# 学 位 論 文 内 容 の 要 旨

博士の専攻分野の名称　博士（理学）　氏名　GARCIA ESCOBAR　Fernando

## 学 位 論 文 題 名

Development of a Catalyst Informatics Pipeline for Heterogeneous Catalyst Design
（不均一系触媒設計のための触媒インフォマティクス・パイプラインの開発）


Catalysts are essential in modern chemistry and industry, enhancing reaction efficiency and enabling processes under milder conditions. They are widely used in applications such as fuel refining, pharmaceuticals, and energy and other raw materials production. Recent advances in Catalyst Design have seen the adoption data-driven methodologies, big data analysis and high-throughput experimentation to accelerate the discovery of efficient catalyst materials. However, the integration of Informatics in Catalyst Design is still in an early stage, with challenges in establishing general workflows and representing catalysts as input Descriptor variables in predictive models. Specifically, the design and selection of Descriptor variables has not yet been generally established.

This thesis proposes a systematic workflow for screening catalyst candidates by designing and selecting compositional Descriptor variables for Multiple Linear, Random Forest, and Support-Vector regression models. The workflow incorporates Descriptor creation and engineering to represent catalysts, proposes the MonteCat algorithm as an efficient Descriptor selection strategy to select for optimal Descriptor variables, incorporates predictive modeling to evaluate catalyst performance for the Water-Gas Shift (WGS) and Oxidative Coupling of Methane (OCM) reactions and proposes potential catalyst candidates with experimental validation for low Temperature OCM catalysts.

Chapter 1 introduces the research context, objectives, and structure of the thesis. Chapter 2 provides a literature review, highlighting the challenges in WGS catalysts and the potential of OCM catalysts. To address these issues, a workflow centered on constructing regression models from literature and experimental data to predict catalyst performance is proposed. Chapter 3 describes the methodology followed in further detail, namely the regression models explored, feature design to represent catalysts in predictive models, feature selection methodologies, the model validation strategy and the inverse prediction approach.

Chapter 4 explores catalyst representation for the Water-Gas Shift (WGS) reaction using a data set from reported literature. Two distinct catalyst representation schemes are explored: one-hot encoding and continuous compositional features calculated from the catalyst compositions. This information, along with reaction conditions are used to construct regression models to predict catalyst performance and search for high conversion catalyst candidates through an inverse prediction.

Chapter 5 introduces the MonteCat algorithm, an optimization strategy inspired by Metropolis-Hastings and Basin Hopping algorithms, to enhance feature selection. MonteCat is used to construct feature subsets from a bank of over 3,000 features, resulting in high-performing predictive models for the OCM reaction. Chapter 6 applies the MonteCat algorithm to propose and experimentally validate OCM catalyst candidates. All tested materials demonstrated improved catalytic performance compared to the bare support.

Chapter 7 concludes with the establishment of a workflow for catalyst performance prediction and a Descriptor search algorithm that successfully proposed and validated catalyst materials for the OCM reaction. Future research should focus on integrating advanced Descriptors with reactivity information to enhance model interpretability and performance further.