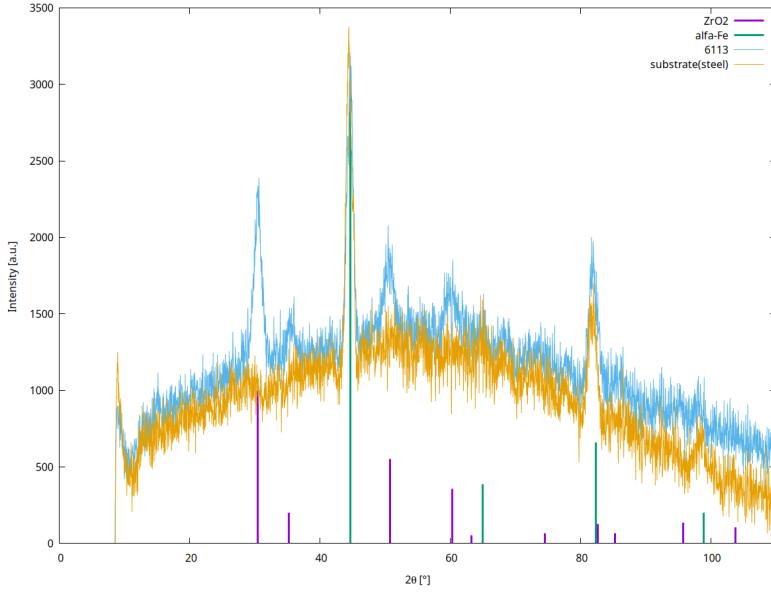


Figure 5.2: XRD spectra of steel substrate (yellow), ZrO₂ (blue), α -Fe (green, idealized) and cubic-ZrO₂ (purple, idealized) [96]

5.1.3 Current Voltage Curves

Figure 5.4a shows I-V curves of an insulating sample, where the abscissa shows the voltage in Volt and the ordinate shows the current in Ampere. Each line represents a I-V measurement at a distinct aluminium contact applied by sputtering through a mask. All of the curves show a max voltage of under 10^{-6} V and around a fifth exhibit hardly any current which is ideal (see figure 5.4d). A single example for a very low conductance measurement in non-logarithmic scale can be seen in figure 5.4g. The conductance (i.e. the gradient at $V=0$) was calculated (see section 4.1) for each measurement. Figure 5.4d shows the distribution of gradients g for the same sample as depicted in 5.4a. Figures 5.4b and 5.4e show measurements of a moderately insulating sample. Most of the I-V curves have a maximum voltage of under 10^{-6} V, but there are some so called pin holes with conductance above the threshold of 10^{-5} V. In figures 5.4a and 5.4b the minimum of some curves is not at 0 V, which means that the function does not cross the origin in non-logarithmic representation. This deviation (see figure 5.4h) looks similar to the deviation due to the photo-currents [100] because of its non-zero current at $V = 0$. Though, the explanation by photo-current is compelling, it does most likely not apply because of ZrO₂'s band gap of 5 eV [101]. A better explanation would be the tunneling effect, whose I-V curves look very similar [102, 103]. Finally, figures 5.4c and 5.4f show a sample where all measurement exhibit relatively high voltages and high calculated conductances. This indicates a overall bad condition of the ZrO₂ layer for insulating. Figure 5.4i shows a I-V curve which can be approximated by Ohm's law.

5.1.4 Preoptimization

Before the optimization with the EMMA algorithm, the boundaries of the input variables were explored. Especially, the DB velocity v_{DB} (i.e. the velocity with which the blade moves and spreads the solution over the sample) was varied and examined during pre-optimization. The slower the DB velocity, the more uniformly the solution evaporated. If the v_{DB} was too slow (less than 1 mm/s), no layer was formed. The reason is that the force exerted by the blade on the liquid is not strong enough to overcome surface tension. This means that behind the blade a meniscus would pull the liquid without leaving behind any solution for the gelling process. Additionally, the temperature while DB T_{DB} affects the evaporation process together with the DB velocity v_{DB} .

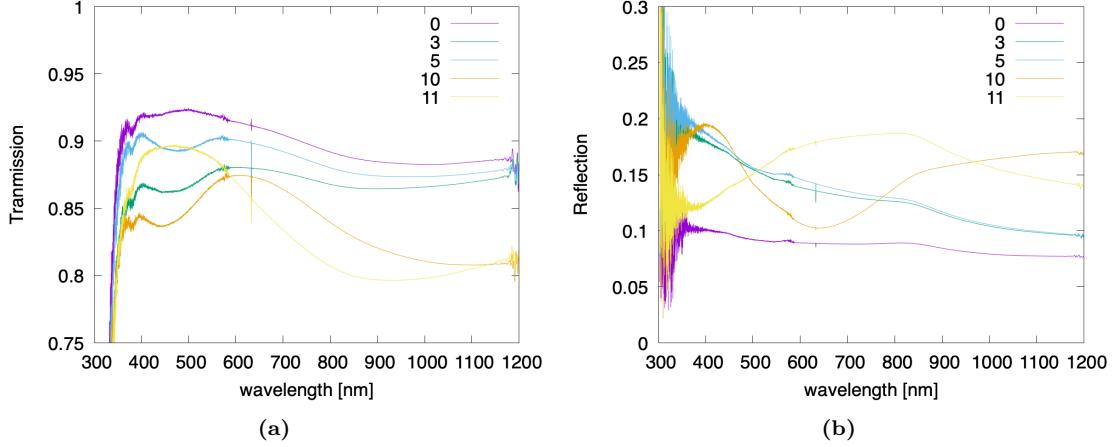


Figure 5.3: UV/Vis/NIR spectra of different ZrO_2 layer counts on glass (a) transmission at 0° and (b) reflection at 45°

5.2 EMMA

A total of 30 recipes (see appendix B) have been investigated in five iterations ($t = 0, \dots, 4$) of the algorithm. Where the first generation encompassed 10 particles and each subsequent generation encompassed 5 particles. The best recipes for each generation can be seen in table 5.1. The experiments for generation 5 where not executed but predictions were already made with the information from the previous generations. The best sample predicted by the algorithm is experiment number 13 with lowest c_{zr} , v_{DB} , T_{DB} , v_{cal} , T_{cal} and second highest λ .

Table 5.1: Global best per generation with experiment numbers and experiment conditions

generation	measurements	enr	c_{zr}	λ	v_{DB}	T_{DB}	v_{cal}	T_{cal}
1	10	1	2	4	10	40	120	300
2	15	5	2	6	10	40	120	300
3	20	2947	4	6	16	80	1080	300
4	25	2405	2	6	10	40	1080	300
5	30	13	2	10	10	40	120	300

The only clear trend from table 5.1 is the layer count, which rises with the generations. The remaining input variables remained more or less the same except for the 3rd generation. It can be seen in table 5.2 that the predicted conductivity $\hat{\gamma}$ (predicted with 5th generation RF; see last column) lowered with each iteration, except for sample 2947, which wasn't predicted but measured. This shows that the easiest part of the algorithm - the selection of the optimum from predicted values - works as expected. Contrarily, the predicted conductivity $\hat{\gamma}$ for each generation's best (predicted by the very generation's RF) does increase with each iteration (see diagonal in table 5.2). This indicates an underestimation of $\hat{\gamma}$ at the beginning and a correction with time. The underestimation probably stems from a lucky selection of initial experiments or a skew in the measurements.

Table 5.2: Predicted conductance measure $\hat{\gamma}$ for each generation by each generation's regression function

enr	1st gen RF	2nd gen RF	3rd gen RF	4th gen RF	5th gen RF
1	1.214185				38.7962
5		4.196626			25.47335
2947			10.9594		10.9594
2405				20.04962	25.47335
13					24.87178

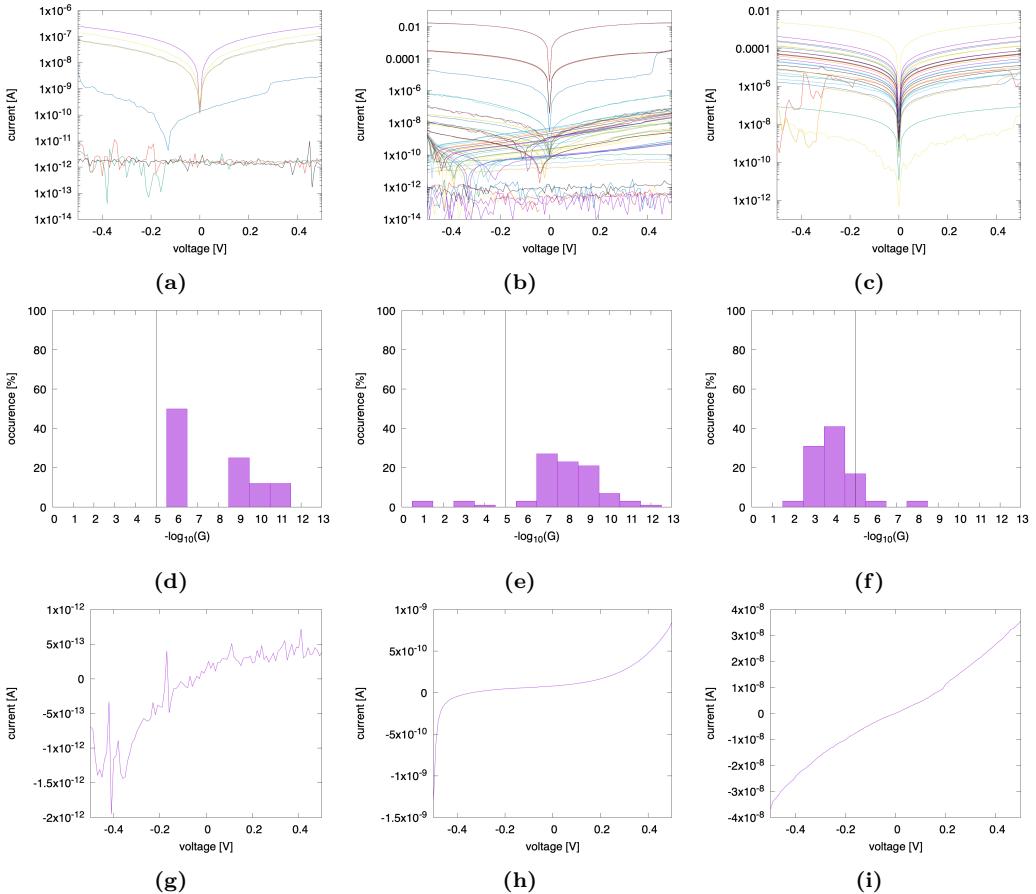


Figure 5.4: Current-voltage curves and the distribution of gradients for three representative samples: insulating well (a),(d),(g), moderately (b),(e),(h), poorly (c),(f),(i); (a)-(c) I-V curves with log transformed current; (d)-(f) frequency histograms of the conductance with pin hole threshold marked as vertical line; (g)-(i) individual I-V measurements

Indeed, the samples with the lowest *optimizands* are among the initial generation (see figure 5.5a, figure 5.5b and appendix B).

In figures 5.5a and 5.5b we can see the two measured main *optimizands*, conductivity γ and pin hole density ρ , of each particle at each generation. The points at generation number 0 indicate particles which were not included to be propagated. Violet lines connect individual particles in time, the blue line connects averages of particles of each generation and yellow lines connect average of only particles which were propagated of each generation. Both γ and ρ were to be minimized and show a clear trend towards low values with increasing generation, indicating that the optimization worked even though the prediction functions (see equations (5.1)–(5.8)) and the chosen samples where not exactly as expected. Neither expected were the measurements for these samples, which might be due to the high measuring error of samples and variation of quality due to uncontrolled independent variables such as room temperature or humidity. Measurements were expected to show clear correlation, which was only observed for T_{cal} . In the scatter plots 5.5c and 5.5d the correlation of calcination temperature T_{cal} with γ and ρ can be seen clearly.

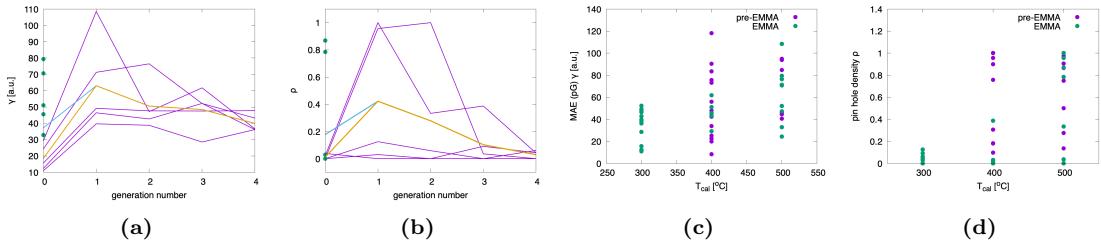


Figure 5.5: (a) calculated conductance measures γ of each particle against the generation number (b) calculated pin hole density ρ of each particle against the generation number (c) calculated conductance measures γ against calcination temperature T_{cal} (d) calculated pin hole density ρ against calcination temperature T_{cal}

Equations (5.1)–(5.4) and (5.5)–(5.8) represent the regression functions at $t = 3$ and $t = 4$, respectively, rounded to 2 significant digits. The expression $h(\lambda - 6)$ translates into layer count only has an influence if larger than 6 and $h(6 - \lambda)$ into layer count influential only if under 6. The first thing to notice is that prediction functions of each generation depend on the same variables. This stems from the fact that the algorithm chooses a single minimal set of basis functions to predict all dependent variables.

$$\hat{\rho}_3 = 0.075 - 0.0014 \cdot v_{cal} + 0.18 \cdot h(6 - \lambda) + 3.9 \cdot 10^{-06} \cdot v_{cal} \cdot T_{cal} \quad (5.1)$$

$$\hat{\gamma}_3 = 43 - 0.097 \cdot v_{cal} + 10 \cdot h(6 - \lambda) + 0.00026 \cdot v_{cal} \cdot T_{cal} \quad (5.2)$$

$$\hat{\lambda}_3 = 9.9 - 0.00064 \cdot v_{cal} - 2.7 \cdot h(6 - \lambda) - 1.3 \cdot 10^{-06} \cdot v_{cal} \cdot T_{cal} \quad (5.3)$$

$$\hat{v}_{cal,3} = -5.2 \cdot 10^{-15} + 0.016 \cdot v_{cal} + 1.3 \cdot 10^{-15} \cdot h(6 - \lambda) + 3.9 \cdot 10^{-21} \cdot v_{cal} \cdot T_{cal} \quad (5.4)$$

$$\hat{\rho}_4 = -0.87 + 0.0047 \cdot T_{cal} - 0.00036 \cdot \lambda \cdot T_{cal} + 0.0024 \cdot h(\lambda - 6) \cdot T_{DB} \quad (5.5)$$

$$\hat{\gamma}_4 = -19 + 0.28 \cdot T_{cal} - 0.022 \cdot \lambda \cdot T_{cal} + 0.16 \cdot h(\lambda - 6) \cdot T_{DB} \quad (5.6)$$

$$\hat{\lambda}_4 = 6.8 - 0.014 \cdot T_{cal} + 0.0018 \cdot \lambda \cdot T_{cal} + 0.0060 \cdot h(\lambda - 6) \cdot T_{DB} \quad (5.7)$$

$$\hat{v}_{cal,4} = 29 - 0.052 \cdot T_{cal} + 0.0011 \cdot \lambda \cdot T_{cal} - 0.011 \cdot h(\lambda - 6) \cdot T_{DB} \quad (5.8)$$

The coefficients in equations (5.1) and (5.2) have the same signs and differ by a factor of roughly 100. This fits the data well since the maxima of ρ and γ are separated by rough factor of 100 (see figure 5.5 and appendix B). The coefficients of the $v_{cal} \cdot T_{cal}$ interaction in equations (5.1) and (5.2) seem low, but the interaction has – considering the minimum value of the interaction of $300 \cdot 120 = 36\,000$ – several orders of magnitude higher influence on ρ_3 and γ_3 than the $h(6 - \lambda)$ term. It is astonishing that the knot of the hinge function for equations (5.1)–(5.4) was chosen so low; basically only including the influence of the lowest layer count $\lambda = 4$. Meaning for equations (5.1) and (5.2), that the lowest layer count produces worse samples which is intuitive, but more than 6 layers don't improve the insulation. Encouragingly, the calcination heating rate v_{cal} (see equation (5.4)) has been predicted perfectly within numerical precision.

The coefficients of equations (5.5) and (5.6) show the same pattern as equations (5.1) and (5.2): identical signs and factor 100. The influence of T_{cal} on the measures of conductance (ρ and γ) is highly interesting. It was expected that calcination temperatures under 400 °C do not suffice to produce compact layers and that therefore the resulting layer does not insulate as well. The original recipe by Hu et al. [91] used 500 °C and thus even 400 °C are low. *Optimizands* $\hat{\rho}$ and $\hat{\gamma}$ increase with higher calcination temperature according to equations (5.5) and (5.6) (see also figures 5.5c and 5.5d), ergo the resistance decreases and the conductance increases with increasing calcination temperature contrary to expectations. Furthermore, the coefficient of the T_{cal} term is the largest of all terms on $\hat{\rho}$ and $\hat{\gamma}$; interestingly also on $\hat{\lambda}$ and \hat{v}_{cal} . The coefficient of the $\lambda \cdot T_{cal}$ interaction is about a tenth in size of the T_{cal} coefficient, but has the extra factor λ (range: 4–12), resulting in the products of coefficients and process variables being in the same order of magnitude for T_{cal} and $T_{cal} \cdot v_{cal}$. It can be noted that the coefficient of the $\lambda \cdot T_{cal}$ interaction always has contrary sign to T_{cal} coefficient in equations (5.5)–(5.8). This negative correlation could hint a compensation of the overestimated influence of T_{cal} on the *optimizands*. Similarly, signs of coefficients of $\lambda \cdot T_{cal}$ and $h(\lambda - 6) \cdot T_{DB}$ correlate negatively in equations (5.5)–(5.8) and the product of coefficients and input variables are again similar in order of magnitude. The process variables T_{cal} and λ appear as basis function (BF)s in generation 3 and 4, whereas v_{cal} only appears in generation 3's RFs and T_{DB} only in generation 4's RFs. This can hint the importance of these variables in the process.

For each generation the MSE was calculated for which only samples from the optimization were used which were available at the time of prediction. That means 15, 20, 25 and 30 samples were used to calculate MSEs at $t = 1, 2, 3, 4$, respectively (compare with table 5.1). The MSEs are 64, 158, 54 and 50 for $t = 1, 2, 3, 4$, respectively. It is interesting that although prediction functions for $t = 3$ predicted v_{cal} perfectly, the combined MSE for $t = 4$ is lower. The high MSE at $t = 2$ can be explained by the prediction functions being only constant values. Apart from the second generation the MSE decreases over time, which indicates that the algorithm works. This decrease in MSE might be attributed to overfitting, though, since prediction and validation were performed on the same data. The MSE for each generation's prediction function was also calculated for pre-optimization (out-of-sample) samples, which are unseen data. The error sank again with each generation (except for the second generation): 102, 118, 58, 50, respectively for $t = 1, 2, 3, 4$. The decrease of MSE with out-of-sample data shows that the regularization method of the MARS algorithm in principle works on investigated samples and that decrease of MSE is not due to overfitting. When comparing the out-of-sample and in-sample MSEs it can be noted that at $t = 1$ the out-of-sample MSE is close to 1.5 fold of in-sample and at $t = 2$ vice versa. This shows the poor prediction ability at the beginning of the optimization which improved with generations. This decrease in validation (out-of-sample) MSE supports the hypothesis stated at the beginning of this section: the measure of conductance was underestimated and the estimate improved over time (see table 5.2).

Even though the main *optimizands* were minimized as required, the RFs were not satisfactory. The two main problems were too many independent variables and too many dependent variables. Both seem closely related and overcomplicate the optimization, but both come with their own implications. Too many independent variables make it harder to distinguish variance due to random error (e.g. unmeasured and uncontrolled variables) from variance due to dependency. The difficulty of identifying meaningful correlation is mainly owed to the curse of dimensionality [81] which makes it hard to collect enough data for each dimension. The events per predictor variable (EPV) (data points per independent variable) are as low as 5 in the here presented optimization. By eliminating three independent variables the EPV could rise to 10, which is stated as rule of thumb for multivariate regression [104]. The results obtained with limited sample number deliver remarkable insight comparing to an EPV of 20–50 stated in the original MARS paper [79] and an EPV of around 20 in the original EMMA paper [77].

The main problem about too many dependent variables specific to this optimization procedure is that the same set of basis functions will be used to predict **all** *optimizands*. This leads to competition between the BFs, as not all dependent variables may depend on the same independent variables. This effect is reinforced by the choice of two independent variables as dependent variables. Independent variables as dependent variables will likely be chosen in BFs, for including them in the RFs is an easy way to reduce the MSE and improve prediction accuracy. These BFs then "take away places" of BFs predicting other dependent variables in the RF. This in turn can lead to wrong predictions and therefore inefficient choice of future samples.

The two independent variables v_{cal} and λ were included as *optimizands* for two reasons: Maximizing the calcination velocity and minimizing the number of layer application iterations leads to minimization of process time. The process time optimization would be better placed in a follow up study. The second reason for including independent variables was to check if the model works. This idea has two flaws: there is only one set of BFs for all *optimizands* and thus additional dependent variables not only complicate the model but make it more difficult to predict the actual dependent variables and chose future data points. Moreover, v_{cal} and λ were only considered with 5% weight. The second flaw is that there are obviously better methods to assess the quality of a model than to make it more complex. Validation methods for sparse data include re-sampling via leave-one-out or k -fold cross validation [105].

5.3 Post-EMMA

In this section the experimental results are analyzed with different methods and compared with the optimization method EMMA which was integrated into the experiment selection process.

5.3.1 ANOVA

The RFs produced by the last iteration of EMMA (see equations 5.5 and 5.6) put a high significance on the calcination temperature T_{cal} with regard to the measures of the conductance. ANOVA is used to double check this.

The null hypothesis is that the results were obtained by pure chance without any dependence of the insulation of a sample on the process variables. A one-way ANOVA was performed for every independent variable on ρ and γ . We used an α value of 0.05 which means that the null hypothesis is rejected if the p-value is under 0.05. Usually, the α -value is chosen to be 0.01 or 0.05 [106, 107]. The choice of a relatively high *alpha*-value is motivated by the rather small amount of data and low events per variable ratio.

For both ρ and λ (for both the EMMA dataset and the whole dataset (=EMMA+pre-EMMA)) the F-values for T_{cal} were in all cases under 0.01, indicating a real influence of T_{cal} on the conductance. Furthermore, both datasets exhibit a p-value < 0.05 for the interaction $T_{cal}:v_{cal}$ on ρ and the whole data set has p-value smaller than 0.01 of v_{cal} on ρ . Whereas v_{cal} on ρ for EMMA dataset has p-value of slightly over 0.05 ANOVA confirms that T_{cal} has influence on both γ and ρ and therefore agrees with EMMA. It is important to mention, that ANOVA assumes categorical data for inputs. Categorical inputs have no order and thus information is lost when using numerical data.

5.3.2 Linear Regression

The data obtained from the EMMA optimisation was used to fit a linear model. The biggest difference between the linear model and EMMA is that the linear regression function will include all input variables, that the BFs are strictly linear and that there are no interaction terms. The coefficients of the λ term have the same sign in equations 5.1, 5.2, 5.5, 5.6 (EMMA RFs) and 5.9, 5.10 and have comparable magnitudes too.

$$\hat{\rho} = -0.022 \cdot c_{zr} - 0.022 \cdot \lambda - 0.0029 \cdot v_{DB} + 0.0058 \cdot T_{DB} - 8.5 \cdot 10^{-5} \cdot v_{cal} + 0.0026 \cdot T_{cal} - 0.76 \quad (5.9)$$

$$\hat{\gamma} = 3.2 \cdot c_{zr} - 0.97 \cdot \lambda - 1.2 \cdot v_{DB} + 0.2 \cdot T_{DB} - 0.014 \cdot v_{cal} + 0.13 \cdot T_{cal} + 7.2 \quad (5.10)$$

All terms in equations 5.9 and 5.10 have the same sign except c_{zr} and the intercept. More particularly, T_{cal} and T_{DB} have positive coefficients, i.e. a negative influence on the resistance and λ , v_{DB} and v_{cal} have negative coefficients, i.e. a positive influence on resistance. An indication that the data is strongly tainted by noise (see also section 5.3.1) is the low R^2 score of 0.41 of $\hat{\rho}$, which is (sad but true) even undercut by the R^2 of linear fit of $\hat{\gamma}$, 0.34. When the linear regression is trained on the whole data set, all terms have the same sign compared to the EMMA data set

trained linear model. v_{cal} and λ are perfectly predicted by linear regression as there is an exact direct proportionality and thus are not displayed here.

	c_{zr}	λ	v_{DB}	T_{DB}	v_{cal}	T_{cal}
$\hat{\rho}$	2.04	46.58	1.15	9.21	13.50	27.52
$\hat{\gamma}$	10.24	7.10	16.46	10.97	7.68	47.55

Table 5.3: Influences of input variables on output variables in percent according to linear regression.

Regarding the coefficients of the linear model, the zirconium concentration in the starting solution c_{zr} and the number of layers applied λ are largest in value, but the values which are factored with these coefficients - i.e. c_{zr} and λ themselves – are the smallest among the input variables. Thus, even though T_{cal} might have a smaller coefficient in the linear RF, the influence on the dependent variable might be larger. In order to compare, the influence of an input variable was calculated by the formula: $I(x_i) = \frac{\bar{x}_i \cdot k_i}{\sum_i \bar{x}_i \cdot k_i}$ with $i = 1, \dots, d$ and where $\bar{x}_i = \frac{(x_{min} + x_{max})}{2}$ is average possible value of the i -th input variable and k_i is the coefficient of the i -th input variable in the linear regression. The biggest influence on $\hat{\rho}$ is λ , followed by T_{cal} . While T_{cal} is the second largest influence on $\hat{\rho}$ as well it is noteworthy that λ contributes the least to $\hat{\gamma}$. That means that according to the linear regression, the number of layers increases the overall resistance of the zirconium ceramics while a lower calcination temperature leads to less defects. Interestingly, the number of layers has the least amount of influence on the number probability of defects, which seems to be an artefact of the noise rather than a process property. The two very different distributions of influence among the input variables can have three reasons: (1) the linear function is not a suitable regression function for the data distribution. (2) the signal to noise ratio is so low that the noise rather than the signal was fitted. (3) the *optimizands* indeed represent different properties of the material.

5.3.3 Grid Search, KRR and SVM

The optimal parameters for each combination of *optimizand* (γ and ρ) and method (KRR and SVM) can be seen in table 5.4. The parameters are often extrema. The hyperparameter γ_{ml} took the smallest possible value (within the grid search) twice. In the polynomial kernel this leads to a lower influence of the input variables with respect to the intercept. The range of the hyperparameter α was 0.01–10 in the grid search. So, 2 and 10 are high values for the regularization parameter α of KRR. The hyperparameter ε specifies the size of the tube within which data does not contribute to the loss and the value 1 is the maximum for ε during this grid search. All these extreme values of hyperparameters (stimulate regularization and indicate noisy and difficult to fit data). The R^2 values were calculated for each of the models using cross validation. Resulting R^2 are all below zero (see table 5.4). A negative R^2 value means that the prediction provided by the model is worse than predicting only the mean of γ and ρ .

Finally, the MAE and MSE were calculated for different data sets (see table 5.5). Each models was trained on the EMMA data set. The first two columns (e) show statistics validated with data the model was trained with (30 data points) and thus, these columns contain the lowest numbers. The next two columns (p) were calculated with data obtained before the EMMA optimization (21 data points). As the pre-optimization data set was mainly used to find appropriate boundaries for the input variables, many data points are outside of the range of the EMMA data set. Thus, prediction for this data set are mainly extrapolations. It must be explicitly noted at this point

method	R^2	kernel	degree	γ_{ml}	α	C	ε
KRR(γ)	-0.50 ± 0.25	poly	1	0.01	2	-	-
KRR(ρ)	-0.45 ± 0.52	rbf	-	0.2	10	-	-
SVM(γ)	-0.37 ± 0.39	poly	2	0.5	-	0.7	1.0
SVM(ρ)	-0.95 ± 0.44	poly	3	0.01	-	0.1	0.1

Table 5.4: R^2 values and best hyperparameters for each *optimizand*-method-combination

that pre-EMMA samples differ from EMMA samples in the stabilization agent and therefore might underperform. The next two columns (c) were produced by using pre-EMMA data for validation, but only keeping points which sit inside the space spanned by the EMMA constraints (5 data points). It should be expected that these statistics should be lower than in the previous two columns. They are only lower for $MARS(\rho)$ and $LR(\rho)$, though. This is likely due to the small sample size. The last two columns show data for predictions of all available data (51 data points). These statistics are lower because of the inclusion of the training set in the test set.

In the self test of the first two columns MARS is the one method to rule them all and provides the lowest metrics. But as soon as unseen data is examined, linear regression excels by its simplicity.

γ	MAE(e)	MSE(e)	MAE(p)	MSE(p)	MAE(c)	MSE(c)	MAE(a)	MSE(a)
MARS	10	171	28	1084	30	1280	17	548
LR	13	250	24	747	34	1327	17	454
KRR	17	548	30	1477	35	2167	23	931
SVM	15	415	26	1050	28	1588	19	677
ρ	MAE(e)	MSE(e)	MAE(p)	MSE(p)	MAE(c)	MSE(c)	MAE(a)	MSE(a)
MARS	0.14	0.03	0.41	0.21	0.39	0.18	0.25	0.11
LR	0.19	0.06	0.30	0.15	0.38	0.13	0.24	0.09
KRR	0.25	0.12	0.38	0.24	0.43	0.29	0.31	0.17
SVM	0.23	0.13	0.41	0.28	0.45	0.33	0.31	0.20

Table 5.5: Comparison of MAE and MSE of different prediction methods for different data sets: EMMA data set (e), pre-EMMA data set (p), pre-EMMA data set within EMMA bounds (c) and complete dataset (a)

5.4 Hindsight is easier than foresight

This section will discuss mistakes, improvements and lessons learned from each phase of this project. Firstly, the Research, exploration and design phase will be discussed in subsection 5.4.1, followed by the Narrowing, pre-optimization and constraint phase reviewd in subsection 5.4.2 and finally the Optimization, modelling and data generation phase examined in subsection 5.4.3.

5.4.1 Research, exploration and design phase

Scrutinize everything! At the beginning of a project it is important to get a “feel”. I took the first recipe as only true recipe and tried to tune it to perfection. The problem was that the base line of that recipe was so low such that improving the recipe still lead to poor performance. This was only noticed this after trying out another recipe. The author’s tip here is to read as much and mostly diverse literature as possible and then try to mimic what ever can be. This will make it possible to acquire “feel” of what works and what does not. This is inspired by a quote from Thomas Torsney-Weir’s data visualization classes: “create as many and different designs as possible”. One should check what works, collect (dis-)advantages of methods and combine the best approaches. This can save you a lot of time.

Moreover, it is important to define the objective of the project as clear and unambiguous as possible. Again, try to reformulate the objective in multiple ways and choose the most fitting. Answer which question should be answered, and establish hypotheses with corresponding null hypotheses. A general experimental design procedure was stated by Cherkassky and Mulier in “Learning from Data” [75]:

1. State the problem
2. Formulate the hypothesis
3. Design the experiment/data generation process

4. Collect the data and perform preprocessing
5. Estimate the model
6. Interpret the model/draw the conclusion

5.4.2 Narrowing, pre-optimization and constraint phase

The selection of input and output variables, data encoding/representation and incorporating a prior knowledge into the design of the learning system is often more critical for an overall success than the design of the learning machine itself. [75] After having decided on a base recipe, it is important to define limits and decrease the number of independent variables. The later is more important due to the curse of dimensionality [75], which states that with increasing dimensionality it get exponentially more difficult to arrive at the same density of data points. By excluding input variables from the model, the complexity of the model can be drastically reduced. This leads to more information which can be extracted from the remaining data, be it in more detail of the dependence or higher confidence interval. A good measure to compare with literature and other methods is the EPV. The EPV in this work is round 5, which is low in comparison with literature consulted [108–115]. Reducing variables is of course no easy task. How does one decide which variables to keep and which to incorporate into the model and which to discard? There is an intuitive (expert) way and a mathematical way. An expert would decide which variables to discard by his expert knowledge about the domain. Mathematical ways include principal component analysis and step-wise regression. Additionally, having wide limits can be an advantage when the data is noisy and the dependence is not too complicated, because by moving samples to the boundaries of the design space, the effect of the noise is reduced [116]. This suggest variable reduction favorable over limit narrowing. There are two opposing principles which should be kept in mind when reducing variables, though: By introducing a lot of variables into a model presupposes that they and their interactions (!) have a statistically significant effect [84]. On the other hand a model can be used to tell if a variable exerts an influence [35]. Another reason to avoid too many variables in a model is that they introduce more parameters to fit and overfitting happens in the regime where to many parameters are fitted with not enough data.

Another step/deed which is well placed in this phase is the estimate the error by reproducing a single sample a certain number of times.

5.4.3 Optimization, modelling and data generation phase

During the optimization the main problems were too many in- and output variables, reproducibility and regularization. Variable selection has already been discussed in the previous section. When working with stochastic algorithms it is especially important to introduce reproducibility via a seed, which determines all subsequent pseudo-random numbers. Finally, during the EMMA optimization the input data were not regularized, because the regularization would be different for different generations. Regularizing with the mean and standard deviation would be either inconsistent across generations or overvalue the distribution of the initial generation. These were the reasons for not including regularization during EMMA optimization. On the other hand the input data could have been regularized to only take values $[-1; 1]$ by defining the regularized variable as $x' = \frac{x - \bar{x}}{x_{\max} - \bar{x}}$ with $\bar{x} = x_{\max} - x_{\min}$.

when adding dep var to model this has two folgen: - beeinflusst the BF and the optimization, but bcs relativ weight is only 0.05 for each, no big beeinflussung on optimization

Chapter 6

Summary and Outlook

The goal of this work was to find a sol-gel doctor blading process resulting in a insulating zirconium ceramic on top of steel. This structure then can be used as substrate for thin layer photovoltaics. The ceramic needs to be homogeneous and defect free in order to insulate properly. A suitable recipe was found [91] and the process variables were optimized via EMMA. The recipe was improved by increasing its stability. The best parameters found by the EMMA optimization were $c_{zr} = 2$, $\lambda = 10$, $v_{DB} = 10$, $T_{DB} = 40$, $v_{cal} = 120$ and $T_{cal} = 300$. After optimization, the data was analyzed with further machine learning methods like kernel ridge regression, support vector regression and linear regression. The comparison of methods showed that EMMA performed well on identifying the important correlation between conductance and process variables. A strong correlation of calcination temperature and conductance in the range of 300–500°C has been detected. Furthermore, the layer count has been verified as an important factor. Next steps include to investigate if the correlation of conductance and calcination temperature are statistically significant or due to small sample size. The aging process could be examined via infra red spectroscopy.

Bibliography

- [1] F. J. Dyson, “Search for artificial stellar sources of infrared radiation,” *Science*, vol. 131, no. 3414, pp. 1667–1668, 1960.
- [2] P. S. Vasekar, A. H. Jahagirdar, and N. G. Dhere, “Photovoltaic characterization of copper-indium-gallium sulfide (cigs2) solar cells for lower absorber thicknesses,” *Thin Solid Films*, vol. 518, no. 7, pp. 1788–1790, 2010. Photovoltaics, solar energy materials and thin films - IMRC 2008, Cancun, Mexico.
- [3] B. Petroleum, “Bp statistical review of world energy 2017,” *Statistical review of world energy*, vol. 65, 2017.
- [4] B. für Geowissenschaften und Rohstoffe, “Daten und entwicklungen der deutschen und globalen energieversorgung,” *Statistical review of world energy*, vol. 65, 2017.
- [5] D. Lübbert and F. Lange, “Uran als kernbrennstoff: Vorräte und reichweite,” *Wissenschaftliche Dienste des deutschen Bundestages, Info-Brief WF VIII G-069/06*, 2006.
- [6] M. Becquerel, “Mémoire sur les effets électriques produits sous l'influence des rayons solaires,” *Comptes rendus hebdomadaires des séances de l'Académie des sciences*, vol. 9, pp. 561–567, 1839.
- [7] W. Smith, “Effect of light on selenium during the passage of an electric current*,” *Nature*, vol. 7, p. 303, 1873.
- [8] W. G. Adams and R. E. Day, “V. the action of light on selenium,” *Proceedings of the Royal Society of London*, vol. 25, no. 171-178, pp. 113–117, 1877.
- [9] A. Einstein, “Über einem die erzeugung und verwandlung des lichtes betreffenden heuristischen gesichtspunkt,” *Annalen der physik*, vol. 4, 1905.
- [10] M. Pagliaro, G. Palmisan, and R. Ciriminna, *Flexible Solar Cells*. Wiley-VCH, 2008.
- [11] T. Markvart and L. Castañer, “Chapter ia-1 - principles of solar cell operation,” in *Solar Cells (Second Edition)* (A. McEvoy, L. Castañer, and T. Markvart, eds.), pp. 3–25, Elsevier, second edition ed., 2013.
- [12] O. Breitenstein, “Understanding the current-voltage characteristics of industrial crystalline silicon solar cells by considering inhomogeneous current distributions,” *Opto-Electronics Review*, vol. 21, no. 3, pp. 259–282, 2013.
- [13] H. Hahn, G. Frank, W. Klingler, A.-D. Meyer, and G. Störger, “Untersuchungen über ternäre chalcogenide. v. über einige ternäre chalcogenide mit chalkopyritstruktur,” *Zeitschrift für anorganische und allgemeine Chemie*, vol. 271, no. 3-4, pp. 153–170, 1953.
- [14] S. Wagner, J. Shay, P. Migliorato, and H. Kasper, “Cuinse2/cds heterojunction photovoltaic detectors,” *Applied Physics Letters*, vol. 25, no. 8, pp. 434–435, 1974.
- [15] L. Kazmerski, F. White, and G. Morgan, “Thin-film cuinse2/cds heterojunction solar cells,” *Applied Physics Letters*, vol. 29, no. 4, pp. 268–270, 1976.

- [16] T. Feurer, P. Reinhard, E. Avancini, B. Bissig, J. Löckinger, P. Fuchs, R. Carron, T. P. Weiss, J. Perrenoud, S. Stutterheim, S. Buecheler, and A. N. Tiwari, “Progress in thin film cigs photovoltaics – research and development, manufacturing, and applications,” *Progress in Photovoltaics: Research and Applications*, vol. 25, no. 7, pp. 645–667, 2017.
- [17] A. McEvoy, L. Castañer, and T. Markvart, eds., *Practical handbook of photovoltaics. Fundamentals and applications(Second Edition)*. Elsevier, second edition ed., 2011.
- [18] K. Mertens, *Photovoltaik: Lehrbuch zu Grundlagen, Technologie und Praxis*. Carl Hanser Verlag GmbH Co KG, 2020.
- [19] U. Rau and H. Schock, “Chapter ic-3 - cu(in,ga)se2 thin-film solar cells,” in *Solar Cells (Second Edition)* (A. McEvoy, L. Castañer, and T. Markvart, eds.), pp. 261–304, Elsevier, second edition ed., 2013.
- [20] J. Hedstrom, H. Ohlsen, M. Bodegard, A. Kyller, L. Stolt, D. Hariskos, M. Ruckh, and H.-W. Schock, “Zno/cds/cu(in,ga)se/sub 2/ thin film solar cells with improved performance,” in *Conference Record of the Twenty Third IEEE Photovoltaic Specialists Conference - 1993 (Cat. No.93CH3283-9)*, pp. 364–371, 1993.
- [21] M. A. Anwar, T. Kurniawan, Y. P. Asmara, W. S. W. Harun, A. N. Oumar, and A. B. D. Nandyanto, “Morphology evaluation of ZrO₂dip coating on mild steel and its corrosion performance in NaOH solution,” *IOP Conference Series: Materials Science and Engineering*, vol. 257, p. 012087, oct 2017.
- [22] K. Kukli, K. Forsgren, M. Ritala, M. Leskelä, J. Aarik, and A. Haårsta, “Dielectric properties of zirconium oxide grown by atomic layer deposition from iodide precursor,” *Journal of The Electrochemical Society*, vol. 148, no. 12, p. F227, 2001.
- [23] R. H. Nielsen and G. Wilfing, *Zirconium and Zirconium Compounds*. John Wiley & Sons, Ltd, 2010.
- [24] H. Yang and P. Jiang, “Large-scale colloidal self-assembly by doctor blade coating,” *Langmuir*, vol. 26, no. 16, pp. 13173–13182, 2010.
- [25] S. Swann, “Magnetron sputtering,” *Physics in Technology*, vol. 19, pp. 67–75, mar 1988.
- [26] P. J. Kelly and R. D. Arnell, “Magnetron sputtering: a review of recent developments and applications,” *Vacuum*, vol. 56, no. 3, pp. 159–172, 2000.
- [27] D. McMullan, “Scanning electron microscopy 1928–1965,” *Scanning*, vol. 17, no. 3, pp. 175–185, 1995.
- [28] M. Kaliva and M. Vamvakaki, “Chapter 17 - nanomaterials characterization,” in *Polymer Science and Nanotechnology* (R. Narain, ed.), pp. 401–433, Elsevier, 2020.
- [29] K. Vernon-Parry, “Scanning electron microscopy: an introduction,” *III-Vs Review*, vol. 13, no. 4, pp. 40–44, 2000.
- [30] G. Schwedt, *Analytische chemie: Grundlagen, Methoden und Praxis*. John Wiley & Sons, 2008.
- [31] E. Göbel, I. Mills, and A. Wallard, *The international system of units (SI)*. Bureau International de Poids et Mesures, 2006.
- [32] A. I. Melker and M. A. Vorobyeva, “Electronic theory of molecule vibrations,” in *Ninth International Workshop on Nondestructive Testing and Computer Simulations* (A. I. Melker, ed.), vol. 6253, pp. 42 – 56, International Society for Optics and Photonics, SPIE, 2006.
- [33] K. B. Beć and C. W. Huck, “Breakthrough potential in near-infrared spectroscopy: Spectra simulation. a review of recent developments,” *Frontiers in chemistry*, vol. 7, p. 48, 2019.
- [34] A. Gontcharov, “No, ai is not statistics,” 2019. Last Accessed 2022-06-21.

- [35] G. Härtler, Statistisch gesichert und trotzdem falsch? Springer Spektrum, 2014.
- [36] J. Howard, “Artificial intelligence: Implications for the future of work,” American Journal of Industrial Medicine, vol. 62, no. 11, pp. 917–926, 2019.
- [37] M. J., M. ML, R. N, and S. CE, “A proposal for the dartmouth summer research project on artificial intelligence,” 1955.
- [38] P. McCorduck and C. Cfe, Machines who think: A personal inquiry into the history and prospects of artificial intelligence. CRC Press, 1982.
- [39] M. A. Apter, “Machines who think: A personal inquiry into the history and prospects of artificial intelligence by pamela mccorduck,” Leonardo, vol. 15, no. 3, p. 242, 1982.
- [40] D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, T. Graepel, T. Lillicrap, K. Simonyan, and D. Hassabis, “Mastering chess and shogi by self-play with a general reinforcement learning algorithm,” 2017.
- [41] F.-H. Hsu, “Ibm’s deep blue chess grandmaster chips,” IEEE Micro, vol. 19, no. 2, pp. 70–81, 1999.
- [42] A. M. Turing, “The imitation game,” Mind, vol. 59, no. 236, pp. 433–460, 1950.
- [43] GPT-3, “A robot wrote this entire article. are you scared yet, human?.” Last Accessed 2022-06-14.
- [44] E. A. M. van Dis, J. Bollen, W. Zuidema, R. van Rooij, and L. Bockting, “Chatgpt: five priorities for research,” 2023.
- [45] J. R. Searle, “Minds, brains, and programs,” Behavioral and Brain Sciences, vol. 3, no. 3, p. 417–424, 1980.
- [46] J. Searle, “I married a computer,” The New York Review of Books, vol. 8, 1999.
- [47] N. Tiku, “The google engineer who thinks the company’s ai has come to life,” The Washington Post, vol. 11, 2022.
- [48] M. Campbell, A. Hoane, and F. hsiung Hsu, “Deep blue,” Artificial Intelligence, vol. 134, no. 1, pp. 57–83, 2002.
- [49] L.-F. Li, X. Wang, W.-J. Hu, N. N. Xiong, Y.-X. Du, and B.-S. Li, “Deep learning in skin disease image recognition: A review,” IEEE Access, vol. 8, pp. 208264–208280, 2020.
- [50] R. C. Deo, “Machine learning in medicine,” Circulation, vol. 132, no. 20, pp. 1920–1930, 2015.
- [51] E. J. Topol, “High-performance medicine: the convergence of human and artificial intelligence,” Nature medicine, vol. 25, no. 1, pp. 44–56, 2019.
- [52] H. Fujiyoshi, T. Hirakawa, and T. Yamashita, “Deep learning-based image recognition for autonomous driving,” IATSS Research, vol. 43, no. 4, pp. 244–252, 2019.
- [53] J. Westermayr, M. Gastegger, M. F. S. J. Menger, S. Mai, L. González, and P. Marquetand, “Machine learning enables long time scale molecular photodynamics simulations,” Chem. Sci., vol. 10, pp. 8100–8107, 2019.
- [54] G. B. Goh, C. Siegel, A. Vishnu, N. O. Hodas, and N. Baker, “Chemception: A deep neural network with minimal chemistry knowledge matches the performance of expert-developed qsar/qspr models,” 2017.
- [55] D. Jha, L. Ward, A. Paul, W.-k. Liao, A. Choudhary, C. Wolverton, and A. Agrawal, “Elementnet: Deep learning the chemistry of materials from only elemental composition,” Scientific reports, vol. 8, no. 1, pp. 1–13, 2018.

- [56] I. H. Sarker, M. H. Furhad, and R. Nowrozy, “Ai-driven cybersecurity: An overview, security intelligence modeling and research directions,” *SN Computer Science*, vol. 2, p. 173, 2021.
- [57] T. M. Andrews, 2017. Last Accessed on 2022-06-03.
- [58] M. L. Littman, I. Ajunwa, G. Berger, C. Boutilier, M. Currie, F. Doshi-Velez, G. Hadfield, M. C. Horowitz, C. Isbell, H. Kitano, K. Levy, T. Lyons, M. Mitchell, J. Shah, S. Sloman, S. Vallor, and T. Walsh, “Gathering strength, gathering storms: The one hundred year study on artificial intelligence (ai100) 2021 study panel report.,” 2021.
- [59] M. V. Koroteev, “Bert: A review of applications in natural language processing and understanding,” 2021.
- [60] X. Liu, Y. Zheng, Z. Du, M. Ding, Y. Qian, Z. Yang, and J. Tang, “Gpt understands, too,” 2021.
- [61] J. Parviainen and M. Coeckelbergh, “The political choreography of the sophia robot: beyond robot rights and citizenship to political performances for the social robotics market,” *AI & SOCIETY*, vol. 36, pp. 715–724, 2021.
- [62] H. Mansourifar and W. Shi, “One-shot gan generated fake face detection,” 2020.
- [63] G. Marcus, E. Davis, and S. Aaronson, “A very preliminary analysis of dall-e 2,” 2022.
- [64] M. Guzdial and M. Riedl, “Toward game level generation from gameplay videos,” 2016.
- [65] last accessed 2022-06-13.
- [66] O. Hassan, “Artificial intelligence, neom and saudi arabia’s economic diversification from oil and gas,” *The Political Quarterly*, vol. 91, no. 1, pp. 222–227, 2020.
- [67] H. Chiroma, U. A. Abdullahi, S. M. Abdulhamid, A. Abdulsalam Alarood, L. A. Gabralla, N. Rana, L. Shuib, I. A. Targio Hashem, D. E. Gbenga, A. I. Abubakar, A. M. Zeki, and T. Herawan, “Progress on artificial neural networks for big data analytics: A survey,” *IEEE Access*, vol. 7, pp. 70535–70551, 2019.
- [68] C. M. Bishop, “Neural networks and their applications,” *Review of scientific instruments*, vol. 65, no. 6, pp. 1803–1832, 1994.
- [69] Y. LeCun, Y. Bengio, et al., “Convolutional networks for images, speech, and time series,” *The handbook of brain theory and neural networks*, vol. 3361, no. 10, p. 1995, 1995.
- [70] J. Westermayr, M. Gastegger, and P. Marquetand, “Combining schnet and sharc: The schnarc machine learning approach for excited-state dynamics,” *The journal of physical chemistry letters*, vol. 11, no. 10, pp. 3828–3834, 2020.
- [71] M. Rupp, “Machine learning for quantum mechanics in a nutshell,” *International Journal of Quantum Chemistry*, vol. 115, no. 16, pp. 1058–1073, 2015.
- [72] C. Rudin, “Least squares and friends,” 2020. last accessed 2022-11-08.
- [73] B. E. Boser, I. M. Guyon, and V. N. Vapnik, “A training algorithm for optimal margin classifiers,” in *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, COLT ’92, (New York, NY, USA), p. 144–152, Association for Computing Machinery, 1992.
- [74] P. H. Winston, *Artificial Intelligence*. Addison-Wesley Longman Publishing Co., Inc., 1992.
- [75] V. Cherkassky and F. Mulier, *Learning from Data*. Wiley, 1998.
- [76] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine learning*, vol. 20, pp. 273–297, 1995.

- [77] L. Villanova, P. Falcaro, D. Carta, I. Poli, R. Hyndman, and K. Smith-Miles, “Functionalization of microarray devices: Process optimization using a multiobjective pso and multiresponse mars modeling,” in *IEEE Congress on Evolutionary Computation*, pp. 1–8, IEEE, 2010.
- [78] J. Kennedy and R. Eberhart, “Particle swarm optimization,” in *Proceedings of ICNN’95-international conference on neural networks*, vol. 4, pp. 1942–1948, IEEE, 1995.
- [79] J. H. Friedman, “Multivariate adaptive regression splines,” *Annals of Statistics*, vol. 19, pp. 1 – 67, March 1991.
- [80] last accessed 2022-10-17.
- [81] J. H. Friedman, “Fitting functions to noisy data in high dimensions,” *Department of Statistics-Stanford University Technical Report*, vol. 101, pp. 1–36, 1988.
- [82] G. Wahba and P. Craven, “Smoothing noisy data with spline functions. estimating the correct degree of smoothing by the method of generalized cross-validation.,” *Numerische Mathematik*, vol. 31, pp. 377–404, 1979.
- [83] G. C. Goodwin and R. L. Payne, “Chapter 6 experiment design,” in *Dynamic System Identification* (G. C. Goodwin and R. L. Payne, eds.), vol. 136 of *Mathematics in Science and Engineering*, pp. 124–174, Elsevier, 1977.
- [84] R. F. Gunst and R. L. Mason, “Fractional factorial design,” *WIREs Computational Statistics*, vol. 1, no. 2, pp. 234–244, 2009.
- [85] K. Vanaja and R. S. Rani, “Design of experiments: Concept and applications of plackett burman design,” *Clinical Research and Regulatory Affairs*, vol. 24, no. 1, pp. 1–23, 2007.
- [86] A. Miller and R. R. Sitter, “Using the folded-over 12-run plackett—burman design to consider interactions,” *Technometrics*, vol. 43, no. 1, pp. 44–55, 2001.
- [87] J. C. Wang and C. F. J. Wu, “A hidden projection property of plackett-burman and related designs,” *Statistica Sinica*, vol. 5, no. 1, pp. 235–250, 1995.
- [88] F. A. C. Viana, “A tutorial on latin hypercube design of experiments,” *Quality and Reliability Engineering International*, vol. 32, no. 5, pp. 1975–1985, 2016.
- [89] U. M. Diwekar and J. R. Kalagnanam, “Efficient sampling technique for optimization under uncertainty,” *AIChE Journal*, vol. 43, no. 2, pp. 440–447, 1997.
- [90] R. A. Fisher et al., “On the ”probable error” of a coefficient of correlation deduced from a small sample.,” *Metron*, vol. 1, pp. 3–32, 1921.
- [91] B. Hu, E. Jia, B. Du, and Y. Yin, “A new sol-gel route to prepare dense al₂o₃ thin films,” *Ceramics International*, vol. 42, no. 15, pp. 16867–16871, 2016. <https://www.sciencedirect.com/science/article/pii/S0272884216312548>.
- [92] R. Plackett and L. Burman, “The design of optimal multifactorial experiments,” *Biometrika*, vol. 33, pp. 305–325, 1946.
- [93] L. Breiman and J. H. Friedman, “Predicting multivariate responses in multiple linear regression,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 59, no. 1, pp. 3–54, 1997.
- [94] D. Carta, L. Villanova, S. Costacurta, A. Patelli, I. Poli, S. Vezzu, P. Scopce, F. Lisi, K. Smith-Miles, R. J. Hyndman, et al., “Method for optimizing coating properties based on an evolutionary algorithm approach,” *Analytical chemistry*, vol. 83, no. 16, pp. 6373–6380, 2011.
- [95] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al., “Scikit-learn: Machine learning in python,” *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

- [96] J. GKatz, “X-ray diffraction powder pattern of metastable cubic zro₂,” *Am. Ceram. Soc.*, vol. 54, p. 531, 1971.
- [97] R. Purohit, S. Saha, and A. Tyagi, “Combustion synthesis of nanocrystalline zro₂ powder: Xrd, raman spectroscopy and tem studies,” *Materials Science and Engineering: B*, vol. 130, no. 1-3, pp. 57–60, 2006.
- [98] D. J. Dumin, “Measurement of film thickness using infrared interference,” *Review of Scientific Instruments*, vol. 38, no. 8, pp. 1107–1109, 1967.
- [99] J. B. de Lima Filho and Ángel Alberto Hidalgo, “Film thickness by interference pattern and optical characterization of polyaniline by spectroscopic ellipsometry,” *Synthetic Metals*, vol. 223, pp. 80–86, 2017.
- [100] A. Pérez-Tomás, A. Lima, Q. Billon, I. Shirley, G. Catalan, and M. Lira-Cantú, “A solar transistor and photoferroelectric memory,” *Advanced Functional Materials*, vol. 28, no. 17, p. 1707099, 2018.
- [101] A. Sinhamahapatra, J.-P. Jeon, J. Kang, B. Han, and J.-S. Yu, “Oxygen-deficient zirconia (zro_{2-x}): A new material for solar light absorption,” *Scientific Reports*, vol. 6, 2016.
- [102] R. M. Feenstra, “Scanning tunneling spectroscopy,” *Surface science*, vol. 299, pp. 965–979, 1994.
- [103] S. Datta, W. Tian, S. Hong, R. Reifenberger, J. I. Henderson, and C. P. Kubiak, “Current-voltage characteristics of self-assembled monolayers by scanning tunneling microscopy,” *Physical Review Letters*, vol. 79, no. 13, p. 2530, 1997.
- [104] E. Vittinghoff and C. E. McCulloch, “Relaxing the rule of ten events per variable in logistic and cox regression,” *American journal of epidemiology*, vol. 165, no. 6, pp. 710–718, 2007.
- [105] N. Kartam, I. Flood, and J. H. Garrett, “Validation and verification,” in *Artificial neural networks for civil engineers: fundamentals and applications*, ASCE, 1997.
- [106] R. Hoffman, “The concept of statistical significance,” 2020. last accessed 2022-11-08.
- [107] T. Sellke, M. J. Bayarri, and J. O. Berger, “Calibration of p-values for testing precise null hypotheses,” *The American Statistician*, vol. 55, no. 1, pp. 62–71, 2001.
- [108] M. A. Ahmed Al-dujaili, S. Jaheel, and H. N. Abbas, “Preparation of ha/β-tcp scaffold and mechanical strength optimization using a genetic algorithm method,” *Journal of the Australian Ceramic Society*, vol. 53, pp. 41–48, 2017.
- [109] V. R. Fernandes, C. M. Vicente, N. Wada, P. S. André, and R. A. Ferreira, “Multi-objective genetic algorithm applied to spectroscopic ellipsometry of organic-inorganic hybrid planar waveguides,” *Optics Express*, vol. 18, no. 16, pp. 16580–16586, 2010.
- [110] A. Kaczmarowski, S. Yang, I. Szlufarska, and D. Morgan, “Genetic algorithm optimization of defect clusters in crystalline materials,” *Computational Materials Science*, vol. 98, pp. 234–244, 2015.
- [111] M. Mahdavi Jafari, S. Soroushian, and G. R. Khayati, “Hardness optimization for al6061-mwcnt nanocomposite prepared by mechanical alloying using artificial neural networks and genetic algorithm,” *Journal of Ultrafine Grained and Nanostructured Materials*, vol. 50, no. 1, pp. 23–32, 2017.
- [112] S. Soltanali, R. Halladj, S. Tayyebi, and A. Rashidi, “Neural network and genetic algorithm for modeling and optimization of effective parameters on synthesized zsm-5 particle size,” *Materials Letters*, vol. 136, pp. 138–140, 2014.
- [113] R. Panwar, V. Agarwala, and D. Singh, “Design and experimental verification of a thin broadband nanocomposite multilayer microwave absorber using genetic algorithm based approach,” *AIP Conference Proceedings*, vol. 1620, no. 1, pp. 406–415, 2014.

- [114] M. F. Schubert, F. W. Mont, S. Chhajed, D. J. Poxson, J. K. Kim, and E. F. Schubert, “Design of multilayer antireflection coatings made from co-sputtered and low-refractive-index materials by genetic algorithm,” *Optics express*, vol. 16, no. 8, pp. 5290–5298, 2008.
- [115] A. Shanaghi, A. Sabour Rouhaghdam, S. A. Tabatabaei, M. Shariatpanahi, A. Rastgoo, A. K. Yousefi, and M. Samadzadeh, “An experimental investigation of corrosion resistance of mild steel by sol-gel process using TiO_2 nanostructure coating and prediction of optimal parameters,” *Proceedings of the Institution of Mechanical Engineers, Part N: Journal of Nanoengineering and Nanosystems*, vol. 227, no. 2, pp. 105–111, 2013.
- [116] A. Giunta, S. Wojtkiewicz, and M. Eldred, “Overview of modern design of experiments methods for computational simulations (invited),” *41st Aerospace Sciences Meeting and Exhibit*, 2003.

Appendix A

Pre-EMMA Experiments

gn	nr	enr	c_{zr}	λ	v_{DB}	T_{DB}	v_{cal}	T_{cal}	γ	ρ
-	146	-	1	10	10	20	2	400	33.82319	0.09804
-	150	-	3	5	10	20	2	400	22.45401	0.02941
-	150	-	3	5	10	20	2	400	8.389750	0.00000
-	151	-	3	2	10	20	2	400	42.35738	0.18182
-	152	-	4	2	10	20	2	400	47.78915	0.18000
-	153	-	4	3	10	20	2	400	19.82546	0.00000
-	154	-	4	4	10	60	2	400	25.19964	0.00000
-	156	-	2	3	10	60	2	400	73.22186	0.75862
-	157	-	2	3	10	60	2	400	83.55425	1.00000
-	158	-	2	3	10	60	2	400	55.98166	0.30612
-	160	-	3	4	10	60	2	400	118.11504	1.00000
-	186	-	2	5	10	20	2	400	75.53455	0.95652
-	187	-	1	10	5	50	2	400	43.48684	0.02222
-	188	-	1	4	0.1	20	2	500	40.46723	0.27586
-	190	-	5	10	0.1	20	2	500	93.74032	0.90476
-	192	-	4	6	14	60	4	500	44.52547	0.13636
-	194	-	5	4	5	20	8	400	90.27868	0.90000
-	195	-	1	5	5	80	2	500	84.67789	0.75000
-	198	-	5	10	5	80	8	500	94.74563	0.86486
-	199	-	4	6	14	60	4	500	93.97734	0.97368
-	201	-	4	6	14	60	4	500	46.08935	0.50000

Table A.1: gn – generation number; nr – experiment number; enr – EMMA serial number

Appendix B

EMMA Experiments

gn	nr	enr	c_{zr}	λ	v_{DB}	T_{DB}	v_{cal}	T_{cal}	γ	ρ
1	207	6113	2	10	20	40	2	500	24.32170	0.00000
1	208	2850	3	8	18	70	18	300	12.38955	0.00000
1	209	5526	3	6	10	50	18	400	29.16378	0.02083
1	211	6554	3	10	16	80	2	500	70.64908	0.78431
1	212	2947	4	6	16	80	18	300	10.95940	0.00000
1	213	8318	3	12	12	80	14	500	79.39680	0.86842
1	214	5648	5	6	10	60	18	400	51.09902	0.03030
1	217	7374	3	10	14	50	10	500	32.86985	0.00000
1	216	2700	5	12	14	60	18	300	15.64646	0.03846
1	218	3956	5	10	20	60	6	400	45.50493	0.02778
2	220	6101	2	4	20	40	2	500	71.25290	0.95652
2	222	7201	2	4	10	40	10	500	108.47050	1.00000
2	223	1462	3	4	12	60	10	300	39.68027	0.12500
2	224	2883	4	4	10	80	18	300	46.44359	0.03030
2	225	1680	5	12	20	70	10	300	49.18166	0.00000
3	226	0001	2	4	10	40	2	300	38.79620	0.05882
3	227	6001	2	4	10	40	2	500	47.17184	0.33333
3	228	6102	3	4	20	40	2	500	76.44246	1.00000
3	229	2884	5	4	10	80	18	300	42.66628	0.00000
3	231	0360	5	12	20	60	2	300	47.64474	0.00000
4	234	4202	3	4	10	40	10	400	61.73920	0.38710
4	235	6105	2	6	20	40	2	500	52.13835	0.03571
4	237	1500	5	12	14	60	10	300	47.53164	0.00000
4	238	1486	3	6	14	60	10	300	28.56022	0.00000
4	239	2923	4	4	14	80	18	300	52.18121	0.09091
5	240	2971	4	8	18	80	18	300	36.17699	0.06061
5	241	2530	3	8	10	50	18	300	36.54716	0.04167
5	242	3077	2	12	16	40	2	400	43.19611	0.00000
5	243	2733	2	10	18	60	18	300	36.16643	0.04348
5	244	2535	4	10	10	50	18	300	47.75640	0.00000

Table B.1: gn – generation number; nr – experiment number; enr – EMMA serial number