
Article:
Yeardley, A. orcid.org/0000-0001-7996-0589, Milton, R., Moghadam, P. orcid.org/0000-0002-1592-0139 et al. (2 more authors) (2022) Active subsets as a tool for structural characterisation and selection of metal-organic frameworks. Chemical Engineering Research and Design, 179. pp. 424-434. ISSN 0263-8762

https://doi.org/10.1016/j.cherd.2022.01.045

Active Subsets as a Tool for Structural Characterisation and Selection of Metal-Organic Frameworks

Aaron S. Yeardley, Robert A. Milton, Peyman Z. Moghadam, Joan Coordiner, Solomon F. Brown*

Department of Chemical and Biological Engineering, The University of Sheffield, Sheffield S10 2TN

Abstract

To date over 80000 metal-organic framework (MOF) structures have been synthesised and only *ca.* 3% of these have had their adsorption capabilities measured for storing oxygen alone. As such, in order to aid the process of producing top-performing MOFs for storing various gases, accurate methods to predict the deliverable capacity of MOFs that have their synthesis method already known is increasingly important. For this purpose, this paper develops a reduced order model (ROM) that can predict the deliverable capacity of synthesised MOFs irrespective to the storage gas across similar gases.

The ROM is constructed by identifying the Active Subspaces through a Sobol' index-based global sensitivity analysis (GSA). The resulting Gaussian Process (GP) regression model efficiently predicts the deliverable capacity given a MOFs pore properties with this reduced dimensional space.

This approach was applied to a practical MOF exploration example by training a ROM with 2745 MOFs storing methane at 30 bar. The ROM was robustly tested and analysed before using it to predict the deliverable capacity of 82221 synthesised MOFs storing oxygen at 30 bar. To ensure validity in the exploration example, the predictions produced from the methane trained ROM were compared to a separate ROM trained using the same MOFs but storing oxygen gas. The methane trained ROM was found to be in agreement with the oxygen trained ROM, and was shown to be a viable tool to identify the top-performing MOF structures for oxygen storage.

Keywords: Metal-Organic Frameworks, Active Subspaces, Reduced Order Modelling, Gaussian Process, Sobol' Indices, Gas Adsorption

1. Introduction

Metal-organic frameworks (MOFs) (Carraro and Gross, 2014; Moghadam et al., 2017b) are a unique class of crystalline porous polymers that allow the design of pore structure and functionality due to their self-assembly synthesis

*Corresponding author.
E-mail address: s.f.brown@sheffield.ac.uk

process where metal building units are bridged by organic ligands. The nature of MOFs excites scientists due to their tailorable structural properties allowing a large number of MOFs to be potentially synthesised (Wilmer et al., 2012; Moghadam et al., 2017b). Consequently, the flexibility of MOFs allows them to be applied in a multitude of industrial settings such as gas storage (Mason et al., 2015; Thornton et al., 2017; Tian et al., 2018) and separation (Li et al., 2009; Bobbitt et al., 2017; Moghadam et al., 2017a).

To date *ca.* 88000 (Moghadam et al., 2017b) structures have been synthesised imposing difficulties in identifying top materials for each purpose. Such large datasets are often analysed through machine learning techniques and these methods have already shown accurate predictions for various gas uptake capacities of MOFs (Fernandez et al., 2013b; Ohno and Mukae, 2016; Fanourgakis et al., 2019). Examples of such techniques include polynomial chaos expansion (Sudret, 2008; Brown et al., 2013), artificial neural networks (Li et al., 2016), and Gaussian Processes (GPs) (Marrel et al., 2009; Yeardley et al., 2020, 2021). These machine learning methods and high-throuput screening (Wilmer et al., 2012; Banerjee et al., 2016; Moghadam et al., 2019) aid in speeding up the process of finding new materials.

Previous studies have shown that machine learning methods enable the discovery and optimal design of MOFs (Moghadam et al., 2019). Fernandez et al. (2013b) compared three machine learning models to predict the storage capacity of MOFs storing methane. The research used 1.3×10^5 hypothetical MOFs at 1, 35 and 100 bar using the pore size, surface area, void fraction and framework density as input variables to multi-linear regression models, decision trees and nonlinear support vector machines. Soon after, Fernandez et al. (2014) furthered the research by introducing the atomic property weighted radial distribution function descriptor as an input variable in addition to the MOFs geometric features. This enables the models to capture the chemical features of a periodic material. The results suggested that more compact MOFs with interatomic ranges from 6 to 9 Å produce higher storage capacities for CO_2 at 0.15 and 1 bar. The first GP applied towards MOFs aided in predicting the methane uptake using the pore properties and the structural relationship as input variables (Ohno and Mukae, 2016). An automatic relevance determination (ARD) kernel calculated the lengthscales to discover the density, volumetric surface area and the void fraction to be the most dominant input variables. The GP surrogate model was then used to successfully identify new MOFs that could outperform them used in the training dataset for the methane uptake at 3.5 MPa. Further work using decision trees and support vector machines (Aghaji et al., 2016) found pore size, void fraction and surface area to be the most important factors when designing MOFs for CO_2 uptake and methane purification. Additionally, research has used the random forest technique to discover new MOFs for separation of xenon and krypton (Simon et al., 2015) and both CO_2 and N_2 uptake (Fernandez and Barnard, 2016). However, the biggest drawback with all the previous research is that only hypothetical MOF data were used to train the machine learning models and the synthesis of top-performing materials is not always straight-forward. For databases containing MOF structures that

are already synthesised Fanourgakis et al. (2019) used the Computation-ready, experimental (CoRE) MOFs (Chung et al., 2014), which have been experimentally synthesised with accurate calculations of their structural features. This research used a random forest algorithm to investigation to predict the methane adsorption in MOFs, discovering a smaller training set size were needed for convergence when using synthetic MOFs compared to the hypothetical MOF databases previously used.

In this study, we have adopted a GP surrogate model that directly emulates a dataset storing experimentally synthesised MOFs with their pore properties computationally measured through simulations (Moghadam et al., 2018). The use of synthesised MOFs has a significant advantage over hypothetical MOFs because knowing the synthesis method can accelerate the production process once top-performing MOFs have been identified (Chung et al., 2014; Thornton et al., 2017). Given the complexity of the data set, a versatile, non-parametric GP approach allows characterisation of the data, while also allowing the semi-analytic evaluation of the Sobol' indices that would otherwise be calculated at significant cost. As such we use the GP surrogate to predict the gravimetric deliverable capacity (GDC) and volumetric deliverable capacity (VDC) of MOFs given their pore properties. Initially, global sensitivity analyses (GSAs) are conducted at various pressures to understand the influence each pore property has on the deliverable capacity at different storage pressures. Then an efficient comparison between MOFs storing oxygen and methane enables the production of a ROM for property prediction.

Previous research studies investigating the design of MOFs identified the important variables without quantifying the impact of each variable. To the authors knowledge only Ohno and Mukae (2016) conducted a sensitivity analysis by calculating qualitative values for the most dominant input variables. Here, our contributions allow the further understanding of MOFs by developing a GP surrogate model that accurately predicts the deliverable capacity of MOFs. Additionally, the model investigates the effect each pore property has on the gaseous uptake of MOFs using GSAs.

Principal component analysis techniques have been widely applied in MOF research (Qiao et al., 2018, 2017; Fernandez et al., 2013a) to assess the interrelationships among MOF descriptors. A principal component analysis allows the dimensionality of multivariate data to be reduced providing a concise representation of data. Qiao et al. (2017) performed a principal component analysis combining four descriptors to make a two-dimensional model to help predict the performance of MOFs for thiol capture. This method of projecting four descriptors into a two-dimensional model was conducted again to aid in separating natural gas (Qiao et al., 2018). In this research, we have developed a stronger technique based on Active Subspaces (Constantine et al., 2014; Constantine and Diaz, 2017) which utilises the GSA to build a reduced order model (ROM). Overall, we present the ROM to be used as a prediction technique which enables fast and efficient material design and discovery for gas storage applications due to the ability to predict

irrespective to whether the MOF is storing oxygen or methane.

The novel approach presented here will allow us to determine: *i).* The most dominant pore properties and their interactions *ii).* The effect pressure has on the most influential pore properties in MOFS *iii).* How the pore properties influence the uptake of oxygen storage in comparison to methane storage and *iv).* A fast and efficient prediction technique for the gravimetric and volumetric deliverable capacity of synthesised MOFs irrespective to the storage gas across similar gases.

2. Mathematical Background

2.1. Gaussian Process Regression

GP regression is a non-parametric machine learning method used to create a blackbox function that predicts the deliverable capacity given a MOFs pore properties. Bayesian conditioning (Williams and Rasmussen, 2006) is used to learn the mean $\left(\bar{f}(\mathbf{x})\right)$ and variance (Σ_f) using a MOF database which includes paired pore properties and measured deliverable capacities as training data $(\mathbf{y} = f(\mathbf{X}) + \mathbf{e})$. Once trained, the GP takes a $(1 \times d)$ row vector as inputs $\mathbf{x}$ to predict output as a Gaussian random variable using the predictive equations

$$y(\mathbf{x}) \sim \mathsf{N}\left[\bar{f}(\mathbf{x}), \Sigma_f + \sigma_e^2\right] \tag{1}$$

where

$$\bar{f}(\mathbf{x}) := k(\mathbf{x}, \mathbf{X})(k(\mathbf{X}, \mathbf{X}) + \sigma_e^2 \mathbf{I})^{-1}\mathbf{y} = k(\mathbf{x}, \mathbf{X})\mathbf{K}^{-1}y \tag{2}$$

$$\Sigma_f := k(\mathbf{x}, \mathbf{x}) - k(\mathbf{x}, \mathbf{X})(k(\mathbf{X}, \mathbf{X}) + \sigma_e^2 \mathbf{I})^{-1}k(\mathbf{X}, \mathbf{x}) = k(\mathbf{x}, \mathbf{x}) - k(\mathbf{x}, \mathbf{X})\mathbf{K}^{-1}k(\mathbf{X}, \mathbf{x}) \tag{3}$$

At the heart of this lies the kernel function $k\colon \mathbb{R}^{i+d} \times \mathbb{R}^{j+d} \to \mathbb{R}^i \times \mathbb{R}^j$, expressing the correlation between responses to input samples of sizes $(i \times d)$ and $(j \times d)$. This work exclusively uses the automatic relevance determination (ARD) kernel (Wipf and Nagarajan, 2007):

$$k(\mathbf{x}', \mathbf{x}) := \sigma_f^2 \exp\left(-\frac{(\mathbf{x} - \mathbf{x}')^\intercal \Lambda^{-2} (\mathbf{x} - \mathbf{x}')}{2}\right) \tag{4}$$

where Λ is a $(d \times d)$ *diagonal* positive definite lengthscale matrix. The learning from training data requires the optimisation $d + 2$ hyperparameters, consisting of Λ, σ_f, and σ_e, through the maximum marginal likelihood $p[\mathbf{y}|\mathbf{X}]$ using the ROMCOMMA software library (Milton and Brown, 2019).

2.2. *Global Sensitivity Analysis*

This works implements a GSA using the variance based Sobol' indices technique (Sobol, 1993, 2001). The calculation of Sobol' indices requires the evaluation of complex multi-dimensional integrals. Here, we shall describe the calculation of Sobol' indices up to evaluating the integrals, which are calculated using the GP regression model resulting in semi-analytic Sobol' indices.

Sobol' indices are calculated by considering a function $y = f(\mathbf{x})$, where $\mathbf{x} \coloneqq [x_1, ..., x_d]$ is a d-dimensional row vector found in the input space, Ω, and y is the model output. Assuming that the inputs are mutually dependent and that $f(\mathbf{x}) \in L^2(\Omega)$ (Sobol, 1993, 2001), then a particular input x_i has a first-order Sobol' index defined by

$$S_{1,i} = \frac{Var\left[E\left[\bar{f}(\mathbf{x})|x_i\right]\right]}{Var\left[\bar{f}(\mathbf{x})\right]} \tag{5}$$

To be able to express the whole effect of an input on the output, the total Sobol' index is

$$S_{T,i} = S_{1,i} + \sum_{j \neq i}^{n} S_{2,ij} + \sum_{j \neq i, k \neq i, j < k}^{n} S_{3,ijk} + ... \tag{6}$$

Therefore, the first-order Sobol' indices measure the contribution to the variance solely attributable to x_i, in contrast, the total Sobol' index of i corresponds to its own contribution including interactions with the other inputs (Saltelli and Homma, 1996; Saltelli et al., 2008).

From here, the partial variances of y are determined through a decomposition method presented by Sobol (1993) which evaluates each term through multi-dimensional integrals. Here, we compute the semi-analytic evaluation of the integrals using the GP regression model's predicted mean in Equation (2). Therefore, we have used GPs as a blackbox function providing an efficient sampling method for $y = f(\mathbf{x})$ in a similar way to that from Chen et al. (2005).

2.3. *Reduced Order Modelling*

The ROM conducts an optimal dimension reduction by locating the Active Subspaces (Constantine et al., 2014; Constantine and Diaz, 2017). The ROM reduces the inputs to a set which is highly relevant to the response through the use of Sobol' indices. It is achieved by finding an orthogonal rotation matrix θ which combines the inputs into a low dimensional subspace. The methodology is achieved by rotating the input basis:

$$\mathbf{x} \coloneqq \mathbf{\Theta}^{\intercal}\mathbf{u} \tag{7}$$

Then by calculating the proportion of response variance ascribable to the first m basis directions of $\mathbf{u}$ through the

Sobol' index, the optimum rotation can be determined by maximising this relevance.

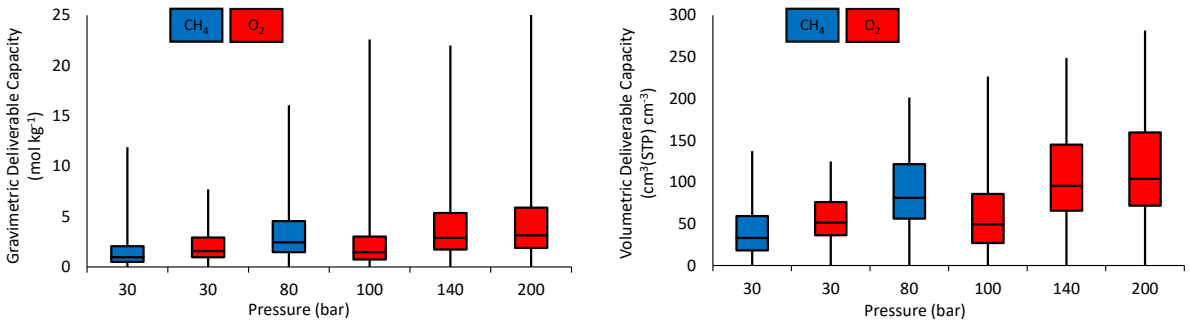
$$S_m(\mathbf{\Theta}) := \frac{Var\left[E\left[\bar{f}(\mathbf{x})|\mathbf{\Theta}\mathbf{x}\right]_m\right]}{Var\left[\bar{f}(\mathbf{x})\right]} \tag{8}$$

Where the dimension $m << d$. Therefore, S_m has to be optimised for $m = 1, ..., d$ in turn, to find the most relevant direction, then the second most relevant, and so on. Essentially, optimization is then a two-stage process: 1) a GP predictor is optimised, then GSA using the predictor rotates the input basis; 2) this rotated basis is then used in a new predictor. Ultimately this yields a predictor that only significantly depends on a dimensionally reduced basis.

3. Method

In this section, we outline how the mathematical tools presented in Section 2 are applied to 2745 structures of synthesised MOFs. Understanding the MOF data is essential in the development of the methodology for creating an efficient predictive technique.

The MOF database used in this study is taken from Moghadam et al. (2018). The force field parameters for oxygen were taken from the TraPPE force field (Zhang and Siepmann, 2006) which is a three-site model. Using the TraPPE model, Moghadam et al. (2018) have obtained excellent agreement with measured oxygen adsorption isotherms for a number of MOFs including UMCM-152, which demonstrates the accuracy of the calculations (Moghadam et al., 2018). In this work, the deliverable capacity is defined as the difference between the amount of gas adsorbed at the storage pressure and the release pressure of 5 bar. For each storage gas and pressure, the descriptive statistics of the deliverable capacity for all 2745 MOF structures are shown in Figure 1 as box and whisker plots. The box in each plot bounds the first and the third quartile showing the median value as the middle line. The whiskers show the minimum and maximum deliverable capacity values. This research uses the deliverable capacity of each MOF as the important output value to analyse.



(a) (b)

Figure 1: Box and whisker plots for the descriptive statistics of the a) gravimetric deliverable capacity and b) volumetric deliverable capacity for all 2745 MOF structures used in this research.

An ideal MOF structure would have a large deliverable capacity, which likely depends on several pore properties including the largest cavity diameter (LCD), volumetric surface area (A_V), gravimetric surface area (A_G), density (ρ), void fraction (VF), and heat of adsorption at 5 bar (H5) relative to the storage gas. The descriptive statistics showing the spread of values for each pore property throughout the 2745 MOF structures are shown in Table 1. Please note, that although the density of each MOF was available, it was not used in calculations due to its dependency with A_G and A_V. Here we focus on using these pore properties and input variables to predict the deliverable capacity of MOFs. Further, these will be analysed using a GSA to understand the influence of each pore property on the deliverable capacity.

Table 1: The pore properties available for use as input variables for a GP.

Pore Property	LCD	A_V	A_G	VF	H5 of O_2	H5 of CH_4	ρ
Units	Å	$m^2\ m^{-3}$	$m^2\ g^{-1}$	-	kJ mol^{-1}	kJ mol^{-1}	g cm^{-3}
Mean	5.70	651	616	0.322	17.5	20.7	N/A
STD	2.09	727	828	0.183	3.63	5.50	N/A
Range	15.4	3096	5203	0.880	31.5	36.2	N/A

The deliverable capacities were available at different pressures; however, they were only measured at 30 bar for both oxygen and methane storage. Therefore, the GSA investigating the deliverable capacity at various pressures was conducted using MOFs storing oxygen. Subsequently, we investigated the effect which changing the storage gas has on the GSA and so kept the storage pressure constant at 30 bar.

GP regression is used to encapsulate the MOF data into a simple framework that can be used to make predictions and to enable the semi-analytic calculation of Sobol' indices. In this research, we are interested in both the gravimetric and volumetric deliverable capacity, and so for each scenario (storage pressure or storage gas), we have to produce two independent GP regression models. The models are regressed using standardised data and so the mean and standard deviations for the input-output data used for training were saved for standardising the test data. GP predictions work by taking a (1×5) row vector of pore property input values to predict the distribution of the output. To do this, the hyperparameters are optimised using the ROMCOMMA software library (Milton and Brown, 2019) where learning is achieved using the MOF database as training data. From there, the predictive equation is used for semi-analytic calculations of Sobol' indices once the GP models have been tested for inaccuracies.

The predicted deliverable capacity of MOFs are tested by comparing them to observed deliverable capacities and to ensure inaccuracies from the GP models are not inherently produced in the GSA. For this reason, 5-fold cross-validation (Hastie, 2009) is used in this research, whereby, the MOF dataset is randomly split into five subsets. Then

four of the five subsets are used as training data, while the remaining is used to test the GP model predictions. This procedure is repeated so that every single sampling point produced is used for testing. Once the GP models have been created, tested, and used for calculation of Sobol' indices, the process can be repeated with rotations and iterations as described in the previous section to produce a ROM.

4. Results

This Section presents the GSA results when comparing both the storage pressure and the storage gas. First, the pressure analysis investigates the impact each pore property has on oxygen storage at four various pressures. Then two GSAs are conducted, as described in the previous Section, to explore the impact that the storage gas has on the deliverable capacity of MOFs at 30 bar. For both studies, the GP models are assessed using cross-validation (Hastie, 2009) to ensure inaccuracies are both carried through to the semi-analytic calculation of Sobol' indices. Finally, the ROM's are built and we present the results to show the predictive capabilities even when the MOFs storage gas is unknown.

4.1. Sensitivity Analysis of Pressure

A GSA of the five pore properties with respect to the deliverable capacity at various pressures has been achieved using a GP regression model. Therefore, confidence in the resulting Sobol' indices depends entirely on the accuracy of the GP predictions. This accuracy was determined using 5-fold cross-validation so that each test point had not been used in training the GP model. The residuals for both the GDC and VDC are presented in Figure 2, comparing the true deliverable capacities to the predicted mean values for all 2745 MOFs storing oxygen at four pressures. The markers in Figure 2 lie close to the black $y = x$ trend line showing good test predictions. Additionally, the standardised diagnostic values are presented in Table 2, quantitatively showing the marks closeness to the trend line through high coefficient of determination values (r^2). The good predictions are given further evidence through low standardised root mean square error (RMSE) values and a predicted distribution shown to be normally distributed as close to 5% of the true standardised deliverable capacity values are outside of the predictions uncertainty distribution. Overall, the high predictive accuracy shown from the GP models enables confidence in the calculation of Sobol' indices.

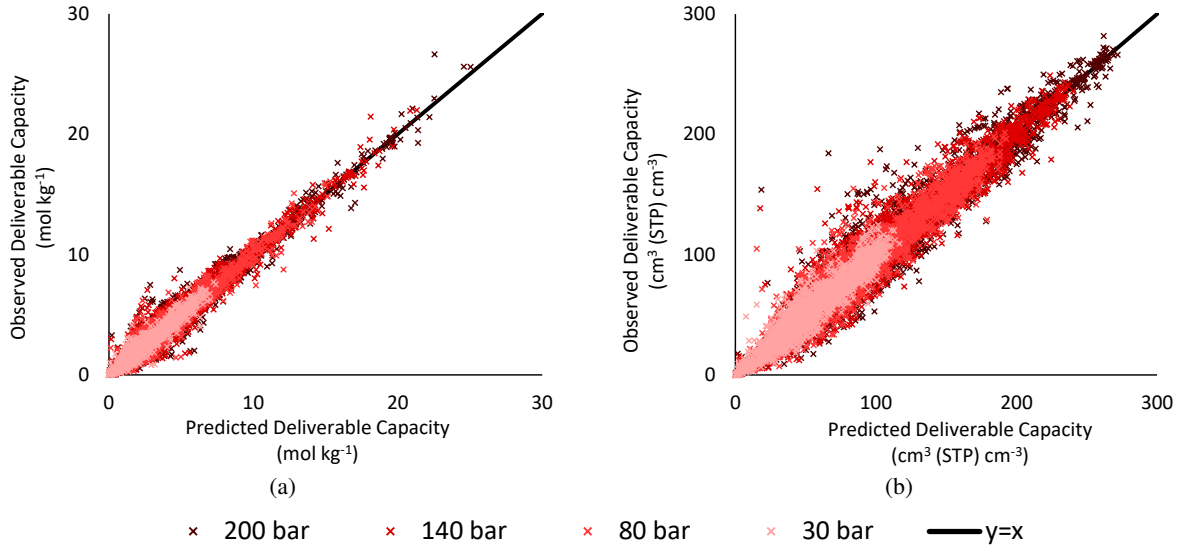


Figure 2: The residuals from testing the GPs predicting a) the gravimetric deliverable capacity and b) the volumetric deliverable capacity of MOFs storing oxygen at various pressures using 5-fold cross-validation.

Table 2: The diagnostic values of the GPs predicting the deliverable capacity of MOFs at various pressures.

Deliverable Capacity	Pressure (bar)	r^2	RMSE	Outliers	Max Error	MOF ref code
GDC	30	0.945	0.234	5.21%	2.53 mol kg^{-1}	NICZAA01
GDC	80	0.960	0.201	5.61%	3.86 mol kg^{-1}	NICZAA01
GDC	140	0.965	0.187	5.50%	4.39 mol kg^{-1}	NICZAA01
GDC	200	0.967	0.181	5.46%	4.64 mol kg^{-1}	NICZAA01
VDC	30	0.912	0.297	5.28%	45.4 cm^3(STP) cm^{-3}	NICZAA01
VDC	80	0.924	0.275	4.99%	89.9 cm^3(STP) cm^{-3}	LIHRIE
VDC	140	0.929	0.266	4.74%	121 cm^3(STP) cm^{-3}	LIHRIE
VDC	200	0.931	0.262	4.55%	136 cm^3(STP) cm^{-3}	LIHRIE

The first GSA of this research investigates the impact each pore property has on the deliverable capacity of synthesised MOFs at different pressures. The calculated Sobol' indices are presented in Figure 4 as total Sobol' indices split to show the first order values at the bottom plus the remaining indices due to interactions with other pore properties at the top. In Figure 3a we can see the impact each pore property has on the GDC, whereas the output for Figure 3b is the VDC. A direct comparison of the storage pressure can be seen for each pore property as the pressure changes from 30 bar on the left to 200 bar on the right as shown by the darkening colours highlighted in Figure 3's legend.

A comparison between Figure 3a and Figure 3b shows the Sobol' index value for gravimetric surface area for the GDC to significantly increase by almost twenty times compared to what it was for the VDC. Therefore, gravimetric surface area is the most dominant pore property impacting the GDC but has negligible impact on the VDC. Whereas, the void fraction's Sobol' indices are four times larger for the VDC compared to the GDC. For either deliverable capacity and all four storage pressures the heat of adsorption at 5 bar and the largest cavity diameter both have first-order Sobol' indices of 0.1 or below and so these pore properties have a negligible impact on MOFs oxygen storage capabilities.

Interestingly, the increasing pressure does causes the impact of the heat of absorption at 5 bar to decrease. However, the importance of the remaining pore properties either increase for the GDC or decrease for the VDC. This may be due to the correlation between outputs or, alternatively, it could simply be a coincidence drawn from the uncertainty in the calculated Sobol' indices. Until further research, the minor changes between pressures are considered to be equivalent and so these findings demonstrate that the influence each pore property has on the MOFs oxygen storage capability in not impacted by the storage pressure.

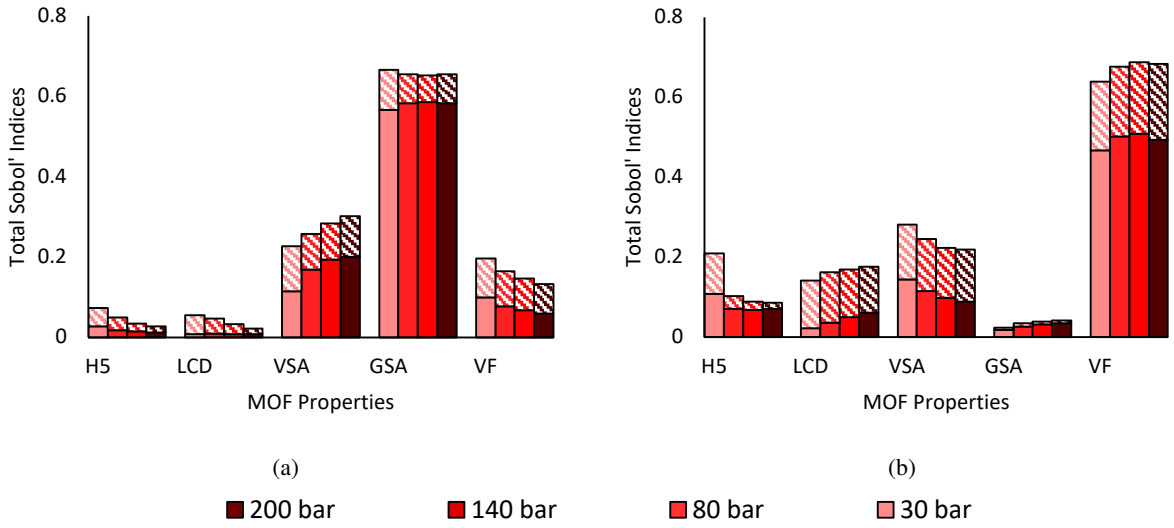


Figure 3: The bar charts showing the calculated total Sobol' indices for each pore property with respect to a) the gravimetric deliverable capacity and b) the volumetric deliverable capacity of MOFs storing oxygen at 30, 80, 140 and 200 bar. Each bar is split with the bottom solid fill corresponding to the first-order Sobol' indices and the top diagonal pattern being the cross effects.

4.2. Sensitivity Analysis of Storage Gas

This Section presents the findings and contributions made when comparing the GSAs made for MOFs storing different gases. Once again, the GP regression model had to be fully validated to ensure trust in the GSAs. Following the 5-fold cross-validation each model had the standardaised error diagnostic values calculated and presented in Table 3.

Interestingly, Table 3 show the maximum prediction error for the GP predicting the VDC of MOFs storing methane to be LIHRIE which is the same MOF causing the largest prediction errors from storing oxygen at 80, 140 and 200 bar as shown in Table 2. Whereas, a GP model for the GDC of methane at 30 bar results in a maximum error for a different MOF structure. Here, Table 2 shows HIFTOG has the largest error of -6.38 mol kg^{-1}. Anomalies such as these often occur in machine learning because the input variables are either outside of the range of the training data used. Alternatively, the input variables could be close to them from training data but have a different output value. Analysis of the MOF structures causing anomalies has shown that the input variables were within the range used for training. Therefore, the MOF is behaving differently to them with similar pore properties. The reason for the difference in the behaviour of these MOF outliers is unclear at this point as there is no common characteristic of these MOFs. In future work, it would be interesting to conduct further chemical analysis to understand why these MOFs behave differently. Instead, this research focuses on a methodology that provides statistical results, validating the GPs showing predictions that enable accurate modelling of synthesised MOFs.

Table 3: The diagnostic values of the GPs predicting the deliverable capacity of MOFs storing oxygen and methane at 30 bar.

Deliverable Capacity	Storage Gas	r^2	RMSE	Outliers	Max Error	MOF ref code
GDC	O_2	0.945	0.234	5.21%	-2.54 mol kg^{-1}	NICZAA01
GDC	CH_4	0.943	0.239	5.90%	-6.38 mol kg^{-1}	HIFTOG
VDC	O_2	0.912	0.297	5.28%	-45.9 cm^3(STP) cm^{-3}	NICZAA01
VDC	CH_4	0.897	0.321	4.92%	-61.1 cm^3(STP) cm^{-3}	LIHRIE

The GSAs conducted for MOFs storing both oxygen and methane at 30 bar produced total Sobol' indices presented in Figure 4, a bar chart similar to Figure 3. The storage gas is compared using the colour of each bar with red representing oxygen and blue representing methane. As expected, both this GSA and the previous GSA calculate identical values for the Sobol' indices of MOFs storing oxygen at 30 bar. Together, this supports the reliability of the results, showing the clear impact each pore property has on the capability of MOFs storing oxygen.

Comparing each GSA demonstrates many similarities between the effects each pore property has depending on the storage gas. Figure 4a shows the gravimetric surface area to dominate the GDC for both storage gases whereas the Sobol' index value for the heat of adsorption at 5 bar increases by more than double when storing methane compared to oxygen for both deliverable capacities. It is important to highlight the first-order Sobol' indices for the heat of adsorption at 5 bar only increases by 24% for GDC and 53% for VDC. This implies that the significant increase ($> 200\%$) in the total Sobol' index is associated with the heat of adsorptions interactions with other pore properties. This phenomenon is clear to see for the largest cavity diameter as the cross effects bar increases significantly from

blue (oxygen storage) to red (methane storage) in Figure 4.

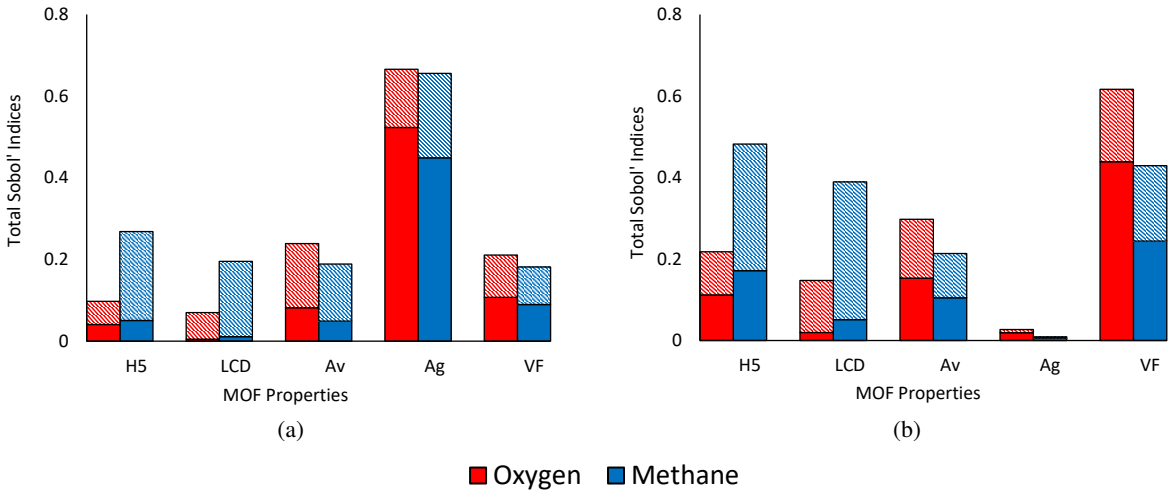


Figure 4: The bar charts showing the calculated total Sobol' indices for each property with respect to a) the gravimetric deliverable capacity and b) the volumetric deliverable capacity of MOFs storing oxygen (red) and methane (blue). Each bar is split with the bottom solid fill corresponding to the first-order Sobol' indices and the top diagonal pattern being the cross effects.

Overall, the GSAs investigating the storage gas successfully quantified the effect each pore property has on both the deliverable capacities of MOFs storing either oxygen or methane. The results show that the interactions between the pore properties have a larger effect on the MOFs capabilities when storing methane compared to oxygen.

It can be seen that the relevance of the pore properties vary depending on whether the deliverable capacity is gravimetric or volumetric. For the GDC, the Sobol' index value for the gravimetric surface area is 66.6% when storing oxygen and 65.6% when storing methane. Whereas, the void fraction dominates the VDC with Sobol' indices of 61.7% when storing oxygen and 42.9% when storing methane. Determining these two as the most important pore properties are consistent with the qualitative findings in the literature (Ohno and Mukae, 2016).

These GSAs have demonstrated the impact of the pore properties has on the MOFs deliverable capacity is the same whether it is storing oxygen or methane. Overall, the pore properties are the input variables used to predict the deliverable capacity, therefore, this provides the potential to predict the deliverable capacity of MOFs given limitations on the training data used, whether the database consists of MOFs storing oxygen or methane.

4.3. ROM Predictor Capabilities

The previous sections have performed GSAs investigating the impact the storage gas and storage pressure has on the effect of each pore property on MOFs deliverable capacity. This study has shown that the MOFs capabilities are largely unaffected by the storage capacity. Therefore, in this Section we have applied the novel GP based ROM

technique to the MOF database, testing the methods capabilities by predicting the deliverable capacity of a MOF storing the opposite gas to that the training database stored. Once again, 5-fold cross-validation was used to ensure any MOF structures used in training the ROM is not used in testing the ROM. For example, a ROM is trained using four folds of MOF structures storing oxygen before it is tested by predicting the deliverable capacity of the remaining MOFs but storing methane.

Individually, four ROMs were trained and tested producing satisfactory prediction results. Figure 5 and Table 4 demonstrate these results in two residual plots and a table showing the standardised diagnostic values. The results are colour coordinated to show the predicted MOFs storage gas. For example, the blue residuals in Figure 5 show the predicted deliverable capacity of MOFs storing methane produced from a ROM trained using MOFs storing oxygen. Having a large correlation coefficient (r^2) and a low percentage of outliers outside two predicted standard deviations provide evidence for our novel ROMs to be used with confidence to predict the deliverable capacity of MOFs storing another gas to what was used for training.

Table 4 highlights the MOF with the maximum error for each ROM showing slight inconsistency with the worst prediction. This is in comparison to previous GP validation tests where Table 3 and Table 2 show NICZAA01 and LIHRIE to cause the worst predictions. Further, Figure 5 clearly shows lower deliverable capacities to be easier to predict as predictions begin to deviate away from the $y = x$ black trend line in both residual plots as the deliverable capacity increases. We believe this could be due to the general population of MOFs used for training being able to store more methane than oxygen on average. Further, the majority of these outliers begin to occur as the heat of absorption for MOFs storing oxygen become closer to the heat of absorption for MOFs storing methane. As the heat of absorption is the only input variable that changes when predicting the uptake of each gas, then it becomes understandable why the ROM prediction capabilities become less accurate when the MOFs have a higher deliverable capacity for methane than oxygen, but a similar heat of absorption. Therefore, as the deliverable capacities of the MOFs increase, the majority have a larger value when storing methane than oxygen, causing minor errors in the ROM when predicting an unknown gas. Hence, there are limitations to using the ROM to predict the deliverable capacity of MOFs at storage pressures greater than 30 bar. At this stage, we believe that as the pressure increases, being able to predict the deliverable capacity irrelevant to the storage gas would become significantly more difficult. Nonetheless, we believe developing this novel technique and testing it on synthesised MOFs storing just oxygen and methane builds on previous MOF prediction studies by being an initial step towards developing a universal model applicable to synthesised MOFs.

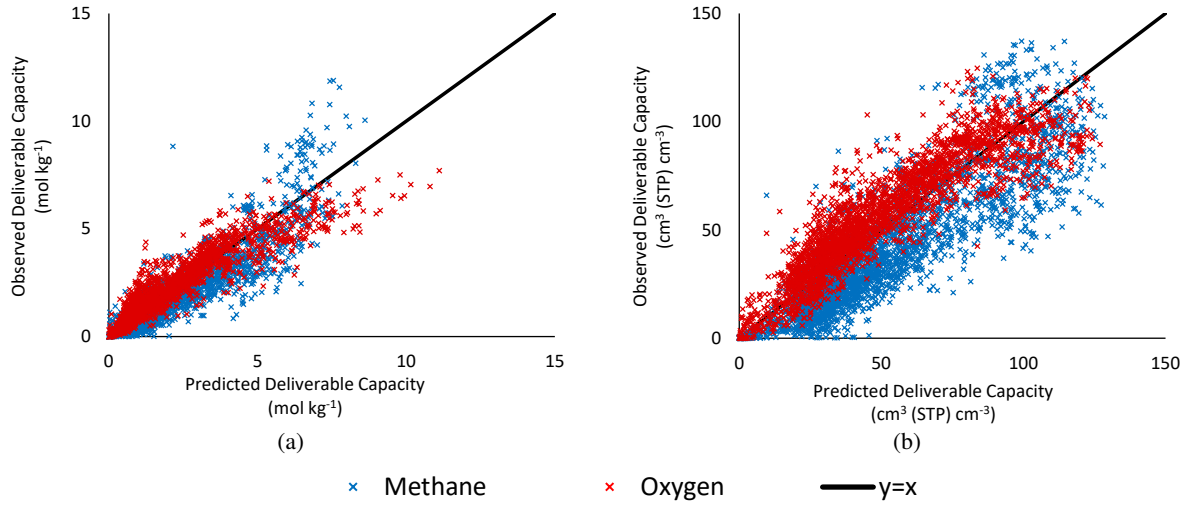


Figure 5: The residuals from testing the ROMs using 5-fold cross-validation to predict a) the gravimetric deliverable capacity and b) the volumetric deliverable capacity, showing the predictions of MOFs storing both oxygen (red) and methane (blue).

Table 4: The diagnostic values of the ROMs predicting the deliverable capacity of MOFs storing oxygen and methane at 30 bar.

Deliverable Capacity	Storage Gas	r^2	RMSE	Outliers	Max Error	MOF ref code
GDC	O_2	0.865	0.373	2.66%	3.83 mol kg^{-1}	XEBHOC
GDC	CH_4	0.826	0.546	4.74%	6.68 mol kg^{-1}	HIFTOG
VDC	O_2	0.829	0.417	1.35%	58.0 cm^3(STP) cm^{-3}	NICZAA01
VDC	CH_4	0.764	0.710	6.34%	65.9 cm^3(STP) cm^{-3}	RELLAW

The ROM has been shown be able to make accurate predictions but more importantly it can also be reduced to a lower amount of dimensions, improving the efficiency. Figure 6 shows the cumulative Sobol' indices for each rotated dimension in the four ROMs. For similarity to previous figures, Figure 6 has the bars coloured depending on the function of the ROM. For example, a ROM trained using MOFs storing methane is used to predict the deliverable capacity of MOFs storing oxygen and therefore the bars for this ROM are coloured in red for oxygen. The cumulative Sobol' indices show how much of the variance is ascribable to i amount of rotated dimensions (U_i). Using Figure 6 we can see how many rotated dimensions each ROM can be reduced to. As expected, when all five dimensions are considered, 100% of the variance of the output is captured by the ROM. The ROM used for predicting the GDC of MOFs storing methane has 80.1% of the variance captured by the primary dimension, U_1. Therefore, one may be confident of making sufficient predictions for a given MOF using just one dimension. Whereas ROMs used for oxygen predictions should provide sufficient predictions using three dimensions. All of the ROMs capture over 90% of its

output variance using four of the five dimensions and so all four models can be confidently reduced to four dimensions. This provides the potential mechanism for Active Subspaces which will provide fewer dimensions making MOF data far easier to understand and explain.

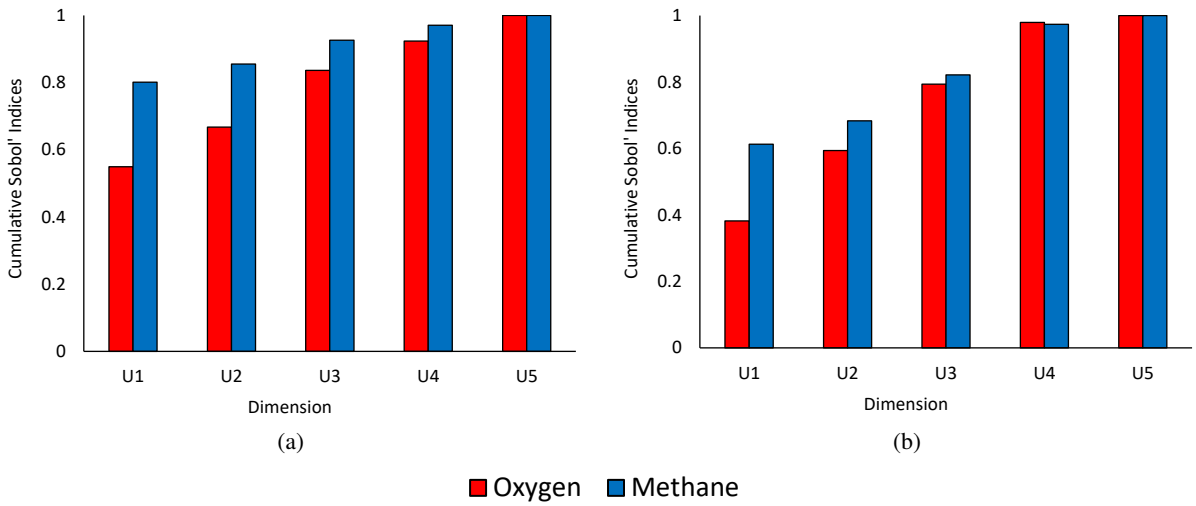


Figure 6: The bar charts showing the cumulative Sobol' indices for each dimension with respect to a) the gravimetric deliverable capacity and b) the volumetric deliverable capacity of MOFs storing oxygen (red) and methane (blue).

4.4. MOF Selection for Oxygen Storage

This section applies the ROM to a practical MOF exploration example. Oxygen storage using MOFs is rarely studied (Moghadam et al., 2018) but provides a significant promise for storage in healthcare, military and industrial applications (DeCoste et al., 2014). For oxygen storage, it is important to both increase storage per unit volume and reduce the storage pressure. Therefore, here we apply the ROM to help predict the top-performing MOFs with respect to the deliverable capacity of oxygen at 30 bar.

This exploration study analysed the Cambridge Structural Database (Moghadam et al., 2017b) to predict the top-performing MOFs for oxygen storage out of the 82221 synthesised MOFs. Predicting the performance of MOFs that have already been synthesised is more beneficial than predicting the performance of hypothetical MOFs because synthesis methods are not straightforward for newly proposed MOFs (Moghadam et al., 2018). After synthesis, measuring the adsorption of MOF structures has only been conducted on *ca.* 3% of the synthesised MOFs (Nazarian et al., 2016). Therefore predicting the deliverable capacity of synthesised structures is hugely important for producing promising MOF structures (Chung et al., 2014; Thornton et al., 2017).

Previously, 2745 MOF structures (Nazarian et al., 2016) out of the 82221 MOFs within the Cambridge Structural Database were used to train the ROMS. These structures were chosen for training because they were screened as

high-performance MOFs and because the deliverable capacity of these synthesised MOFs was calculated using computational simulations (Moghadam et al., 2018). From this, two sets of data were used to train the ROMs, one using MOFs storing oxygen and the other used the same MOFs but storing methane. These models provided a belief that they can be used to make satisfactory predictions irrelevant to a MOFs storage gas across similar gases. Thus, in this section we predict the top-performing MOFs for oxygen storage using a ROM trained from MOFs storing methane. The results are then compared to that from a ROM trained on the correct storage gas data.

Table 5 compares the top five MOF structures for oxygen storage predicted from our novel machine learning method. Here, the five promising structures were chosen because of their high predicted values for both gravimetric and volumetric deliverable capacity at 30 bar storage. However, due to limited volume in oxygen storage tanks, the VDC was prioritised over the GDC when needed. Each MOF had the deliverable capacity's mean value and the 95% confidence error predicted as shown in Table 5. Here, the ROM trained using 2745 MOFs storing methane, predicted the top MOF for oxygen storage at 30 bar to be VIKDID with a VDC of 118 ± 15.0 cm^3(STP) cm^{-3}. Additionally, the predicted GDC was high at 7.64 ± 0.655 mol kg^{-1} making this a very promising MOF for oxygen storage. A comparison to MOF structures with known deliverable capacities from the training data found just two MOFs similar to the five predicted by the ROM. Table 5 shows these known MOFs to have similar pore properties to the five predicted MOFs giving confidence in the predictions. Hence, it is clear that the ROM is predicting top-performing MOFs from structures with pore property values consistent to what have been found in previous research (Moghadam et al., 2018).

Additionally, Figure 7 compares the ten predicted top-performing MOFs with the distribution from the training MOF database. The training MOF database is the 2745 MOF structures that Moghadam et al. (2018) previously screened and analysed (through grand canonical Monte Carlo simulations) as top-performing MOF structures. Therefore, it is remarkable that top-performing MOFs from ROM predictions on the entire Cambridge Structural Database consistently have a predicted deliverable capacity within the limits or above the synthesised MOFs previously used for training. This result enables the use of MOFs with high storage capabilities that have already been synthesised. However, the close boundary to previously screened and simulated MOFs may hint towards the limit of synthesised MOFs to date. Clearly, with so many MOFs structures available to model, synthesise and then test, predicting top-performing MOFs for so many different uses will always be a challenge. The extent to whether it is possible to quantify all synthesised and non-synthesised MOFs is still unknown. Nonetheless, we believe that this new method provides a step towards quickly assessing capabilities of readily available MOFs for oxygen and methane storage.

Further, Figure 7 compares the predicted deliverable capacities from the ROM trained using MOFs storing methane and a ROM trained using MOFs storing oxygen. This comparison shows the reliability in using a methane trained

ROM to predict the oxygen deliverable capacity due to the closeness between the two results. Figure 7 shows the oxygen trained predictions (red markers) to always be within the 95% error bars predicted by the methane trained ROM (in blue) for these top-performing MOFs. Therefore, these results provide confidence that the novel machine learning method can be used to predict the deliverable capacity of MOFs irrelevant to the storage gas across similar gases. This is am important discovery due to the limitations involved in measuring the adsorption of so many MOF structures for so many gases.

Table 5: The top five MOFs for oxygen storage identified by a ROM with comparison to two MOF structures with computationally calculated oxygen deliverable capacities.

MOF ref code	LCD (Å)	A_V ($m^2\ m^{-3}$)	A_G ($m^2\ g^{-1}$)	VF (-)	GDC (mol kg^{-1})	VDC (cm^3(STP) cm^{-3})
Predicted MOFs						
VIKDID	7.45	2280	3557	0.748	7.64 ± 0.655	118 ± 15.0
HEFDUT	9.94	2127	2661	0.732	6.57 ± 0.498	117 ± 13.2
NIMJUP	8.05	2289	2972	0.726	6.60 ± 0.488	115 ± 13.0
EJEKEK	7.62	2197	3155	0.703	7.22 ± 0.516	113 ± 13.4
VODSUC	7.70	2159	3176	0.687	7.39 ± 0.518	113 ± 13.4
Computationally Measured MOFs						
BEPROF	8.00	2120	3013	0.700	6.95	114
MATVEJ	8.53	2116	2956	0.690	6.27	101

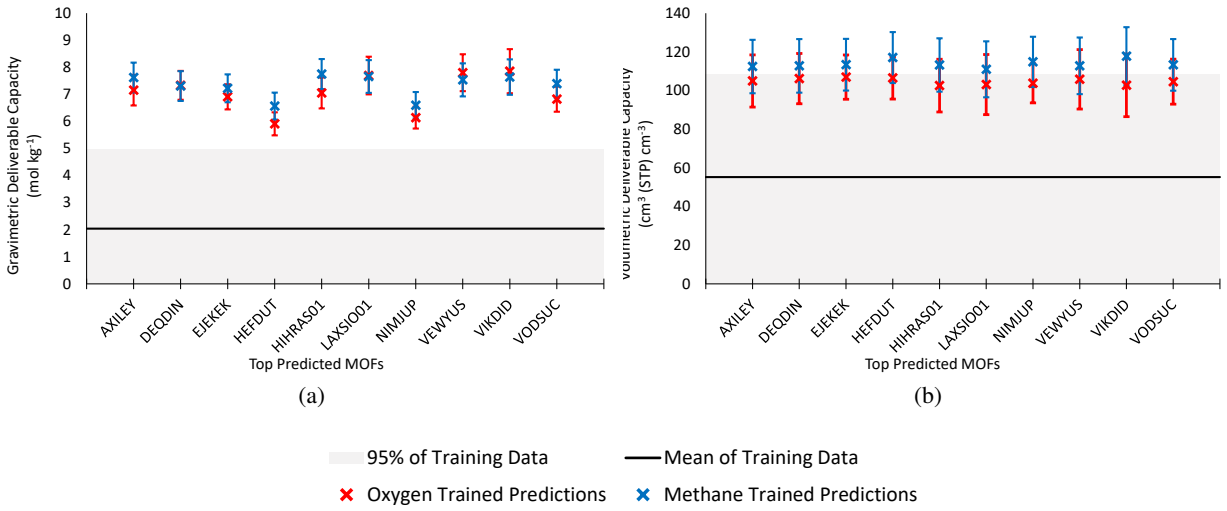


Figure 7: The predicted gravimetric (a) and volumetric (b) deliverable capacities at 30 bar storage for the ten most promising MOFs in comparison to the distribution of the training data. As predicted by a ROM trained using MOFs storing oxygen (red) and methane (blue)

5. Conclusion

With *ca.* 88000 structures of synthesised metal-organic frameworks (MOFs) available, identifying top materials for gaseous storage is difficult. Therefore, this paper investigates the pore properties affecting the gaseous deliverable capacity of synthesised MOFs. This paper has conducted two investigations using global sensitivity analyses (GSAs) before developing a reduced order model (ROM) prediction technique that results in faster more efficient material design and discovery for gas storage applications. A dataset of experimentally synthesised MOFs provides the information for learning producing a pathway to conduct GSAs discovering the most dominant pore properties to be accurately synthesised when designing MOFs. The GSAs was achieved by developing a Gaussian Process (GP) surrogate model enabling the semi-analytic calculation of Sobol' indices. Consequently, care was taken to validate the surrogate model ensuring inaccuracies were not carried onto the GSA.

The first GSA confirmed that the surface area and void fraction are the two most dominant pore properties while the heat of adsorption at 5 bar and the largest cavity diameter have a negligible effect on the deliverable capacity. Additionally, further GSAs conducted during this investigation has shown that the pore properties Sobol' indices are the same for different storage pressures. Hence, pressure has little effect on changing the impact each pore property has on the deliverable capacity of MOFs. The second study of GSAs explored the effect the storage gas has on the influence of the pore properties by comparing oxygen storage to methane storage. Once again, the findings were consistent with the literature and the previous study showing the gravimetric surface area and void fraction to be the dominant pore properties. Further, it has shown that the pore properties interactions between themselves increase

significantly when storing methane instead of oxygen. This aspect of the research showed the first-order Sobol' indices to be similar for MOFs storing either gas and so suggests that the impact of the pore properties is irrelevant to the storage gas.

This hypothesis was tested by training a ROM using MOF data storing one gas and then tested the ROM using MOFs storing the other gas. Importantly, our results provided evidence for good predictor capabilities irrelevant to the MOF storage gas shown by the standardised diagnostic values. Additionally, further analysis of the ROMs showed the capability to reduce the dimensions with confidence from five to four or three rotated dimensions. The ROM was applied to a practical MOF exploration example, to help identify the top-performing MOFs for oxygen storage at 30 bar. Here, the methane trained ROM was found to be in agreement with the oxygen trained ROM, predicting VIKDID as the top-performing MOF structure for oxygen storage with a VDC of 118 ± 15.0 cm^3(STP) cm^{-3} and a gravimetric deliverable capacity of 7.64 ± 0.655 mol kg^{-1}.

These findings provide a potential mechanism to improve on commonly used principal component analyses by using Active Subspaces instead of linear combinations. Future investigations are necessary to further understand the applicability of these kinds of conclusions to other storage gases. This would require improved understanding of correlated inputs and further additional data would be required that includes MOFs storing various other gases to further develop and confirm these initial findings. This work has only compared oxygen storage to methane storage which are similar in nature in the way that they interact with the MOF framework. The adsorption in mainly dominated by geometric properties and van der Waals forces, whereas the electrostatic forces are week. Therefore, the next stage would be to investigate whether the ROM can be used to successfully make predictions for MOFs storing any gas or whether there is a select list of similar storage gases. For example, water and carbon dioxide are not considered similar to oxygen and methane as the polar and quadrupolar adsorbates have larger interactions from the surface chemistry. Hypothetically, numerous ROM's could be created dependent on the type of storage gas which could be classified based on these findings and future research. We believe that apart from looking at pore properties future research should also investigate the chemical and mechanical properties of MOFs.

Our novel contributions from this research can be summarised as follows:

- We have successfully predicted the GCD and VDC using an extensive database of synthesised MOF providing a basis for selecting appropriate structures for given applications.
- The predictions have been used to calculate a GSA to identify the most dominant pore properties and quantify their interactions depending on the application. Understanding how the storage gas or storage pressure impacts the behaviour of the pore properties enables researchers to consider the MOFs proposed usage before synthesising new materials.

- We presented a novel methodology for an Active Subset ROM which successfully predicts the gaseous uptake of MOFs for similar gas molecules.
- We have applied a ROM on a practical MOF exploration example, discovering the methane trained ROM to be in agreement with the oxygen trained ROM predicting VIKDID to be the top-performing MOF structure for oxygen storage at 30 bar.

Overall, this study provides vital information regarding the importance of the pore properties for synthesised MOFs. It has provided an efficient method to predict the deliverable capacity of MOFs. Scientists wanting to predict and search for promising MOFs used for storing any gas can adopt the ROM.

References

Aghaji, M.Z., Fernandez, M., Boyd, P.G., Daff, T.D., Woo, T.K., 2016. Quantitative Structure–Property Relationship Models for Recognizing Metal Organic Frameworks (MOFs) with High CO2Working Capacity and CO2/CH4Selectivity for Methane Purification. European Journal of Inorganic Chemistry 2016, 4505–4511. doi:10.1002/ejic.201600365.

Banerjee, D., Simon, C.M., Plonka, A.M., Motkuri, R.K., Liu, J., Chen, X., Smit, B., Parise, J.B., Haranczyk, M., Thallapally, P.K., 2016. Metal-organic framework with optimally selective xenon adsorption and separation. Nature Communications 7, 1–7. doi:10.1038/ncomms11831.

Bobbitt, N.S., Mendonca, M.L., Howarth, A.J., Islamoglu, T., Hupp, J.T., Farha, O.K., Snurr, R.Q., 2017. Metal-organic frameworks for the removal of toxic industrial chemicals and chemical warfare agents. Chemical Society Reviews 46, 3357–3385. doi:10.1039/c7cs00108h.

Brown, S., Beck, J., Mahgerefteh, H., Fraga, E.S., 2013. Global sensitivity analysis of the impact of impurities on CO2 pipeline failure. Reliability Engineering and System Safety 115, 43–54. URL: http://dx.doi.org/10.1016/j.ress.2013.02.006, doi:10.1016/j.ress.2013.02.006.

Carraro, M., Gross, S., 2014. Hybrid materials based on the embedding of organically modified transition metal oxoclusters or polyoxometalates into polymers for functional applications: A review. Materials 7, 3956–3989. doi:10.3390/ma7053956.

Chen, W., Jin, R., Sudjianto, A., 2005. Analytical Variance-Based Global Sensitivity Analysis in Simulation-Based Design Under Uncertainty. Journal of Mechanical Design 127, 875. doi:10.1115/1.1904642.

Chung, Y.G., Camp, J., Haranczyk, M., Sikora, B.J., Bury, W., Krungleviciute, V., Yildirim, T., Farha, O.K., Sholl, D.S., Snurr, R.Q., 2014. Computation-ready, experimental metal-organic frameworks: A tool to enable high-throughput screening of nanoporous crystals. Chemistry of Materials 26, 6185–6192. doi:10.1021/cm502594j.

Constantine, P.G., Diaz, P., 2017. Global sensitivity metrics from active subspaces. Reliability Engineering and System Safety 162, 1–13. URL: http://dx.doi.org/10.1016/j.ress.2017.01.013, doi:10.1016/j.ress.2017.01.013, arXiv:1510.04361.

Constantine, P.G., Dow, E., Wang, Q., 2014. Active subspace methods in theory and practice: Applications to kriging surfaces. SIAM Journal on Scientific Computing 36, A1500–A1524. doi:10.1137/130916138, arXiv:1304.2070.

DeCoste, J.B., Weston, M.H., Fuller, P.E., Tovar, T.M., Peterson, G.W., LeVan, M.D., Farha, O.K., 2014. Metal-organic frameworks for oxygen storage. Angewandte Chemie - International Edition 53, 14092–14095. doi:10.1002/anie.201408464.

Fanourgakis, G.S., Gkagkas, K., Tylianakis, E., Klontzas, E., Froudakis, G., 2019. A Robust Machine Learning Algorithm for the Prediction of Methane Adsorption in Nanoporous Materials. Journal of Physical Chemistry A 123, 6080–6087. doi:10.1021/acs.jpca.9b03290.

Fernandez, M., Barnard, A.S., 2016. Geometrical Properties Can Predict CO2 and N2 Adsorption Performance of Metal-Organic Frameworks (MOFs) at Low Pressure. ACS Combinatorial Science 18, 243–252. doi:10.1021/acscombsci.5b00188.

Fernandez, M., Boyd, P.G., Daff, T.D., Aghaji, M.Z., Woo, T.K., 2014. Rapid and accurate machine learning recognition of high performing metal organic frameworks for CO2 capture. Journal of Physical Chemistry Letters 5, 3056–3060. doi:10.1021/jz501331m.

Fernandez, M., Trefiak, N.R., Woo, T.K., 2013a. Atomic property weighted radial distribution functions descriptors of metal-organic frameworks for the prediction of gas uptake capacity. Journal of Physical Chemistry C 117, 14095–14105. doi:10.1021/jp404287t.

Fernandez, M., Woo, T.K., Wilmer, C.E., Snurr, R.Q., 2013b. Large-scale quantitative structure-property relationship (QSPR) analysis of methane storage in metal-organic frameworks. Journal of Physical Chemistry C 117, 7681–7689. doi:10.1021/jp4006422.

Hastie, T., 2009. The elements of statistical learning : data mining, inference, and prediction. Springer series in statistics. 2nd ed. ed., Springer, New York.

Li, J.R., Kuppler, R.J., Zhou, H.C., 2009. Selective gas adsorption and separation in metal-organic frameworks. Chemical Society Reviews 38, 1477–1504. doi:10.1039/b802426j.

Li, S., Yang, B., Qi, F., 2016. Accelerate global sensitivity analysis using artificial neural network algorithm: Case studies for combustion kinetic model. Combustion and Flame 168, 53–64. doi:10.1016/j.combustflame.2016.03.028.

Marrel, A., Iooss, B., Laurent, B., Roustant, O., 2009. Calculations of Sobol indices for the Gaussian process metamodel. Reliability Engineering and System Safety 94, 742–751. doi:10.1016/j.ress.2008.07.008.

Mason, J.A., Oktawiec, J., Taylor, M.K., Hudson, M.R., Rodriguez, J., Bachman, J.E., Gonzalez, M.I., Cervellino, A., Guagliardi, A., Brown, C.M., Llewellyn, P.L., Masciocchi, N., Long, J.R., 2015. Methane storage in flexible metal-organic frameworks with intrinsic thermal management. Nature 527, 357–361. URL: http://dx.doi.org/10.1038/nature15732, doi:10.1038/nature15732.

Milton, R.A., Brown, S.F., 2019. ROMCOMMA. URL: https://github.com/C-O-M-M-A/rom-comma.

Moghadam, P.Z., Islamoglu, T., Goswami, S., Exley, J., Fantham, M., Kaminski, C.F., Snurr, R.Q., Farha, O.K., Fairen-Jimenez, D., 2018. Computer-aided discovery of a metal-organic framework with superior oxygen uptake. Nature Communications 9, 1–8. URL: http://dx.doi.org/10.1038/s41467-018-03892-8, doi:10.1038/s41467-018-03892-8.

Moghadam, P.Z., Ivy, J.F., Arvapally, R.K., Dos Santos, A.M., Pearson, J.C., Zhang, L., Tylianakis, E., Ghosh, P., Oswald, I.W., Kaipa, U., Wang, X., Wilson, A.K., Snurr, R.Q., Omary, M.A., 2017a. Adsorption and molecular siting of CO2, water, and other gases in the superhydrophobic, flexible pores of FMOF-1 from experiment and simulation. Chemical Science 8, 3989–4000. doi:10.1039/c7sc00278e.

Moghadam, P.Z., Li, A., Wiggin, S.B., Tao, A., Maloney, A.G., Wood, P.A., Ward, S.C., Fairen-Jimenez, D., 2017b. Development of a Cambridge Structural Database Subset: A Collection of Metal-Organic Frameworks for Past, Present, and Future. Chemistry of Materials 29, 2618–2625. doi:10.1021/acs.chemmater.7b00441.

Moghadam, P.Z., Rogge, S.M., Li, A., Chow, C.M., Wieme, J., Moharrami, N., Aragones-Anglada, M., Conduit, G., Gomez-Gualdron, D.A., Van Speybroeck, V., Fairen-Jimenez, D., 2019. Structure-Mechanical Stability Relations of Metal-Organic Frameworks via Machine Learning. Matter 1, 219–234. URL: https://www.sciencedirect.com/science/article/pii/S2590238519300062, doi:10.1016/J.MATT.2019.03.002.

Nazarian, D., Camp, J.S., Sholl, D.S., 2016. A Comprehensive Set of High-Quality Point Charges for Simulations of Metal-Organic Frameworks. Chemistry of Materials 28, 785–793. doi:10.1021/acs.chemmater.5b03836.

Ohno, H., Mukae, Y., 2016. Machine Learning Approach for Prediction and Search: Application to Methane Storage in a Metal-Organic Framework. Journal of Physical Chemistry C 120, 23963–23968. doi:10.1021/acs.jpcc.6b07618.

Qiao, Z., Xu, Q., Cheetham, A.K., Jiang, J., 2017. High-Throughput Computational Screening of Metal-Organic Frameworks for Thiol Capture. Journal of Physical Chemistry C 121, 22208–22215. doi:10.1021/acs.jpcc.7b07758.

Qiao, Z., Xu, Q., Jiang, J., 2018. High-throughput computational screening of metal-organic framework membranes for upgrading of natural gas. Journal of Membrane Science 551, 47–54. URL: https://doi.org/10.1016/j.memsci.2018.01.020, doi:10.1016/

j.memsci.2018.01.020.

Saltelli, A., Homma, T., 1996. Importance measures in global sensitivity analysis of model output. Reliab. Eng. Sys. Safety 52, 1–17.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., 2008. Global Sensitivity Analysis. The primer ed., John Wiley & Sons Ltd.

Simon, C.M., Mercado, R., Schnell, S.K., Smit, B., Haranczyk, M., 2015. What Are the Best Materials to Separate a Xenon/Krypton Mixture? Chemistry of Materials 27, 4459–4475. doi:10.1021/acs.chemmater.5b01475.

Sobol, I.M., 1993. Sensitivity analysis for nonlinear mathematical models. Mathematical Modelling Computational Experiments 1, 407–414. doi:10.18287/0134-2452-2015-39-4-459-461., arXiv:arXiv:1305.4373v1.

Sobol, I.M., 2001. Global sensitivity indices for nonlinear mathematical models. Review. Mathematics and Computers in Simulation , 271–280.

Sudret, B., 2008. Global sensitivity analysis using polynomial chaos expansions. Reliability Engineering and System Safety 93, 964–979. doi:10.1016/j.ress.2007.04.002.

Thornton, A.W., Simon, C.M., Kim, J., Kwon, O., Deeg, K.S., Konstas, K., Pas, S.J., Hill, M.R., Winkler, D.A., Haranczyk, M., Smit, B., 2017. Materials Genome in Action: Identifying the Performance Limits of Physical Hydrogen Storage. Chemistry of Materials 29, 2844–2854. doi:10.1021/acs.chemmater.6b04933.

Tian, T., Zeng, Z., Vulpe, D., Casco, M.E., Divitini, G., Midgley, P.A., Silvestre-Albero, J., Tan, J.C., Moghadam, P.Z., Fairen-Jimenez, D., 2018. A sol-gel monolithic metal-organic framework with enhanced methane uptake. Nature Materials 17, 174–179. doi:10.1038/NMAT5050.

Williams, C.K.I., Rasmussen, C.E., 2006. Gaussian processes for machine learning. The MIT Press, London. URL: http://www.gaussianprocess.org/gpml/chapters/.

Wilmer, C.E., Leaf, M., Lee, C.Y., Farha, O.K., Hauser, B.G., Hupp, J.T., Snurr, R.Q., 2012. Large-scale screening of hypothetical metal-organic frameworks. Nature Chemistry 4, 83–89. doi:10.1038/nchem.1192.

Wipf, D.P., Nagarajan, S., 2007. A New View of Automatic Relevance Determination, in: Proceedings of the 20th International Conference on Neural Information Processing Systems, NIPS'07, Curran Associates Inc.. pp. 1625–1632. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.143.8009{&}rep=rep1{&}type=pdf.

Yeardley, A.S., Bellinghausen, S., Milton, R.A., Litster, J.D., Brown, S.F., 2021. Efficient global sensitivity-based model calibration of a high-shear wet granulation process. Chemical Engineering Science 238, 116569. URL: https://doi.org/10.1016/j.ces.2021.116569, doi:10.1016/j.ces.2021.116569.

Yeardley, A.S., Bugryniec, P.J., Milton, R.A., Brown, S.F., 2020. A study of the thermal runaway of lithium-ion batteries: A Gaussian Process based global sensitivity analysis. Journal of Power Sources 456, 228001. URL: https://doi.org/10.1016/j.jpowsour.2020.228001, doi:10.1016/j.jpowsour.2020.228001.

Zhang, L., Siepmann, J.I., 2006. Direct calculation of Henry's law constants from Gibbs ensemble Monte Carlo simulations: Nitrogen, oxygen, carbon dioxide and methane in ethanol. Theoretical Chemistry Accounts 115, 391–397. doi:10.1007/s00214-005-0073-1.