Universitetet i Stavanger

DET TEKNISK-NATURVITENSKAPELIGE FAKULTET

MASTEROPPGAVE

Studieprogram/spesialisering: Msc. Computational Engineering	Vårsemesteret, 2021 Åpen / ~~Konfidensiell~~
Forfatter: Joachim Alexander Worum	
Fagansvarlig: Aksel Hiorth Veileder(e): Pål Østebø Andersen (UiS), Mayooran Kiritharan (ConocoPhillips)	
Tittel på masteroppgaven: Causal Inference of Interwell Connectivity Engelsk tittel: Causal Inference of Interwell Connectivity	
Studiepoeng: 30	
Emneord: Causal inference analysis, Interwell connectivity estimation, Strength of connectivity estimation, Hypothesis testing, Classification.	Sidetall: ……99…… + vedlegg/annet: ……31…… Stavanger, ……15.06.2021…….. dato/år

Causal Inference of Interwell Connectivity

Joachim Alexander Worum

June 2021

Abstract

Interwell connectivity is a measure of the degree to which information is translated between wells in a reservoir. The translated information may include changes in pressure, fluid flow, or other physical properties. This connectivity is separated into internal connectivities (through the reservoir) and external connectivities (through the infrastructure). Understanding these interwell connectivities may improve the overall understanding of the reservoir, aid in optimizing oil and gas production, and avoid new and undesired connectivities. We seek to determine if causal inference can be applied to aid in the detection of interwell connectivities within the oil and gas field, Eldfisk, operated by ConocoPhillips. Eldfisk is a mature field with a complex network of injection and production wells and a large and connected infrastructure of oil and gas installations. We examined thirteen wells, where eight are water injection wells, and five are oil and gas producing wells. Between these, there are three injection to injection well connectivities and five injection to production well connectivities, for a total of eight well-pair connectivities. These connectivities are known to ConocoPhillips and have been identified through various methods, including manual comparison of pressure responses between wells. The connectivity we seek to identify is a pressure response between a stimulating well and a target well through the reservoir. Data is obtained from Bottom Hole Pressure sensors where this is available and Tubing Head Pressure sensors otherwise. The pressure data is split into smaller datasets that target periods when a stimulating well is experiencing a change in pressure. These periods are identified by an on/off indicator, binary data indicating whether a well is opened or closed. Every well is given status as a stimulating well for each period identified by the on/off-indicator. All other wells are considered as target wells during these periods. Two different causal inference methods were applied to the datasets; Peter Clark Momentary Conditional Independence + (PCMCI+) and Temporal Causal Discovery Framework (TCDF). PCMCI+ is a method that employs partial correlation to infer causality, and TCDF employs convolutional neural networks. We found that the PCMCI+ algorithm was able to identify seven out of eight well-pairs with known connectivities. However, it was less precise, with approximately one out of five identified connectivities being true. The TCDF-algorithm identified three out of eight well-pairs with known connectivities. This model was more precise, with approximately one out of four identified connectivities being true. We also found that the overall complexity of the original dataset was effectively reduced by splitting it into smaller datasets. These datasets specifically targeted periods when the wells experienced changes in their pressure profiles. External interferences from a field-wide water supply were not effectively reduced and were a cause of a majority of false connectivities. Group-wide interference from wells sharing the same production installation was found to be effectively reduced. Using tubing head pressure data when bottom hole pressure data was unavailable, was found to be an effective substitute.

Acknowledgement

Throughout the development of this thesis, I have received a great deal of guidance, positive critique, and helpful advice. This has been an invaluable help in shaping this thesis.

I would like to thank my two supervisors, Pål Østebø Andersen and Mayooran Kiritharan. I am deeply grateful for their guidance and encouraging words, as well as all the time they have dedicated to assisting me. Their continued support has been a great help in shaping the thesis to what it is today.

I wish to show my appreciation to ConocoPhillips for providing me with the thesis subject, materials, and all data used in this thesis.

I would like to thank my family and friends for their continued support and encouraging words during this time.

Finally, a special thanks to my son, Levi Worum, for all of the happy distractions he has given me.

Contents

List of Tables

List of Figures

Nomenclature

List of symbols

t, τ	Time indices
T	End of time index
n, N	Number of observations
i, j	Variable indices
δ	Difference between two values
r	Spearman correlation coefficient
θ, β	Coefficients
ϵ	Stationary random component
Δ	Difference operator
y	Response variable of Multivariate Linear Regression
x	Predictor variable
$\mathbf{x}$	Vector of predictor variables
p	Number of predictors
X_t,Y_t	Elements of time series
$\mathbf{X}$,$\mathbf{Y}$,$\mathbf{Z}$	Sets of time series data
$\perp\!\!\!\perp$	Independence operator
$\not\perp\!\!\!\perp$	Dependence operator
H_0,H_1	Null and alternative hypothesis
d	Dependence measure
$\mathbf{d}$	Set of dependency measurements
μ_d	Mean of a set of dependency measurements
t_S	Test statistic for the Student's t-test
σ	Standard deviation
$\mathbf{D}$	Set of differences
$\mathbf{R}$	Set of ranks of variables
$\mathbf{R}^+, \mathbf{R}^-$	Sets of ranks corresponding to positive and negative differences

W	Wilcoxons ranked sum test statistic
α	Significance threshold, related to a p-value
$\mathbf{e_X}, \mathbf{e_Y}$	Sets of residuals for $\mathbf{X}$ and $\mathbf{Y}$ regressed on some variable.
$\rho_{XY \cdot Z}, \rho$	Partial correlation coefficient of $\mathbf{e_X}$ and $\mathbf{e_Y}$.
$\mathbf{S}$	Two dimensional dataset containing a set of time series variables. Contains a time index, t, and a variable index, i.
a^{ji}	Attention score for the causal link from $\mathbf{X}^i$ to $\mathbf{X}^j$.
$\mathbf{a}^j$	Set of attention scores a^{ji} for $i \in N$
γ^j	Truncation threshold for the attention scores.
$\mathbf{A}^j$	Truncated attention scores from $\mathbf{a}^j$
$\mathbf{P}^j$	Set of potential causes for $\mathbf{X}^j$
$\mathbf{C}^j$	Set of validated causes for $\mathbf{X}^j$
s	Hyperparameter of the TCDF-algorithm called *significance*, controls the conversion rate for potential causes to validated causes.
D_R^i	Difference in training loss between first and last epoch of the randomly permuted time-series.
D_r^i	Difference in training loss between the first and last epoch of the real time-series.
K	Hyperparameter for the kernel size of the TCDF-algorithm.
L_{max}	Maximum number of possible links for a single variable generated by the PCMCI+ algorithm.
μ^T, μ^F	Mean partial correlation or attention score of the True connectivity and False connectivity distributions.
U	Test statistic for the Mann-Whitney U-test.

Abbreviations

AARD	Average Absolute Relative Deviation
ADF	Augmented Dickey-Fuller
AD-DSTCNs	Attention-based Dilated Depthwise Separable Temporal Convolutional Networks
ANN	Artificial Neural Network
BP	Back Propagation Neural Network
BHP	Bottom Hole Pressure
CNN	Convolutional Neural Network
CI	Conditional Independence
DAG	Directed Acyclic Graph

EFAST	Extended Fourier Amplitude Sensitivity Test
FN	False Negative
FP	False Positive
GOR	Gas/Oil Ratio
IQR	Interquantile Range
LSTM	Long Short-Term Memory
MLR	Multivariate Linear Regression
PCMCI+	Peter Clark Momentary Conditional Independence +
PIVM	Permutation Importance Validation Method
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
RSS	Residual Sum of Squares
TCDF	Temporal Causal Discovery Framework
THP	Tubing Head Pressure
TN	True Negative
TP	True Positive
ParCorr	Partial Correlation

Chapter 1

Introduction

Interwell connectivity is commonly referred to when discussing the translation of information from a stimulating well A to a target well B. The information being translated may be pressure, fluid flow, or some other type of information. The main idea is that the two wells are connected in some regard. Interwell connectivity can be broken down into three main parts; the information being translated, the degree to which the information is translated, and the time lag of the translation.

In an oil and gas field, the topic of interwell connectivity is an important one, as it may be a cause of sub-optimal production rates from a well due to interference from another well. Understanding these relationships between the wells may help to better understand the reservoir as a whole and aid in future well-planning and optimization of already producing wells.

In this thesis, the topic is to identify interwell connectivity through causal inference by the models; Peter Clark Momentary Conditional Independence + (PCMCI+) introduced by Runge (2020), and Temporal Causal Discovery Framework (TCDF) introduced by Nauta et al. (2019). The two models were applied to pressure data of wells from the mature Eldfisk field in the North Sea. Eldfisk is a highly complex system of wells with both internal (through the reservoir) and external (through the infrastructure) connectivities. Pressure changes between well-pairs is the translated information that we seek to identify. The full time-series data is split into multiple smaller datasets for each stimulating well, and a hypothesis test is used to determine if there is statistically significant evidence of interwell connectivity between a given well-pair.

This topic has previously been explored with different methods, such as; pressure pulse testing proposed by Johnson et al. (1966) to identify a lagged response in a target well. Heffer et al. (1997) applied a Spearman rank correlation on flow rates to infer geomechanical processes and interwell connectivity. Using Multivariate Linear Regression, Dinh (2009) describe interwell connectivity by the fluctuations in the Bottom Hole Pressure. Mata (2010) identifies periods when a well experience a rate change and then search a selected group of wells for a correlating peak to infer connectivity. Deep learning models have also been attempted for both synthetic and real data (Panda and Chopra 1998; Du et al. 2020; Cheng et al. 2020).

1.1 Overview of thesis structure

Chapter 1.2: Previous work on the inference of interwell connectivity is examined. Different inference methods are included in this examination, from simpler methods such as correlation matching to more complex methods employing deep learning. The objectives of the thesis are listed at the end.

Chapter 2: Theory related to interwell connectivity, and as no set definition of interwell connectivity has been found, we explain our definition of the term. We describe some elements of time series analysis such as hypothesis testing, statistical tests, and probability value and give detailed explanations of PCMCI+ and TCDF.

Chapter 3: Gives a short introduction to the Eldfisk field the data is collected from and the type of data that is used (pressure and on/off-indicators). There is an overview of the well-pairs with known connectivities presented. Additionally, a section explains what types of external interferences are present in the data and a section that highlights the connectivity between the well-pairs.

Chapter 4: Detailed description of the methodology used to generate the final results. The chapter is divided into the main sections of data selection, data extraction, preprocessing of data, model application, classification, and strength of connectivity measure.

Chapter 5: Results presented in sections corresponding to the different tests performed. These results are presented in tables, boxplots, and short descriptions that examine the results' numerical and visual representation.

Chapter 6: Detailed interpretation of the results. Divided into sections corresponding to each of the various tests and models applied. Focus on the sensitivity, specificity, and precision of the models. The final classification is presented as Directed Acyclic Graphs.

Chapter 7: Conclusions are drawn from the development of the thesis. There is also a section describing possible future work that may improve the precision of the models.

Appendix A: Lists all tables with descriptive statistics used to generate boxplots in the Results chapter.

Appendix B: Figures of known connectivities, along with a short description for each.

Appendix C: Lists raw output from the PCMCI+ and TCDF-algorithms from the method demonstration in Chapter 2.

1.2 Literature review

In the following, we will examine different methods used to infer interwell connectivity by time-series data, both in terms of connectivity strength and time lag.

1.2.1 Cross-correlation

Johnson et al. (1966) propose a method of pulse testing to identify connectivity between well-pairs in a reservoir. The method involves applying a series of pulses in a stimulating well and then recording the pressure changes in a responding well using a very sensitive differential pressure gauge. The recorded data is in the form of a time series of pressure readings, $\{Y_t, t \in T\}$, where t is some point in time of the full length of the recorded period, T. Generating a unique series of pressure pulses in a stimulating well, the pressure response in a target well and be distinguished from other noise in the reservoir over some time. The pulses are generated by changing the flow in the stimulating well, which translates to a pressure wave transported through the reservoir. The response is measured by the amplitude and time lag of the pulses. They mention that reservoirs with low transmissibility are more difficult to test due to longer time lags and weaker response amplitude.

Heffer et al. (1997) applied a Spearman rank correlation analysis on monthly flow rate data of injector/producer well pairs. Spearman rank correlation is a measure of the monotonic relationship between two variables (Spearman 1987). The Spearman correlation coefficient, r, is calculated as seen in (1.1):

$$r = 1 - \frac{6 \sum_{i=1}^{n} \delta_i^2}{n(n^2 - 1)} \tag{1.1}$$

Where n is the number of samples, and δ_i is the difference in ranks for the two variables being measured. The rank represents the numerically ordered values (ascending or descending) in some variable. The Spearman correlation coefficient is returned in the range $[-1, 1]$, where -1 indicates an inverse monotonic relationship (negative correlation), 1 indicates a direct monotonic relationship (positive correlation), and 0 indicates no monotonic relationship (no correlation). The analysis of Heffer et al. (1997) involves selecting pairs of injector/producer wells and computing the Spearman correlation coefficient of the injection and production rates over a suitable time range. Once a peak correlation has been observed, extract the time lag and the correlation coefficient. The intent was to infer geomechanical processes and interwell connectivity. Interestingly, they did not find a time lag in the correlation between the well-pairs. They suggest that the missing time lag is due to the injection rate changing the rock strain throughout the field. They conclude that the Spearman rank correlation confirms communication between producer/injector pairs. Additionally, they found that geomechanical processes indeed influence the correlations. It should be noted here that the correlation coefficient will only hold information about the strength and direction of the relationship, and it cannot determine dependency (Mukaka 2012).

The Spearman rank correlation was also applied by Jansen and Kelkar (1997). They highlight some challenges correlating the production data of well-pairs, namely the non-stationarity and non-linear nature of the data. When the data is non-stationary, the mean, variance, or covariance (or any combination of these) of a time series may change over time, i.e., the data distribution is not constant over time. Non-stationarity is problematic because many statistical tools assume that the time series data is stationary (Manuca and Savit 1996). The Augmented Dickey-Fuller (ADF) test introduced by Dickey and Fuller (1979) can identify non-stationarity in a time series. ADF assumes that the time series can be written on the autoregressive form (1.2):

$$Y_t = \theta Y_{t-1} + \epsilon_t \tag{1.2}$$

Y_t is an element of a time series, θ is a coefficient, and ϵ is a stationary random component. For $|\theta| < 1$, the time series converges to stationarity, for $|\theta| \geq 1$ the time series do not converge and is non-stationary. Dickey and Fuller (1979) estimates the test statistics for the estimate of θ, which enables a one-sided hypothesis test to be stated as; $|\theta| = 1$ as the null hypothesis, and $|\theta| < 1$ for the alternative hypothesis. Critical values of the test statistic depend on the true form of the autoregressive time series and sample size. As such, one should choose to test for stationarity, level (mean) stationarity, or trend stationarity. Commonly, visual inspection is used to determine which type of stationarity to test for (Kočenda and Černý 2015). Cheung and Lai (1995) calculated critical values for different autoregressive forms and sample sizes of a time series.

To deal with the non-stationary nature of production data, Jansen and Kelkar (1997) introduce Wavelet transformation, which allows for decomposing time series data into frequency and smoothed components. Cross-correlation coupled with the Wavelet transform was found to be capable of inferring interwell connectivity. Another method of data transformation that is shown to transform a non-stationary time series to a stationary one effectively is differencing (Nason 2006). When differencing is applied to a time series, the difference is taken as shown in (1.3).

$$\Delta Y_t = Y_t - Y_{t-1} \tag{1.3}$$

Y_t is an element of a time series, and Δ indicates the difference operator.

Similar to previous work, Soeriawinata and Kelkar (1999) applied the Spearman rank coefficient to obtain interwell connectivity. Additionally, they account for the superposition effect caused by the fact that a single producer can be affected by multiple injectors. They introduce constructive and destructive terms to describe the effect of multiple injectors on a single producer's signal. They applied a method that searched for and accepted the injectors with a cross-correlation to a single producer higher than a set arbitrary threshold. Once a group was defined, the combined correlation had to be greater than a set threshold to be approved else the entire group was rejected. They show that the method identified strong connectivities between groups of wells on both synthetic and real data.

Ramakrishnan and Raghuraman (2004) showed that it was possible to correlate a pressure pulse from a stimulating well to a responding well over large distances. A measure of the time lag can be obtained by searching peak correlation over a time range between the well-pairs. They show that the time lag is related to the interwell permeability by estimating the diffusion constant.

1.2.2 Linear regression

Albertoni and Lake (2003) applies a Multivariate Linear Regression (MLR) model with oil production rate and water injection rate as input to infer connectivity. An MLR model is a method to describe the linear relationship between a response variable and predictor variables, as shown in (1.4).

$$y = \beta^0 + \sum_{i=1}^{p} \beta^i x^i + \epsilon \tag{1.4}$$

Where y is a response variable, $x^i \in \mathbf{x}$ are predictor variables, β is the coefficient weights, p is the number of predictors, and ϵ is a random error term (James et al. 2013). The input variable x^i can be related to injection rates or BHP, and the output variable y can be related to the production rate or the BHP for a responding well. The coefficient weights, β^i, act as a quantification of the connectivity between a given injector and the producer. They also apply a diffusivity filter on each injector's injection rate, defined by a diffusivity constant dependent on the medium and distance between injector and producer. This diffusivity filter accounts for the time lag and attenuation expected between the effect of the injector and the producers' response. The model was tested on synthetic and real data, using a homogeneous setting, a heterogeneous setting, and a sealing fault on the synthetic data. To determine how well the model performed, they applied a tracer test on the synthetic reservoir data and compared it to the model by the injector/producer pairs' MLR weights. The model's performance was also visually interpreted by looking at the symmetry of the weights between the injector/producer pairs. There are several assumptions associated with the method. These include; no new wells added to the model during training, constant Bottom Hole Pressure (BHP) in the producers, constant Gas/Oil Ratio (GOR), and constant effective permeability. They conclude that the model can quantify the connectivity between the injector/producer pairs using only injection and production rates as inputs. The results from the model were also useful in determining sections of differing permeability and sealing faults.

Yousef, Gentil, et al. (2006) relates the interwell connectivity problem to a Capacitance Resistance Model (CRM). CRM is a non-linear multivariate regression model that takes injection rates and, if available, BHP as input and production rates as output. This model applies capacitance and resistance (compressibility and transmissibility) to obtain a more accurate model than Albertoni and Lake (2003). They found that this model could better capture the attenuation and time lag between the injector/producer pairs. Sayarpour et al. (2009) introduced three solutions for the CRM, one for the volume of the entire field, one for each producer's drainage volume, and one for the drainage volume of each producer/injector pair. The CRM was further developed in (Yousef, Jensen, et al. 2008), which introduces two plots; log-log and a Lorenz-style flow capacity/storativity plot. These techniques show that the model can identify geological features in the reservoir, such as fractures and high permeability zones.

Jamali and Ettehadtavakkol (2017) highlights some challenges with the CRM in (Yousef, Gentil, et al. 2006), especially the model's scalability and its application to mature, large-scale oil fields. These challenges include long computational times and difficulties in obtaining a convergent solution. They mention several assumptions associated with CRM. These include; no changes to the well settings and slightly compressible fluids. Their proposed solution to some of these challenges is to break up the full production period into smaller periods with their own connectivity map. They argue that major changes to the field, such as new or shut-in wells, will have a major effect on the connectivity map. To reduce the computational power required, they show that the radius of influence from an injection well eventually converges and decreases in connectivity error

flattens.

In the study (Dinh 2009), they present a model in which they describe the interwell connectivity using fluctuations in the BHP instead of the production/injection rate. This work is a continuation of the work done by Albertoni and Lake (2003), and the same principle of an MLR model is used. This model sets the BHP from all injectors as input data and the producers' BHP as the output. They also found no need to use the diffusivity filters introduced by Albertoni and Lake (2003), as the results they obtained were near perfect for a homogeneous synthetic reservoir. Assumptions on the well conditions include a constant production and injection rate. When the total production rate is lower than the total injection rate, the results remain the same for different constant production rates. This indicates that the connectivity weights do not depend on the production/injection rates but on the reservoir conditions. However, they found that once the total production rate was set equal to the injection rate, the model reduced significantly in performance.

In the report (Jensen et al. 2007), they identified and improved certain aspects of the MLR model proposed by Albertoni and Lake (2003). They found that collinearity between injectors and non-stationarity in the production data caused negative connectivity weights to occur. The collinearity was addressed by introducing ridge regression on the MLR coefficients. They also found that the MLR coefficients were heavily dependent on the assumption of a constant BHP. When the BHP changed, there were drastic differences in the coefficients. They suggest adding BHP data to the analysis, or the selection of data is such that one obtains a near-constant BHP.

In the thesis by Mata (2010), they employ a search algorithm to identify periods when one well is experiencing increased BHP while the other wells in the region remain unchanged. The pressure change was then treated as a pulse test, which they used to identify responding wells. To identify the time lag between the stimulating pressure pulse and the responding well, they calculated the time derivatives of the pressure for the stimulating and responding well. They noted that the time derivatives of the pressure pulse became a series of unique peaks. From (Ramakrishnan, Thambynayagam, et al. 2006), by cross-correlating the time derivatives of the pressure in the stimulating and responding wells, they could reliably extract the time lag. These pulses were easily identified in some well-pairs. In others, the responding signal was too weak to be identified as a response. For a shorter time scale, the method of cross-correlating the pressure pulse could be used. Finally, they applied a field-wide solution to CRM obtained from (Sayarpour et al. 2009) for longer time scales. The methods were tested on both synthetic and real data. In conclusion, they found that the field-wide CRM solution did not apply to the real data, as they could not identify large enough periods in which the assumptions of CRM were satisfied. However, for the shorter time scale, the use of pressure pulses, and cross-correlation between well-pairs, they could extract time lags consistent with well-test analysis.

1.2.3 Deep learning

In (Panda and Chopra 1998), they applied an Artificial Neural Network (ANN) using injection rates and production rates as input and output. In its most basic form, ANN

consists of an input layer, a hidden layer, and an output layer. For each layer, there are several processing units called neurons. These accept inputs and generate outputs that depend on some activation function. The hidden layer contains weights for each neuron, which will be updated to minimize the output error. Panda and Chopra (1998) applied the method to synthetic data, and the trained model was accepted when the accuracy was greater than 99%. From the trained model, they were able to identify geological features in the reservoir, such as localized permeability trends, faults, and pinch outs.

Du et al. (2020) examines the connectivity between wells in a reservoir, using simulated data of one injection well and four producing wells. They set up the reservoir as four equal quadratic blocks with different average permeabilities and a production well in each block. Two machine learning methods were used; Back Propagation Neural Network (BP) and Convolutional Neural Network (CNN). The structure of a BP is very similar to an ANN. The only difference is that the BP will feed the error terms back through the various layers of its structure and update its weights. The error terms are calculated as the difference between a predicted and an actual value. The CNN structure consists of an input layer, a convolutional layer, a pooling layer, a fully connected layer, and an output layer. The convolutional layer consists of a weight matrix that is smaller than the input data. The weight matrix is slid over the input data, and the dot product between the inputs and the weight matrix is calculated, thus reducing the input size. The pooling layer is simple downsampling of the convoluted input, which further reduces the problem's dimensionality. The input, output, and fully connected layers all perform similarly to a traditional ANN. In (Du et al. 2020), the BP and CNN were trained using oil production, injection pressure, and water cut in the input layer. The output layer had four nodes, which corresponded to each of the four blocks' average permeability. To quantify connectivity between the injector and the producers, they define a connectivity factor as the predicted permeability for each block over the sum of predicted permeabilities. They conclude that the CNN model had a more accurate prediction with an Average Absolute Relative Deviation (AARD) of 15.35%.

Cheng et al. (2020) applied a Long Short-Term Memory (LSTM) model with an Extended Fourier Amplitude Sensitivity Test (EFAST) on a synthetic dataset. A unique feature of LSTM, compared to a traditional ANN or BP, is the memory layer. This memory layer enables the algorithm to predict on input with very large time dependencies while keeping a constant error rate, thus avoiding exponential growth or decay on the error gradients (Hochreiter and Schmidhuber 1997). The memory layer contains memory cells, input and output gates, and a forget gate (Sak et al. 2014). EFAST is a global sensitivity analysis, which quantitatively expresses how a model's output reacts to the change in the input parameters (Saltelli et al. 1999). Cheng et al. (2020) argues that an LSTM model is a good fit for the non-linear and complex relationship one can find between producing and injecting wells. An LSTM model was built for each of the producers, using all injectors as inputs. They had one hidden layer with 20 nodes and a single output node for the predicted production. The activation functions were logistic sigmoid and hyperbolic tangent. Additionally, they used a dropout layer in order to avoid overfitting the data. Once the LSTM model was trained to a satisfactory accuracy, they applied the EFAST analysis to obtain the weights that each of the injectors contributed to the total predicted production for each producer. Using the LSTM-EFAST method, they obtained a good fit between the synthetic production data and the predicted pro-

duction data, with a Root Mean Square Error (RMSE) of 0.02.

1.2.4 Knowledge gaps

No studies were found to employ Tubing Head Pressure (THP) to infer interwell connectivity. THP is a pressure reading from the top of the well, usually installed upstream of the choke valve. Due to its proximity to the choke valve and, consequently, the production separator, it is expected to contain a higher degree of variance than the BHP. BHP data is not always readily available for all wells in older fields, and production rate data is usually not sampled at a similarly high frequency as the BHP data. In this thesis, we will employ BHP data where this is available and THP otherwise. Inclusion of THP allows for including a higher number of wells, which do not have BHP data, in the search for interwell connectivity.

No studies were found to treat the inference of interwell connectivity as a classification problem, demonstrate the model's ability to identify unknown connectivities and report a metric of the predictive power of the model they employ. In this thesis, emphasis is on classification and reporting the success rate of the classification.

A few studies found inferring interwell connectivity on a mature, large-scale oil and gas field. However, most of these studies had strict assumptions, such as requiring a constant BHP or a constant production/injection rate. In the thesis by Mata (2010), the field examined is of large scale and mature. However, they perform an initial clustering/-grouping of the wells to reduce possible well-pair combinations. In this thesis, a highly complex oil and gas field will be analysed, no assumptions are made on the BHP/THP data, and there is no initial filtering to reduce the number of possible well-pair combinations.

1.3 Objectives

- Classify selected well pairs consisting of a causing well and a target well into two categories: *connectivity* and *no connectivity*. Classification performed by the methods PCMCI+ and TCDF.
- Obtain a measure of the strength of connectivity for wells that are classified with *connectivity*.

Chapter 2

Theory

2.1 Interwell connectivity

Interwell connectivity is a common term used to describe a translation of information from one well to another in a defined open/closed system. However, we have not found a single definition to describe or quantify this term. This term can be divided into three parts; the information being translated, the degree of connectivity, and the time lag between the stimulating signal and the response.

The information being translated have commonly been the production rates and injection rates, i.e., the fluid flow (Albertoni and Lake 2003; Cheng et al. 2020; Heffer et al. 1997; Jamali and Ettehadtavakkol 2017; Jansen and Kelkar 1997; Soeriawinata and Kelkar 1999; Yousef, Gentil, et al. 2006). BHP has also been used by identifying a pressure pulse from a stimulating well to a responding well (Dinh 2009; Mata 2010; Ramakrishnan and Raghuraman 2004). An advantage of setting fluid flow as input/output is the direct relation to injection rates and production rates. However, there are usually no continuous flow measurements for each well, and it is instead measured on a weekly to monthly well test interval. On the other hand, BHP measurements can be obtained down to time scales of minutes or seconds. This thesis will use pressure as the translated information and apply BHP and THP as the input and output data due to its smaller time scales.

The degree of the response is defined by the model employed to identify connectivity. Using Spearman rank correlation analysis, the correlation coefficients have been used as a quantitative measure of the connectivity (Jansen and Kelkar 1997; Soeriawinata and Kelkar 1999). With MLR models, the weight coefficients were used to define the interwell connectivity (Albertoni and Lake 2003; Dinh 2009). CRM takes advantage of the MLR algorithm, and the connectivity was related to the weight coefficients (Yousef, Gentil, et al. 2006). In (Du et al. 2020), using a CNN to predict the average permeability between wells, they defined an interwell connectivity factor as a ratio of the predicted permeability between the wells. Cheng et al. (2020) employed an LSTM-EFAST model; the strength of the response was related to the sensitivity analysis (EFAST). The PCMCI+ algorithm provides p-values to evaluate the uncertainty in the connection and t-statistics to measure the connectivity strength. The TCDF algorithm does not quantify the strength of the connection. In this thesis, we define the connectivity strength by using the t-statistics from the PCMCI+ algorithm.

The time lag is defined as the time it takes for the information being initiated by a stimulating well to be received in a responding well. This definition is common for all of the models we have studied in this thesis. However, different methods have been used to obtain the time lag. In Spearman rank correlation analysis, this has been obtained by fixing the stimulating well in time and searching for a peak correlation coefficient over a time range in the responding well (Mata 2010; Soeriawinata and Kelkar 1999). Using CRM, the time lag was obtained through an iterative process of testing various lags and optimizing the connectivity weights (Yousef, Gentil, et al. 2006). Through the PCMCI+ algorithm, the time lag is identified by applying a test of conditional independence and identifying the time lag with the highest dependence. A distinction is made between lagged and instantaneous (contemporaneous) time steps. Here, an instantaneous connection means that the temporal dependency between two variables occurs on a smaller time scale than the sampling interval of the data. Conversely, a lagged connection means that the temporal dependency occurs on a larger time scale than the sampling interval.

2.2 Time series analysis

In time series analysis, we aim to obtain some new information from time series data. Where time series data are some time-ordered measurements $X_t \in \mathbf{X}$, and t is a time index. In this thesis, the existence (or non-existence) of a relationship between two time series, $\mathbf{X}$ and $\mathbf{Y}$, is what we are looking to prove and quantify.

2.2.1 Hypothesis testing

Hypothesis testing is the idea of disproving a statement made regarding a set of observations or variables. These statements should be simple and specific (Banerjee et al. 2009). One such statement is that *there is connectivity between well A and well B*, this statement can be clearly defined as either true or false. By rewriting the statement as *Well A is dependent of Well B*, it can be defined in mathematical terms as (2.1).

$$\mathbf{X} \not\perp\!\!\!\perp \mathbf{Y} \tag{2.1}$$

Where, $\mathbf{X}$ and $\mathbf{Y}$, represent the wells A and B, respectively, and $\not\perp\!\!\!\perp$ is the dependence operator. To prove categorically that is true, all possible observations of the variables in (2.1) must be known and shown to be true, which is an unfeasible amount of data to process. Instead, we look to disprove that it is true, which is a much simpler task as it only requires a single observation where the statement is false. We separate the statement into two; a null hypothesis H_0, and an alternative hypothesis H_1, as shown in (2.2). Where the null hypothesis is assumed to be true until proven otherwise.

$$\begin{aligned} H_0 &: \quad \mathbf{X} \not\perp\!\!\!\perp \mathbf{Y} \\ H_1 &: \quad \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \end{aligned} \tag{2.2}$$

Where $\perp\!\!\!\perp$ is the independence operator. We can define a *measure of dependency*, such as partial correlation from the PCMCI+ algorithm or attention score from the TCDF-algorithm, which is easier to quantify and prove, as compared to the more ambiguous term of *connectivity*. Partial correlation and attention score will be further explained in

sections 2.3.1 and 2.3.2, which details the PCMCI+ and TCDF-algorithms.

Statistical tests

For any given measurement of dependency, d, we must make an assumption of at which level d signifies *dependence* or *independence*. This assumption may simply be that $d \geq 0$ signifies *dependence*. However, for a series of dependency measurements, $d_t \in \mathbf{d}$, on real data, it is unlikely that all of these measurements will be ≥ 0. Instead, we aim to disprove that $\mu_d \geq 0$, where μ_d is the mean value of $\mathbf{d}$. The hypothesis test (2.2) may be rewritten as (2.3).

$$\begin{aligned} H_0 : &\quad \mu_d \geq 0 \\ H_1 : &\quad \mu_d < 0 \end{aligned} \tag{2.3}$$

To disprove that $\mu_d \geq 0$, a statistical test such as Student's t-test or Wilcoxon Signed Rank test may be used.

Student's t-test is a method that compares the mean of two variables and estimates the difference between them (Livingston 2004). The t-test is a parametric test that assumes the underlying data is normally distributed. Additionally, there is an assumption that the variance in both variables should be approximately equal. To perform a statistical test using the t-test method on (2.1), we use μ_d as a mean of one variable, 0 as a fixed value. Using a single variable against a fixed value is also known as a one-sample hypothesis test, as we only test one variable. Using two variables, we refer to it as two-sample hypothesis tests. The one-sample t-test can then be calculated as shown in (2.4).

$$t_S = \frac{\mu_d - 0}{\sigma / \sqrt{n}} \tag{2.4}$$

t_S is the t-statistic, σ is the standard deviation of the observations, and n is the number of observations or measurements. The calculated t-statistic can then be used to quantify the difference between μ_d and 0.

Wilcoxon Signed-Rank test is a method that compares the median of two variables (Wilcoxon 1945). This method is a non-parametric test that does not assume that the underlying data is normally distributed. For tests on a two-sample hypothesis, the Wilcoxon Signed-Rank test requires that the two samples are paired, meaning that they are dependent somehow. Calculation of the Wilcoxon test statistic, W-statistic, for (2.1) is performed by the following steps:

1. Calculate the difference: $\mathbf{D} = d_t - 0$ for $d_t \in \mathbf{d}$. If $d_t = 0$, calculate the mean instead of taking the difference for this sample.
2. Rank all samples in $\mathbf{D}$ on their absolute value: $\mathbf{R} = rank(|\mathbf{D}|)$
3. Calculate the sum of all positive and negative ranks: $\sum \mathbf{R}^+$ and $\sum \mathbf{R}^-$. The sign of the rank is found by the corresponding sample in $\mathbf{D}$.
4. The test statistic is found by taking the smallest value of the summed ranks, $W = min(\sum \mathbf{R}^+, \sum \mathbf{R}^-)$.

This calculation is for small sample sizes ($n < 20$), for larger sample sizes, the test statistic can be approximated to a normal distribution (Bellera et al. 2010). We use this method for all one-sample hypothesis tests where the underlying data follows a non-normal distribution.

Mann-Whitney U-test is a test that compares the difference between two variables (Mann and Whitney 1947). This test is a non-parametric test that does not assume the underlying data is normally distributed. Instead, it assumes that the two variables have distributions of the same shape. Unlike the Wilcoxon Signed Rank test, this test can be applied to two variables of unequal size. The U-statistic is calculated as shown in (2.5).

$$U^i = \sum \mathbf{R}^i - \frac{n^i(n^i+1)}{2}, \quad for \quad i \in \{1,2\} \tag{2.5}$$

All samples from both groups, 1 and 2, are ranked against each other in a single set, $\mathbf{S}$. The corresponding ranks are split into their respective groups, $\mathbf{R}^i$. Here, n^i is the number of samples in each group. Calculation of U^i for $i \in \{1,2\}$ returns two values for U, and it is the smaller of the two that is used as the test statistic. We use this method for all two-sample hypothesis tests where the underlying data follows a non-normal distribution.

Probability value

The t-statistic and the W-statistic can be used directly to quantify the difference between two variables. However, it is impractical only to report the test statistics, as they do not follow a bounded range and vary widely depending on which test is used. Instead, the probability value (or p-value) can be used. The p-value is obtained from the probability distribution of the respective test statistics. If the obtained test statistic is unlikely to occur under the assumption of the null hypothesis, the corresponding p-value will be low. To determine if the null hypothesis should be rejected, a significance threshold, α, needs to be set. Typically a low value such as $\alpha = 0.05$ or $\alpha = 0.01$ is used. If p-value $< \alpha$, we reject the null hypothesis, as it is unlikely for the test-statistic to occur given the assumption that the null hypothesis is true. Instead, we accept the alternative hypothesis as the more likely case to be true.

2.2.2 Classification metrics

When some analysis is performed on labeled data, classification metrics may be used to evaluate the performance of the models employed. Labeled data means that the data can be separated into categories, such as labeling a well-pair with *connectivity* or *no connectivity*. When there are only two labels to classify, we call this binary classification, and a confusion matrix with its corresponding metrics is an appropriate method for evaluation of the models (Cheung and Lai 2009).

The confusion matrix consists of grouping wrong and correct classifications by the model into four groups of universal labels; True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). In this thesis, we set *connectivity* as Positive, and *no connectivity* as Negative. Classifying a well-pair with known connectivity with the label *connectivity* is counted towards TP, classifying the same well-pair with the label *no connectivity* is counted towards FN. Similarly, for a well-pair with no known connectivity, correct classification of *no connectivity* is counted towards TN, and classifying the same

well-pair with *connectivity* is counted towards FP.

Sensitivity, specificity, and precision are three metrics that can be calculated from the confusion matrix. Sensitivity is a ratio of the correctly classified positives against all positives in the data, as shown in (2.6). This metric tells us how accurate the model is at correctly classifying *connectivity*.

$$sensitivity = \frac{TP}{TP + FN} \tag{2.6}$$

Specificity is similar to sensitivity, except it calculates the ratio for the negatives, as shown in (2.7). This metric tells us how accurate the model is at classifying *no connectivity*.

$$specificity = \frac{TN}{TN + FN} \tag{2.7}$$

Precision is the ratio of all correctly classified positives against all classified positives, as shown in (2.8). This metric indicates how precise the model is at classifying the label we set as positives.

$$precision = \frac{TP}{TP + FP} \tag{2.8}$$

2.3 Causal Inference

Similar to correlation analysis, causal inference aims to measure the relationship between two or more variables. However, causal inference determines both the strength of the relationship as well as the direction of causation.

An important aspect of causal inference models is the identification of a confounder. From Reichenbach's common cause principle: If there is a correlation between two random variables, **X**, and **Y**, and a direct causal relationship can be excluded, then there must be a third variable, **Z**, which causally influences both (Hofer-Szabó et al. 1999). This relationship is visualized by a Directed Acyclic Graph (DAG) in Figure 2.1.

2.3.1 Peter Clark Momentary Conditional Independence +

The PCMCI+ algorithm introduced by Runge (2020) is a complete algorithm for inference of causal relationships. The algorithm takes an arbitrary number of time series variables as inputs (limited by computational power costs) in a single dataset, and by Conditional Independence (CI) tests, returns a complete set of inferred causal relations for the time series variables. There are two main components to the method, the PC1 algorithm (a modification of the PC algorithm by Spirtes and Glymour (1991), and the MCI algorithm (Runge et al. 2019)).

(Runge 2020), provides a a function, `run_bivci()`, that plots the lagged correlations to visually determine how to set the maximum time lag, τ_{max}, for the PCMCI+ algorithm. This function applies the chosen independence test and searches over a specified time range. Setting a large maximum time lag does not degrade the model or make it

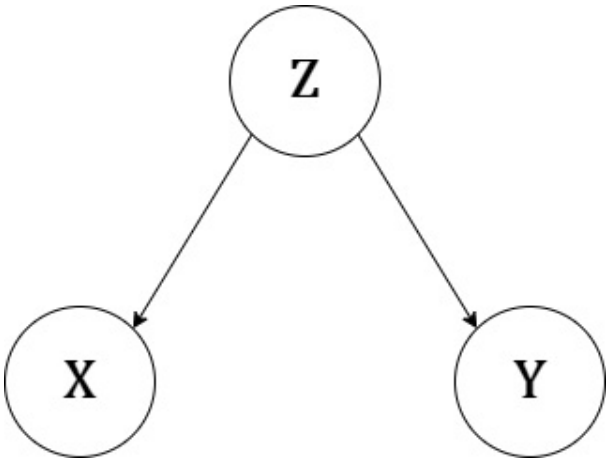


Figure 2.1: Directed Acyclic Graph visualizes Reichenbach's common cause principle. Nodes indicate variables and arrows indicate a direct causal link. Here, **Z**, is the confounding variable.

overfit. This step is done to avoid using excessive computational power. In Figure 2.2, an example of the generated plot by the `run_bivci()` function is shown, using three wells and the ParCorr independence test. In this example, peak partial-correlation occurs for all well-pairs when $\tau = 0$.

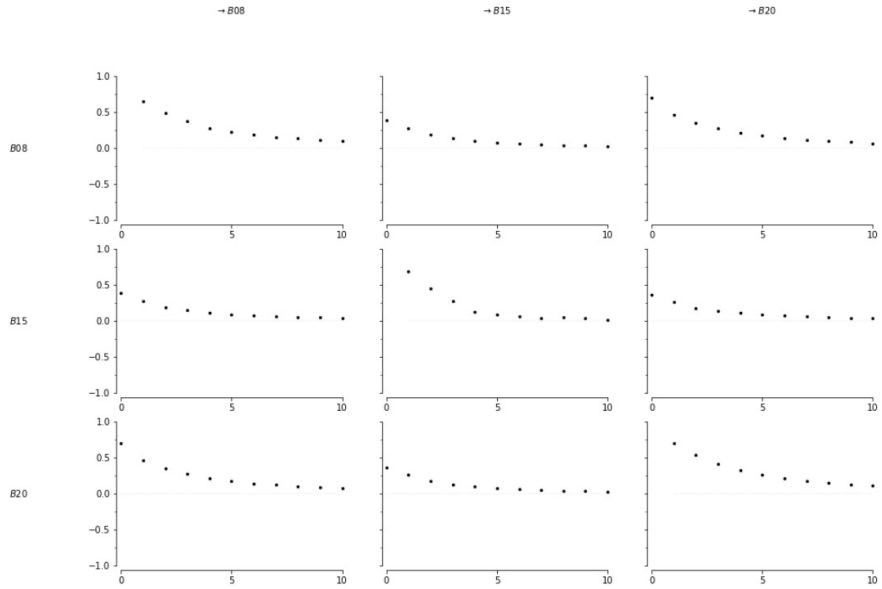


Figure 2.2: Unmodified plot generated by the `run_bivci()` function from the Tigramite package by Runge (2020). The x-axis represent the time lag, the y-axis represent the partial correlation measure.

Conditional independence

The topic of conditional independence is a fundamental one in causal discovery (Dawid 1979), and it is a core part of the PCMCI+ algorithm. Conditional independence can be expressed as (2.9):

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} | \mathbf{Z} \tag{2.9}$$

Here, $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ are random variables, $\perp\!\!\!\perp$ is the independence operator. The relationship in (2.9) is represented visually by the DAG in Figure 2.1, and we say that $\mathbf{X}$ is independent of $\mathbf{Y}$ given $\mathbf{Z}$. By independence, we mean that no information from $\mathbf{X}$ can be related to $\mathbf{Y}$. There are different methods used to test for conditional independence. We will describe the method of Partial Correlation testing, which is the method used in this thesis for the PCMCI+ algorithm.

Partial correlation

Partial correlation testing is a method where the linear dependence between two variables is measured, given that the influence of all other variables has been eliminated. The following is a short description of how to estimate partial correlation using the variables from (2.9); we look to estimate the partial correlation between the variables $\mathbf{X}$ and $\mathbf{Y}$:

1. Perform simple linear regression of $\mathbf{X}$ onto $\mathbf{Z}$ and return the residuals, $\mathbf{e_X}$.
2. Perform simple linear regression of $\mathbf{Y}$ onto $\mathbf{Z}$ and return the residuals, $\mathbf{e_Y}$.
3. Calculate Pearson's correlation coefficient between the residuals, $\mathbf{e_X}$ and $\mathbf{e_Y}$.

The correlation coefficient from step (3) in the above is known as the partial correlation and can be noted as $\rho_{XY \cdot Z}$ (Pellet and Elisseeff 2007). We shorten the partial correlation notation to ρ.

The PCMCI+ workflow

1. **Input:** The PCMCI+ algorithm is given an input of a single dataset (2.10), where N is the number of variables, and T is the end of the time series. The dataset is formatted as a two-dimensional array, with variables in the columns and time steps in the rows. In our case, this dataset contains columns of individual wells' pressure data, and each row represents a time step of 5 minutes.

$$\mathbf{S} = \{X_t^i \,|\, t \in T, i \in N\} \tag{2.10}$$

2. **Algorithm:**
   (a) A fully connected DAG is initialized (excluding contemporaneous dependencies), on all X_t^i for $t \in [1, \tau_{max}]$ ($t = 0$ is excluded in this step) and $i \in N$. Here, τ_{max} is the maximum time lag to search for a link. We set this to $\tau_{max} = 100$ which translates to a maximum time lag of 8 hours and 20 minutes. The node pair dependencies (links) are ordered by time (2.11), i.e., past nodes can only be linked forward in time (Figure 2.3a).

$$X_{t-1}^i \to X_t^i \tag{2.11}$$

(b) The PC1-algorithm is applied to the time series data: A test of conditional independence is applied to every link (2.12).

$$\rho^{ij} = \text{parcorr}(X^i_{t-\tau_{max}}, X^j_t) \tag{2.12}$$

For all $i, j \in [1, N]$ and $t \in [1, \tau_{max}]$. Here, i and j, represent two separate variables from the input (or two separate wells), N is the number of variables used as inputs for the model (in this case, the number of wells), parcorr() is the partial correlation function, and ρ^{ij} is the partial correlation between i and j. Every link generates a test statistic and corresponding p-value, by formulating a hypothesis test as (2.13) and applying Student's t-test.

$$\begin{aligned} H_0 : &\quad X^i_{t-\tau_{max}} \perp\!\!\!\perp X^j_t \\ H_1 : &\quad X^i_{t-\tau_{max}} \not\perp\!\!\!\perp X^j_t \end{aligned} \tag{2.13}$$

For all links that returned a p-value less than a set significance threshold, α, the null hypothesis is rejected, and the test statistic and p-values are stored and sorted. These links, $(X^i_{t-\tau_{max}}, X^j_t)$, were calculated as a subset of dimension zero, $\mathbf{S}_0$, i.e., no other nodes were included in the conditioning of the given link. The test statistic and p-value for each link are then updated by iteratively increasing the number of nodes included in the subset. The node added is the one with the highest test statistic from the previous iteration. Thus, a subset of dimension one, $\mathbf{S}_1$, includes the given link, $(X^i_{t-\tau_{max}}, X^j_t)$, and the highest scoring node from the previous iteration, $max(\mathbf{S}_0)$. Updated test statistics and p-values are calculated from the new subset, $\mathbf{S}_1$. The iterations stop once the dimension of the subset covers all of the remaining nodes, $\mathbf{S}_n$, setting the time-lagged links between the nodes as shown in Figure 2.3b.

(c) A new DAG is initialized and linked on all contemporaneous node pairs (2.14).

$$(X^i_t, X^j_t) \quad for \quad i \neq j \tag{2.14}$$

The MCI test is applied to all of the lagged links identified in step (b) and all contemporaneous links to identify false positives from step 1 and removing independent contemporaneous links. Unlike the algorithm in step (b), the MCI-algorithm does not iterate over different subsets for the conditioning phase. Instead, it takes the final subset, $\mathbf{S}_n$, from step (b) and performs a single iteration for each node pair. The MCI-algorithm returns a test statistic and p-value for all of the node pairs. The contemporaneous links are left undirected, as shown in Figure 2.3c.

(d) The undirected contemporaneous links are given direction by applying a majority or conservative rule. This determination of direction is achieved by retesting conditional independence for the contemporaneous links on both variables $\mathbf{X}^i$ and $\mathbf{X}^j$, Figure 2.3d.

3. **Output:** The results of the PCMCI+ algorithm is matrices of dimensions $[N_{cause}, N_{target}, \tau_{max}+1]$. There are two matrices of note, the value matrix (`val_matrix`) and the p-value matrix (`p_matrix`). The value matrix contains values from the conditional independence tests, in this case it is the partial correlation measure. The p-value matrix contains p-values corresponding to the partial correlation measures.

Runge et al. (2019) mentions that these can be directly interpreted as the uncertainty (p-values) in the links and the strength (partial correlations) of the links. In this case, the value matrix correspond to the strength of connectivity between all wells

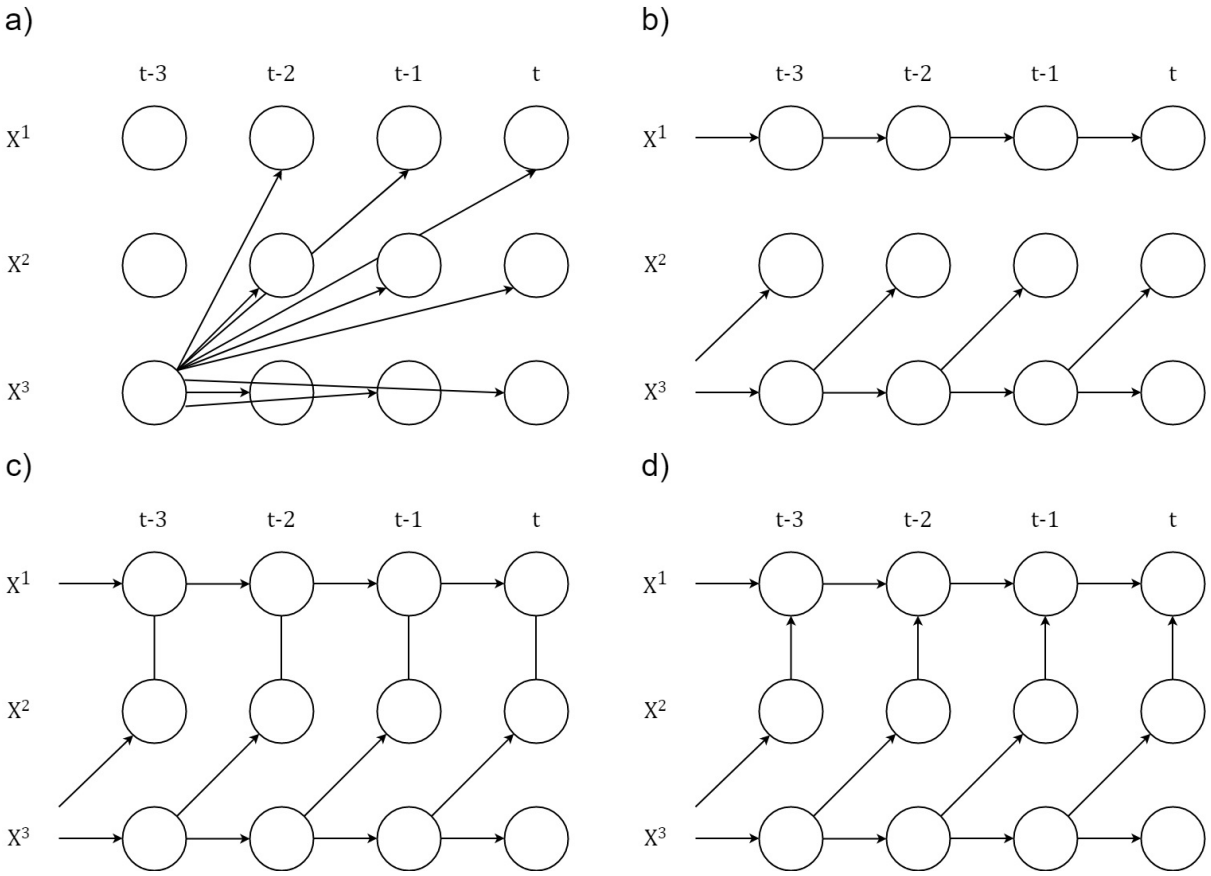


Figure 2.3: Visualization of the PCMCI+ algorithm. $\mathbf{X}^i$ for $i \in \{1, 2, 3\}$ each represent time-series data. Circles are single datapoints from each time series $\mathbf{X}^i$, ordered horizontally by the time index t. Arrows are directed dependencies and lines are undirected dependencies. a) For simplicity, a single node is visualized. Fully connected and time-lagged forward in time. b) Directionally linked nodes on time-order. $\mathbf{X}^1$ is shown to be fully independent of $\mathbf{X}^2$ and $\mathbf{X}^3$. $\mathbf{X}^2$ is dependent on the time lagged $\mathbf{X}^3$. c) Contemporaneously dependent nodes linked without direction using the MCI test. $\mathbf{X}^2$ and $\mathbf{X}^1$ found to be contemporaneously linked. d) Directed contemporaneous links have been identified between $\mathbf{X}^1$ and $\mathbf{X}^2$. Adapted from (Runge et al. 2019).

Tuning parameters associated with the PCMCI+ algorithm:

- The conditional independence test used is a partial correlation test, implemented by the function `parcorr()`. This method tests for linear dependence between the variables input to the model and returns a partial correlation measure in the range $[-1, 1]$. The strength of a link between two variables can be related to the returned partial correlation measure for the respective variables. Other options can handle non-linear data. However, those were not tested in this thesis.

- Selection of min/max time lag, τ_{min}/τ_{max}: This parameter sets the limit at which the algorithm will search for conditional independence. We will set $\tau_{min} = 0$, and τ_{max} will be selected such that it at least encompass all time lags of known connectivities between the wells in this thesis. By increasing τ_{max}, the model does not degrade in its ability to infer dependence. Instead, it is an increased computational power cost.

- Significance threshold, α: Sets the threshold in which we reject the null hypothesis for independence between two variables. Lower values mean that we fail to reject the null hypothesis for independence more often. We will select the significance threshold of the full-scale test by evaluating the results from the small-scale test.

- Contemporaneous collider rule and Conflict resolution: These parameters control whether the PCMCI+ algorithm is order-independent in terms of the input variables. We chose the default values for these parameters, ensuring that the PCMCI+ algorithm is order-independent.

- Reset lagged links: Options are `True` or `False`. When the parameter is set to `True`, it can improve the detection of lagged causality links by considering all lagged links during the second phase (MCI) of the PCMCI+ algorithm. When `False`, it will only consider lagged links identified and restricted by the significance threshold in the first phase (PC). We set this parameter to `True`, as this can only improve the model, though at a higher computational power cost.

Method demonstration

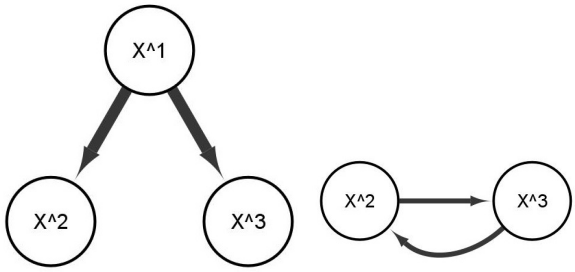
To demonstrate that the method works, we generate a simple synthetic time-series with three variables, (X_t^1, X_t^2, X_t^3), where the variables, X_t^2 and X_t^3, are dependent on the variable, X_t^1, with a time-delay of 3 and 5 steps, as shown in (2.15). The data was generated by the `var_process()` function provided in the Tigramite package by Runge (2020). Each variable can be considered as the pressure time-series seen in a given well, and the dependent variables can be considered as connected wells. This synthetic time-series also reflects Figure 2.1. In Table 2.1, the parameters used for the demonstration is shown.

$$\begin{aligned} X_t^1 &= 0.7X_{t-1}^1 + \epsilon \\ X_t^2 &= 0.7X_{t-1}^2 + 0.4X_{t-3}^1 + \epsilon \\ X_t^3 &= 0.7X_{t-2}^3 + 0.4X_{t-5}^1 + \epsilon \end{aligned} \tag{2.15}$$

Table 2.1: Parameters for the PCMCI+ algorithm used for the method demonstration.

CI test	**Min/max time delay**	**Significance threshold**	**Contemp. collider rule**	**Conflict resolution**	**Reset lagged links**
ParCorr	0/10	0.01	Majority	True	False

The results we obtained for the demonstration is shown in (2.16). The method was fully accurate in classifying links and relatively accurate in estimating the links' strength (coefficients).



(a) DAG of (2.16). (b) DAG of (2.17).

Figure 2.4: Figures a) and b) show DAG illustrations for the method demonstration of the PCMCI+ algorithm. Loops of self-causation were removed. Illustrations generated by Cytoscape (Shannon et al. 2003)

$$\begin{aligned} \hat{X}_t^1 &= 0.613X_{t-1}^1 \\ \hat{X}_t^2 &= 0.614X_{t-1}^2 + 0.375X_{t-3}^1 \\ \hat{X}_t^3 &= 0.508X_{t-2}^3 + 0.329X_{t-5}^1 \end{aligned} \tag{2.16}$$

What if the synthetic time-series contains a hidden confounder? In (2.17), the results of the PCMCI+ algorithm is shown with the variable X_t^1 hidden from the model. Here, we see that the method inaccurately infers some dependency between the variables X_t^2 and X_t^3 on a zero time delay. In Figure 2.4, the corresponding DAG for (2.16) and (2.17) is shown. Self-causation loops were removed, the width of the links

$$\begin{aligned} \hat{X}_t^2 &= 0.696X_{t-1}^2 + 0.186X_t^3 \\ \hat{X}_t^3 &= 0.560X_{t-2}^3 + 0.186X_t^2 \end{aligned} \tag{2.17}$$

2.3.2 Temporal Causal Discovery Framework

The TCDF algorithm is a new approach to causal inference introduced in (Nauta et al. 2019), which attempts to employ a neural network to predict causal relationships between time series while keeping it interpretable. This method is fully separate from the PCMCI+ algorithm mentioned above, which tests for conditional independence by simple linear regression.

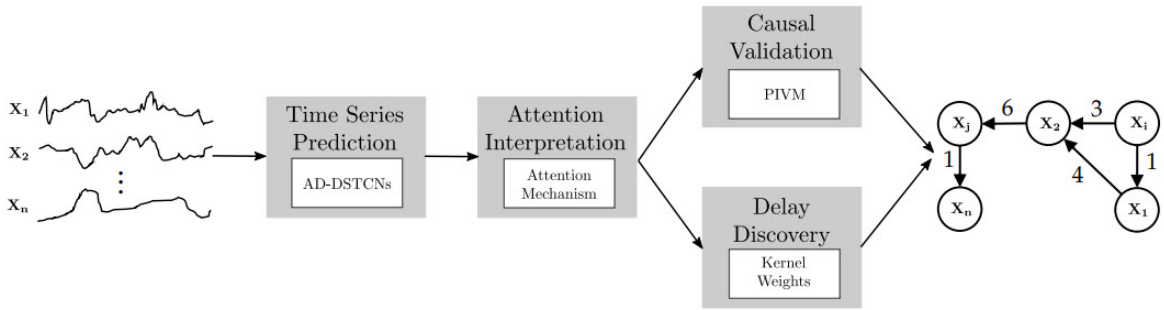


Figure 2.5: TCDF algorithm overview. From (Nauta et al. 2019)

In Figure 2.5, an overview of the TCDF algorithm is presented. TCDF can be divided into four main parts; time series prediction using a modified CNN approach, attention interpretation, causal validation, and delay discovery.

The TCDF workflow

1. **Input:** The TCDF-algorithm is given an input of a single dataset (2.18) where N is the number of variables, and T is the end of the time series. The dataset is in the format of a .csv file. In our case, this dataset contains columns of individual wells' pressure data.

$$\mathbf{S} = \{X_t^i \mid t \in T, i \in N\} \tag{2.18}$$

2. **Algorithm:**

   (a) In the time series prediction part, they present an architecture they call Attention-based Dilated Depthwise Separable Temporal Convolutional Networks (AD-DSTCNs). The input for the model is N-number of time series, $\mathbf{X}^i$, and a separate CNN is applied to each time series, $\mathbf{X}^i$, to predict $\hat{\mathbf{X}}^j$ (2.19).

$$\hat{\mathbf{X}}^j, \mathbf{a}^j = \text{CNN}(\mathbf{X}^i) \quad for \quad i, j \in [1, N] \tag{2.19}$$

   To determine which time series in $\mathbf{S}$ influences the predicted time series, $\hat{\mathbf{X}}^j$, they apply an attention mechanism. Each predicted time series, $\hat{\mathbf{X}}^j$, will have an accompanying attention vector, $\mathbf{a}^j$, containing attention scores a^{ji} as shown in (2.19). Where a^{ji} can be seen as the weight that the CNN has given the time series, $\mathbf{X}^i$, to predict time series, $\hat{\mathbf{X}}^j$. The attention scores are initialized as 1, and they are increased or decreased depending on whether the CNN uses $\mathbf{X}^i$ to predict $\hat{\mathbf{X}}^j$. The attention scores are updated after each training epoch. A high attention score, $a^{ji} > 1$, implies that there is a causal link from $\mathbf{X}^i$ to $\mathbf{X}^j$.

   (b) Attention interpretation is the next part of the algorithm and deals with processing the attention scores, $\mathbf{a}^j$. After training on the CNNs, the attention scores are truncated as shown in (2.20):

$$\mathbf{A}^j = \begin{cases} a^{ji} & if \quad a^{ji} > \gamma^j \\ 0 & otherwise \end{cases} \tag{2.20}$$

   The idea is to set a lower limit on the attention scores that implies a causal link. The threshold, γ^j, is obtained by a search algorithm and a set of rules applied to the attention scores. These rules include: $\gamma^j \geq 1$, number of potential causes can not be greater than 50% of N time series, and there must be two or more time series as potential causes. They introduce potential causes as $\mathbf{P}^j$, where the time series $\mathbf{X}^i$ is added to $\mathbf{P}^j$ as a potential cause to $\mathbf{X}^j$ given a set of rules applied to their attention scores. The relational rules are given for $A^{ji} \in \mathbf{A}^j$ and $A^{ij} \in \mathbf{A}^i$:

      i. $A^{ji} = 0$ and $A^{ij} = 0$: No correlation between $\mathbf{X}^i$ and $\mathbf{X}^j$.
      ii. $A^{ji} = 0$ and $A^{ij} > 0$: $\mathbf{X}^i$ is added to $\mathbf{P}^j$.

iii. $A^{ji} > 0$ and $A^{ij} = 0$: $\mathbf{X}^j$ is added to $\mathbf{P}^i$.

iv. $A^{ji} > 0$ and $A^{ij} > 0$: $\mathbf{X}^j$ is added to $\mathbf{P}^i$ and $\mathbf{X}^i$ is added to $\mathbf{P}^j$.

(c) In the causal validation step, they introduce the Permutation Importance Validation Method (PIVM). This method seeks to identify which potential causes, $\mathbf{P}^j$, are true causes, $\mathbf{C}^j$, of the time series $\mathbf{X}^j$. The true causes, $\mathbf{C}^j$, are identified by comparing two differences in prediction losses. The first difference in prediction loss D_r^i is obtained by calculating the difference in loss from the first and last epoch (2.21).

$$D_r^i = RMSE(\mathbf{e}_{first}^i) - RMSE(\mathbf{e}_{last}^i) \quad for \quad \mathbf{X}^i \in \mathbf{P}^j \tag{2.21}$$

$\mathbf{e}_{first}^i$ and $\mathbf{e}_{last}^i$ is the residuals from the first and last epoch of training on the time series i. The second difference in prediction loss D_R^i, is obtained by first randomly permuting the values (2.22) and then training on these permuted values as in (2.19) and calculating the difference as in (2.21).

$$\mathbf{X}_R^i = \text{rand}(\mathbf{X}^i) \quad for \quad \mathbf{X}^i \in \mathbf{P}^j \tag{2.22}$$

By randomly permuting $\mathbf{X}^i \rightarrow \mathbf{X}_R^i$, temporal precedence has been removed, and one expects to see a degradation in the predictability of causation for $\mathbf{X}^i$ onto $\mathbf{X}^j$. In essence, if the loss in the second prediction is significantly greater than the first prediction, it implies that $\mathbf{X}^i$ was a true cause of $\mathbf{X}^j$. They add a hyperparameter: *significance*, $s \in [0, \infty]$, used as a tuning parameter to accept a prediction loss as significant. The s-parameter controls to what degree we allow potential causes to be converted to validated causes. A lower value means we allow fewer conversions from potential to validated causes. This conversion can be observed by the expression (2.23):

$$\mathbf{X}^i \in \mathbf{C}^j \quad if \quad D_R^i \leq D_r^i \cdot s \tag{2.23}$$

(d) The final step of the algorithm is delay discovery, an estimate of the time lag between the true cause, $\mathbf{X}^i$ onto $\mathbf{X}^j$. They found that the kernel size in the CNN can be interpreted as the time lag for causation. Here, the kernel is a matrix of dimensions one by K, where $K \in [1, N]$, traverses over the input and calculates the dot-product of the variables within the kernel filter. By traversing backward through the CNN, following the largest kernel weights, the algorithm can identify which time step in $\mathbf{X}^i$ had the greatest effect on a given time step in $\mathbf{X}^j$.

3. **Output:** The results of the TCDF-algorithm are reported as a printout of all validated causes, $\mathbf{C}^j$, and their corresponding targets, $\mathbf{X}^j$, along with a time delay. To extract the attention scores, the variable `allscores` from `runTCDF.main()` can be extracted. This variable is in the format of a Python dictionary, containing all attention scores for the possible combinations of variable pairs. The target variable is set as `keys()` to the Python dictionary, and the causing variable is set as a value to the corresponding keys.

Hyperparameters associated with the TCDF-algorithm:

- Epochs: Number of epochs to run the CNNs. Epochs refer to the number of cycles the algorithm passes the entire dataset through a neural network. More epochs will reduce the training error; however, it will also lead to more overfitting. Default: 1000.

- Learning rate: Specifies the learning rate of the CNNs. Lower learning rates mean that the algorithm will take longer to fit the data, and more epochs are required. Higher learning rates will allow the algorithm to fit the data faster. When the learning rate is too high, there is a possibility that the algorithm does not converge on the minimum of the loss function. Default: 0.01.

- Hidden layers: Determines the complexity of the CNNs. More hidden layers mean a higher complexity, and the algorithm will fit the input data to the output flexibly. Default: 0.

- Optimizer: Determines which algorithm to use for gradient descent optimization. Default is the "Adam" optimizer, proposed by Kingma and Ba (2014) and shown to perform well compared to other optimization methods. We will only use the "Adam" method in this thesis.

- Significance: This is a special parameter for the TCDF-algorithm and controls the degree to which we allow potential causes to be converted to validated causes. Higher values mean that more potential causes are converted to validated causes. We will set this parameter to 1.0 to have more flexibility to interpret the attention scores of the model.

- Kernel size and dilation coefficient: The kernel size determines how much of the input time series the sliding kernel sees at each step. The kernel size relates to the maximum time delay the algorithm can find by $\tau_{max} = K - 1$ where τ_{max} is the maximum time delay. Dilation coefficient is a parameter used in the CNNs to modify the step size of the sliding kernel. Nauta et al. (2019) recommend setting this equal to the kernel size parameter. We will set both of these parameters to 100, such that the maximum time lag to search is 8 hours and 20 minutes.

Method demonstration

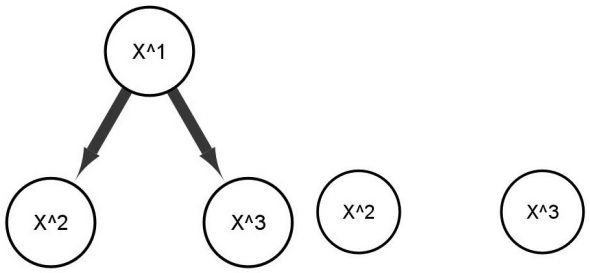
To demonstrate that the method works, we apply the same synthetic time-series as used in the demonstration of the PCMCI+ method in section 2.3.1. For readability, we repeat the synthetic time series (2.15) in (2.24). In Table 2.2, the hyperparameters used for the demonstration is shown.

$$\begin{aligned} X_t^1 &= 0.7X_{t-1}^1 + \epsilon \\ X_t^2 &= 0.7X_{t-1}^2 + 0.4X_{t-3}^1 + \epsilon \\ X_t^3 &= 0.7X_{t-2}^3 + 0.4X_{t-5}^1 + \epsilon \end{aligned} \tag{2.24}$$

In (2.25), the results of running the TCDF-algorithm on (2.24) is shown. Every dependency was correctly identified. This method does not return the strength of connectivity. As such, the links are shown without a coefficient modifier.

Table 2.2: Hyperparameters for the TCDF-algorithm used for the method demonstration.

Dilation coefficient	**Epochs**	**Hidden layers**	**Kernel size**	**Learning rate**	**Optimizer**	**Significance**
10	100	0	10	0.01	Adam	1.0



(a) DAG of (2.25). (b) DAG of (2.26).

Figure 2.6: Figures a) and b) show DAG illustrations for the method demonstration of the TCDF-algorithm. Loops of self-causation were removed. Illustrations generated by Cytoscape.

$$\begin{aligned} \hat{X}_t^1 &= X_{t-1}^1 \\ \hat{X}_t^2 &= X_{t-1}^2 + X_{t-3}^1 \\ \hat{X}_t^3 &= X_{t-2}^3 + X_{t-5}^1 \end{aligned} \tag{2.25}$$

Similarly to the PCMCI+ method demonstration, we hide the variable X_t^1 from the TCDF-algorithm. The results are shown in (2.26), and we find that the TCDF-algorithm is able to accurately ignore the dependencies induced by the hidden confounder, X_t^1. In Figure 2.6, the results from (2.25) and (2.26) is shown as DAG illustrations.

$$\begin{aligned} \hat{X}_t^2 &= X_{t-1}^2 \\ \hat{X}_t^3 &= X_{t-2}^3 \end{aligned} \tag{2.26}$$

Chapter 3

Data

The data we use for this thesis is obtained from the Eldfisk field in the North Sea, provided by ConocoPhillips. Eldfisk is a mature oil and gas field that employs waterflooding as pressure support. Production started in 1979, and since then, there have been more than 90 wells drilled (including workovers on existing wells). The field can be divided into two parts; North and South. The Eldfisk Complex is located in the southern region, which includes the production installations Eldfisk 2/7-Alpha and 2/7-Sierra. In the northern region, approximately six kilometers north of the Eldfisk Complex, the production installation Eldfisk 2/7-Bravo is located. Eldfisk complex in the southern region also provides injection water and gas lift by the utility installation Eldfisk 2/7-Ester. Both the southern and northern production installations are dependent on this delivery of water injection and gas lift.

ConocoPhillips has identified connectivity between wells over a long period. Identification of connectivities helps to understand the reservoir better, optimize oil and gas production, and avoid new and unexpected connectivities. Different methods have been employed to obtain this information, such as observing the pressure response between wells or comparing water cut for a producing well before and after initiation of a stimulating water injector. A more costly method, though more accurate, is tracer testing by injecting tracer chemicals in an injection well and sampling target wells to identify which of these are affected by the tracer chemical. Knowledge of interwell connectivities is especially important as new wells are introduced, which may open new fluid flow paths.

3.1 Data descriptions

Pressure data from a well can usually be read from two sources, the BHP and the THP. BHP is measured by pressure gauges deep in the well and is used to read the pressure in the reservoir. The THP gauge is located at the top of the well tubing and measures pressure from the well upstream of the choke valve. The THP is often seen to have a large amount of variance due to well slugging and other factors, such as changes in the pressure downstream the well. We refer to well slugging here as an accumulation of oil/water in the well, which builds up under pressure and is then released up the well; this causes an oscillating pressure profile to occur with periods of low and high pressure. From our initial tests, we found that the THP is affected by many external factors and variance, compared to the BHP, which causes the pressure reading to contain more noise, which

in turn translates to more false signals in terms of detecting or not detecting connectivity.

The on/off indicators are binary data that signal an open well, 1, or a closed well, 0. Most are triggered by a change in the automatic flow valve. This valve usually controls the opening and closing of a well. It is possible to open or close a well using other valves. However, this is a rare occurrence as it causes the THP gauge to lose its reading on the well pressure. Some of the on/off indicators, especially on newer wells, have sampling intervals of seconds and minutes instead of only registering when a change is made.

The wells used for the models are a mix of wells with known connectivities, summarized in Table 3.1. In total, there are eight known well-pairs with connectivity, mixed as five injector-producer connectivities and three injector-injector connectivities. The injection wells are water injectors, and the production wells all produce a mix of oil, water, and gas. There were two known well-pairs with producer-producer connectivity in the Eldfisk field. However, there was not enough data for these to be included in this thesis. The degree of connectivity is not established for any of the well-pairs. The time lag is visually estimated by comparing pressure trends between the two wells for each well-pair. For the well-pair A07/THP and A20/BHP, we could not identify a period where the pressure response between the two wells was strong enough to be visually represented in a figure. This connectivity has previously been identified by a chemical injection method (a tracer) in the injection well A07/THP. When the injected chemical was identified in the well A20/BHP, it was established that there is connectivity between the two wells. We include this well pair in our analysis even though the pressure response have not been identified visually, to see if the models we tested in this thesis can identify this connectivity.

Table 3.1: Summary table of the wells used for the models. Known connectivities are listed as a single link from the stimulating well to the target well, however, causation may go both ways. Time lag is the visually estimated time for a stimulus from a stimulating well to be observed as a pressure response in a target well.

Stimulating well			**Target well**			
Well	**Pressure data**	**Type**	**Type**	**Pressure data**	**Well**	**Time lag (hours)**
A-04	BHP	Injector	Producer	THP	A-01	5
A-07	THP	Injector	Producer	THP	A-18	3
A-07	THP	Injector	Producer	BHP	A-20	Unknown
B-08	BHP	Injector	Injector	THP	B-15	6
B-20	BHP	Injector	Injector	THP	B-15	2
S-02	BHP	Injector	Producer	BHP	A-03	4
S-17	BHP	Injector	Producer	THP	A-12	2
S-18	BHP	Injector	Injector	BHP	B-20	1

3.2 Connectivity examples

In Figure 3.1, the pressure response between the stimulating injection well B15/THP and the target well B20/BHP can be seen. There is a rapid increase and subsequent decrease of pressure from B15/THP in the middle of the shut-in period. This change in pressure is attributed to the THP-data and maintenance work. The THP can be isolated from the BHP and increased/decreased without affecting the BHP. This behavior of the THP adds additional noise to the data. The time delay of the pressure response is approximately 2 hours.

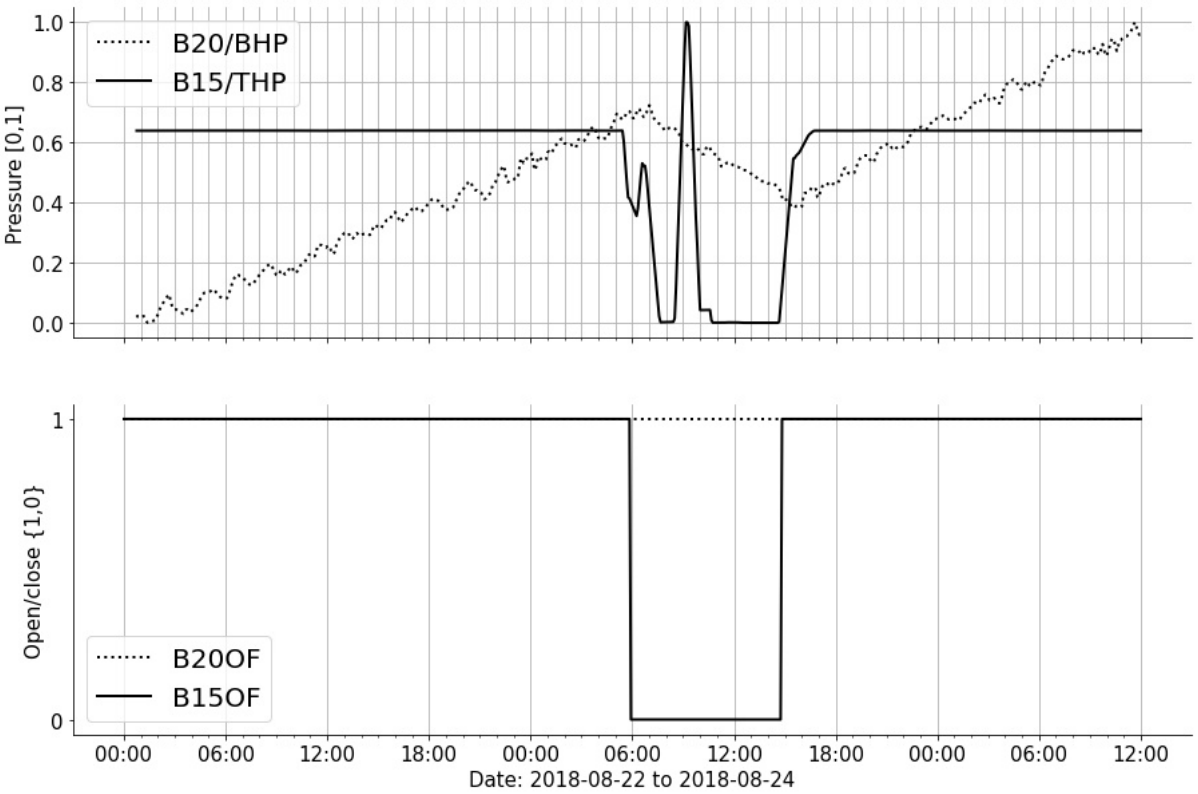


Figure 3.1: Zoomed in view of a pressure response between B15/THP and B20/BHP. Pressure is normalized on a [0,1]-range within this period.

In Figure 3.2, the pressure response between the stimulating injection well S02/BHP and the target production well A03/BHP is shown. There is a clear response seen in the responding well A03/BHP as the pressure in the stimulating well S02/BHP rapidly reduces. The time delay of the pressure response is approximately 4 hours. The reciprocal pressure response of A03/BHP as the stimulating well and S02/BHP as the responding well was not identified visually.

Additional short descriptions and figures of the remaining well-pair connectivities can be found in Appendix B.

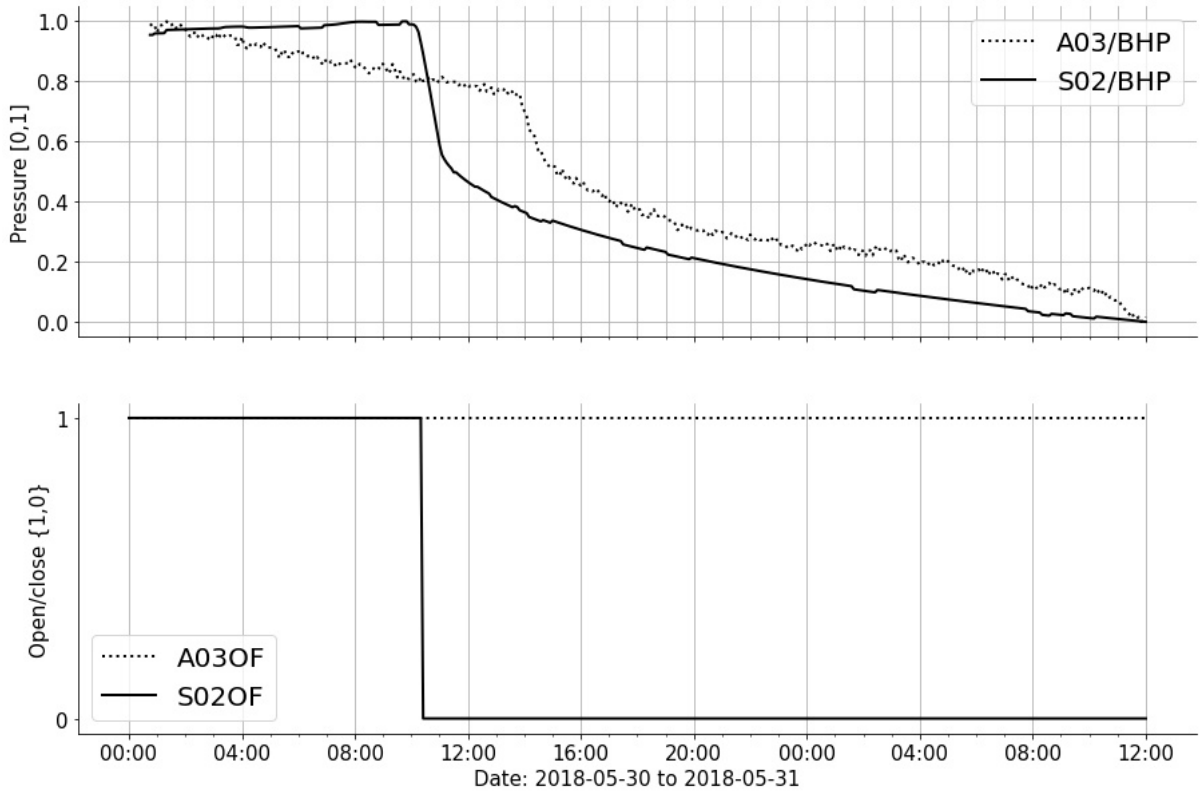


Figure 3.2: Zoomed in view of a pressure response between S-02 and A-03. Pressure is normalized on a [0,1]-range within this period. Data is smoothed with a moving average of 10 time steps.

3.3 External interferences

For a mature field such as the Eldfisk field, several external connectivities exist between the wells. These external interferences include the infrastructure and day-to-day operations, which interfere with the pressure readings.

Production wells are usually always open to a production separator with a set choke setting. Depending on the optimization of the wells, the choke setting may sometimes be changed from its standard setting. However, it is mostly kept as stable as possible. Stable production to a production separator suggests that the THP pressure, and to a lesser extent, the BHP, is dependent on the pressure in the production separator for each production installation. The production installations of Eldfisk Complex share a single production separator, and Eldfisk 2/7-Bravo has its own production separator. We consider this as external connectivity limited to each respective group of Eldfisk Complex and Eldfisk 2/7-Bravo.

The injection wells are, similarly to the production wells, kept as stable as possible. They are reliant on a shared water source, suggesting that the pressure of the injection wells depends on this shared water supply during these stable periods. We consider this as external connectivity spanning the whole field.

Both the injection and production wells may experience a simultaneous shut-in, which occurs when the emergency shut-down logic is triggered for the respective production installations. This external connectivity may affect a single production installation, or the whole field, depending on the severity of the underlying reason to why the shut-down logic was triggered.

The pressure profile of a single well may change significantly from time to time due to maintenance. During these periods, a well is shut-in, usually in periods of hours and sometimes days. We consider these periods as the most likely to identify internal connectivities, as few other wells are shut in simultaneously. However, they may be shut in consecutively from the first well as part of a larger maintenance campaign. Consecutive shut-ins may be a source of false signals, as it will appear as if these wells are affected by the previous wells' pressure change.

Chapter 4

Methodology

4.1 Data selection

For the type of data, we chose pressure data. We use pressure data because it is sampled in time frames of seconds and minutes instead of production rates, which are only sampled approximately once every month. Specifically, BHP where this is available and THP otherwise. While injection rates are sampled similarly to pressure data, we have five known connectivities between injector-producer well-pairs. To ensure that analysis can be performed on time frames of minutes and hours, we chose not to use injection rates coupled with production rates, as this would require analysis to be performed on monthly time frames. Previous work using injection/production rate data has solved this by either simply analysing connectivity on monthly time frames or having production rates on higher sampling rates. Some have used injection rates in combination with pressure data. We did not attempt to link injection rate data to pressure data. The system is considered open, and we do not attempt to create a complete model that can describe fully the changes a receiving well is experiencing.

There are several sources of external interferences which we attempt to filter out. This filtering is done by specifically targeting periods when the pressure profile of a well changes significantly. Due to this targeting of smaller periods, many smaller datasets are generated from one original dataset. To compensate, we aggregate the data for each well by identifying periods when a single well changes open/close configuration (well A). We obtain multiple smaller datasets for which we can apply the algorithms PCMCI+ and TCDF. Each of these multivariate time-series represents the stimulating well A acting on the receiving wells B. This splitting of the original dataset is a similar approach as the one employed by Mata (2010), in which they identified periods where one well experiences a rate change and then search a selected group of wells for a correlating peak. We differ in that we use the open/close signals of the wells to identify periods when a stimulating well is changing pressure profile. We do not identify groups, clusters of wells to examine. Instead, we allow all wells to be examined if they do not open or close during an identified period.

Generally, under the assumption that there can be connectivity between injector to injector and injector to producer well-pairs, one should obtain at least one well-pair with known connectivity and a well-pair with no known connectivity to verify that the models can identify connectivity and no connectivity. Additionally, depending on the extent

of external interferences, there should be two well pairs for each of these interferences for comparison. There is a field-wide water supply and an emergency shutdown logic per production installation as interfering variables for this thesis. Since we only select wells from a single field, the field-wide water supply will be difficult to identify as an interfering variable. The group-wide emergency shutdown logic may be easier to identify as an interfering variable, as there are wells with and without known connectivity from each oil platform available.

- Wells were selected to ensure that a mix of known connectivities and no known connectivities were added to the dataset. Additionally, it was ensured that there was a mix of wells from the oil platforms: 2/7-Alpha, 2/7-Bravo, and 2/7-Sierra.
- Two types of data from the wells were extracted; pressure and an on/off indicator. For the pressure data, we use BHP where this is available, THP otherwise. The on/off indicator is a binary time series indicating when a well is shut-in or open.

4.2 Data extraction

The data is measured by sensors located within each well. There are no uniform sampling intervals for all of the sensors. The raw pressure data had sampling intervals ranging from every few seconds to every few minutes. Most of the on/off indicators do not have any specific sampling interval, some are triggered by the opening/closing of the well, and a few have sampling intervals every few seconds. The time range was selected as three years from 2018 to 2020, and a data point interval of every five minutes was used. Five-minute intervals were selected because much of the maintenance performed on the wells are usually completed within a few hours. Thus, we found that this interval was a good balance between collecting enough information from most maintenance periods and adding too many data points.

Data was extracted from PI ProcessBook 2015 R2, Version 3.6.0.82, a software that enables real-time sensor data, such as pressure sensors, to be extracted and visualized. Extraction was done in two ways: The first was using Microsoft Excel, Version 16.0.13929.20206, and a PI ProcessBook add-in; this method was used on the pressure data and enabled a more convenient way of downloading and resampling the raw data. Resampling was done as a time-weighted average, illustrated in Figure 4.1. The other method was to download the raw data directly to a .txt file, and using a custom function in the programming language Python (Van Rossum and Drake 2009), converted it to a .csv file with the correct date/time format. We used the second method on the on/off indicators because we did not find a good method of forward-filling (the previous data point is copied forward in time until a new data point is met, applied in step 4.3 and illustrated in Figure 4.1) the missing data using the PI ProcessBook add-in for Excel. After extracting the needed data, we had one .csv file containing all of the pressure data and separate .csv files for each of the on/off indicators for each well.

- Pressure data obtained from PI ProcessBook by Excel with a PI ProcessBook add-in. On/off indicators obtained by downloading raw data from PI ProcessBook. All data stored as .csv files.

- The time range of the extracted data was set to 01.01.2018 00:00 until 31.12.2020 23:59.

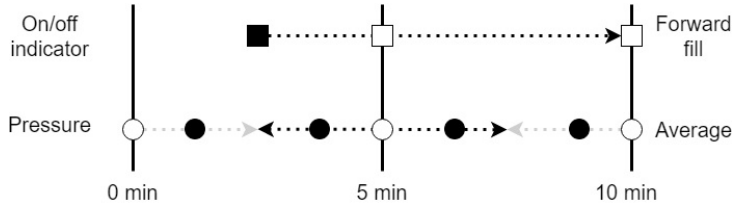


Figure 4.1: Illustration of the use of the forward-fill method on the on/off indicator data and time-weighted average on the pressure data. Filled square and circle represent existing data points for on/off indicator and pressure data, respectively. Empty squares and circles represent new data points obtained by forward-fill and time-weighted average.

4.3 Preprocess data

The .csv files are loaded into a Python script for further work. We use the Pandas package (McKinney 2010; Reback et al. 2021) for handling the data in Python. From the data extraction in step 4.2, a few data points end up as erroneous entries, i.e., an error message from the Excel PI ProcessBook add-in was posted for the erroneous cells. There are various reasons why some data points may end up erroneous, and one common reason is the sensor signal stopping for a short period. These erroneous data points were relatively few and could be identified as text instead of numbers in the given cells. We wrote a Python function to merge all of the .csv files from step 4.2 and remove these erroneous data points from the dataset. Rows containing erroneous data were removed by searching for any row containing non-numeric values.

Since the on/off indicator data has varying sampling intervals, there will be gaps in the data once merged on the same time range and interval as the pressure data. These gaps were forward filled using the Pandas function `df.ffill()`. The first data point of the on/off indicators does not usually start immediately after the selected period. As such, we are left with a short period without on/off data at the beginning of the dataset due to the forward-fill method. These rows of missing data were removed.

To identify all periods in the dataset when a change in a single well is occurring, we applied a filter (4.1) based on the on/off indicators, illustrated in Figure 4.2. The on/off indicator columns are binary where, 1 = *open* and 0 = *closed*. This filter is a boolean mask that creates a sliding window of two data points, calculates the mean, and returns `True` if this sliding window equals 0.5. The filter allows us to identify the dates when the binary columns switch between 0 and 1, i.e., we identify when a change in well configuration occurs.

Boolean filter: `rolling(2).mean() == 0.5`. If `True`, signifies that a well changed from open to close, or close to open. If `False`, no changes to the well occurred. (4.1)

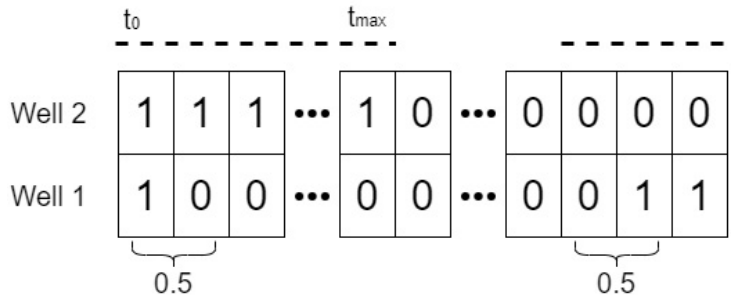


Figure 4.2: Illustration of the boolean mask applied to a dataframe. *Well 1* represent the stimulating well, *Well 2* the responding well. From the *Well 1* row, every time the well changes a open/close configuration the on/off indicator switches between 0 and 1. t_0 and t_{max} is the start and end time of the chosen time window where we allow *Well 1* to not have made any changes. After t_{max}, all data is dropped until a new change occurs. Collected data is represented by the dashed lines.

The wells are, in many cases, shut in simultaneously or consecutively. To ensure that the models do not interpret these shut-ins as interwell connectivity, we had to filter these interferences out. This filtering was done by identifying a time window before and after a change in the stimulating well and include these periods in the datasets, illustrated in Figure 4.2 as dashed lines from t_0 to t_{max}. If a target well experienced a change in configuration during these time windows, they were removed from the dataset. Early attempts were tested using a strict rule of removing the full dataset if "any change to any target well" occurred, but it was too limiting, and we could not identify enough periods for each stimulating well to obtain good results. Instead, we apply a less strict rule of removing only the target well experiencing changes. These configuration changes were identified similarly, as previously mentioned, by using a rolling mean on the data column. The target wells removed in this period will still be used as stimulating wells for the same period in a new dataset, illustrated in Figure 4.3. This solution solves two challenges of the data: (1) the mentioned simultaneous/consecutive shut-ins, and (2) a single response in a third well, *Well B*, should only be caused by either *Well A1* or *Well A2* (Figure 4.3). We argue that this solution works because, over time, the real causation of the response in *Well B* will accumulate multiple datasets with the causation well present. The false causation of the response will only be present during the specific configuration of both *Well A1* and *Well A2* changing configuration.

From the above, we had to decide on the length of the time window before and after the stimulating well changes configuration. The time window size will directly affect how many wells are included in a given dataset for a stimulating well. As the window size increases, fewer target wells will be included as some of them will experience a change in configuration during this time window. When the window size decreases, more wells will be included; however, there will be a shorter silent period before and after a change has occurred in the stimulating well, which will become a source of additional noise in the dataset.

To determine the optimal window size, we iterated over multiple sizes and noted how many data points a few chosen causation/target well-pairs could obtain. Figure 4.4 show the reduction in data points as the window size increases. To keep the period before a

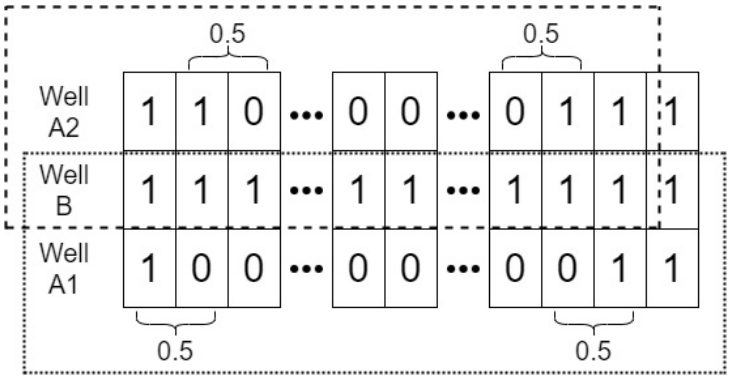


Figure 4.3: Illustration of the well selection for each dataframe. *Well A1* represent the stimulating well acting the responding *Well B*. *Well A2* represent the second well experiencing changes during the identified period, and consequently, acts as a stimulating well. The dotted window is the resulting dataset after removal of *Well A2*. The dashed window is the resulting dataset when *Well A2* is the stimulating well acting on the responding *Well B* with *Well A1* removed.

change occurred as silent as possible from other interfering wells, we balanced the length of the time window with the number of samples we were able to return for all target wells and the general performance of the model. In Figure 4.5, we can see the attention scores of the TCDF-algorithm change from a homogeneous setup to a more heterogeneous one as the time window increases. This change is shown for the specific case of the stimulating well B20/BHP acting on a selection of target wells. For the time window after a change, we set this to be the same as the length before the change; additionally, if two changes in the stimulating well occur within a short time interval, ex. When a well is closed and opened within a short period. We ignore the second change (the opening), as this will already be contained within the time window of the first change (the closing), albeit with a shorter time window.

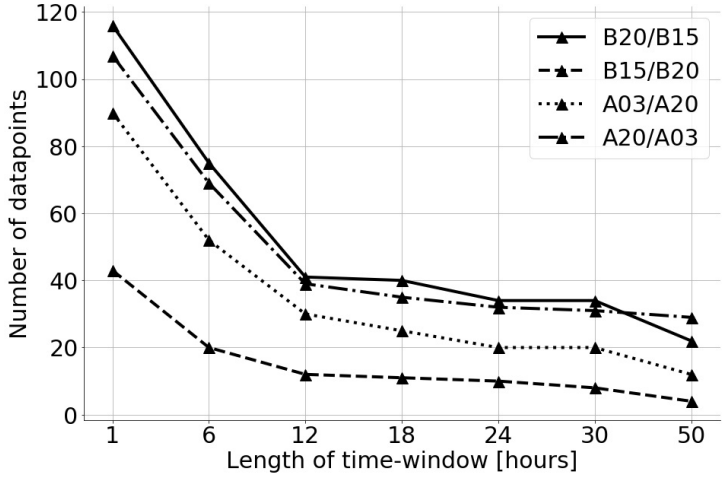
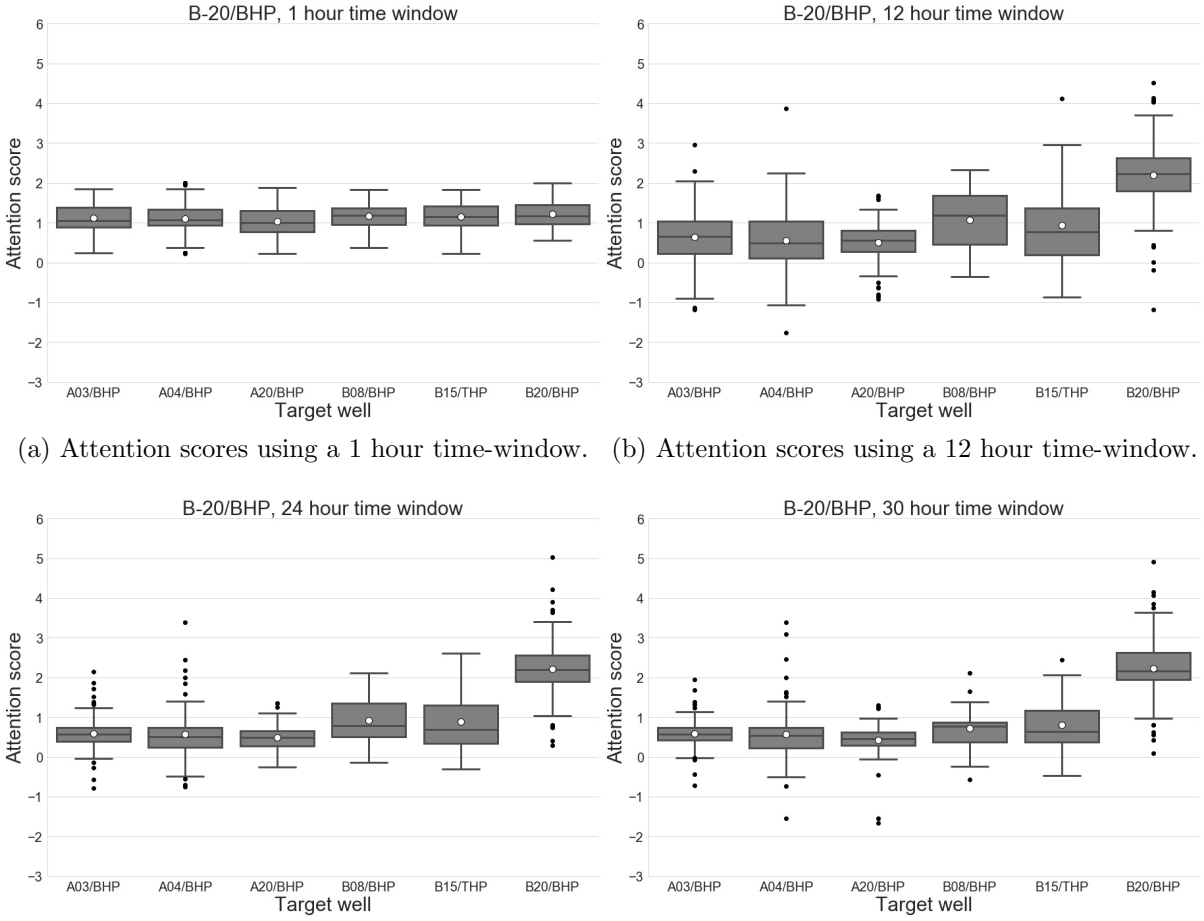


Figure 4.4: Plot of the number of datapoints obtained for a few causation/target well-pairs, as the time window length is increased. The x-axis show the length of the time window in one direction, a full dataset contains a time window of similar length in both directions.



(a) Attention scores using a 1 hour time-window.

(b) Attention scores using a 12 hour time-window.

(c) Attention scores using a 24 hour time-window.

(d) Attention scores using a 30 hour time-window.

Figure 4.5: Figures a) to d) show the overall performance of the TCDF-algorithm for different time-window lengths prior and subsequent to a change in the stimulating well. In this case, B-20 was used as the stimulating well. The attention scores move from a more homogeneous setting in a) to a more heterogeneous setting in b) and c), by d) the setting returns to a homogeneous setting. White dots represent the mean, and black line inside box is the median, or 50% quantile. Upper and lower bound of the boxes is the 75% and 25% quantile (Q1 and Q3), respectively. The whiskers are calculated as: $Q1 - IQR \cdot 1.5$ and $Q3 + IQR \cdot 1.5$, where IQR is the Inter Quantile Range, or $Q3 - Q1$. Black dots represent outliers in the data.

Additional noise reduction in the datasets was attempted by applying a smoothing function (rolling mean) on the data. We found that especially the THP could have much variance. We attempted different window sizes for the rolling mean function, from one (no smoothing) up to 100 data points. We found a slight reduction in the variance of the attention scores when the size of the rolling window was set to ten data points. There was no, or minimal, reduction of variance in the attention scores when the window size increased beyond ten data points. Additionally, the first datapoints of the dataset would be removed until a full window was filled. Therefore, we aimed to keep the rolling window as small as possible to avoid unnecessary reductions of the already small datasets. The rolling window size was set to ten data points.

The last part of the preprocessing was to normalize the data. Normalization is technically unnecessary for the PCMCI+ model as it already has a built-in normalization procedure. However, the TCDF model employs CNNs, requiring normalization of the data to perform optimally (Sola and Sevilla 1997; Jayalakshmi and Santhakumaran 2011), especially considering that there is a wide range between the pressure readings when using both BHP and THP. To keep all our data similar for both models, we used the `MinMaxScaler()` function from the scikit-learn library (Pedregosa et al. 2011) in Python to normalize the data to a $[0, 1]$ range.

- Erroneous data is removed by searching for non-numeric cells in the dataframe.
- Data is resampled to five-minute intervals. Forward-fill on the on/off indicators was used to fill the data gaps.
- A boolean mask is applied as a filter to identify any period when a well changes open/close configuration. If a well does not change open/close configuration after a given time window, rows are removed until a new change occurs.
- Target wells experiencing changes in the same period as the stimulating well is removed from the dataset.
- A moving average of ten data points is used and the data is normalized to a $[0, 1]$ range.
- Individual .csv files were created for each well, and each period the well changed open/close configuration.

4.4 Analyse data

Since we do not have any quantifiable measures for the interwell connectivities in our oil and gas field, we decided to run three different tests. These were designed to give us a view of the models' performance and validate the generality of the models. The three tests: small-scale baseline, small-scale modified, and full-scale modified.

The small-scale baseline test is used to get a rough idea of how the models perform against a single, large dataset containing all of the information from the period. The assumption is that the results will reflect a high degree of external interferences, which will produce false connectivities. Though, the models may be able to identify well-pairs that do not have connectivities, thus reducing the search-space for manual identification of connectivities.

The small-scale modified test is used to train and optimize the models against a few known connectivities from the field. The idea is to verify that the models can identify known connectivities by calibrating the hyperparameters and parameters of the TCDF and PCMCI+ algorithms. The same hyperparameters and parameters obtained in this small-scale test will be used in the full-scale test.

The full-scale modified test is used to verify whether the models can identify known connectivities that the models have not seen before. Additionally, more wells in the data suggest that there will be more noise, and a less clear separation between *no connectivity* and *connectivity* is expected.

4.4.1 Small-scale baseline

This test was used to view how the models performed with minimal modifications to the dataset. The dataset for the small-scale baseline test was set up as seen in Table 4.1. The A-wells (A-03, A-04, and A-20) are situated far away from the B-wells (B-08, B-15, and B-20). There is no known connectivity between the A -and B-wells. There are known connectivities between the stimulating wells B-08 and B-20 acting on the target well B-15. The B-wells were also specifically chosen as these are three injection wells from the same installation, meaning that they have a strong connection through external sources, and as a consequence, it will be difficult for the models to separate these wells.

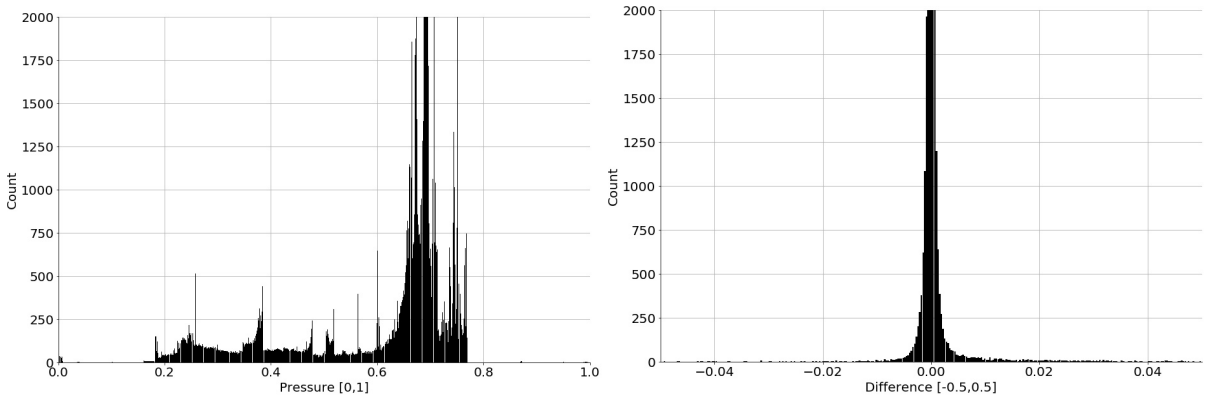
The modifications applied to this dataset were to normalize the sampling intervals for the data points to five-minute intervals and standardize the range of the pressure data to a $[0, 1]$-range.

Table 4.1: Setup of dataset for the baseline test.

Time range	**Sampling interval**	**Data sets**	**Length (rows)**	**Wells**	**Known connectivities**
01.01.18 00:00 - 31.12.20 23:59	5 min	1	315 528	A-03, A-04, A-20 B-08, B-15, B-20	B-08, B-15, B-20

PCMCI+

PCMCI+ assumes stationarity in the data, and this was tested by the Augmented Dickey-Fuller (ADF) test (Dickey and Fuller 1979). This test returned a p-value of 1.0E-7 with a test statistic of -10.3, which suggests that the data is stationary (Cheung and Lai 1995). However, a comparison of the histogram with and without differencing (1.3) the data can be seen in Figure 4.6. These histograms show that the unmodified data reaches a hard stop as it approaches 0.8, with a few outliers. Additionally, the data looks to be spreading towards a (normalized) pressure of 0.2. We attempt to transform the data such that it approximates more to a normal distribution. This transformation was done by differencing the data (1.3), which calculates the difference between each time step and uses these values as the data instead (Nason 2006). After differencing the data once, the ADF-test returned a p-value of 1.0E-7 with a test statistic of -67.4. A lower test statistic indicates a higher degree of stationarity. More importantly, the data has a more even spread over the mean value. For that reason, the data was modified by differencing, even if the ADF-test indicates the unmodified data is already stationary.



(a) Unmodified data of B20/BHP. *bins*=2000. n=315 528.

(b) Modified data of B20/BHP by differencing. *bins*=2000. n=315 527.

Figure 4.6: Figures (a) and (b) illustrates the difference between the unmodified and modified data of B20/BHP. The *Count*-range has been cut short for a better visual representation.

We applied the function, `run_bivci()`, provided by the Tigramite package (Runge 2020) for Python to search for the maximum time delay, τ_{max}. The range $[0, 100]$ was searched, and the correlation was found to peak very early, after one to two time steps. Since no indications of lagged correlation were found beyond one to two time steps, a maximum time delay of two time steps would be sufficient. However, increasing the maximum time delay is purely a computational cost; the maximum time delay was set to 100 time steps. This maximum time delay was selected to validate that the model encompassed the time lag of all known connectivities. General parameters controlling the direction of causality was set to the default values, except for *Reset lagged links* which was set to `True`, as seen in Table 4.2. This parameter may improve the detection of lagged causality links by reconsidering all lagged links in the MCI-phase. The only cost of applying this parameter is that it requires additional computational power.

Table 4.2: Tuning parameters for the PCMCI+ algorithm used for the small-scale baseline test.

CI test	Min/max time delay	Significance threshold	Contemp. collider rule	Conflict resolution	Reset lagged links
ParCorr	0/100	0.01	Majority	True	True

TCDF

The TCDF algorithm does not perform a test on unseen data when training the model. Instead, it performs a test on the permuted time series when validating any potential cause. To enable any verification that tuning performed on the hyperparameters was improving or degrading the model, a function, `test()`, was added to the model. This function tests the trained model on unseen data and returns the MSE. Additionally, ten-fold cross-validation was performed (Kohavi 1995), with a train/test-split of 60% training data and 40% test data for each fold. The hyperparameters were tuned by grid search, and we found the optimal setting to be as listed in Table 4.3. Interestingly, the *kernel size* parameter, along with its accompanying *dilation coefficient* parameter, was found to provide no benefit beyond the first couple of time steps, suggesting that this model strongly favors the first few time steps in its inference. The *significance* parameter was adjusted up to 1.0 such that it no longer affects the outcome of the validation of the causes.

Table 4.3: Hyperparameters for the TCDF-algorithm used for the small-scale baseline test. These were the defaults set by Nauta et al. (2019).

Dilation coefficient	Epochs	Hidden layers	Kernel size	Learning rate	Optimizer	Significance
10	800	0	10	0.1	Adam	1.0

4.4.2 Small-scale modified

For this test, we apply modifications to the dataset as described in step 4.3. These modifications include applying a boolean filter to identify periods when a stimulating well changes open/close configuration. Any target well that experience changes (open/close) during these identified periods are removed from that specific dataset. The length of the time window before and after a change in open/close configuration for the target well generally performed optimally in the 18 to 24-hour range. We selected a 24-hour time window before and after an identified change in well configuration. The same wells as in the small-scale baseline test in step 4.4.1 are used to identify connectivity within a group of wells. Since we are applying a boolean filter on every single well, the resulting data will be split into individual datasets for each stimulating well. The setup of the datasets can be seen in Table 4.4.

PCMCI+ and TCDF are applied to each multivariate time-series, and only the cause-link results of the stimulating well acting on target wells are extracted from each model.

We wrote a Python script to automate this process. This method of extracting cause and effect is somewhat different from Runge et al. (2019), Runge (2020), and Nauta et al. (2019), which applied their respective models on the full system, i.e., extracting cause and effect from all variables simultaneously. We tested this method in the small-scale baseline test in step 4.4.1. However, we were aware that the overall complexity and connectivity through external sources would be a major source of false signals and noise. Without controlling for this interference acting on the system, we did not find any meaningful way of extracting information and identifying true connectivity from the model.

Table 4.4: Setup of datasets for the small-scale modified test. The data is split into multiple smaller datasets.

Time range	**Sampling interval**	**Data sets**	**Length (rows)**	**Wells**	**Known connectivities**
01.01.18 00:00 - 31.12.20 23:59	5 min	430	567	A-03, A-04, A-20 B-08, B-15, B-20	B-08, B-15, B-20

PCMCI+

The PCMCI+ algorithm, as mentioned in the small-scale baseline test, assumes stationary data. In the small-scale baseline test, it was found that the unmodified data is stationary by the ADF-test. However, for this test, the selection of periods when the wells experience large changes causes most of the generated datasets to contain non-stationary data. A second-order difference was found to generate stationarity in the data. This transformation is shown by the example in Table 4.5, of the well B15/THP for a specific period in the data, and a first-order difference is not enough to transform the non-stationary data to stationary. Though higher-order differences can further improve the ADF-statistic, we selected the lowest order necessary to obtain a p-value less than 0.01.

Parameters of the PCMCI+ algorithm were kept similar to those in the baseline test, shown again in Table 4.6. If the p-value generated by the PCMCI+ algorithm was higher than the chosen significance threshold, the corresponding partial correlation measure was removed. For each generated range of partial correlations, the mean value was taken.

Table 4.5: ADF-statistic and p-values of different orders of differencing-transformations. Obtained from the period of 2020-07-12 to 2020-07-14 of the well B15/THP.

	Unmodified	**1st-order**	**2nd-order**	**3rd-order**
ADF-statistic	1.089	-2.628	-7.498	-9.843
p-value	0.995	0.087	1.0E-7	1.0E-7

Table 4.6: Tuning paremeters for the PCMCI+ algorithm used for the small-scale modified test.

CI test	**Min/max time delay**	**Significance threshold**	**Contemp. collider rule**	**Conflict resolution**	**Reset lagged links**
ParCorr	0/100	0.01	Majority	True	True

TCDF

Tuning the TCDF-algorithm was performed slightly differently for this test than the baseline test because, in this test, the main dataset has been split into multiple, smaller datasets. With few data points to train on and each dataset representing a very small portion of the full period, we took an approach of training on a random selection of the datasets. First, the training was performed by selecting a random 50% of the datasets as training sets and using the previously mentioned 60/40 train/test split on each dataset. Hyperparameter tuning was performed by setting an MSE threshold of 0.1 and checking how many of the datasets the model could obtain an MSE less than 0.1. The idea of the MSE threshold was to ensure that the model generalized enough to train and learn on most of the datasets while accepting that some datasets would not provide enough information for the model to learn. We did not remove any of the datasets that had a higher MSE than the selected threshold, the reason being that we only wanted to use this as an indicator and avoid adding a layer of subjectivity to the data selection.

With the hyperparameters listed in Table 4.7, we were able to obtain the highest number of accepted trained models with corresponding datasets. We note that the hyperparameter for the time steps (Dilation coefficient and Kernel size) had a weak to no effect on the overall performance of the models. We found that adding hidden layers to the CNNs quickly degraded the performance of the models, suggesting that a simpler model is sufficient for inference of connectivity with our datasets. Learning rates were tested in the range $[0.1, 0.001]$. Higher learning rates, up to 0.1, were observed to cause oscillations in the training error. Lower learning rates, down to 0.001, were effective. However, the model obtained similar stable training loss using a higher learning rate of 0.05. The number of training epochs was reduced from the default value of 1000 epochs to 400 epochs. The Significance parameter was set to 1.0.

Table 4.7: Hyperparameters, including MSE threshold, for the TCDF-algorithm used for the small-scale modified test. *Test MSE* is the value of the MSE threshold used to accept/reject a trained model with its corresponding dataset.

Dilation coefficient	**Epochs**	**Hidden layers**	**Kernel size**	**Learning rate**	**Optimizer**	**Significance**	**Test MSE**
100	400	0	100	0.05	Adam	1.0	<0.1

4.4.3 Full-scale modified

For the full-scale test, we apply the models on all wells with known connectivities within the same field, intending to test the generality of the models and identify their limitations. The wells included in this test introduces a higher complexity in the form of; injection/producer well-pairs, a group of S-wells that have connectivities with both the A-wells and B-wells and more wells with THP data. It is expected that the added complexity will cause the models to become less accurate with the inference of connectivity. The methodology is otherwise the same as the one employed in the small-scale test, and the hyperparameters of the TCDF-algorithm will not be changed.

4.5 Classification

PCMCI+

The PCMCI+ algorithm provides a form of classification by selecting cause-links based on a significance threshold, α, given a p-value obtained from hypothesis testing on the partial correlations. The Student's t-test is used for this hypothesis test. Classification by the Student's t-test works by itself during the baseline test, as there is only one value obtained for the full dataset. However, every single dataset will generate a partial correlation measure during the small-scale and full-scale tests, which enables a second hypothesis test. The second hypothesis is formed by defining *connectivity*, as all partial correlation measures greater than or equal to some threshold, and *no connectivity* otherwise. The threshold is identified as a positive value, defined by a clear separation in the distribution of all generated partial correlations. This threshold selection is made by visual inspection. The Wilcoxon signed-rank test is used to calculate the test statistic and accompanying p-value for the hypothesis test. This statistical test is used as there is no assumption of normality in the data, and it can work on few data points.

TCDF

The same principle as mentioned for the PCMCI+ algorithm is applied to the results from the TCDF-algorithm. However, a threshold of the attention score is instead used. The threshold is, in this case, set to 1, which is the value the attention scores are initialized.

4.6 Strength of connectivity

PCMCI+

The PCMCI+ algorithm already provided an output of the estimated strength and confidence of the inferred causality with the partial correlation and p-values, respectively. As such, we applied the partial correlation measure directly as an indication of the strength of connectivity.

TCDF

The TCDF-algorithm does not provide any direct means of quantifying the strength of connectivity. The natural choice for quantifying this measure became the attention scores; however, these are not bounded. Instead, we used the p-values, which are derived from the attention scores. The p-values are bounded in the [0,1]-range, which allows us to measure and visualize the strength on a relational scale within a bounded range. From the tests done on the datasets, it was noted that a strong link was given a fairly low number, in the range of 0.1 and lower. A base 10 log-transform was applied to the p-values in the DAG illustrations. Visualizations of the strength were made by mapping the log-transformed p-values linearly from zero to the significance threshold selected for the data. Unmodified p-values are presented in tables.

4.7 Overview of workflow

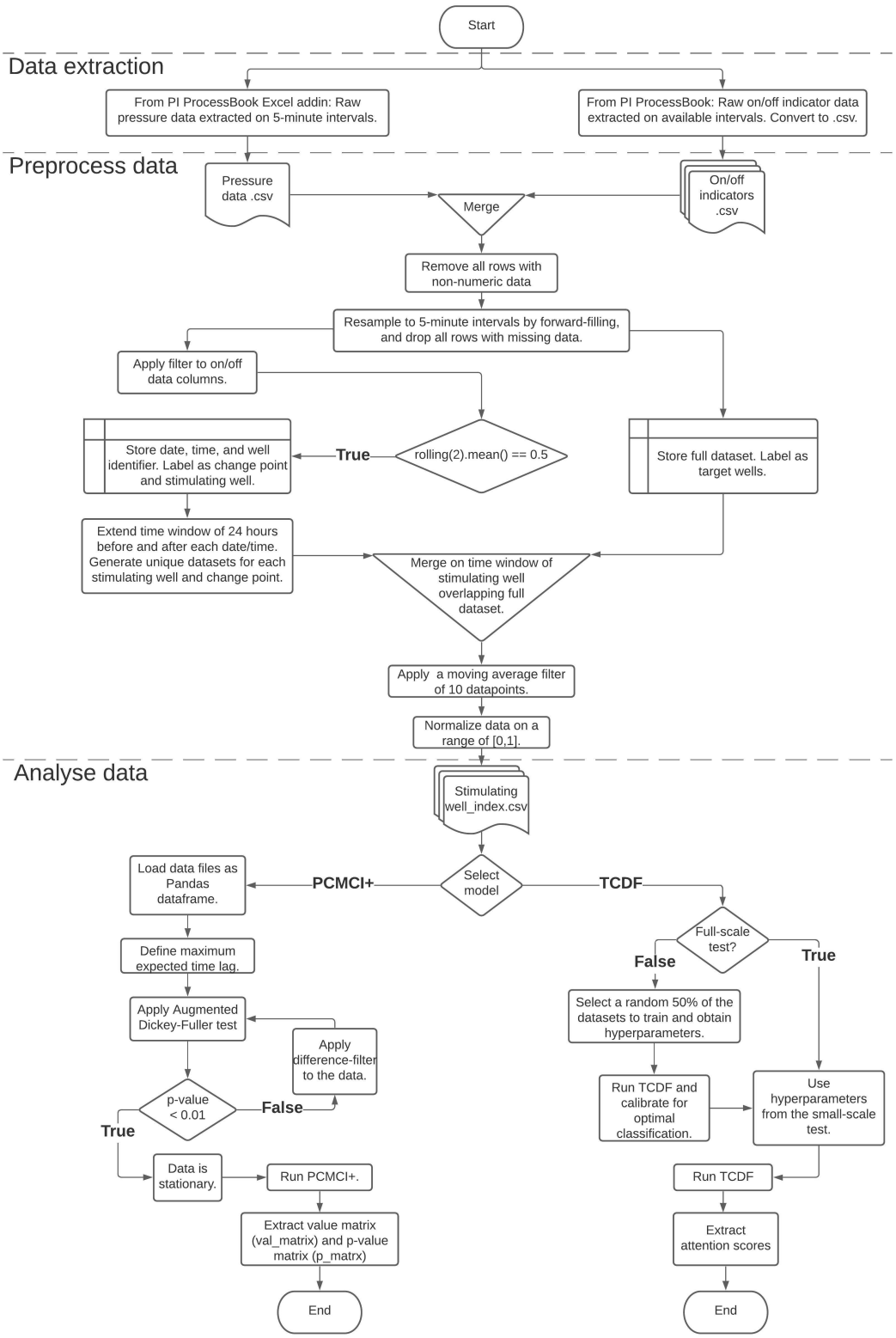


Figure 4.7: Workflow of the Data extraction, Preprocess data, and Analyse data steps in the methodology.

Chapter 5

Results

5.1 Small-scale baseline

PCMCI+

The report produced by the PCMCI+ algorithm lists almost 30 links for each of the six wells. Each link represents the estimated partial correlation between two given wells, for each time lag from τ_{min} to τ_{max}. Since the model returns a unique partial correlation measure for each time lag, the number of possible links for each single well is given by $L_{max} = N \cdot (\tau_{max} - \tau_{min}) - 1$. Thus, the maximum number of links for a single well is, in this case, 599 links. These links represent the predicted connectivity between the respective wells.

We have listed some of the identified links, or connectivities, in Table 5.1. In this table, we selected only the strongest time-delay causation-link from each well pair. Consequently, only time-delays of zero and one time steps appear in the causation links. The strength of the links, measured by partial correlation, r, ranges from 0.709 for the strongest link to -0.036 for the weakest link. It is interesting to note that all causal links are either on time delay 0 or 1. We can also observe that the self-causation links with a time delay of 1 are prevalent as the strongest link for all wells, except for the stimulating well B20/BHP acting on the target well B08/BHP. The p-value, which measures the statistical significance of each link, was not reported in Table 5.1 because it is controlled by the significance threshold parameter and is less than 0.01 for all links.

Table 5.1: Results obtained by applying the PCMCI+ algorithm applying minimal modifications to the dataset, and extracting the strongest cause-link for each well-pair. *Stimulating well* represent the stimulating well acting on the *Target well*. *Time delay* is the identified time delay for each cause-link. *Partial correlation*, show the relative strength of the cause-links.

Stimulating well: A03/BHP			**Stimulating well: A04/BHP**		
Target well	**Partial correlation**	**Time delay**	**Target well**	**Partial correlation**	**Time delay**
A03/BHP	0.299	1	A03/BHP	-0.036	1
A04/BHP	-0.048	1	A04/BHP	0.473	1
A20/BHP	0.154	0	A20/BHP	-0.305	0
B08/BHP	-0.186	0	B08/BHP	0.139	0
B15/THP	-0.092	0	B15/THP	0.274	0
B20/BHP	-0.125	0	B20/BHP	0.196	0
Stimulating well: A20/BHP			**Stimulating well: B08/BHP**		
Target well	**Partial correlation**	**Time delay**	**Target well**	**Partial correlation**	**Time delay**
A03/BHP	0.154	0	A03/BHP	-0.186	0
A04/BHP	-0.305	0	A04/BHP	0.139	0
A20/BHP	0.341	1	A20/BHP	-0.225	0
B08/BHP	-0.225	0	B08/BHP	0.455	1
B15/THP	-0.191	0	B15/THP	0.34	0
B20/BHP	-0.179	0	B20/BHP	0.709	0
Stimulating well: B15/THP			**Stimulating well: B20/BHP**		
Target well	**Partial correlation**	**Time delay**	**Target well**	**Partial correlation**	**Time delay**
A03/BHP	-0.092	0	A03/BHP	-0.125	0
A04/BHP	0.274	0	A04/BHP	0.196	0
A20/BHP	-0.191	0	A20/BHP	-0.179	0
B08/BHP	0.34	0	B08/BHP	0.709	0
B15/THP	0.57	1	B15/THP	0.342	0
B20/BHP	0.342	0	B20/BHP	0.417	1

TCDF

From the TCDF-algorithm, a report is produced, which includes the training loss, potential causes, validated causes, and the time delay for the validated causes. Additionally, the test loss and attention score were extracted from the TCDF-algorithm. From Table 5.2, it is shown that the algorithm identified a validated cause (connectivity) for every target well; these are all previous time-steps of the target itself. There were three wells with a second validated cause. Training and test loss are shown to be relatively close to each other, suggesting that the training did not overfit too much. Additionally, the training and test loss is comparative for each of the six wells.

Table 5.2: Results obtained from analysis using TCDF-algorithm with default hyperparameters and minimal modifications to the dataset. *Target*, indicates the responding well. *Potential causes* and *Validated causes*, indicates stimulating wells acting on the *Target*. *Train loss* and *Test loss* is calculated as the MSE. *Time delay* is measured in the given sampling interval of the dataset, i.e., a time delay of 2 is equivalent to 10 minutes. *Attention score* is the coefficient-type value produced by the CNNs.

Target	**Train loss**	**Test loss**	**Potential causes**	**Validated causes**	**Time delay**	**Attention score**
A03/BHP	8.0E-06	1.9E-05	A03/BHP	A03/BHP	1	2.87
A04/BHP	1.6E-05	3.2E-05	A04/BHP	A04/BHP	1	3.60
A20/BHP	8.0E-06	8.0E-06	A20/BHP	A20/BHP	1	2.77
			B08/BHP	B08/BHP	0	1.06
B08/BHP	6.0E-06	8.0E-06	B08/BHP	B08/BHP	1	4.83
			B15/THP	B15/THP	0	1.00
B15/THP	1.1E-05	1.5E-05	B15/THP	B15/THP	1	3.45
B20/BHP	1.1E-05	1.2E-05	B20/BHP	B20/BHP	1	4.82
			B15/THP	B15/THP	0	1.01

5.2 Small-scale modified

The results for the small-scale modified test are obtained from a total of 430 datasets, each containing 567 rows of pressure data. In total, there were 243.810 rows of pressure data analysed. Each row of each dataset contains a varying number of wells (variables), from a maximum of six wells to a minimum of two wells. The number of wells in each dataset depends on how many target wells did not experience changes in their on/off-indicator during the period. It is noted that the number of data points for the two models, PCMCI+ and TCDF, are not equal. They vary from a high of 89 data points from the well B20/BHP acting on the wells A03/BHP and A20/BHP to a low of 5 data points from the well B15/THP acting on the well B20/BHP. The number of data points for self-causation represents how many datasets were obtained and used for each stimulating well to generate the attention scores.

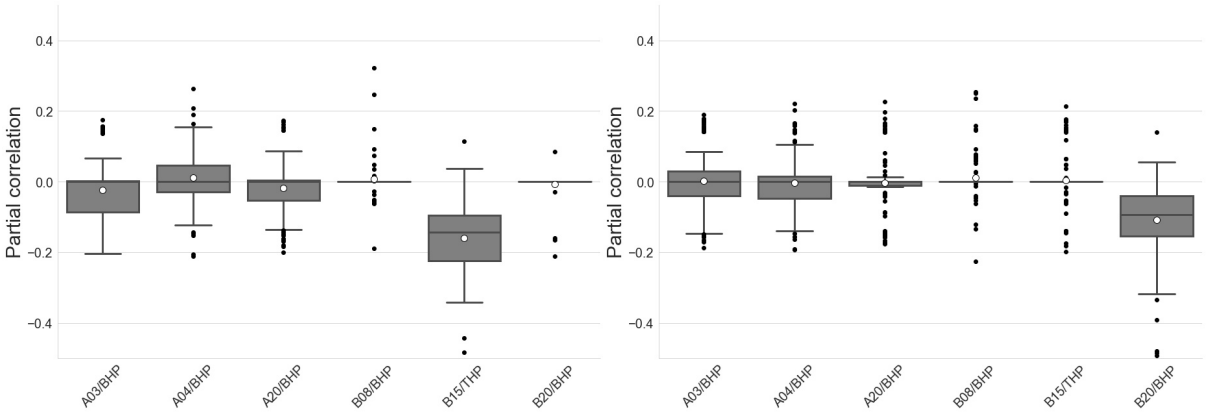
The results from the PCMCI+ algorithm are listed in Tables A.1-A.6, and the TCDF-algorithm in Tables A.8-A.13 found in Appendix A. We refer to numbers from these tables in this text, however, for readability of the thesis the tables were put in the appendix.

PCMCI+

The mean and median of self-causation are shown to be relatively similar to the rest of the well-pairs. Interestingly, all self-causation have negative mean and median partial correlations. The mean ranges from -0.159 to -0.010, and the median ranges from -0.143 to -0.012. The mean and median of the stimulating wells acting on the other target wells range from -0.096 to 0.097 and -0.160 to 0.146, respectively. The well-pair B15/THP acting on B20/BHP have the highest negative partial correlation, with a mean of -0.096 and median of -0.160; this is a well-pair with known connectivity. However, this well-pair only has five data points, which is a low number representing a distribution. The reciprocal link, B20/BHP acting on B15/THP, has 33 data points, and a weak positive partial correlation is seen, with a mean of 0.014 and a median of 0.008. The second well-pair with known connectivity, B08/BHP and B15/THP, has a weak positive partial correlation for both links.

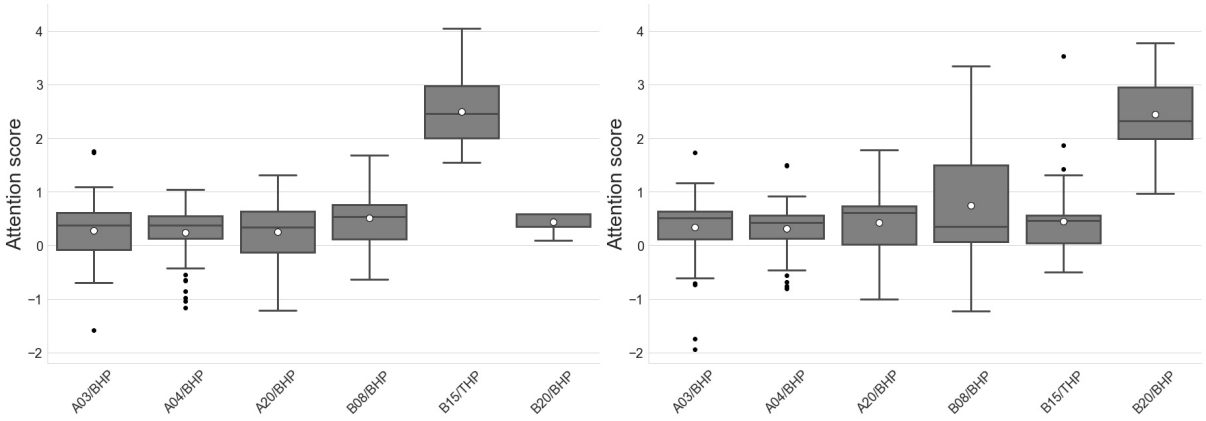
The standard deviation ranges from a low of 0.093 to a high of 0.119 (excluding self-causation). Self-causation has a comparatively similar standard deviation from 0.045 to 0.111. These are relatively high values considering that the mean and median are close to zero for most wells.

Visual inspection of the boxplots in Figure 5.1 show that there is some separation to be seen between the boxplots, though it is difficult to pinpoint any specific well-pair with good separation. There are many outliers seen in all boxplots, and in many cases, the boxplots did not generate properly. However, the boxplots corresponding to self-causation are shown to have few outliers and are separated from the target wells.



(a) Stimulating well: B15/THP (b) Stimulating well: B20/BHP

Figure 5.1: Figures a) and b) show the generated partial correlations of the small-scale test for the wells B15/THP and B20/BHP, visualized as boxplots.



(a) Stimulating well: B15/THP (b) Stimulating well: B20/BHP

Figure 5.2: Figures a) and b) show the generated attention scores of the stimulating wells B15/THP and B20/BHP for the small-scale test, visualized as boxplots.

TCDF

The mean and median are shown to be consistently higher for self causation, compared to the stimulating well acting on the other target wells (Figure 5.2). The mean of self causation is in the range 2.447 to 2.873, and the median is within a range of 2.320 to 2.973. The mean of the stimulating well acting on other targets has a wider range of -0.177 to 0.744. We note that the lowest reported mean of -0.177 belongs to the stimulating well B08/BHP acting on B15/THP, which has known connectivity.

From visual inspection of the generated attention scores, it is observed that there is no clear separation between the attention scores, which could easily be interpreted as *connectivity* or *no connectivity*. This observation excludes self causation, which shows a clear separation from the target wells. The stimulating well B15/THP acting on B20/BHP is shown to be missing the upper quantile whisker; this is a consequence of the very low sample size of 5 data points for this well pair. It is noted that there are few outliers shown in the data. However, there a several outliers seen in the link between the stimulating well B15/THP acting on the target well A04/BHP. The attention scores of the stimulating well B20/BHP acting on the target well B08/BHP is shown to stretch significantly more than other well-pairs. This well-pair do not have a known connectivity. However, this well-pair share the same production installation and the same water supply.

5.3 Full-scale modified

There were 879 datasets obtained for the full-scale test, each with 567 rows of data for a total of 498.393 rows of pressure data. The number of datasets for each stimulating well was in the range of 36 for S18/BHP to 111 for B20/BHP. The well pair with the fewest number of data points were again B15/THP acting on B20/BHP.

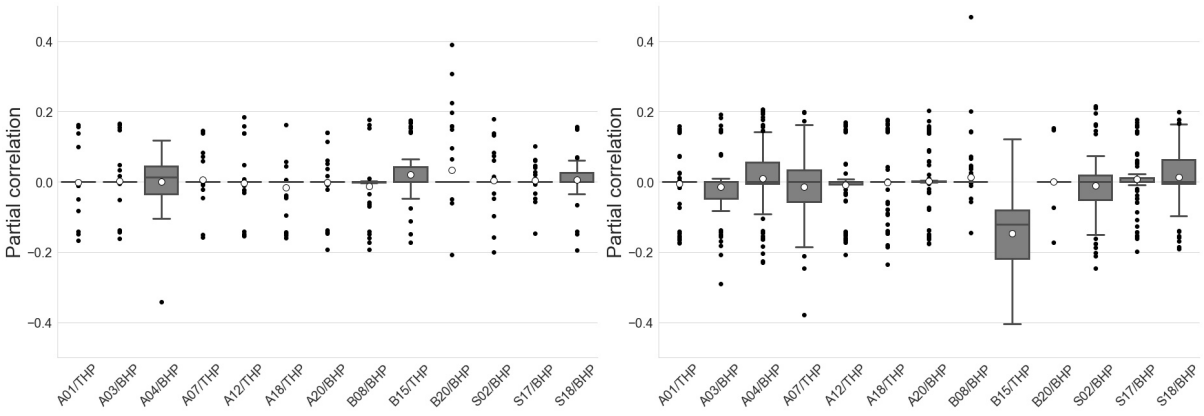
The results from the PCMCI+ algorithm are listed in Tables A.15-A.27, and the TCDF-algorithm in Tables A.30-A.42 found in Appendix A. We refer to numbers from these tables in this text, however, for readability of the thesis the tables were put in the appendix.

PCMCI+

Similar to the small-scale modified test, the generated partial correlations' mean value is relatively close to zero. For self-causation, the mean ranges from -0.147 to 0.001, and the standard deviation ranges from 0.050 to 0.128. Except for one well, A04/BHP, all wells had a negative mean partial correlation for self causation. For stimulating wells acting on target wells, the mean ranges from -0.077 to 0.119, and the standard deviation ranges from 0.062 to 0.167. The range of means for stimulating wells acting on target wells is similar in magnitude to the range of means for self causation. However, it shifted slightly upwards to be more symmetrical over zero. The standard deviation range for stimulating wells acting on target wells has increased in magnitude compared to standard deviations for self causation.

In Figure 5.3, boxplots have been generated for the partial correlations of the stimulating wells A04/BHP and B15/THP. When comparing the minimum and maximum of the generated partial correlations in Figure 5.3, it is seen that there is a very tight range from -0.200 to 0.200, with a few outliers closer to -0.400 and 0.400. For the stimulating well A04/BHP in Figure 5.3a, there is one target well, B20/BHP, which separates itself from the other target wells in the figure. Here, B20/BHP is shown to have a range of partial correlations, which are shifted higher into the positive range, with only a few data points with negative partial correlations. This well pair, A04/BHP acting on B20/BHP, which do not have known connectivity, also happen to be the well pair with the highest mean of partial correlations. For the stimulating well B15/THP, there is a less pronounced separation to be seen for any target well. In Figure 5.3b, the target well B08/BHP can be seen to have a slight shift of the minimum and maximum range higher into the positive range, as compared to the other target wells. This well pair, B15/THP acting on B08/BHP, has known connectivity.

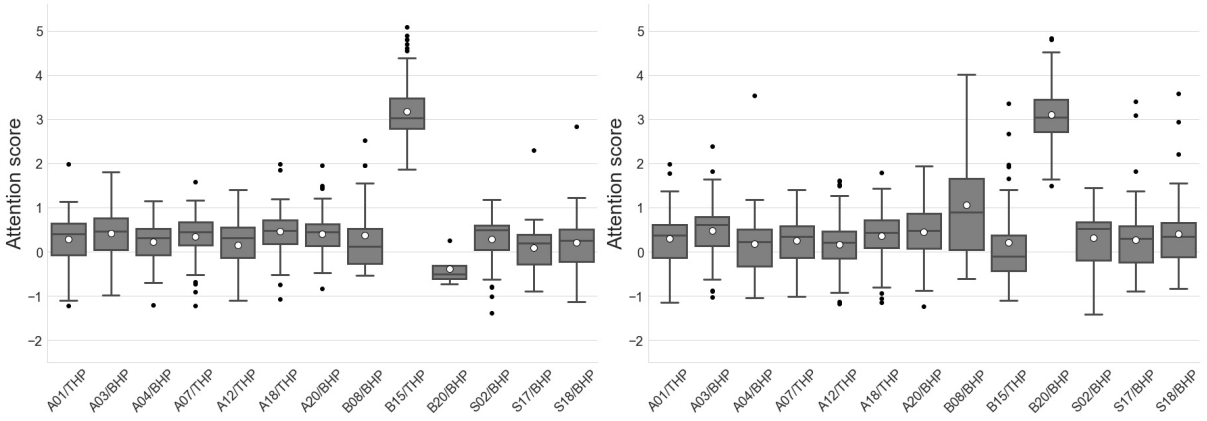
From Figure 5.3, it is also shown that the boxplots consist mostly of outliers, and the body of the boxplots does not form in most cases, which is similar to the small-scale test. However, it is found to occur more often in this full-scale test.



(a) Stimulating well: A04/BHP

(b) Stimulating well: B15/THP

Figure 5.3: Figures a) and b) show the generated partial correlations of the full-scale test, visualized as boxplots. Zoomed in on the range -0.4 to 0.4.



(a) Stimulating well: B15/THP (b) Stimulating well: B20/BHP

Figure 5.4: Figures a) and b) show the generated attention scores of the full-scale test, visualized as boxplots.

TCDF

Similar to the small-scale modified test, the mean attention score is consistently higher for self-causation. However, compared to the small-scale modified test, these mean attention scores of self-causation are approximately 0.5 points higher, with most laying in the 3.0 range. The mean attention scores are generally higher in the full-scale test compared to the small-scale modified test. This change was most evident by the stimulating well A03/BHP, which had most of the mean attention scores close to zero, while in the full-scale test, most have been shifted upwards approximately 0.3 points. The stimulating well S18/BHP has a distinctly narrow range on the attention scores, with the highest standard deviation being 0.601.

Visual inspection of Figure 5.4, show that there is no clear separation between the attention scores, which could be used to clearly identify *connectivity* and *no connectivity*. In the few occurrences when a mean attention score is distinctly higher than the rest of the target wells, the range of the attention scores are notably higher as well, ex. Well B20/BHP acting on B08/BHP. This well-pair do not have a known connectivity; however, it was the only well-pair with an IQR that stretches beyond an attention score of 1. The stimulating well B15/THP acting on the target well B20/BHP, a well-pair with known connectivity, is shown to have a distinct separation with the attention scores collecting well below zero. However, there are very few ($n = 5$) data points for this well-pair.

Chapter 6

Discussion

6.1 Small-scale baseline

PCMCI+

The baseline test of the PCMCI+ algorithm identifies the prevalence of self-causation by the stimulating wells. Additionally, it can be seen that there is very little variance in the number of time-steps identified as the strongest link. For self-causation, the strongest link is consistently after one time step (5 minutes), and any other link is on a contemporaneous or, in two cases, at one time step. We note that there are a few links that can be attributed to known connectivities. These are: B08/BHP acting on B15/THP and the reciprocal connectivity, as well as B20/BHP acting on B15/THP, listed in Table 6.1. However, their partial correlations (their strength) are not strong enough to pick out through classification, considering that we have a range of generated partial correlations in the range $[-0.305, 0.709]$, as shown in Figure 6.1.

Table 6.1: Table of true connectivities identified by the PCMCI+ algorithm in the small-scale baseline test.

Stimulating well	**Target well**	**Time delay**	**Partial correlation**
B15/THP	B08/BHP	0	0.340
B15/THP	B20/BHP	0	0.342

We attempted a subjective classification by visual inspection of a DAG illustration. The threshold for the partial correlation was set to be symmetric on -0.3 and 0.3, meaning correlations above or below the threshold are taken as a true link, illustrated in Figure 6.2. We note that the true positive links between B08/BHP acting on B15/THP, and B20/BHP acting on B15/THP, were identified. However, there are very strong links between B08/BHP acting on B20/BHP and A04/BHP acting on A20/BHP (including their reciprocals). These strong links lead us to believe that the model picks up external interference acting on the system, such as the same emergency shut-down logic or increase/decrease of source pressure. We find that more work needs to be done on the dataset to more clearly separate false and true connectivities, especially by reducing the interference from external sources.

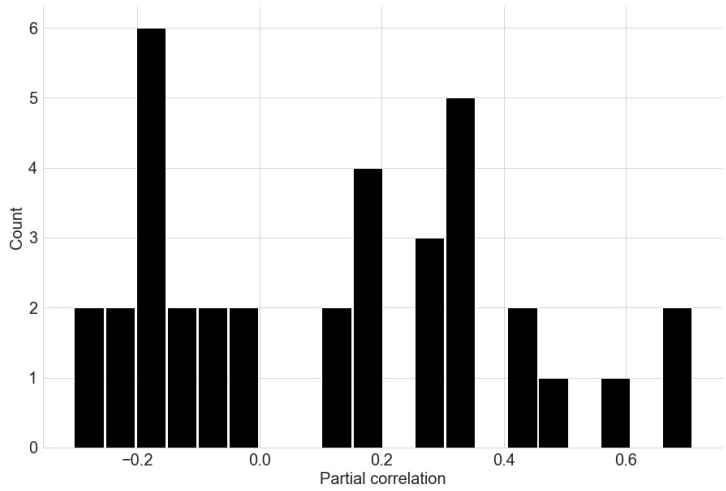


Figure 6.1: Histogram of the generated partial correlations for the PCMCI+ baseline test. $n = 36$, $bins = 20$.

The PCMCI+ algorithm does provide a method of classification; however, it is based on the set threshold, α, on the *p-value*. We found that this method still generated too many false positive links given a threshold of $\alpha < 0.01$. One option is to set the threshold lower, though this is mostly a subjective matter, and it does not resolve the main issue of the dataset containing external interferences, which in turn causes false connectivities to be identified.

The PCMCI+ algorithm provides a convenient solution by setting the partial correlation value as the degree of connectivity. Partial correlation values closer to zero are considered weak, and higher or lower values are considered stronger. The degree of connectivity for the identified links is illustrated in Figure 6.2 by the width of the arrows.

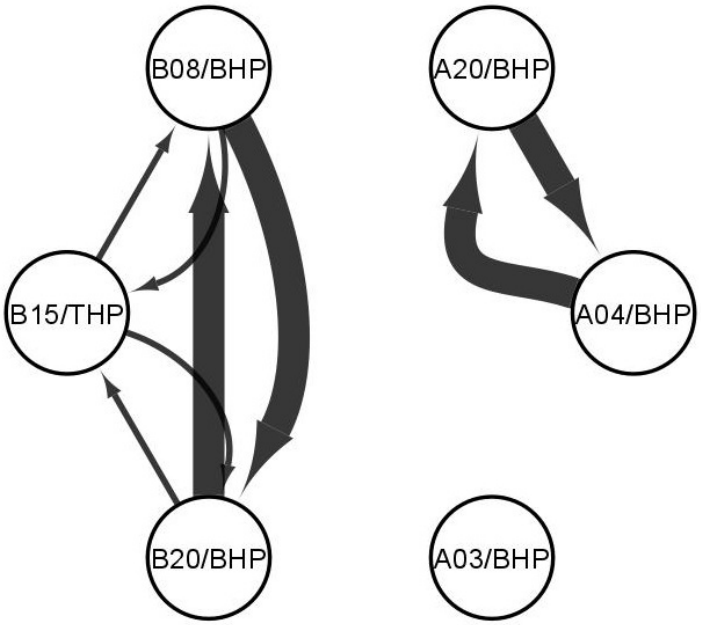


Figure 6.2: DAG of the generated partial correlations for the PCMCI+ baseline test. Subjective selection of links to maximize true positive links. Partial correlation greater than 0.3 or less than -0.3 were accepted as true links. The width of the arrows is a visual representation of the strength of connectivity.

TCDF

In the baseline test, we did not make any modifications to the dataset. The results are shown in Table 5.2 and illustrated in Figure 6.3 is the inferred connectivities by the TCDF-algorithm. The TCDF-algorithm identified nine validated causes, where six of these are attributed to self-causation. There were three validated causes for the well-pair B15/THP acting on B20/BHP, B15/THP acting on B08/BHP, and B08/BHP acting on A20/BHP. The link between the B-wells is true connectivities, while the indicated connectivity of B08/BHP acting on A20/BHP is false connectivity. However, we should note that these validated connectivities are very weak, the strongest link being 1.06 and the two others were 1.01 and 1.00. Considering that the attention scores are initialized at 1.00, we must consider that these links may have just happened by chance. Additionally, they were generated on a zero time lag, suggesting that the identified connectivity occurs instantaneously. If we ignore self-causation, the model did not pick up any strong links between the wells, unlike the PCMCI+ algorithm that identified stronger links within the B-wells.

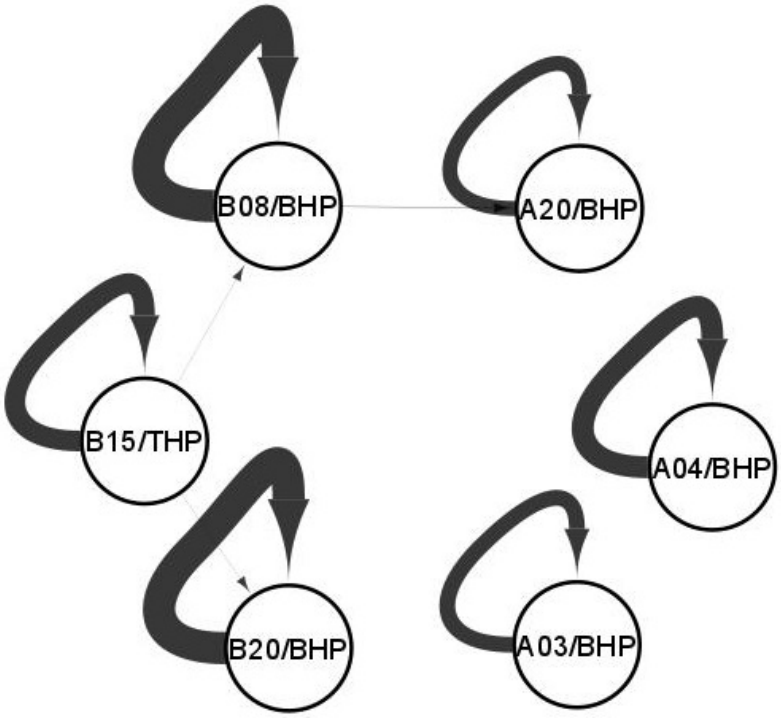


Figure 6.3: Visualization of the interwell connectivities listed in Table 5.2. The width of the arrows (strength of connectivity) is calculated from a normalized range of the attention scores.

6.2 Small-scale modified

In the small-scale modified test, the data was modified to reduce the overall noise and control the effect external factors had on the data. To infer interwell connectivity by the TCDF or PCMCI+ algorithms, we need to understand at what level their respective causation measures signify *connectivity* versus *no connectivity*. When dividing the original

dataset, we obtained many smaller datasets, each representing a specific stimulating well. The number of datasets for each stimulating well ranges from 105 to 35, meaning that there are varying degrees of uncertainty in the well-pairs' connectivity measures. Since the number of datasets has increased from one large dataset to 430 smaller ones, a hypothesis test on the causation measures was applied. There are 30 possible links between all well-pairs (excluding self-causation), four known connectivities, and 26 links with no known connectivities. Thus, by classifying all links with *connectivity*, the baseline precision is calculated to be 13.3%.

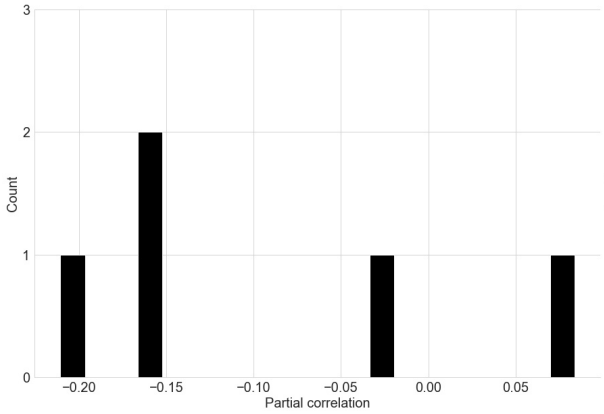
PCMCI+

The main assumption made regarding the well pairs is that a partial correlation of zero signifies *no connectivity*, and a positive partial correlation signifies *connectivity*, i.e., when the pressure is reduced in a well, the connecting well also reduces some of its pressure. There was an expectation that the mean and median values of the partial correlations for well-pairs with known connectivities would lean towards the positive end. However, relatively small movements of the mean and median are beyond zero and into the positive side. In the Results section, we also mentioned that the well-pairs with known connectivities show both positive and negative mean partial correlations, suggesting that there will be some links that will not be identified properly by this assumption.

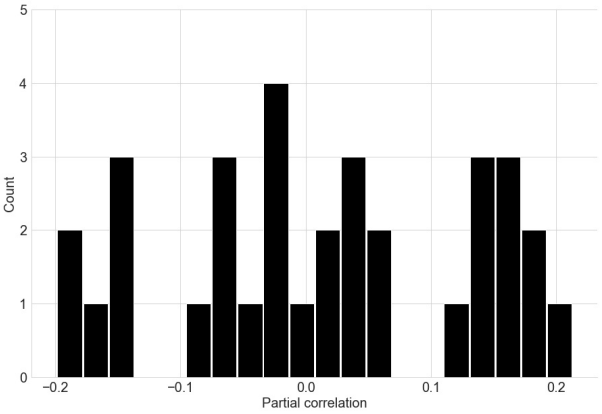
During the PCMCI+ algorithm, a hypothesis test is applied to filter the partial correlation measures, i.e., a form of classification has already been performed on each dataset. It was considered to use the same statistical test as employed by the PCMCI+ algorithm, the Student's t-test, and while it has been shown that it can be used for small samples sizes $n \leq 5$ (De Winter 2013), it is still affected by a non-normal distribution when the sample size is small (Lumley et al. 2002). In Figure 6.4, the distribution is shown for three different levels of sample sizes; the stimulating well B15/THP acting on the target well B20/BHP ($n = 5$), the stimulating well B20/BHP acting on the target well B15/THP ($n = 33$), and the stimulating well B20/BHP acting on the target well A20/BHP ($n = 75$). It is not easy to prove that the generated partial correlations follow a normal distribution in all cases. In the two largest sample sizes, Figure 6.4b and 6.4c, three distinct groups can be seen. In Figure 6.4b, the group centered around a partial correlation of 0.15 is seen to have some approximation to a normal distribution; however, this may be coincidental.

In Figure 6.5, a comparison of all generated partial correlations are shown. There is a clear formation of three separate groups, which was similarly identified in Figure 6.4. We find that this is sufficient evidence that the generated partial correlations do not support an assumption of a normal distribution. For that reason, we elected to apply Wilcoxon's signed-rank test for the hypothesis test, which does not assume a normal distribution (Wilcoxon 1945) and works well with few samples (Bellera et al. 2010). Wilcoxon signed-rank test is calculated by the Scipy package (Virtanen et al. 2020) in Python. We chose the clear separation of positive partial correlations, at approximately $\rho = 0.15$ in Figure 6.5, as the breakpoint for the hypothesis test.

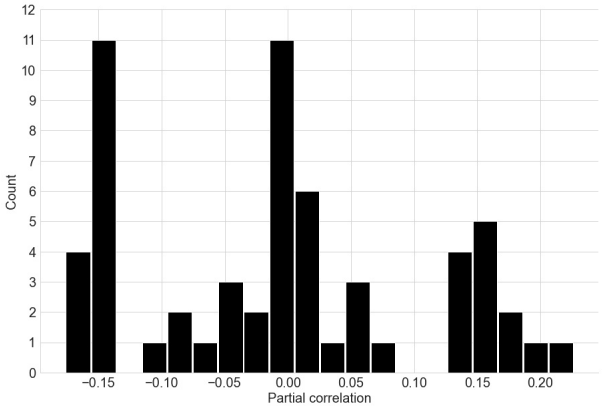
The hypothesis test is formulated as:



(a) Stimulating well: B15/THP.
Target well: B20/BHP.
$n = 5$, $bins = 20$.



(b) Stimulating well: B20/BHP.
Target well: B15/THP.
$n = 33$, $bins = 20$.



(c) Stimulating well: B20/BHP.
Target well: A20/BHP.
$n = 75$, $bins = 20$.

Figure 6.4: Figures (a) to (c) show histograms of the generated partial correlations for different sample sizes.

$$\begin{aligned} H_0 : &\quad \rho^{ji} \geq 0.15 \\ H_1 : &\quad \rho^{ji} < 0.15 \end{aligned} \tag{6.1}$$

Where H_0 is the null hypothesis, H_1 is the alternative hypothesis, and ρ^{ji} is the mean partial correlation for the stimulating well, j, acting on the target well, i. The hypothesis test is set up with the assumption that all well-pairs have *connectivity*, and the alternative is *no connectivity*.

In Table 6.2, classification by the generated p-values have been performed. This classification assumes that both links between a well-pair must be identified to classify them, i.e., we do not set both links equal to each other given that one of them has been classified. With this classification, we obtain a precision of 19.1% when sensitivity is 100% at $\alpha = 10^{-8}$. All p-values corresponding to their respective well pairs can be found in Appendix A.1.1.

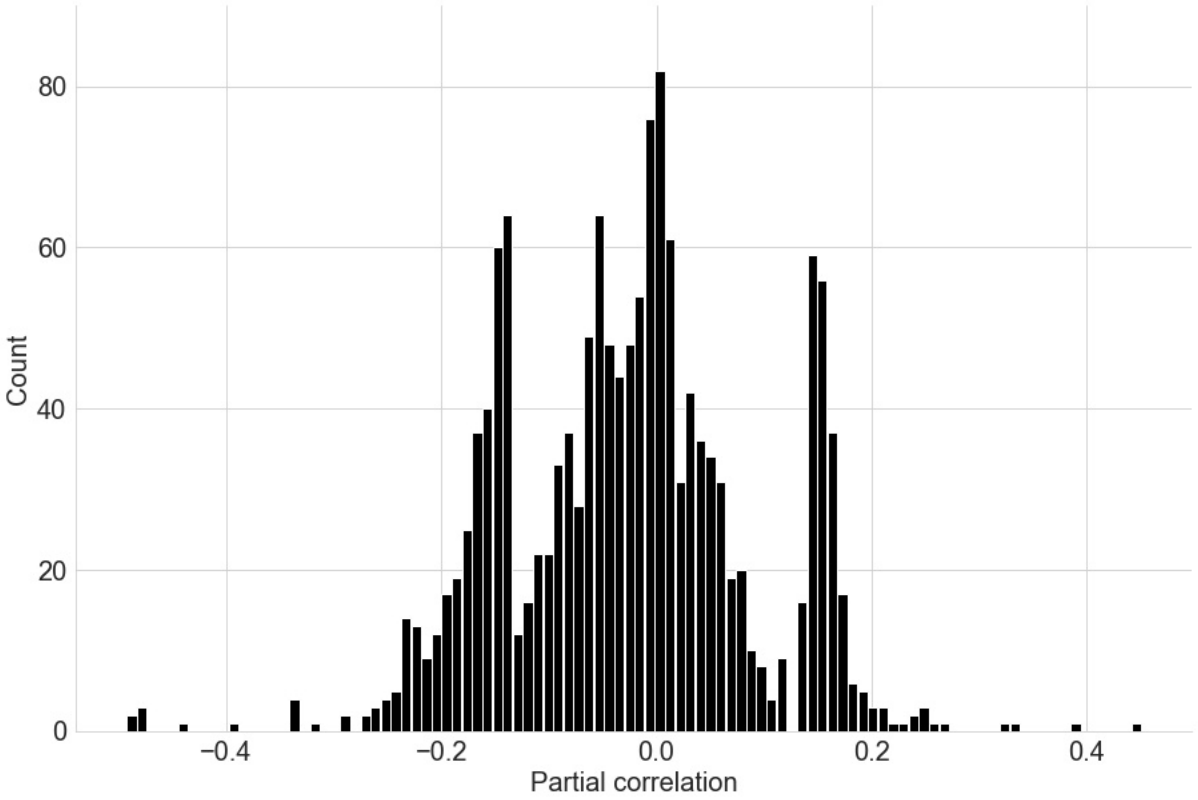


Figure 6.5: Histogram of all generated partial correlations. Partial correlations of self-causation have been removed.
$n = 2574$, $bins = 100$.

Classifying by an assumption that both links within a well-pair are equal, the precision improved to 28.6% with a more reasonable significance threshold of $\alpha = 10^{-4}$.

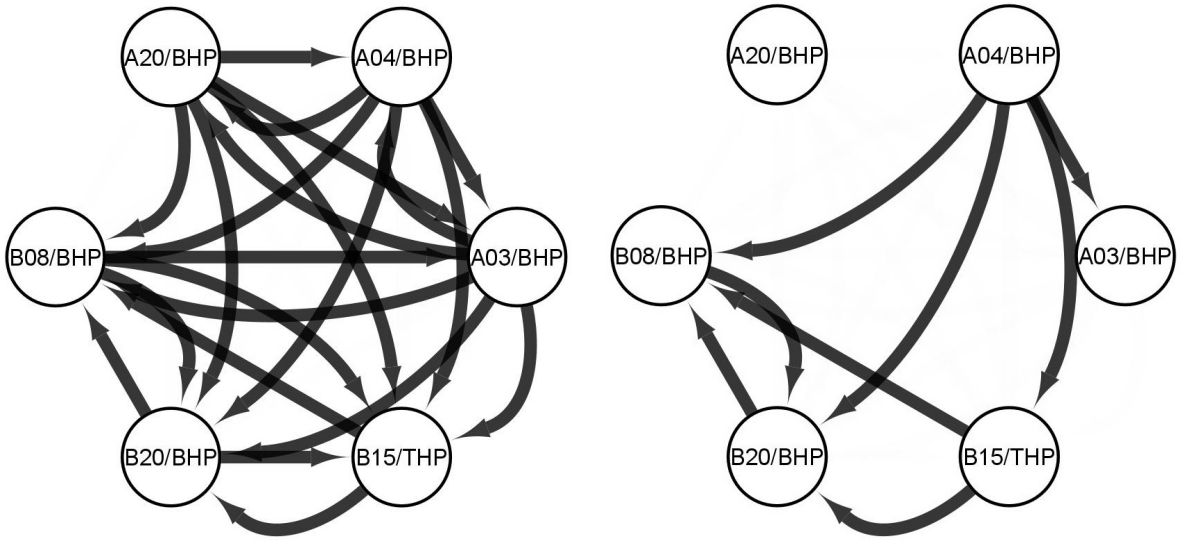
Overall, we found the performance to be poor through this classification method with a precision of 19.1% when sensitivity is 100%, which is a minor increase from the baseline of 13.3%. For completeness, we will apply the algorithm to the full-scale test, using the same parameters used in this small-scale test to obtain a sensitivity of 100%. However, we do not expect that the precision of the model will improve.

In Figure 6.6a, a DAG is presented on the assumption that all identified links are unique, and both links within a well-pair must be classified. This classification requires a significance threshold on the p-values of $\alpha = 10^{-8}$ to obtain a 100% sensitivity. Interestingly, the B-wells only link within their group (except for one link from B08/BHP to A03/BHP). The A-wells do not show a similar pattern and instead link to both the A and B groups. The greatest source of false connectivities is shown to be generated by the A-wells. In Figure 6.6b, the DAG illustrates the classification with the assumption that a single link represents the two-way connectivity of the well-pair. This classification requires a significance threshold on the p-values of $\alpha = 10^{-4}$ to obtain a 100% sensitivity. The well A04/BHP is shown to be the greatest source of false connectivities in this figure, with four links shown to be classified with *connectivity*. Moving from $\alpha = 10^{-8}$ to $\alpha = 10^{-4}$, a clear pattern appears in that almost every link for the injection wells (A04/BHP, B08/BHP, B15/THP, B20/BHP) is generated when $\alpha = 10^{-4}$. Further re-

Table 6.2: Sensitivity and specificity varied by a significance threshold, α, on the p-values, for classification of the small-scale modified dataset using the PCMCI+ algorithm. P-values used to generate this table can be found in Table A.7.

Significance, α	$\mathbf{10^0}$	$\mathbf{10^{-2}}$	$\mathbf{10^{-4}}$	$\mathbf{10^{-5}}$	$\mathbf{10^{-7}}$
Sensitivity	0.0000	0.2500	0.5000	0.7500	1.0000
Specificity	1.0000	0.9615	0.7692	0.5385	0.3462
Precision	N/A	0.5000	0.2500	0.2000	0.1905

ductions of the significance threshold to $\alpha = 10^{-8}$, the links for the producing wells (A03/BHP, A20/BHP) are generated, implying that the injection wells and production wells have different levels to their influence on target wells, i.e., the injection wells have a stronger effect on their surrounding wells compared to the production wells.

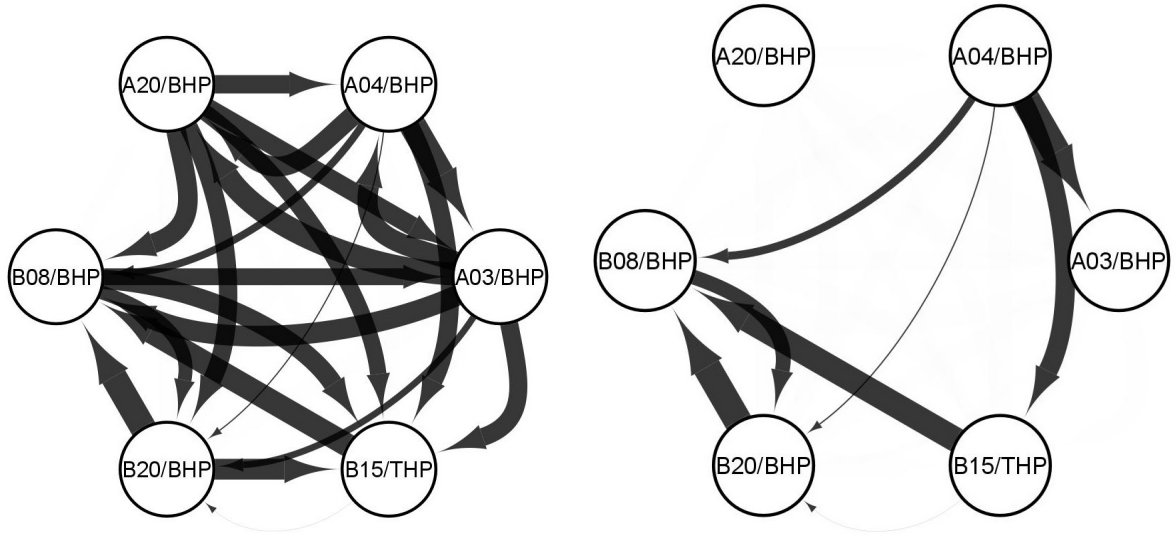


(a) $\alpha = 10^{-8}$ (b) $\alpha = 10^{-4}$

Figure 6.6: Figures (a) and (b) show DAGs of the classification on two different assumptions: (a) assumes that both links of a well pair must be identified, (b) assumes that a single link represent a two-way connectivity of the well-pair. Illustrations generated in Cytoscape.

In Figure 6.7 the same classification is shown as in Figure 6.6, however, the links have been given weight proportional to the partial correlation value generated for each link by the PCMCI+ algorithm. The weighting is represented by an increase (higher correlation) or decrease (lower correlation) of the width of the arrows. The weighting is relative, with the widest arrow given to the link with the highest partial correlation and the smallest arrow given to the link with the lowest partial correlation.
From Figures 6.7a and 6.7b, no clear pattern can be seen that support the idea that the generated partial correlations represent the true strength of connectivity, as seen by the link from the stimulating well B15/THP to the target well B20/BHP. A link that is known to be relatively strong is given the lowest weighting. Additionally, the strength of connectivity between the stimulating well A04/BHP and the B-wells is shown to have a

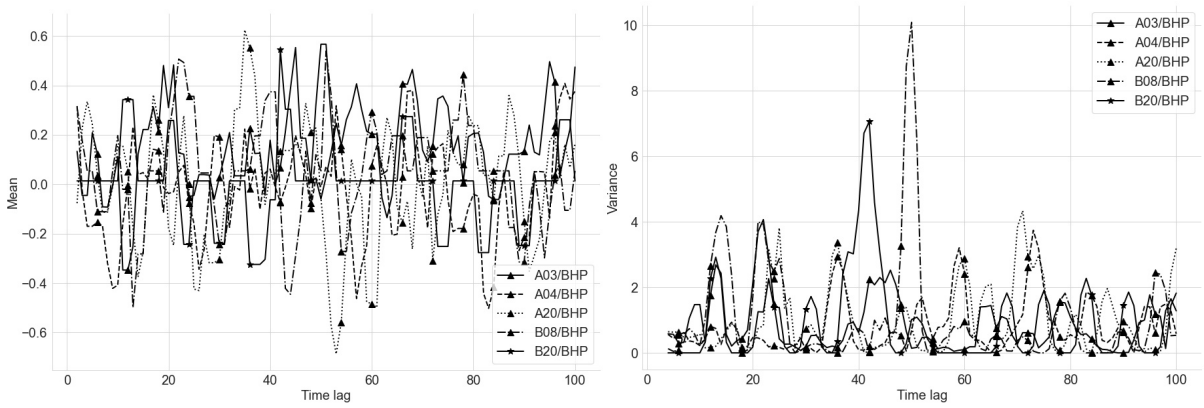
wide range of different strengths.



(a) $\alpha = 10^{-8}$ (b) $\alpha = 10^{-4}$

Figure 6.7: Figures (a) and (b) show DAGs of the classification from Figure 6.6, with weighting applied on the links. Illustrations generated in Cytoscape.

From a further investigation of the generated partial correlations, specifically, the full range of time lags. We found that the mean value of the partial correlations does not change significantly, even during periods of large changes in well pairs with known connectivities. Instead, there is a significant increase in variance, highlighted in Figure 6.8, which shows a comparison of a rolling mean and a rolling variance window over the whole time range. For the rolling variance, a second-order difference was first applied; this was found to improve the separability of the wells. Figure 6.8, show one of the more clear examples of how the separability of the data improved with this transformation. Other well-pairs had less of an improvement. Classification using these findings may be attempted in future work.



(a) Rolling mean.

(b) Rolling variance.

Figure 6.8: Figures (a) and (b) show a rolling mean and variance window on the lagged partial correlations for the stimulating well B15/THP. Window size = 3. Time lag = steps of 5 min intervals. Markers shown every 6 steps (every 30 min).

TCDF

The attention scores are, as mentioned in chapter 2.3.2, initialized as 1 and increased or reduced if the CNN has given it attention or not. From the boxplots in Figure 5.2, it is evident that in our datasets, self-causation is a major factor, as self-causation is consistently given a mean attention score greater than 2. By actively seeking periods when the stimulating well is either trending upwards or downwards, we acknowledge that this is a consequence of how we have selected our datasets, and the behavior can be described as autoregressive. Self-causation as the main driver of causality is also evident from our results in the baseline test, as shown in Table 5.2, every potential and validated cause is directly linked to self-causation. In the results from Tables A.8-A.13, not a single well is given a mean attention score greater than 1 (excluding self-causation). The closest well pair, B20/BHP acting on B08/BHP, obtains an attention score of 0.744.

The hypothesis test is formulated as:

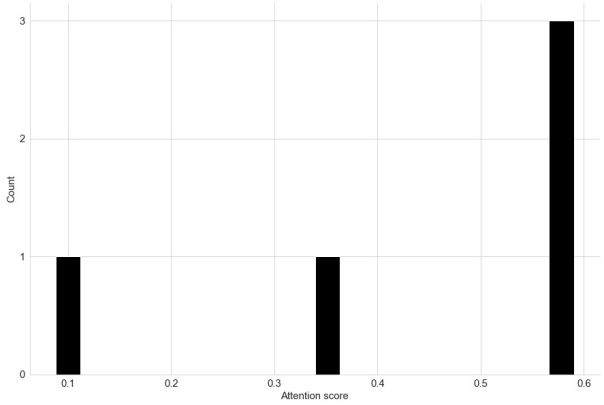
$$\begin{aligned} H_0 &: \quad a_{ji} \geq 1 \\ H_1 &: \quad a_{ji} < 1 \end{aligned} \tag{6.2}$$

Where H_0 is the null hypothesis, H_1 is the alternative hypothesis, and a_{ji} is the attention score for the stimulating well, j, acting on the target well, i. The way we set up the hypothesis test is to set the null hypothesis with the assumption that all well-pairs have *connectivity*, and the alternative is *no connectivity*. There are certain limitations in these datasets for the statistical test, mainly sample size and undetermined distribution of the attention scores. This limitation is especially the case for the stimulating well B15/THP acting on the well B20/BHP, with as few as five data points.

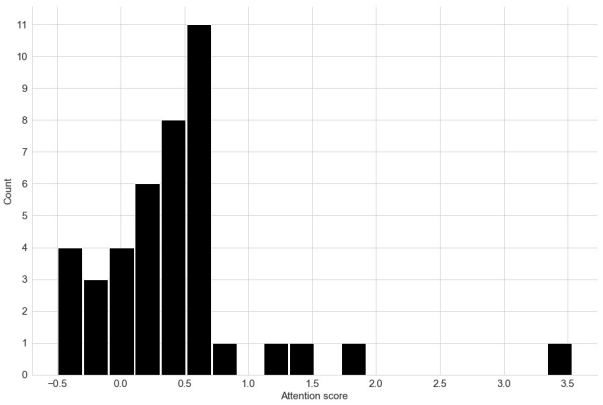
The statistical test, Student's t-test, assumes a normal distribution in the data, an assumption that is difficult to prove with the limited number of samples in our datasets, as shown by the three examples in Figure 6.9. Here, a very small sample size ($n = 5$) is shown for the stimulating well B15/THP acting on B20/BHP, a medium sample size ($n = 41$) is shown for the stimulating well B20/BHP acting on B15/THP, and a large sample size ($n = 89$) is shown for the stimulating well B20/BHP acting on A20/BHP. It is clear from the three histograms that the attention scores do not approximate a normal distribution given the obtained sample sizes. The Wilcoxon signed-rank test (Wilcoxon 1945) is instead used, which is a non-parametric statistical test without the assumption of normality and works well with few data points (Bellera et al. 2010). We use the Scipy package (Virtanen et al. 2020) in Python to perform the Wilcoxon statistical test.

For this test, the significance level is initially set to $\alpha = 0.01$. Well-pairs with a p-value lower than 0.01 are classified with *no connectivity*, according to the alternative hypothesis, and well-pairs with higher p-values are classified with *connectivity*, according to the null hypothesis. We illustrate this classification with a DAG in Figure 6.10. All p-values corresponding to their respective well pairs can be found in Appendix A.1.2.

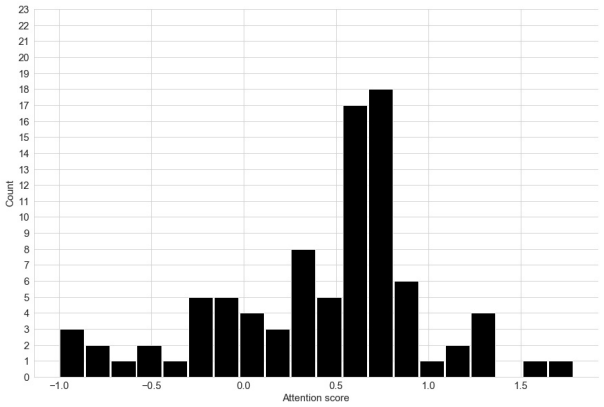
From Figure 6.10, there are two wells classified with *connectivity*, as shown with the arrows, these are; B15/THP acting on B20/BHP, and B20/BHP acting on B08/BHP. One of the connectivities (B15/THP on B20/BHP) is correctly identified; however, the reciprocal link, B20/BHP acting on B15/THP, is not identified. The second identified connectivity, B20/BHP on B08/BHP, is false connectivity. Additionally, the algorithm



(a) Stimulating well: B15/THP.
Target well: B20/BHP.
$n = 5$, $bins = 20$.



(b) Stimulating well: B20/THP.
Target well: B15/BHP.
$n = 41$, $bins = 20$.



(c) Stimulating well: B20/THP.
Target well: A20/BHP.
$n = 89$, $bins = 20$.

Figure 6.9: Figures (a) to (c) show histograms of the generated attention scores for different sample sizes.

did not identify connectivity for the stimulating well B08/BHP acting on the target well B15/THP.

Classification by the p-values was tested by varying the significance level, and the results can be seen in Table 6.3. Sensitivity reflects the correct identification of *connectivity*, and specificity reflects the correct identification of *no connectivity*. The total number of links is 30 for this dataset, which excludes links of self-causation and includes reciprocal links. There are four links of known connectivity, and consequently, 26 links of no known connectivity. From Table 6.3, it is shown that there is a steady increase in the sensitivity as the significance level is reduced from 10^{-1} to 10^{-3}, further reductions required a significant increase in the number of false connectivities.

Selecting the significance level when sensitivity is 100% from Table 6.3, the corresponding DAG can be seen in Figure 6.11a. From the DAG, it is noted that the B-wells are only linking with each other by two and three links, and no links from the B-wells to the A-wells can be seen. The A-wells have four or five links each, and they are linking

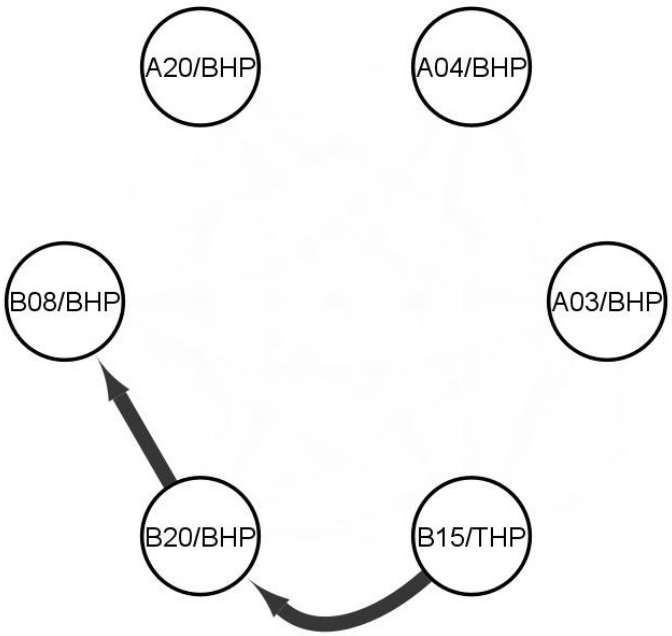


Figure 6.10: DAG of the classification by the p-values. A hypothesis test was applied to the estimated attention score means for each stimulating/target well-pair with a significance level of $\alpha = 0.01$. Self-causation links were manually removed. Illustration generated in Cytoscape.

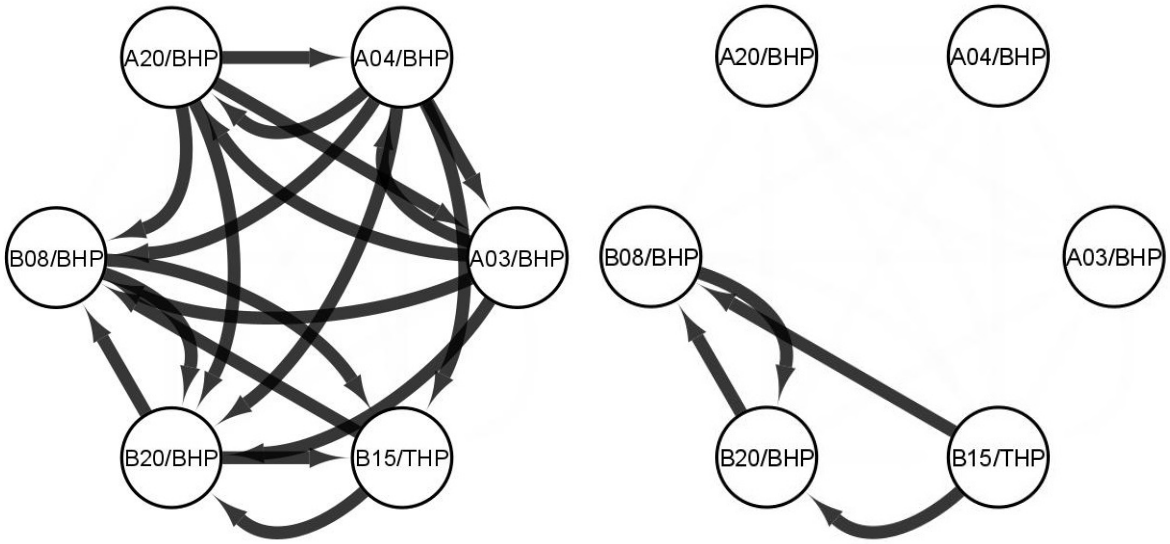
Table 6.3: Sensitivity, specificity, and precision, varied by a significance threshold, α, on the p-values, for classification of the small-scale modified dataset using the TCDF-algorithm.

Significance, α	$\mathbf{10^{-1}}$	$\mathbf{10^{-2}}$	$\mathbf{10^{-3}}$	$\mathbf{10^{-6}}$	$\mathbf{10^{-9}}$
Sensitivity	0.0000	0.2500	0.5000	0.7500	1.0000
Specificity	1.0000	0.9615	0.9231	0.6923	0.4231
Precision	N/A	0.5000	0.5000	0.2727	0.2105

within the A-wells group and to the B-wells. When $\alpha = 10^{-9}$, and the sensitivity is 100%, the precision is calculated to be 21.1%, suggesting that for every true connectivity identified, three false connectivities are identified. This precision assumes that each well-pair connectivity and its reciprocal connectivity need to be identified. One could also assume that a single link between a well pair implies that the reciprocal link also exists (but is yet identified by the model). By the latter assumption, all links can be identified at a more reasonable significance threshold of $\alpha = 10^{-3}$, as shown in Figure 6.11b. The precision obtained for the latter assumption is 67.7%.

Applying a base 10 log-transform on the p-values, the modified DAGs from Figure 6.11 can be seen in Figure 6.12. Since the strength of connectivity is set as a relational representation and based on the set significance threshold, the two Figures 6.12a and 6.12b, show different widths on the arrows for the same well-pairs.

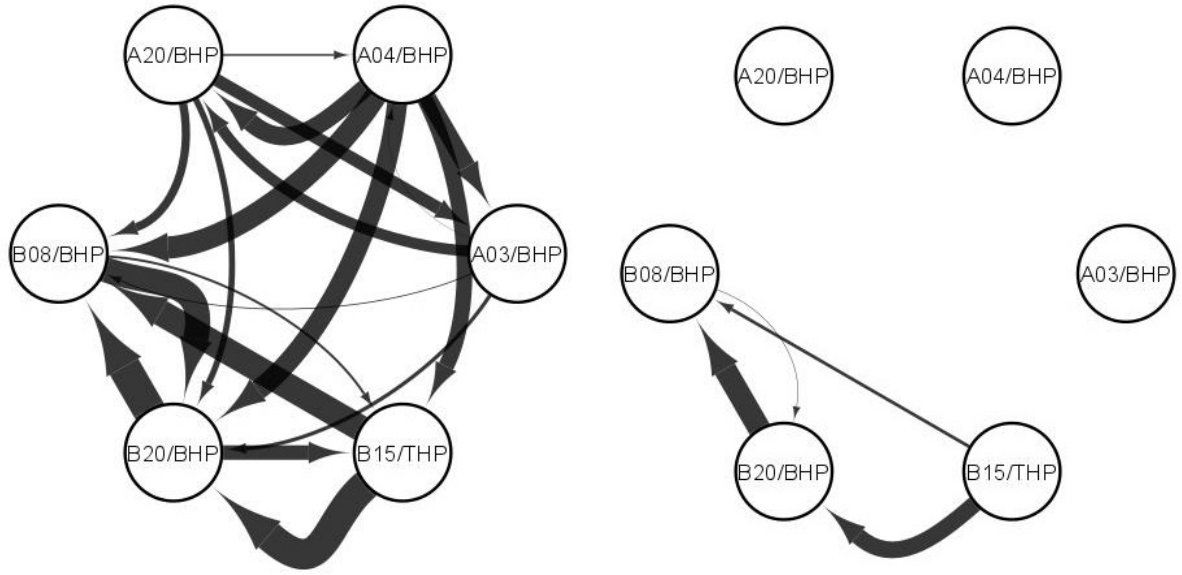
In Figure 6.12a, we find that the injection wells (B-wells and A04/BHP) are shown to generally have a stronger relationship compared to the production wells (A03/BHP



(a) $\alpha = 10^{-9}$. (b) $\alpha = 10^{-3}$.

Figure 6.11: Figures (a) and (b) show DAGs of the classification on two different assumptions: (a) assumes that both links of a well pair must be identified, (b) assumes that a single link represent the full connectivity of the well pair. Illustrations generated in Cytoscape.

and A20/BHP). The stimulating well A04/BHP is also shown to have little variance in its connectivity strength to all other wells. This low variance is likely a consequence of the significance threshold being set to a very low value of 10^{-9} and the amount of known false connectivities added to the result. The classification shown in Figure 6.12b have a more reasonable significance threshold of 10^{-3}; however, the strongest link is shown to be the stimulating well B20/BHP acting on the target well B08/BHP, which is false connectivity. These figures show that there is likely still a somewhat high degree of external connectivity acting on the injection wells.



(a) $\alpha = 10^{-9}$. (b) $\alpha = 10^{-3}$.

Figure 6.12: Figures (a) and (b) show DAGs of the classification from Figure 6.11 with the width of the arrows representing the strength of connectivity. Illustrations generated in Cytoscape.

6.3 Full-scale modified

In this test, the aim was to identify all known well-pair connectivities from the field. There are a total of 156 possible links between the 13 wells, excluding self-causation. In these 156 possible links, there are eight unique well-pair connectivities (16 connectivities if the reciprocal is included). By classifying all links with *connectivity*, we obtain a baseline precision of 10.3%. We examine different scenarios of the performance of the two algorithms PCMCI+ and TCDF. However, the final performance of the models will be a product of the parameters and hyperparameters obtained in the small-scale test.

6.3.1 PCMCI+

Moving from the small-scale modified test to the full-scale modified one, we found the distributions very similar, even though the number of wells has increased from six to thirteen. In Figure 6.13, the distribution of all partial correlations is shown for the full-scale test. Similar to the small-scale test, the distribution seems divided into three groups centered close to a partial correlation of -0.2, 0.0, and 0.2. Considering the similarity between the distributions of the small-scale and full-scale test, we find that the same hypothesis test and breakpoint at $r = 0.15$ from the small-scale test can be used for the full-scale test.

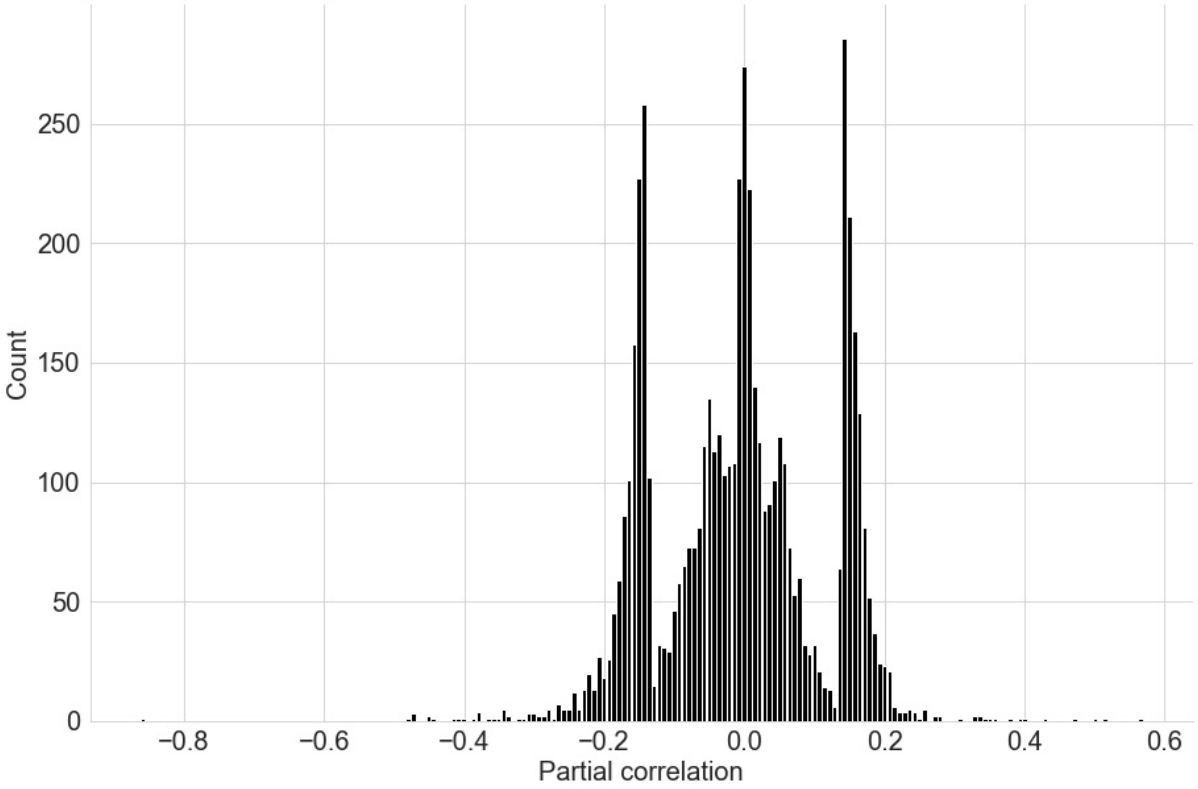


Figure 6.13: Histogram of all generated partial correlations for the full-scale test. Partial correlations of self-causation have been removed. $n = 11414$, $bins = 200$

In Table 6.4, the sensitivity, specificity, and precision have been listed for different significance thresholds. If a single link between a well-pair is classified with *connectivity* or

no connectivity, the reciprocal link is given the same classification. We use this assumption as it was found to provide the highest precision when all known connectivities had been identified in the small-scale test. All p-values corresponding to their respective well pairs can be found in Appendix A.2.1.

Table 6.4: Sensitivity and specificity varied by a significance threshold, α, on the p-values, for classification of the full-scale modified dataset using the PCMCI+ algorithm. P-values used to generate this table can be found in Tables A.28 and A.29.

Significance, α	$\mathbf{10^0}$	$\mathbf{10^{-1}}$	$\mathbf{10^{-3}}$	$\mathbf{10^{-4}}$	$\mathbf{10^{-10}}$
Sensitivity	0.0000	0.1250	0.3750	0.8750	1.0000
Specificity	1.0000	0.9857	0.7714	0.4857	0.0143
Precision	N/A	0.5000	0.1579	0.1628	0.1039

From Table 6.4, a maximum precision of 50.0% is obtained when $\alpha = 10^{-1}$, however, this is with only two true connectivities identified, for the well-pair B15/THP and B20/BHP. We ignore this result as it does not provide any new information and instead examine the second-highest maximum precision of 16.3% obtained with $\alpha = 10^{-4}$. This significance threshold also happens to be the same that the small-scale modified test used to obtain its maximum precision.

Further examination of the classified links, we found that producer/producer links generated twelve false connectivities. Since we do not have any producer/producer connectivities in the dataset, we remove these links to improve the precision of the model. With the producer/producer links removed, the precision improved to 18.9%. It is noted that the precision of 18.9% in this full-scale test is very close to the precision of 19.1% obtained in the small-scale test, suggesting that the model's performance is robust to an increase in the number of wells introduced.

In Figure 6.14, the final classification can be seen, where links have been given a weight based on the mean partial correlation for the respective well-pairs. The width of the links is relative, from the maximum to the minimum mean partial correlation in this test. No clear pattern emerges from Figure 6.14, which could further improve the classification.

A few wells appear to be major sources of false signals; A04/BHP, A07/THP, and S18/BHP. In Table 6.5, all true and false signals from Figure 6.14 have been listed. All three wells are injection wells which may provide some explanation. However, other injection wells such as B08/BHP and B15/THP have few false signals. Comparatively, the production wells generated fewer false signals overall, in part due to the removal of all producer/producer links.
In Table 6.5, the precision for each individual well is also shown. Two wells stand out: the injection well B15/THP, which has relatively high precision, and the production well A03/BHP, which has a precision of zero. For the rest of the wells, it is found that, generally, the injection wells have a lower precision ($< 25.0\%$), and the production wells have a higher precision ($\geq 25.0\%$). This difference in precisions suggests a higher degree of interference affecting the models' ability to infer true connectivity between the injection wells, specifically the shared supply of water injection. We conclude that the shared water

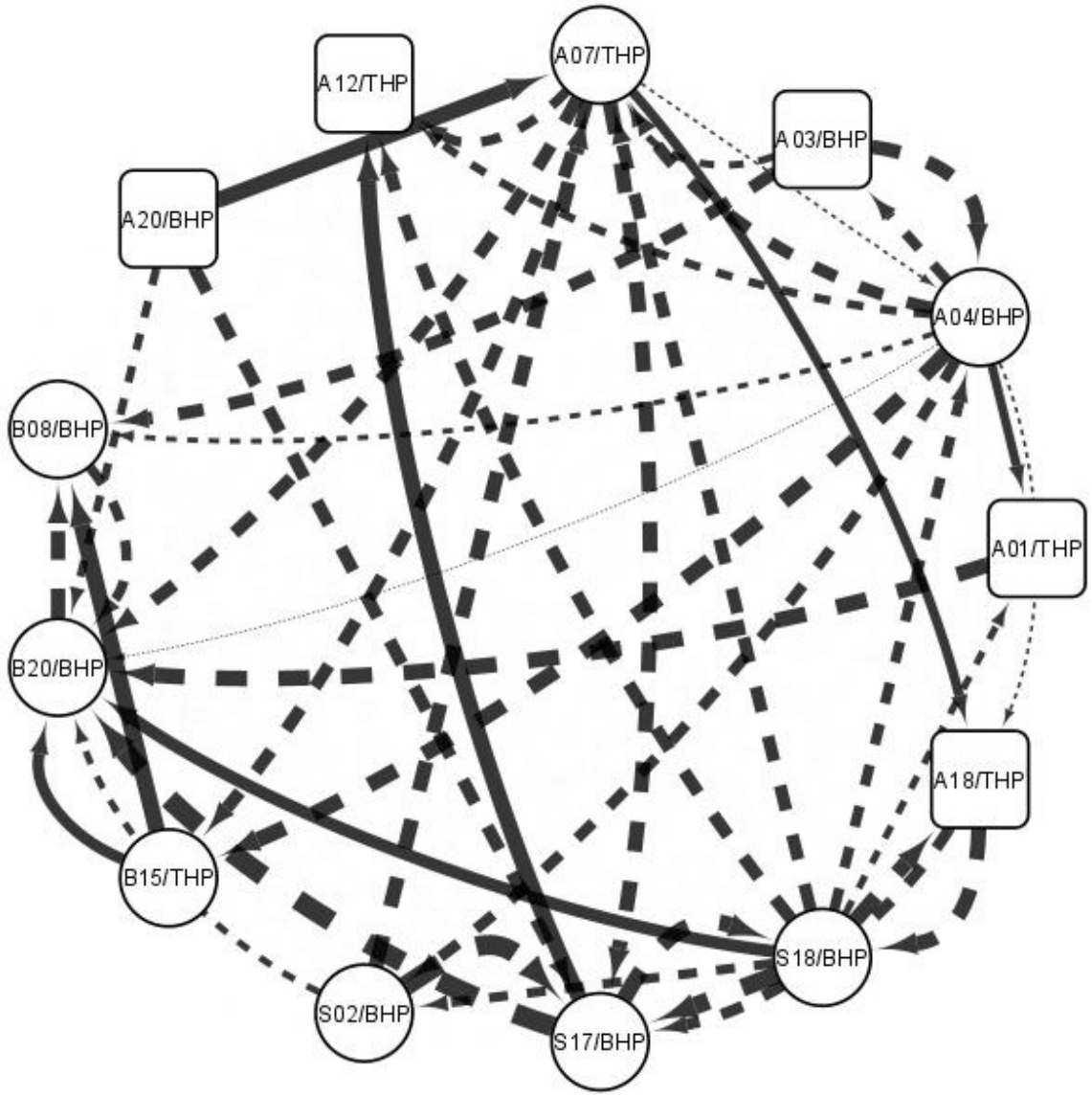


Figure 6.14: DAG of the full-scale test. True connectivities illustrated by a solid arrow, false connectivities by a dashed arrow. Injection wells illustrated by a circle, and production wells by a rounded square.

supply is still affecting the injection wells enough to generate many false connectivities. More work needs to be done to the datasets in order to filter out this type of interference to improve the precision of the model.

We had previously assumed that the partial correlation measure generated by the PCMCI+ algorithm could be directly applied for the strength of connectivity. We also assume that true connectivity will generally return higher partial correlation measures compared to false connectivity. In Figure 6.15, all links from Figure 6.14 have been plotted in a histogram to show the distribution for *True connectivity* and *False connectivity*. Links labeled with *True connectivity* in Figure 6.15, does not appear to consistently return higher partial correlation measures, as we assumed. Since the links labeled with *True connectivity* are spread over a relatively narrow range and contained within the range of links labeled with *False connectivity*. We do not find that there is enough evi-

Table 6.5: Distribution of true and false connectivities for the various wells. Precision is calculated for all true and false connectivities from each well. First section of wells are injectors, and the second section is producers.

Well	**As stimulating**		**As target**		
	False	**True**	**False**	**True**	**Precision**
A04/BHP	8	1	3	0	8.0%
A07/THP	5	1	4	1	18.0%
B08/BHP	1	0	3	1	20.0%
B15/THP	0	2	2	0	50.0%
B20/BHP	1	0	7	2	20.0%
S02/BHP	3	0	2	0	0.0%
S17/BHP	2	1	5	0	12.0%
S18/BHP	7	1	2	0	10.0%
	False	**True**	**False**	**True**	**Precision**
A01/THP	1	0	1	1	33.0%
A03/BHP	3	0	1	0	0.0%
A12/THP	0	0	3	1	25.0%
A18/THP	1	0	2	1	25.0%
A20/BHP	2	1	0	0	33.0%

dence to suggest that the partial correlation measure is representative of the strength of connectivity. To further verify this claim that the two distributions are equal, we apply a hypothesis test on the partial correlation measures from Figure 6.15. The hypothesis test is formulated as (6.3), and we set a significance threshold of $\alpha = 0.05$.

$$\begin{aligned} H_0 &: \quad \mu^T = \mu^F \\ H_1 &: \quad \mu^T \neq \mu^F \end{aligned} \tag{6.3}$$

Where μ^T and μ^F is the mean partial correlation of the *True connectivity* and *False connectivity* distributions, in Figure 6.15. Using Mann-Whitney U-test, we obtain a p-value of 0.55. Since $0.55 > 0.05$, we do not have enough evidence to reject the null hypothesis and conclude that the partial correlation measures in this test do not represent the strength of connectivity.

We examine any notable difference between THP and BHP as a source of data similar to what we did above, with the strength of connectivity measures. The same hypothesis test as in (6.3) is performed, except we compare the distribution of partial correlations generated by any well-pair with one or two wells using THP-data against any well-pair that do not contain THP-data. The resulting p-value was 0.66, suggesting that we do not have enough evidence to reject the null hypothesis and conclude that there is no difference between the distribution means of BHP and THP data.

For the group- and field-wide external connectivities, we examine Figure 6.14. Interestingly, B-wells do not have links to A- or S-wells; however, there are links from A- and S-wells to B-wells, suggesting that the field-wide external connectivity (from the shared

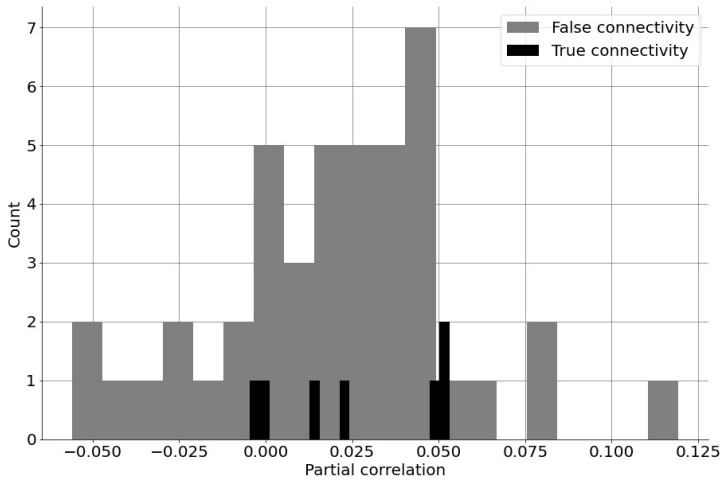
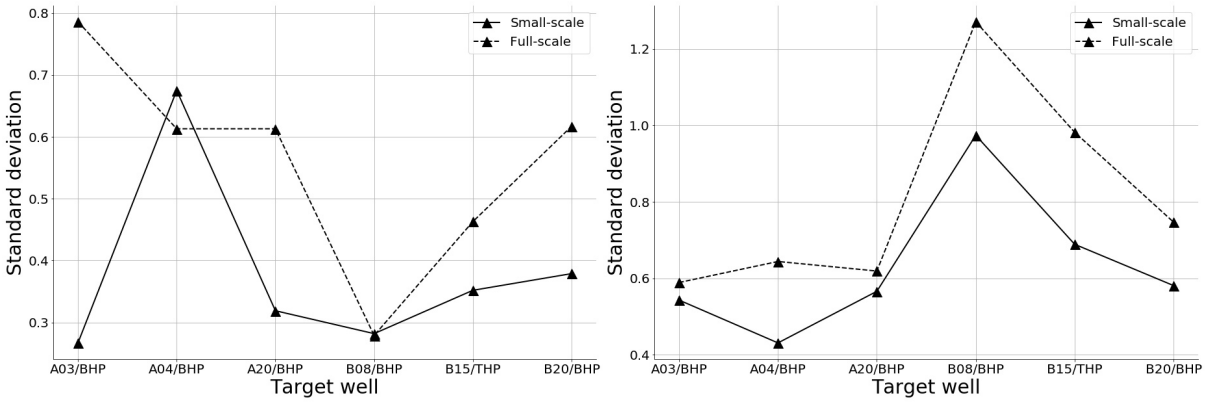


Figure 6.15: Histogram of the mean partial correlation distribution for the links labeled as *True connectivity* and *False connectivity*. *bins*=20.

water supply through the injection wells) have not been effectively removed. This claim is further supported by Table 6.5, which show that specifically the A- and S-injection wells; A04/BHP, A07/THP, S18/BHP, and to a lesser extent, S02/BHP and S17/BHP, produce the highest number of false signals as stimulating wells. For the group-wide external connectivity, we expect a high number of links identified between well-pairs from the same group, i.e., the A-, B-, or S-wells should link within their respective groups. We did not find that this was the case, at least not to the same extent as the field-wide external connectivity.

6.3.2 TCDF

In Figure 5.4, we note that there is a less clear separation between the targets' attention scores compared to the small-scale modified test. There is a general widening of the attention scores; an example is given as a comparison of the standard deviation for the stimulating wells A04/BHP and B20/BHP in Figure 6.16. Additionally, several of the attention score ranges are stretching upwards to and above 1. This stretching will cause an increase in false connectivities. It is uncertain why an increase in the number of wells would cause the attention scores to widen, as the TCDF-algorithm trains a separate CNN for each input time series. There are five producers and eight injectors in the full-scale test and two producers and four injectors for the small-scale test, meaning that the producer/injector distribution is relatively similar for the two tests. A difference between these two tests is the inclusion of four additional wells using THP as the data source. We have previously mentioned that this data contains more noise and is more prone to external interferences. This inclusion of THP data may be why more noise is experienced in this test, and a widening of the attention score ranges is seen.



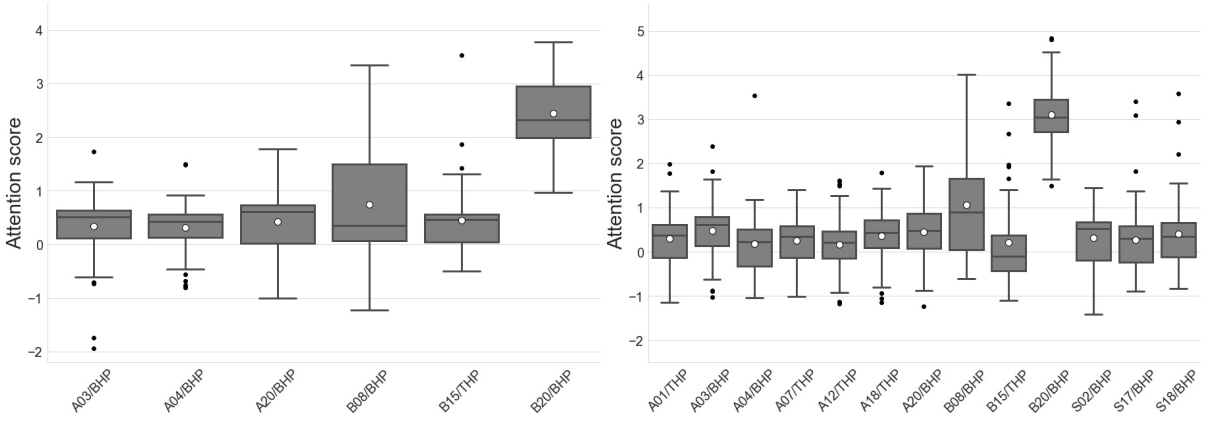
(a) Stimulating well A04/BHP.

(b) Stimulating well B20/BHP.

Figure 6.16: Figures (a) and (b) show a comparison of the standard deviation for the ranges of the small-scale and full-scale tests.

A comparison can be drawn of the stimulating well B20/BHP from the small-scale and full-scale test, shown in Figure 6.17. The target well B15/THP is shown to have an increased range in the full-scale test compared to the small-scale test; additionally, it has been shifted further down towards zero. Essentially, the conditions for inference of connectivity have worsened for the B20/BHP and B15/THP well-pair. The target well A04/BHP can also be seen to have been both widened and shifted downwards, an improvement towards inference of no connectivity. The target wells A03/BHP and A20/BHP have undergone less change, with the whiskers slightly widened. Finally, the target well B08/BHP has been slightly shifted upwards, and the range of the upper whisker has been widened, while the lower whisker has been shortened, which worsens the inference of connectivity.

We make another comparison, this time with the stimulating well A04/BHP from the small-scale and full-scale test, as shown in Figure 6.18. A04/BHP had a distinctly narrow range for all of the target wells in the small-scale test. In the full-scale test, these ranges

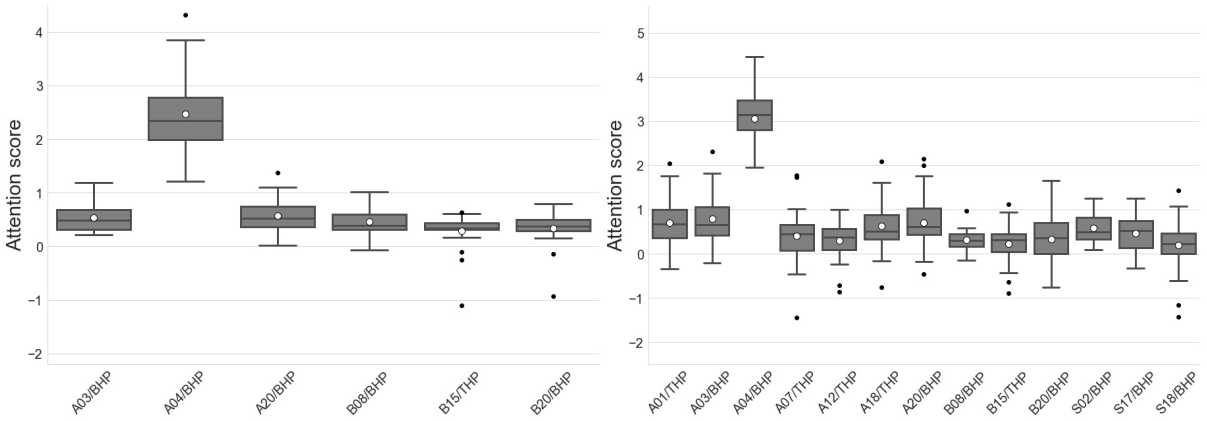


(a) Stimulating well B20/BHP from the small-scale modified test.

(b) Stimulating well B20/BHP from the full-scale modified test.

Figure 6.17: Figures (a) and (b) show boxplots of the stimulating well B20/BHP.

have been widened considerably. The target wells A03/BHP, and A20/BHP can be seen to stretch upwards and closer to an attention score of 1, worsening the conditions for inference of no connectivity for these well pairs. The target well B20/BHP has undergone the biggest change, while the mean value of the attention scores is practically unchanged (0.335 and 0.337), the standard deviation increased from 0.379 to 0.617, which again will worsen the conditions for inference of no connectivity between the well pair.



(a) Stimulating well A04/BHP from the small-scale modified test.

(b) Stimulating well A04/BHP from the full-scale modified test.

Figure 6.18: Figures (a) and (b) show boxplots of the stimulating well A04/BHP.

When moving from a dataset of few, less complex wells to more wells with higher complexity, there seem to be both degradations and improvements towards the classification of connectivity/no connectivity. We found more cases in which the conditions for correct inference of connectivity/no connectivity were worsened, and few cases improved.

The same classification approach as in the small-scale test will be applied, using a hypothesis test on the attention scores. In Table 6.6, the sensitivity, specificity and precision is listed for the full-scale test. These results are without the assumption that an iden-

tified connectivity signifies that the reciprocal link exists, mentioned in the small-scale test. The maximum precision was 15.8%, with a sensitivity of 56.3%. However, using the significance threshold from the small-scale test, $\alpha = 10^{-3}$, the precision reduces to 14.3% with a sensitivity of 18.8%. These results can easily be improved with the assumption mentioned above. All p-values corresponding to their respective well pairs can be found in Appendix A.2.2.

Table 6.6: Sensitivity, specificity and precision varied by a significance threshold, α, on the p-values, for classification of the full-scale modified dataset using the TCDF-algorithm.

Significance, α	$\mathbf{10^0}$	$\mathbf{10^{-3}}$	$\mathbf{10^{-5}}$	$\mathbf{10^{-8}}$	$\mathbf{10^{-11}}$
Sensitivity	0.0000	0.1875	0.5625	0.7500	1.0000
Specificity	1.0000	0.8714	0.6571	0.3000	0.0786
Precision	N/A	0.1429	0.1579	0.1091	0.1103

In Table 6.7. the sensitivity, specificity, and precision are listed with the assumption of a single link representing the two-way connectivity between the well pairs. While the maximum precision only had a marginal improvement, the model was able to identify all connectivities when the significance threshold was lowered to 10^{-5}. When the significance threshold was set equal to the small-scale test, $\alpha = 10^{-3}$, we obtain a precision of 17.7% with a sensitivity of 37.5%.

Table 6.7: Sensitivity, specificity and precision varied by a significance threshold, α, on the p-values, for classification of the full-scale modified dataset using the TCDF-algorithm. Assumption of single link representing two-way connectivity.

Significance, α	$\mathbf{10^0}$	$\mathbf{10^{-2}}$	$\mathbf{10^{-3}}$	$\mathbf{10^{-4}}$	$\mathbf{10^{-5}}$
Sensitivity	0.0000	0.1250	0.3750	0.7500	1.0000
Specificity	1.0000	0.9143	0.8000	0.6000	0.4714
Precision	N/A	0.1429	0.1765	0.1765	0.1778

After examining the connectivities identified by the model, it was noted that a large number of false connectivities were occurring between producer/producer well-pairs. There are known connectivities between producers in the Eldfisk field. However, there are no known connectivities between the producers in this dataset. As such, we do not attempt to identify this type of relationship, and all producer/producer links were manually removed, reducing the total number of possible links from 156 to 136, and the baseline precision increased to 11.8%. With the producer/producer links removed, the sensitivity increased to 37.5%, specificity increased to 86.7%, and precision increased to 27.3%. The final result is shown in Figure 6.19; unfortunately, there was only one additional well-pair, A04/BHP and A01/THP, identified at this significance threshold.

While we would have liked to see a larger portion of the known connectivities to be identified by the model, achieved by lowering the significance threshold below $\alpha = 10^{-3}$, we find it difficult to justify such as an action. Simply by the fact that the full-scale test

was implemented to verify that all parameters calibrated by the small-scale test could be used to identify new connectivities. There may be some justification for lowering the significance threshold, given that the precision in Table 6.7 holds a surprisingly stable level even when we change the significance threshold from $\alpha = 10^{-3}$ to $\alpha = 10^{-5}$. However, if the TCDF-algorithm performs similarly for several different levels of α, this would have to be further tested in future work. We conclude that the TCDF-algorithm is effective at identifying interwell connectivity with a precision of 27.3%.

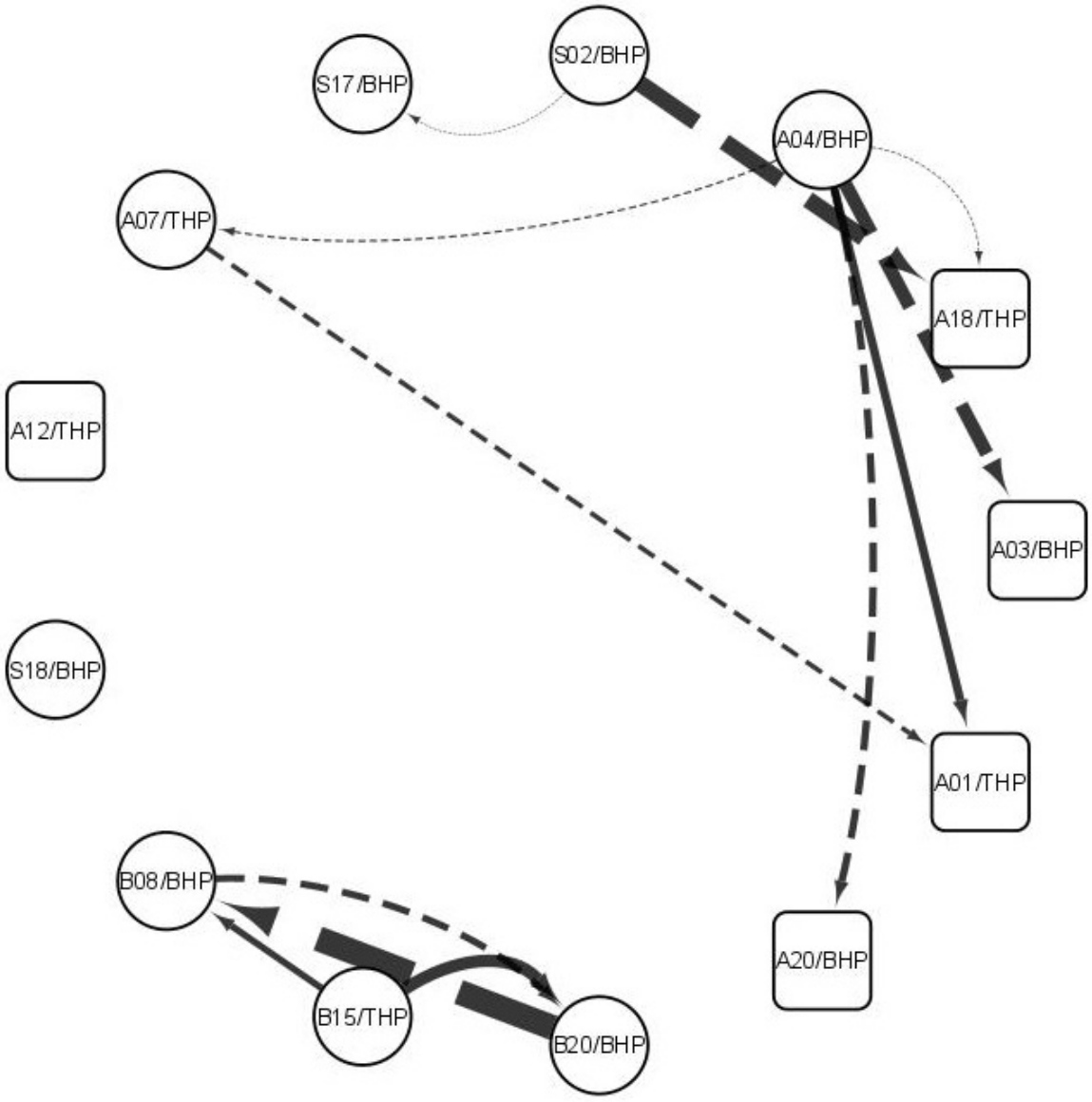


Figure 6.19: DAG of the full-scale test by the TCDF-algorithm. True connectivities illustrated by a solid arrow, false connectivities by a dashed arrow. Injection wells illustrated by a circle, and production wells by a rounded square.

In Table 6.8, we examine the distribution of true and false connectivities. The producing wells seem to be a large source of error, especially considering that twelve links were already removed, which were producer-producer links. More work needs to be done before the producing wells can be applied effectively for causal inference of interwell connectivity.

Table 6.8: Distribution of true and false connectivities for the various wells. First section of wells are injectors, and the second section is producers.

Well	**As stimulating**		**As target**		
	False	**True**	**False**	**True**	**Precision**
A04/BHP	3	1	0	0	25.0%
A07/THP	1	0	1	0	0.0%
B08/BHP	1	0	1	1	33.3%
B15/THP	2	0	0	0	100.0%
B20/BHP	1	0	1	1	33.3%
S02/BHP	2	0	0	0	0.0%
S17/BHP	0	0	1	0	0.0%
S18/BHP	0	0	0	0	0.0%
	False	**True**	**False**	**True**	
A01/THP	0	0	1	1	50.0%
A03/BHP	0	0	1	0	0.0%
A12/THP	0	0	0	0	0.0%
A18/THP	0	0	2	0	0.0%
A20/BHP	0	0	1	0	0.0%

To evaluate if the inferred strength of connectivity was indicative of the actual strength of connectivity, we again examined the distributions of the true and false connectivities, illustrated in Figure 6.20. We note that the three links of *True connectivity* are shown to be very close to the center of the distribution of *False connectivity*. The TCDF-algorithm generated very few links at a significance threshold of $\alpha = 10^{-3}$, and a visual inspection makes it very clear that there is little difference between the two distributions.

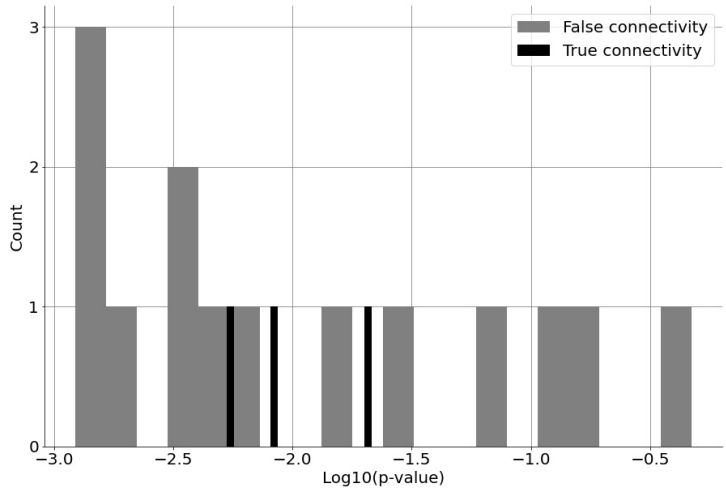


Figure 6.20: Histogram of the mean partial correlation distribution for the links labeled as *True connectivity* and *False connectivity*. *bins*=20.

To verify that the two distributions are indeed similar, we formulate a hypothesis test as seen in (6.4).

$$\begin{aligned} H_0 : &\quad \mu^T = \mu^F \\ H_1 : &\quad \mu^T \neq \mu^F \end{aligned} \tag{6.4}$$

Where μ^T and μ^F is the mean of the *True connectivity* and *False connectivity* distributions, in Figure 6.20. Using Mann-Whitney U-test, we obtain a p-value of 0.71. Since $0.71 > 0.05$, we do not have enough evidence to reject the null-hypothesis, and conclude that the base 10 log-transformed p-values in this test do not represent strength of connectivity.

Chapter 7

Conclusion

- Both models, PCMCI+ and TCDF, required significant changes to the original dataset to be applied effectively. Without modifications, identified connectivities were predominantly caused by external connectivities.
- The PCMCI+ algorithm identified and classified seven out of eight well-pairs with known connectivities, giving a sensitivity of 87.5%. The precision obtained was 18.9%, meaning that approximately one of five identified links are true connectivities. It was further found that the mean partial correlations obtained did not reflect the strength of connectivity between the identified links.
- The TCDF-algorithm, was able to identify and classify three out of eight well-pairs with known connectivities, giving a sensitivity of 37.5%. The precision obtained was 27.3%, meaning that approximately one of four identified links are true connectivities. It was further found that the base 10 log-transformed p-values obtained did not reflect the strength of connectivity between the identified links.
- The complexity of the original dataset was reduced effectively by splitting it into smaller datasets, which specifically targeted periods of change in the stimulating wells. The external connectivity of a common pressure source for the injection wells was not effectively reduced. The external connectivity of a shared production installation for the wells was effectively reduced.
- Using THP as a data source where BHP was unavailable was found to be an effective substitute. No notable difference was found for the results of the THP data compared to the BHP data.

7.1 Future work

- It was found that the size of the time window controlling the number of data points for each well-pair was a limiting factor. Larger window sizes provide more data points per dataset; however, it reduces the number of datasets. Applying a more dynamic selection of window sizes, fit for each well-pair, may improve the number of total data points.
- The PCMCI+ algorithm showed signs of improving by calculating the variance as a rolling window on the second-order difference of the full time-lagged partial

correlations. By applying an appropriate classification technique on the variance data, the results from the PCMCI+ algorithm may improve.

- Inclusion of other variables may improve the performance of the models. We highlight; the pressure data of the shared water supply for the water injectors and pressure data downstream of the THP. Other variables include; bottom hole temperature data, injection rates, and production rates if sampled on higher rates such as seconds or minutes.

Bibliography

Albertoni, A. and Lake, L. W (2003). "Inferring interwell connectivity only from well-rate fluctuations in waterfloods". In: *SPE Reservoir Evaluation & Engineering* 6.1, pp. 6–16. DOI: 10.2118/83381-PA.

Banerjee, A., Chitnis, U. B., Jadhav, S. L., Bhawalkar, J. S., and Chaudhury, S (2009). "Hypothesis testing, type I and type II errors". In: *Industrial psychiatry journal* 18.2, p. 127. DOI: 10.4103/0972-6748.62274.

Bellera, C. A., Julien, M., and Hanley, J. A (2010). "Normal approximations to the distributions of the Wilcoxon statistics: accurate to what N? Graphical insights". In: *Journal of Statistics Education* 18.2, pp. 1–17. DOI: 10.1080/10691898.2010.11889486.

Cheng, H., Vyatkin, V., Osipov, E., Zeng, P., and Yu, H (2020). "LSTM Based EFAST Global Sensitivity Analysis for Interwell Connectivity Evaluation Using Injection and Production Fluctuation Data". In: *IEEE Access* 8, pp. 67289–67299. DOI: 10.1109/ACCESS.2020.2985230.

Cheung, Y. W. and Lai, K. S (1995). "Lag order and critical values of the augmented Dickey–Fuller test". In: *Journal of Business & Economic Statistics* 13.3, pp. 277–280. DOI: 10.2307/1392187.

— (2009). "A systematic analysis of performance measures for classification tasks". In: *Information processing & management* 45.4, pp. 427–437. DOI: 10.1016/j.ipm.2009.03.002.

Dawid, A. P (1979). "Conditional independence in statistical theory". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 41.1, pp. 1–15. DOI: 10.1111/j.2517-6161.1979.tb01052.x.

De Winter, J. C (2013). "Using the Student's t-test with extremely small sample sizes". In: *Practical Assessment, Research, and Evaluation* 18.10. DOI: 10.7275/e4r6-dj05.

Dickey, D. A. and Fuller, W. A (1979). "Distribution of the estimators for autoregressive time series with a unit root". In: *Journal of the American statistical association* 74.366a, pp. 427–431. DOI: 10.1080/01621459.1979.10482531.

Dinh, A. V (2009). "Interwell Connectivity Tests in Waterflood Systems". PhD thesis. 660 Parrington Oval, Norman, OK 73019, USA: University of Oklahoma.

Du, S., Wang, R., Wei, C., Wang, Y., Zhou, Y., Wang, J., and Song, H (2020). "The connectivity evaluation among wells in reservoir utilizing machine learning methods". In: *IEEE Access* 8, pp. 47209–47219. DOI: 10.1109/ACCESS.2020.2976910.

Heffer, K. J., Fox, R. J., McGill, C. A., and Koutsabeloulis, N. C (1997). "Novel techniques show links between reservoir flow directionality, earth stress, fault structure and geomechanical changes in mature waterfloods". In: *SPE Journal* 2.2, pp. 91–98. DOI: 10.2118/30711-PA.

Hochreiter, S. and Schmidhuber, J (1997). "Long short-term memory". In: *Neural computation* 9.8, pp. 1735–1780. DOI: 10.1162/neco.1997.9.8.1735.

Hofer-Szabó, G., Rédei, M., and Szabó, L. E (1999). "On Reichenbach's common cause principle and Reichenbach's notion of common cause". In: *The British Journal for the Philosophy of Science* 50.3, pp. 377–399. DOI: 10.1093/bjps/50.3.377.

Jamali, A. and Ettehadtavakkol, A (2017). "Application of capacitance resistance models to determining interwell connectivity of large-scale mature oil fields". In: *Petroleum Exploration and Development* 44.1, pp. 132–138. DOI: 10.1016/S1876-3804(17)30017-4.

James, G., Witten, D., Hastie, T., and Tibshirani, R (2013). *An introduction to statistical learning*. Springer, New York, NY. DOI: 10.1007/978-1-4614-7138-7.

Jansen, F. E. and Kelkar, M. G (1997). "Application of wavelets to production data in describing inter-well relationships". In: *SPE Annual Technical Conference and Exhibition*. DOI: 10.2118/38876-MS.

Jayalakshmi, T. and Santhakumaran, A (2011). "Statistical normalization and back propagation for classification". In: *International Journal of Computer Theory and Engineering* 3.1. Retrieved from: ijcte.org, pp. 1793–8201.

Jensen, J. L., Lake, L. W., Al-Yousef, A., Weber, D., Liang, X., Edgar, T. F., Demiroren, N., and Kaviani, D (2007). *Interwell Connectivity and Diagnosis Using Correlation of Production and Injection Rate Data in Hydrocarbon Production*. Tech. rep. United States. DOI: 10.2172/924621.

Johnson, C. R., Greenkorn, R. A., and Woods, E. G (1966). "Pulse-testing: a new method for describing reservoir flow properties between wells". In: *Journal of Petroleum Technology* 18.12, pp. 1599–1604. DOI: 10.2118/1517-PA.

Kingma, D. P. and Ba, J (2014). *Adam: A method for stochastic optimization*. Retrieved from: arXiv:1412.6980.

Kočenda, E. and Černý, A (2015). *Elements of time series econometrics: An applied approach*. ISBN: 9788024623535. Charles University in Prague, Karolinum Press.

Kohavi, R (1995). "A study of cross-validation and bootstrap for accuracy estimation and model selection". In: *Ijcai* 14.2, pp. 1137–1145. DOI: 10.5555/1643031.1643047.

Livingston, E. H (2004). "Who was student and why do we care so much about his t-test?" In: *Journal of Surgical Research* 118.1, pp. 58–65. DOI: 10.1016/j.jss.2004.02.003.

Lumley, T., Diehr, P., Emerson, S., and Chen, L (2002). "The importance of the normality assumption in large public health data sets". In: *Annual review of public health* 23.1, pp. 151–169. DOI: 10.1146/annurev.publhealth.23.100901.140546.

Mann, H. B. and Whitney, D. R (1947). "On a test of whether one of two random variables is stochastically larger than the other". In: *The annals of mathematical statistics* 18.1, pp. 50–60. DOI: 10.1214/aoms/1177730491.

Manuca, R. and Savit, R (1996). "Stationarity and nonstationarity in time series analysis". In: *Physica D: Nonlinear Phenomena* 99.2-3, pp. 134–161. DOI: 10.1016/S0167-2789(96)00139-X.

Mata, C (2010). "Analytical Methods for Establishing Interwell Connectivity in a North Sea Field". MA thesis. Exhibition Rd, South Kensington, London SW7 2BU, Storbritannia: Imperial College London.

McKinney, Wes (2010). "Data Structures for Statistical Computing in Python". In: *Proceedings of the 9th Python in Science Conference*. Ed. by Stéfan van der Walt and Jarrod Millman, pp. 56–61. DOI: 10.25080/Majora-92bf1922-00a.

Mukaka, M. M (2012). "A guide to appropriate use of correlation coefficient in medical research". In: *Malawi medical journal* 24.3, pp. 69–71. DOI: 10.2307/1422689.

Nason, G. P (2006). *Statistics in volcanology*. Geological Society of London. Chap. Stationary and non-stationary time series. DOI: 10.1144/IAVCEI001.

Nauta, M., Bucur, D., and Seifert, C (2019). "Causal discovery with attention-based convolutional neural networks". In: *Machine Learning and Knowledge Extraction* 1.1, pp. 312–340. DOI: 10.3390/make1010019.

Panda, M. N. and Chopra, A. K (1998). "An integrated approach to estimate well interactions". In: *SPE India Oil and Gas Conference and Exhibition*. DOI: 10.2118/39563-MS.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12. Retrieved from: jmlr.org, pp. 2825–2830.

Pellet, J. P. and Elisseeff, A (2007). *Partial correlation and regression-based approaches to causal structure learning*. Tech. rep. Retrieved from: researchgate.net. IBM, North Castle, NY, USA, Research.

Ramakrishnan, T. S. and Raghuraman, B (2004). "A Method for Continuous Interpretation of Permanent Monitoring Pressure Data". In: *SPE Annual Technical Conference and Exhibition*. DOI: 10.2118/90910-MS.

Ramakrishnan, T.S., Thambynayagam, R.K.M., Tilke, P.G., and B, Raghuraman (2006). *Methods and Apparatus for Remote Real Time Oil Field Management*. US Patent No. 7,096,092. DC: U.S. Patent and Trademark Office.

Reback, Jeff, McKinney, Wes, jbrockmendel, Bossche, Joris Van den, Augspurger, Tom, Cloud, Phillip, Hawkins, Simon, gfyoung, Sinhrks, Roeschke, Matthew, Klein, Adam, Petersen, Terji, Tratner, Jeff, She, Chang, Ayd, William, Naveh, Shahar, patrick, Garcia, Marc, Schendel, Jeremy, Hayden, Andy, Saxton, Daniel, Jancauskas, Vytautas, Gorelli, Marco, Shadrach, Richard, McMaster, Ali, Battiston, Pietro, Seabold, Skipper, Dong, Kaiqi, chris-b1, and h-vetinari (Apr. 2021). *pandas-dev/pandas: Pandas 1.2.4*. Version v1.2.4. DOI: 10.5281/zenodo.4681666.

Runge, J (2020). "Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets". In: *Proceedings of Machine Learning Research* 124. Retrieved from: proceedings.mlr.press, pp. 1388–1397.

Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., and Sejdinovic, D (2019). "Detecting and quantifying causal associations in large nonlinear time series datasets". In: *Science Advances* 5.11. DOI: 10.1126/sciadv.aau4996.

Sak, H., Senior, A., and Beaufays, F (2014). *Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition*. Retrieved from: arXiv:1402.1128.

Saltelli, A., Tarantola, S., and Chan, K. S (1999). "A quantitative model-independent method for global sensitivity analysis of model output". In: *Technometrics* 41.1, pp. 39–56. DOI: 10.1080/00401706.1999.10485594.

Sayarpour, M., Zuluaga, E., Kabir, C. S., and Lake, L. W (2009). "The use of capacitance–resistance models for rapid estimation of waterflood performance and optimization". In: *Journal of Petroleum Science and Engineering* 69.3-4, pp. 227–238. DOI: 10.1016/j.petrol.2009.09.006.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., ..., and Ideker, T (2003). "Cytoscape: a software environment for integrated models of biomolecular

interaction networks". In: *Genome research* 13.11, pp. 2498–2504. DOI: 10.1101/gr.1239303.

Soeriawinata, T. and Kelkar, M (1999). "Reservoir management using production data". In: *SPE mid-continent operations symposium*. DOI: 10.2118/52224-MS.

Sola, J. and Sevilla, J (1997). "Importance of input data normalization for the application of neural networks to complex industrial problems". In: *IEEE Transactions on nuclear science* 44.3, pp. 1464–1468. DOI: 10.1109/23.589532.

Spearman, C (1987). "The proof and measurement of association between two things". In: *The American journal of psychology* 100.3/4, pp. 441–471. DOI: 10.2307/1422689.

Spirtes, P. and Glymour, C (1991). "An algorithm for fast recovery of sparse causal graphs". In: *Social science computer review* 9.1, pp. 62–72. DOI: 10.1177/089443939100900106.

Van Rossum, Guido and Drake, Fred L. (2009). *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace. ISBN: 1441412697.

Virtanen, Pauli, Gommers, Ralf, Oliphant, Travis E., Haberland, Matt, Reddy, Tyler, Cournapeau, David, Burovski, Evgeni, Peterson, Pearu, Weckesser, Warren, Bright, Jonathan, van der Walt, Stéfan J., Brett, Matthew, Wilson, Joshua, Millman, K. Jarrod, Mayorov, Nikolay, Nelson, Andrew R. J., Jones, Eric, Kern, Robert, Larson, Eric, Carey, C J, Polat, İlhan, Feng, Yu, Moore, Eric W., VanderPlas, Jake, Laxalde, Denis, Perktold, Josef, Cimrman, Robert, Henriksen, Ian, Quintero, E. A., Harris, Charles R., Archibald, Anne M., Ribeiro, Antônio H., Pedregosa, Fabian, van Mulbregt, Paul, and SciPy 1.0 Contributors (2020). "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nature Methods* 17, pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

Wilcoxon, F (1945). "Individual comparisons by ranking methods". In: *Biometrics Bulletin* 1.6, pp. 80–83. DOI: 10.2307/3001968.

Yousef, A., Gentil, P., Jensen, J., and Lake, L (2006). "A Capacitance Model To Infer Interwell Connectivity From Production and Injection Rate Fluctuations". In: *SPE Reservoir Evaluation & Engineering* 9, pp. 630–646. DOI: 10.2118/95322-PA.

Yousef, A. A., Jensen, J. L., and Lake, L. W (2008). "Integrated interpretation of interwell connectivity using injection and production fluctuations". In: *Mathematical Geosciences* 41.1, pp. 81–102. DOI: 10.1007/s11004-008-9189-x.

Appendix A

Tables of descriptive statistics

A.1 Small-scale modified

A.1.1 PCMCI+

Descriptive statistics

Table A.1: Descriptive statistics of the stimulating well A03/BHP acting on the target wells. *count* is the number of datapoints (partial correlations) for the respective wellpairs.

	A03/BHP	**A04/BHP**	**A20/BHP**	**B08/BHP**	**B15/THP**	**B20/BHP**
count	62	28	21	38	43	22
mean	-0.035	0.011	0.025	0.014	0.000	-0.061
std	0.045	0.120	0.075	0.129	0.119	0.121
min	-0.154	-0.209	-0.139	-0.221	-0.188	-0.287
25%	-0.055	-0.064	-0.006	-0.067	-0.117	-0.165
50%	-0.027	0.008	0.000	0.024	-0.004	-0.062
75%	-0.003	0.140	0.053	0.143	0.124	0.038
max	0.053	0.184	0.163	0.189	0.192	0.144

Table A.2: Descriptive statistics of the stimulating well A04/BHP acting on the target wells. *count* is the number of datapoints (partial correlations) for the respective wellpairs.

	A03/BHP	**A04/BHP**	**A20/BHP**	**B08/BHP**	**B15/THP**	**B20/BHP**
count	15	34	15	15	18	13
mean	0.048	-0.010	0.009	-0.056	0.000	0.097
std	0.074	0.062	0.051	0.094	0.117	0.170
min	-0.143	-0.180	-0.143	-0.170	-0.153	-0.195
25%	0.017	-0.046	-0.003	-0.145	-0.131	0.012
50%	0.046	-0.012	0.024	-0.060	0.018	0.146
75%	0.075	0.035	0.040	-0.008	0.069	0.164
max	0.165	0.106	0.061	0.159	0.174	0.390

Table A.3: Descriptive statistics of the stimulating well A20/BHP acting on the target wells. *count* is the number of datapoints (partial correlations) for the respective wellpairs.

	A03/BHP	**A04/BHP**	**A20/BHP**	**B08/BHP**	**B15/THP**	**B20/BHP**
count	32	32	62	32	37	29
mean	0.004	0.001	-0.054	0.009	-0.009	0.002
std	0.130	0.107	0.054	0.124	0.111	0.130
min	-0.201	-0.200	-0.233	-0.203	-0.189	-0.164
25%	-0.035	-0.065	-0.074	-0.081	-0.075	-0.144
50%	0.011	0.005	-0.048	-0.002	-0.028	0.023
75%	0.039	0.067	-0.022	0.123	0.060	0.146
max	0.452	0.169	0.087	0.237	0.214	0.165

Table A.4: Descriptive statistics of the stimulating well B08/BHP acting on the target wells. *count* is the number of datapoints (partial correlations) for the respective wellpairs.

	A03/BHP	**A04/BHP**	**A20/BHP**	**B08/BHP**	**B15/THP**	**B20/BHP**
count	55	66	56	91	37	10
mean	-0.007	-0.006	-0.023	-0.112	0.007	-0.022
std	0.128	0.094	0.105	0.088	0.109	0.112
min	-0.199	-0.217	-0.176	-0.476	-0.209	-0.172
25%	-0.143	-0.058	-0.140	-0.160	-0.053	-0.126
50%	-0.003	-0.004	-0.015	-0.094	0.020	-0.005
75%	0.142	0.046	0.023	-0.051	0.087	0.025
max	0.182	0.251	0.166	0.027	0.169	0.164

Table A.5: Descriptive statistics of the stimulating well B15/THP acting on the target wells. *count* is the number of datapoints (partial correlations) for the respective wellpairs.

	A03/BHP	**A04/BHP**	**A20/BHP**	**B08/BHP**	**B15/THP**	**B20/BHP**
count	52	58	48	19	75	5
mean	-0.034	0.015	-0.028	0.028	-0.159	-0.096
std	0.114	0.101	0.109	0.116	0.098	0.122
min	-0.205	-0.211	-0.201	-0.190	-0.483	-0.211
25%	-0.143	-0.046	-0.139	-0.044	-0.224	-0.166
50%	-0.041	0.007	-0.008	0.007	-0.143	-0.160
75%	0.055	0.066	0.033	0.061	-0.095	-0.029
max	0.175	0.263	0.173	0.323	0.114	0.084

Table A.6: Descriptive statistics of the stimulating well B20/BHP acting on the target wells. *count* is the number of datapoints (partial correlations) for the respective wellpairs.

	A03/BHP	**A04/BHP**	**A20/BHP**	**B08/BHP**	**B15/THP**	**B20/BHP**
count	75	74	59	27	33	105
mean	0.002	-0.004	-0.007	0.043	0.014	-0.108
std	0.116	0.093	0.113	0.114	0.119	0.111
min	-0.187	-0.192	-0.176	-0.225	-0.199	-0.493
25%	-0.085	-0.071	-0.118	-0.026	-0.057	-0.155
50%	-0.001	-0.003	-0.002	0.059	0.008	-0.094
75%	0.142	0.030	0.058	0.086	0.139	-0.040
max	0.190	0.220	0.227	0.255	0.213	0.140

P-values

Table A.7: Generated p-values for every well-pair combination of the small-scale modified test using the PCMCI+ algorithm. Rows represent the stimulating wells and columns represent the target wells.

	A03/ BHP	**A04/ BHP**	**A20/ BHP**	**B08/ BHP**	**B15/ THP**	**B20/ BHP**
A03/BHP	7.58E-12	1.86E-05	1.81E-05	5.84E-06	1.60E-07	4.77E-07
A04/BHP	3.05E-04	3.65E-07	6.10E-05	1.22E-04	1.45E-04	4.55E-01
A20/BHP	1.56E-05	4.62E-06	7.58E-12	1.44E-05	7.20E-07	6.62E-05
B08/BHP	2.91E-08	7.66E-12	4.60E-10	1.19E-16	3.28E-07	3.91E-03
B15/THP	1.39E-09	5.19E-10	6.43E-09	6.45E-04	5.28E-14	6.25E-02
B20/BHP	1.86E-11	3.89E-13	4.75E-10	4.13E-04	1.47E-05	5.84E-19

A.1.2 TCDF

Descriptive statistics

Table A.8: Descriptive statistics of the stimulating well A03/BHP acting on the target wells. *count* is the number of datapoints (attention scores) for the respective wellpairs.

	A03/BHP	**A04/BHP**	**A20/BHP**	**B08/BHP**	**B15/THP**	**B20/BHP**
count	62	47	32	46	51	42
mean	2.861	-0.116	0.348	-0.082	-0.07	-0.155
std	0.639	0.517	0.382	0.532	0.41	0.462
min	1.836	-1.564	-0.429	-1.853	-1.198	-1.379
25%	2.415	-0.434	0.055	-0.277	-0.268	-0.447
50%	2.678	-0.012	0.412	-0.004	-0.016	-0.149
75%	3.262	0.198	0.565	0.302	0.209	0.169
max	4.675	0.75	1.007	0.694	0.68	0.618

Table A.9: Descriptive statistics of the stimulating well A04/BHP acting on the target wells.

	A03/BHP	**A04/BHP**	**A20/BHP**	**B08/BHP**	**B15/THP**	**B20/BHP**
count	24	35	26	17	25	19
mean	0.535	2.471	0.572	0.465	0.289	0.337
std	0.267	0.675	0.319	0.282	0.352	0.379
min	0.22	1.213	0.019	-0.068	-1.101	-0.935
25%	0.316	1.991	0.358	0.308	0.315	0.284
50%	0.482	2.343	0.526	0.39	0.338	0.377
75%	0.683	2.773	0.75	0.595	0.435	0.504
max	1.19	4.317	1.374	1.015	0.638	0.796

Table A.10: Descriptive statistics of the stimulating well A20/BHP acting on the target wells.

	A03/BHP	**A04/BHP**	**A20/BHP**	**B08/BHP**	**B15/THP**	**B20/BHP**
count	36	45	61	38	50	40
mean	0.363	0.265	2.467	0.443	0.229	0.329
std	0.541	0.372	0.484	0.32	0.284	0.311
min	-1.168	-0.707	1.524	-0.289	-0.632	-0.19
25%	-0.018	0.094	2.057	0.256	0.071	0.157
50%	0.462	0.316	2.472	0.36	0.289	0.277
75%	0.661	0.511	2.663	0.686	0.411	0.436
max	1.56	1.001	3.605	1.349	0.687	1.396

Table A.11: Descriptive statistics of the stimulating well B08/BHP acting on the target wells.

	A03/BHP	**A04/BHP**	**A20/BHP**	**B08/BHP**	**B15/THP**	**B20/BHP**
count	79	71	78	91	44	19
mean	0.304	0.037	0.379	2.873	-0.177	0.211
std	0.606	0.482	0.532	0.722	0.395	0.768
min	-2.032	-1.1	-0.762	1.045	-1.39	-1.133
25%	0.05	-0.226	-0.034	2.164	-0.367	-0.061
50%	0.319	0.008	0.435	2.973	-0.223	0.142
75%	0.638	0.368	0.763	3.508	0.075	0.225
max	1.705	1.687	1.739	4.422	0.914	2.835

Table A.12: Descriptive statistics of the stimulating well B15/THP acting on the target wells.

	A03/BHP	**A04/BHP**	**A20/BHP**	**B08/BHP**	**B15/THP**	**B20/BHP**
count	64	61	63	22	75	5
mean	0.277	0.236	0.25	0.508	2.493	0.442
std	0.558	0.484	0.525	0.604	0.552	0.223
min	-1.588	-1.164	-1.218	-0.637	1.545	0.087
25%	-0.081	0.123	-0.136	0.112	2	0.353
50%	0.369	0.376	0.34	0.534	2.462	0.587
75%	0.608	0.544	0.631	0.759	2.976	0.591
max	1.76	1.044	1.308	1.686	4.052	0.591

Table A.13: Descriptive statistics of the stimulating well B20/BHP acting on the target wells.

	A03/BHP	**A04/BHP**	**A20/BHP**	**B08/BHP**	**B15/THP**	**B20/BHP**
count	89	81	89	29	41	105
mean	0.342	0.318	0.426	0.744	0.45	2.447
std	0.543	0.431	0.566	0.974	0.689	0.581
min	-1.937	-0.811	-1.001	-1.229	-0.497	0.972
25%	0.12	0.13	0.022	0.066	0.046	1.993
50%	0.505	0.426	0.605	0.353	0.457	2.32
75%	0.639	0.562	0.732	1.496	0.566	2.953
max	1.727	1.491	1.785	3.35	3.535	3.771

P-values

Table A.14: Generated p-values for every well-pair combination of the small-scale modified test using the TCDF-algorithm. Rows represent the stimulating wells and columns represent the target wells.

	A03/ BHP	**A04/ BHP**	**A20/ BHP**	**B08/ BHP**	**B15/ THP**	**B20/ BHP**
A03/BHP	1.00E+00	1.20E-09	4.81E-07	1.76E-09	2.57E-10	8.24E-09
A04/BHP	1.72E-05	1.00E+00	2.29E-05	2.10E-04	6.15E-06	6.59E-05
A20/BHP	1.13E-06	2.77E-09	1.00E+00	1.00E-07	3.78E-10	3.03E-08
B08/BHP	1.83E-12	2.84E-13	2.50E-12	1.00E+00	3.81E-09	1.11E-03
B15/THP	4.79E-11	5.84E-12	7.04E-12	1.83E-03	1.00E+00	2.16E-02
B20/BHP	1.20E-15	2.61E-14	2.18E-13	6.64E-02	2.96E-06	1.00E+00

A.2 Full-scale modified

A.2.1 PCMCI+

Descriptive statistics

Table A.15: Descriptive statistics of the stimulating well A01/THP acting on the target wells.

	A01/THP	**A03/BHP**	**A04/BHP**	**A07/THP**	**A12/THP**	**A18/THP**
count	51	21	21	25	20	12
mean	-0.043	-0.019	-0.039	0.021	0.004	0.047
std	0.085	0.104	0.103	0.117	0.160	0.088

A20/BHP	**B08/BHP**	**B15/THP**	**B20/BHP**	**S02/BHP**	**S17/BHP**	**S18/BHP**
8	21	33	18	23	26	27
-0.042	0.012	-0.007	0.046	-0.001	0.007	0.030
0.106	0.110	0.118	0.129	0.098	0.114	0.107

Table A.16: Descriptive statistics of the stimulating well A03/BHP acting on the target wells.

	A01/THP	**A03/BHP**	**A04/BHP**	**A07/THP**	**A12/THP**	**A18/THP**
count	25	66	27	26	18	21
mean	-0.018	-0.031	0.035	-0.019	0.002	-0.030
std	0.105	0.052	0.138	0.150	0.105	0.096
A20/BHP	**B08/BHP**	**B15/THP**	**B20/BHP**	**S02/BHP**	**S17/BHP**	**S18/BHP**
19	32	41	17	29	25	44
0.022	0.024	-0.016	-0.064	-0.009	-0.077	-0.033
0.113	0.135	0.119	0.114	0.116	0.095	0.118

Table A.17: Descriptive statistics of the stimulating well A04/BHP acting on the target wells.

	A01/THP	**A03/BHP**	**A04/BHP**	**A07/THP**	**A12/THP**	**A18/THP**
count	13	19	42	12	16	13
mean	-0.005	0.005	0.001	0.021	-0.011	-0.054
std	0.115	0.103	0.077	0.106	0.119	0.100
A20/BHP	**B08/BHP**	**B15/THP**	**B20/BHP**	**S02/BHP**	**S17/BHP**	**S18/BHP**
16	16	19	12	14	15	21
-0.006	-0.035	0.046	0.119	0.012	0.009	0.013
0.103	0.118	0.115	0.165	0.111	0.062	0.093

Table A.18: Descriptive statistics of the stimulating well A07/THP acting on the target wells.

	A01/THP	**A03/BHP**	**A04/BHP**	**A07/THP**	**A12/THP**	**A18/THP**
count	23	18	14	50	23	16
mean	-0.045	0.009	-0.056	-0.096	0.002	0.000
std	0.110	0.112	0.128	0.128	0.135	0.120
A20/BHP	**B08/BHP**	**B15/THP**	**B20/BHP**	**S02/BHP**	**S17/BHP**	**S18/BHP**
25	20	33	16	27	16	25
-0.024	-0.039	0.025	0.027	0.022	0.041	0.034
0.114	0.097	0.129	0.121	0.112	0.121	0.107

Table A.19: Descriptive statistics of the stimulating well A12/THP acting on the target wells.

	A01/THP	**A03/BHP**	**A04/BHP**	**A07/THP**	**A12/THP**	**A18/THP**
count	24	31	31	36	70	27
mean	0.014	-0.030	0.045	0.019	-0.045	0.037
std	0.111	0.114	0.088	0.120	0.088	0.091
A20/BHP	**B08/BHP**	**B15/THP**	**B20/BHP**	**S02/BHP**	**S17/BHP**	**S18/BHP**
31	29	45	23	37	43	34
0.002	-0.012	-0.030	0.009	0.008	0.020	0.017
0.101	0.116	0.105	0.121	0.114	0.115	0.121

Table A.20: Descriptive statistics of the stimulating well A18/THP acting on the target wells.

	A01/THP	**A03/BHP**	**A04/BHP**	**A07/THP**	**A12/THP**	**A18/THP**
count	26	28	28	30	24	75
mean	-0.019	0.001	0.010	0.009	0.034	-0.014
std	0.112	0.102	0.138	0.133	0.139	0.050
A20/BHP	**B08/BHP**	**B15/THP**	**B20/BHP**	**S02/BHP**	**S17/BHP**	**S18/BHP**
18	22	39	19	31	23	27
0.015	0.016	-0.005	0.042	-0.022	0.013	0.042
0.110	0.111	0.123	0.109	0.110	0.132	0.115

Table A.21: Descriptive statistics of the stimulating well A20/BHP acting on the target wells.

	A01/THP	**A03/BHP**	**A04/BHP**	**A07/THP**	**A12/THP**	**A18/THP**
count	13	25	25	30	28	15
mean	0.018	0.033	0.008	0.051	0.018	0.049
std	0.105	0.128	0.107	0.122	0.107	0.109
A20/BHP	**B08/BHP**	**B15/THP**	**B20/BHP**	**S02/BHP**	**S17/BHP**	**S18/BHP**
62	28	35	19	36	31	34
-0.052	0.012	0.022	-0.011	0.006	0.026	-0.047
0.055	0.114	0.114	0.148	0.084	0.127	0.124

Table A.22: Descriptive statistics of the stimulating well B08/BHP acting on the target wells.

	A01/THP	**A03/BHP**	**A04/BHP**	**A07/THP**	**A12/THP**	**A18/THP**
count	38	54	59	55	48	35
mean	-0.050	0.019	-0.017	0.021	0.012	0.011
std	0.118	0.119	0.110	0.125	0.136	0.134
A20/BHP	**B08/BHP**	**B15/THP**	**B20/BHP**	**S02/BHP**	**S17/BHP**	**S18/BHP**
58	96	36	13	51	46	60
-0.020	-0.106	0.002	0.003	0.000	-0.004	0.014
0.108	0.091	0.113	0.121	0.091	0.113	0.129

Table A.23: Descriptive statistics of the stimulating well B15/THP acting on the target wells.

	A01/THP	**A03/BHP**	**A04/BHP**	**A07/THP**	**A12/THP**	**A18/THP**
count	26	43	56	54	36	30
mean	-0.014	-0.028	0.013	-0.020	-0.021	-0.005
std	0.118	0.115	0.121	0.131	0.124	0.132
A20/BHP	**B08/BHP**	**B15/THP**	**B20/BHP**	**S02/BHP**	**S17/BHP**	**S18/BHP**
40	20	79	4	50	42	55
0.006	0.053	-0.147	0.014	-0.016	0.013	0.020
0.124	0.123	0.106	0.163	0.113	0.109	0.104

Table A.24: Descriptive statistics of the stimulating well B20/BHP acting on the target wells.

	A01/THP	**A03/BHP**	**A04/BHP**	**A07/THP**	**A12/THP**	**A18/THP**
count	45	68	68	66	58	37
mean	-0.033	0.011	0.020	0.027	-0.007	0.006
std	0.122	0.123	0.150	0.119	0.122	0.128
A20/BHP	**B08/BHP**	**B15/THP**	**B20/BHP**	**S02/BHP**	**S17/BHP**	**S18/BHP**
61	26	34	111	62	60	77
-0.015	0.036	0.006	-0.103	0.016	-0.004	0.030
0.118	0.107	0.117	0.104	0.101	0.125	0.117

Table A.25: Descriptive statistics of the stimulating well S02/BHP acting on the target wells.

	A01/THP	**A03/BHP**	**A04/BHP**	**A07/THP**	**A12/THP**	**A18/THP**
count	27	34	22	34	35	29
mean	0.009	-0.002	-0.012	0.040	0.020	0.018
std	0.104	0.106	0.123	0.148	0.115	0.106
A20/BHP	**B08/BHP**	**B15/THP**	**B20/BHP**	**S02/BHP**	**S17/BHP**	**S18/BHP**
32	19	32	19	63	17	35
0.012	-0.011	-0.004	-0.023	-0.126	0.082	0.000
0.119	0.098	0.191	0.159	0.087	0.113	0.110

Table A.26: Descriptive statistics of the stimulating well S17/BHP acting on the target wells.

	A01/THP	**A03/BHP**	**A04/BHP**	**A07/THP**	**A12/THP**	**A18/THP**
count	24	30	30	35	27	25
mean	-0.014	-0.022	-0.003	0.002	0.050	-0.036
std	0.110	0.119	0.094	0.115	0.111	0.124
A20/BHP	**B08/BHP**	**B15/THP**	**B20/BHP**	**S02/BHP**	**S17/BHP**	**S18/BHP**
31	30	43	27	29	77	42
-0.015	-0.005	0.025	0.083	-0.011	-0.087	0.050
0.110	0.116	0.126	0.097	0.104	0.109	0.139

Table A.27: Descriptive statistics of the stimulating well S18/BHP acting on the target wells.

	A01/THP	**A03/BHP**	**A04/BHP**	**A07/THP**	**A12/THP**	**A18/THP**
count	15	16	18	15	20	10
mean	-0.022	0.021	0.006	0.023	0.025	0.045
std	0.122	0.097	0.167	0.127	0.118	0.132
A20/BHP	**B08/BHP**	**B15/THP**	**B20/BHP**	**S02/BHP**	**S17/BHP**	**S18/BHP**
21	17	21	16	12	11	36
0.012	-0.070	-0.044	0.022	-0.002	0.064	-0.031
0.104	0.103	0.121	0.108	0.096	0.099	0.094

P-values

Table A.28: Generated p-values for all well-pair combinations with the A-wells as target wells. Obtained from the full-scale modified test using the PCMCI+ algorithm. Rows represent the stimulating wells and columns represent the target wells.

	A01/ THP	**A03/ BHP**	**A04/ BHP**	**A07/ THP**	**A12/ THP**	**A18/ THP**	**A20/ BHP**
A01/THP	5.15E-10	9.54E-06	9.54E-07	5.25E-06	7.08E-04	2.44E-03	7.81E-03
A03/BHP	8.34E-07	1.64E-12	1.52E-03	1.62E-04	5.34E-05	4.77E-06	3.36E-04
A04/BHP	1.71E-03	1.26E-04	1.65E-08	4.88E-04	3.05E-04	4.88E-04	3.05E-05
A07/THP	3.34E-06	7.63E-05	6.10E-04	7.56E-10	3.49E-04	5.80E-04	4.17E-07
A12/THP	3.02E-05	3.41E-06	1.77E-05	1.01E-05	3.56E-13	6.02E-05	3.41E-06
A18/THP	2.35E-05	5.86E-06	8.17E-05	5.31E-05	5.68E-04	5.28E-14	5.34E-04
A20/BHP	2.44E-04	1.20E-04	1.13E-06	2.83E-04	4.58E-05	8.36E-03	7.58E-12
B08/BHP	1.35E-07	1.11E-07	1.15E-10	1.64E-07	4.55E-06	4.53E-05	2.72E-10
B15/THP	1.67E-05	4.82E-08	6.97E-09	1.46E-09	1.79E-06	5.79E-05	5.73E-07
B20/BHP	4.83E-08	6.26E-10	1.89E-08	1.54E-09	2.22E-09	8.87E-06	3.36E-10
S02/BHP	1.53E-05	1.77E-06	3.34E-05	6.90E-04	7.77E-06	6.53E-06	1.11E-05
S17/BHP	1.19E-07	1.36E-05	3.18E-06	1.47E-06	1.33E-04	8.17E-06	3.41E-06
S18/BHP	1.22E-04	6.10E-05	7.69E-03	4.27E-03	4.83E-04	6.45E-02	4.10E-05

Table A.29: Generated p-values for all well-pair combinations with the B and S-wells as target wells. Obtained from the full-scale modified test using the PCMCI+ algorithm. Rows represent the stimulating wells and columns represent the target wells.

	B08/ BHP	**B15/ THP**	**B20/ BHP**	**S02/ BHP**	**S17/ BHP**	**S18/ BHP**
A01/THP	6.68E-05	2.98E-06	4.75E-03	2.38E-07	7.04E-05	7.37E-05
A03/BHP	1.59E-04	5.45E-08	1.53E-05	7.23E-06	1.19E-07	3.17E-08
A04/BHP	4.27E-04	3.92E-03	7.33E-01	6.10E-04	6.10E-05	6.68E-06
A07/THP	3.81E-06	1.27E-04	3.36E-03	2.91E-05	1.68E-03	8.80E-05
A12/THP	1.60E-05	1.41E-08	2.10E-05	8.41E-07	4.77E-07	1.25E-05
A18/THP	5.25E-05	1.20E-06	5.34E-05	4.53E-06	4.75E-04	1.78E-04
A20/BHP	6.13E-05	9.76E-06	4.20E-04	3.03E-07	1.68E-04	9.67E-07
B08/BHP	1.78E-17	1.31E-06	4.64E-03	1.57E-09	4.28E-08	1.72E-09
B15/THP	1.21E-03	1.15E-14	3.75E-01	3.17E-09	4.24E-07	6.62E-10
B20/BHP	1.07E-04	3.46E-06	6.15E-20	1.14E-10	1.72E-09	4.89E-10
S02/BHP	7.63E-06	8.57E-06	9.65E-04	5.17E-12	1.74E-02	2.39E-06
S17/BHP	1.13E-05	1.11E-05	5.32E-03	4.80E-06	2.46E-14	1.34E-04
S18/BHP	1.53E-05	1.34E-05	4.27E-04	9.77E-04	2.44E-02	1.68E-07

A.2.2 TCDF

Descriptive statistics

Table A.30: Descriptive statistics of the stimulating well A01/THP acting on the target wells.

	A01/THP	**A03/BHP**	**A04/BHP**	**A07/THP**	**A12/THP**	**A18/THP**
count	51	26	33	38	26	22
mean	3.119	0.672	0.282	0.221	0.656	0.863
std	0.665	0.550	0.483	0.438	0.553	0.759
A20/BHP	**B08/BHP**	**B15/THP**	**B20/BHP**	**S02/BHP**	**S17/BHP**	**S18/BHP**
11	34	43	35	36	39	40
0.432	0.212	0.240	0.110	0.394	0.048	0.233
0.440	0.404	0.313	0.375	0.480	0.403	0.472

Table A.31: Descriptive statistics of the stimulating well A03/BHP acting on the target wells.

	A01/THP	**A03/BHP**	**A04/BHP**	**A07/THP**	**A12/THP**	**A18/THP**
count	34	66	45	45	29	33
mean	0.613	3.422	0.293	0.490	0.599	0.744
std	0.411	0.910	0.601	0.500	0.392	0.460
A20/BHP	**B08/BHP**	**B15/THP**	**B20/BHP**	**S02/BHP**	**S17/BHP**	**S18/BHP**
30	44	50	41	44	47	56
0.667	0.248	0.252	0.362	0.490	0.226	0.375
0.402	0.582	0.426	0.544	0.468	0.475	0.502

Table A.32: Descriptive statistics of the stimulating well A04/BHP acting on the target wells.

	A01/THP	A03/BHP	A04/BHP	A07/THP	A12/THP	A18/THP
count	26	24	42	20	30	26
mean	0.700	0.786	3.063	0.404	0.295	0.622
std	0.572	0.595	0.613	0.719	0.423	0.585
A20/BHP	**B08/BHP**	**B15/THP**	**B20/BHP**	**S02/BHP**	**S17/BHP**	**S18/BHP**
26	17	25	19	20	21	29
0.705	0.309	0.226	0.335	0.578	0.468	0.189
0.613	0.279	0.463	0.617	0.330	0.425	0.586

Table A.33: Descriptive statistics of the stimulating well A07/THP acting on the target wells.

	A01/THP	A03/BHP	A04/BHP	A07/THP	A12/THP	A18/THP
count	36	30	22	50	36	34
mean	0.647	0.502	0.171	3.319	0.138	0.549
std	0.737	0.582	0.322	0.871	0.418	0.576
A20/BHP	**B08/BHP**	**B15/THP**	**B20/BHP**	**S02/BHP**	**S17/BHP**	**S18/BHP**
36	25	38	27	31	28	35
0.556	0.115	0.143	0.191	0.275	0.268	0.150
0.606	0.299	0.545	0.387	0.447	0.531	0.625

Table A.34: Descriptive statistics of the stimulating well A12/THP acting on the target wells.

	A01/THP	A03/BHP	A04/BHP	A07/THP	A12/THP	A18/THP
count	44	44	55	56	70	44
mean	0.577	0.350	0.493	0.546	2.943	0.534
std	0.555	0.587	0.530	0.564	0.574	0.601
A20/BHP	**B08/BHP**	**B15/THP**	**B20/BHP**	**S02/BHP**	**S17/BHP**	**S18/BHP**
46	46	60	38	53	55	56
0.518	0.277	0.413	0.260	0.274	0.557	0.499
0.612	0.448	0.644	0.384	0.466	0.972	0.616

Table A.35: Descriptive statistics of the stimulating well A18/THP acting on the target wells.

	A01/THP	A03/BHP	A04/BHP	A07/THP	A12/THP	A18/THP
count	35	40	49	54	39	75
mean	0.806	0.419	0.465	0.422	0.588	2.792
std	0.769	0.667	0.556	0.415	0.703	0.500
A20/BHP	**B08/BHP**	**B15/THP**	**B20/BHP**	**S02/BHP**	**S17/BHP**	**S18/BHP**
26	35	57	40	45	45	56
0.679	0.289	0.137	0.145	0.332	0.145	0.318
0.588	0.478	0.476	0.522	0.535	0.454	0.528

Table A.36: Descriptive statistics of the stimulating well A20/BHP acting on the target wells.

	A01/THP	**A03/BHP**	**A04/BHP**	**A07/THP**	**A12/THP**	**A18/THP**
count	19	32	41	47	33	23
mean	0.623	0.567	0.267	0.259	0.582	0.795
std	0.778	0.599	0.585	0.510	0.632	0.879
A20/BHP	**B08/BHP**	**B15/THP**	**B20/BHP**	**S02/BHP**	**S17/BHP**	**S18/BHP**
62	34	46	36	47	47	50
2.988	0.326	0.040	0.298	0.335	0.087	0.254
0.603	0.429	0.709	0.505	0.576	0.462	0.495

Table A.37: Descriptive statistics of the stimulating well B08/BHP acting on the target wells.

	A01/THP	**A03/BHP**	**A04/BHP**	**A07/THP**	**A12/THP**	**A18/THP**
count	79	78	70	75	78	65
mean	0.489	0.573	0.341	0.395	0.299	0.632
std	0.646	0.483	0.548	0.480	0.496	0.501
A20/BHP	**B08/BHP**	**B15/THP**	**B20/BHP**	**S02/BHP**	**S17/BHP**	**S18/BHP**
77	96	44	19	66	67	79
0.645	3.147	0.352	0.669	0.358	0.253	0.374
0.600	0.714	0.618	0.860	0.625	0.475	0.692

Table A.38: Descriptive statistics of the stimulating well B15/THP acting on the target wells.

	A01/THP	**A03/BHP**	**A04/BHP**	**A07/THP**	**A12/THP**	**A18/THP**
count	67	63	59	62	69	56
mean	0.278	0.427	0.227	0.346	0.158	0.467
std	0.603	0.579	0.491	0.524	0.513	0.524
A20/BHP	**B08/BHP**	**B15/THP**	**B20/BHP**	**S02/BHP**	**S17/BHP**	**S18/BHP**
61	21	79	5	56	60	73
0.407	0.379	3.175	-0.378	0.288	0.089	0.205
0.521	0.883	0.700	0.382	0.536	0.532	0.582

Table A.39: Descriptive statistics of the stimulating well B20/BHP acting on the target wells.

	A01/THP	**A03/BHP**	**A04/BHP**	**A07/THP**	**A12/THP**	**A18/THP**
count	91	87	78	82	90	77
mean	0.296	0.485	0.176	0.261	0.169	0.366
std	0.614	0.589	0.644	0.533	0.568	0.572
A20/BHP	**B08/BHP**	**B15/THP**	**B20/BHP**	**S02/BHP**	**S17/BHP**	**S18/BHP**
86	27	40	111	76	77	93
0.451	1.062	0.209	3.097	0.320	0.278	0.400
0.619	1.271	0.982	0.747	0.634	0.791	0.747

Table A.40: Descriptive statistics of the stimulating well S02/BHP acting on the target wells.

	A01/THP	**A03/BHP**	**A04/BHP**	**A07/THP**	**A12/THP**	**A18/THP**
count	40	40	35	43	46	36
mean	0.478	0.556	0.306	0.414	0.430	0.826
std	0.526	0.590	0.438	0.517	0.604	0.692
A20/BHP	**B08/BHP**	**B15/THP**	**B20/BHP**	**S02/BHP**	**S17/BHP**	**S18/BHP**
41	30	39	34	63	28	45
0.613	0.351	0.247	0.273	3.067	0.566	0.355
0.520	0.391	0.338	0.408	0.794	0.668	0.661

Table A.41: Descriptive statistics of the stimulating well S17/BHP acting on the target wells.

	A01/THP	**A03/BHP**	**A04/BHP**	**A07/THP**	**A12/THP**	**A18/THP**
count	52	46	45	49	56	44
mean	0.660	0.587	0.367	0.336	0.268	0.590
std	0.687	0.697	0.340	0.561	0.488	0.481
A20/BHP	**B08/BHP**	**B15/THP**	**B20/BHP**	**S02/BHP**	**S17/BHP**	**S18/BHP**
52	39	52	39	35	77	56
0.627	0.288	0.262	0.259	0.342	3.018	0.590
0.637	0.428	0.433	0.439	0.730	0.755	0.875

Table A.42: Descriptive statistics of the stimulating well S18/BHP acting on the target wells.

	A01/THP	**A03/BHP**	**A04/BHP**	**A07/THP**	**A12/THP**	**A18/THP**
count	27	25	22	23	26	22
mean	0.451	0.473	0.449	0.328	0.587	0.466
std	0.573	0.425	0.442	0.554	0.369	0.384
A20/BHP	**B08/BHP**	**B15/THP**	**B20/BHP**	**S02/BHP**	**S17/BHP**	**S18/BHP**
27	18	27	19	15	16	36
0.604	0.362	0.368	0.400	0.551	0.498	3.036
0.503	0.280	0.394	0.376	0.332	0.391	0.601

P-values

Table A.43: Generated p-values for all well-pair combinations with the A-wells as target wells. Obtained from the full-scale modified test using the TCDF-algorithm. Rows represent the stimulating wells and columns represent the target wells.

	A01/ THP	**A03/ BHP**	**A04/ BHP**	**A07/ THP**	**A12/ THP**	**A18/ THP**	**A20/ BHP**
A01/THP	1.00E+00	1.70E-02	1.93E-06	6.25E-08	2.60E-02	2.74E-01	2.92E-03
A03/BHP	4.68E-05	1.00E+00	4.26E-08	8.27E-07	1.41E-02	1.38E-03	6.27E-02
A04/BHP	5.35E-03	3.16E-02	1.00E+00	1.02E-01	1.44E-06	2.76E-02	6.87E-02
A07/THP	2.85E-03	1.80E-03	4.00E-05	1.00E+00	2.11E-07	2.09E-03	6.39E-04
A12/THP	3.76E-01	7.82E-04	8.61E-06	2.11E-06	1.00E+00	4.53E-02	2.25E-02
A18/THP	3.45E-02	7.94E-05	1.12E-07	6.14E-09	7.30E-03	1.00E+00	1.20E-03
A20/BHP	2.00E-02	1.82E-04	7.80E-08	1.85E-09	9.81E-03	1.51E-01	1.00E+00
B08/BHP	2.56E-06	2.27E-06	6.44E-11	3.83E-10	7.20E-13	1.71E-08	4.48E-05
B15/THP	7.87E-09	2.83E-08	1.26E-11	9.66E-08	5.41E-12	9.18E-10	5.52E-09
B20/BHP	1.38E-09	9.51E-09	5.40E-13	8.74E-11	6.42E-15	5.20E-11	2.77E-10
S02/BHP	5.33E-03	1.86E-02	3.47E-07	2.11E-06	6.14E-05	3.36E-01	1.32E-04
S17/BHP	2.45E-03	7.76E-03	3.63E-09	7.72E-08	2.48E-09	6.61E-02	1.55E-04
S18/BHP	2.66E-03	7.83E-05	3.49E-05	4.36E-05	1.84E-05	4.38E-04	1.45E-03

Table A.44: Generated p-values for all well-pair combinations with the B and S-wells as target wells. Obtained from the full-scale modified test using the TCDF-algorithm. Rows represent the stimulating wells and columns represent the target wells.

	B08/ BHP	**B15/ THP**	**B20/ BHP**	**S02/ BHP**	**S17/ BHP**	**S18/ BHP**
A01/THP	1.83E-07	3.16E-08	1.91E-07	3.30E-05	3.59E-08	2.32E-07
A03/BHP	4.08E-09	3.58E-09	1.13E-07	7.08E-08	4.03E-09	9.10E-09
A04/BHP	2.10E-04	3.23E-05	4.19E-04	6.19E-03	8.29E-04	1.30E-05
A07/THP	6.15E-06	7.92E-08	5.50E-06	4.74E-06	5.54E-06	1.71E-05
A12/THP	2.98E-09	6.26E-10	3.87E-08	1.13E-08	7.06E-04	3.84E-07
A18/THP	1.24E-07	1.30E-10	1.14E-07	1.19E-08	4.00E-08	3.04E-08
A20/BHP	2.00E-07	2.36E-07	8.40E-08	1.13E-06	1.28E-08	8.84E-10
B08/BHP	1.00E+00	4.85E-05	7.88E-03	5.24E-08	1.63E-11	5.71E-09
B15/THP	4.81E-03	1.00E+00	3.98E-02	4.03E-09	4.54E-10	3.45E-11
B20/BHP	2.82E-01	2.02E-03	1.00E+00	1.77E-10	3.60E-10	4.02E-10
S02/BHP	6.19E-06	1.13E-07	7.45E-07	1.00E+00	1.14E-02	5.28E-05
S17/BHP	1.31E-07	6.97E-10	3.90E-07	4.52E-06	1.00E+00	9.96E-04
S18/BHP	9.82E-05	4.97E-05	6.59E-05	3.46E-02	6.73E-04	1.00E+00

Appendix B

Data descriptions and figures

A-04 & A-01

Figure B.1 show the pressure response between the stimulating production well A01/THP and the target injection well A04/BHP. The maximum time delay is approximately 1 hour. We were not able to visually identify a similar type of pressure response with A04/BHP as the stimulating well and A01/THP as the target well.

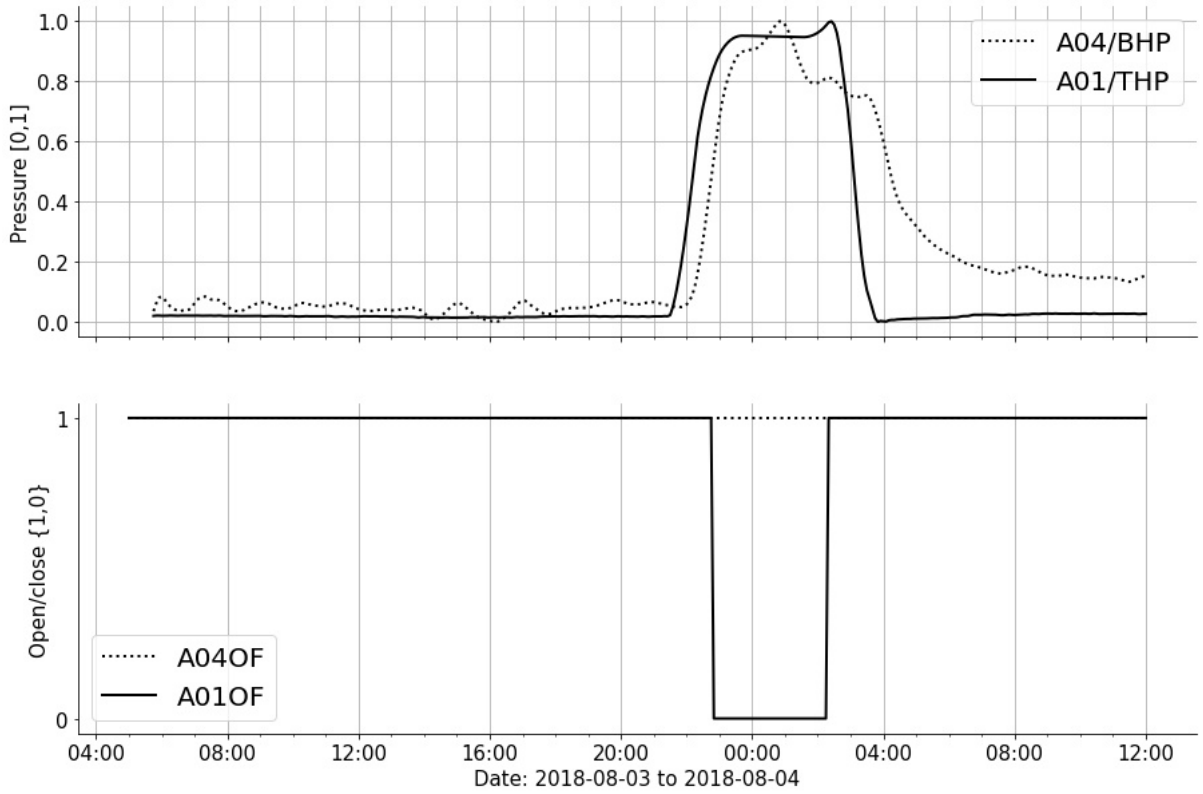


Figure B.1: Zoomed in view of a pressure response between A01/THP and A04/BHP. Pressure is normalized on a [0,1]-range within this period. Data is smoothed with a moving average of 10 time steps.

A-07 & A-18

In Figure B.2, the pressure response between the stimulating injection well A07/THP and the target production well A18/THP can be seen. It is uncertain whether the pressure in A18/THP responds with a time delay of approximately 3 hours, or a near instantaneous response when A07/THP reaches a certain pressure threshold. We were not able to identify a reciprocal link with A18/THP as the stimulating well.

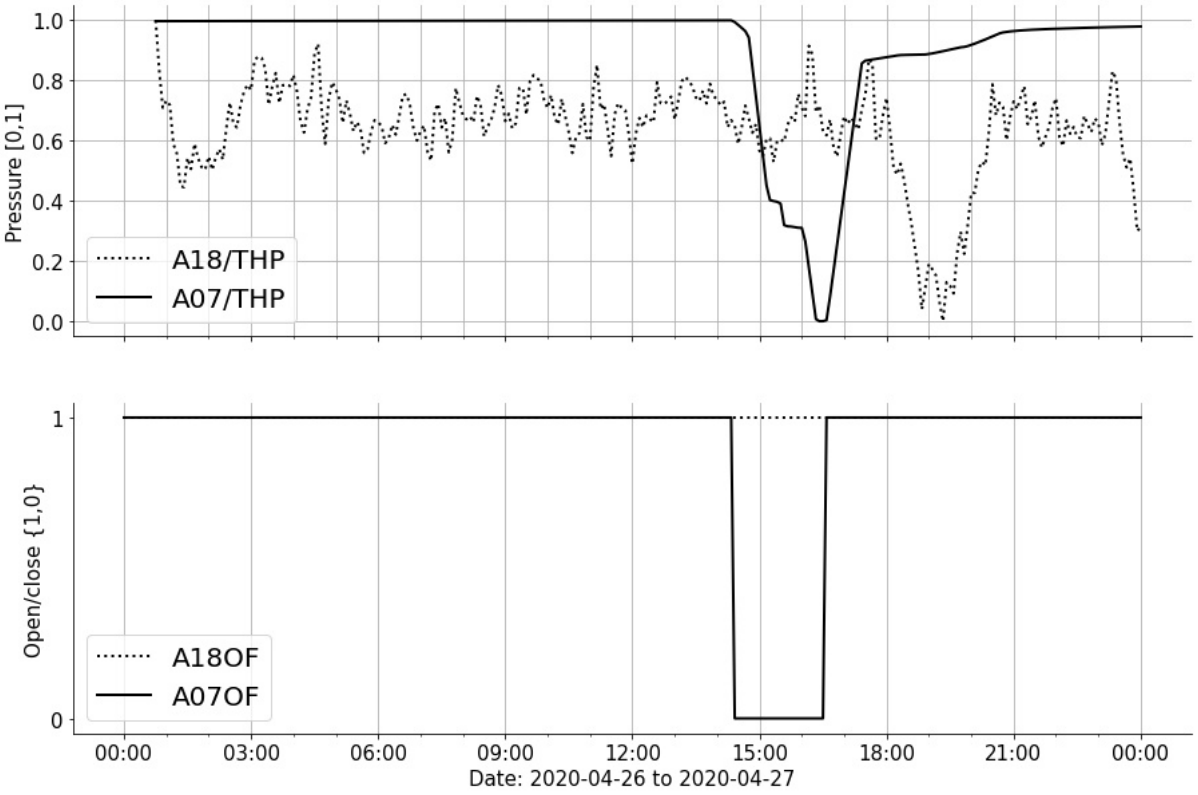


Figure B.2: Zoomed in view of a pressure response between A07/THP and A18/THP. Pressure is normalized on a [0,1]-range within this period. Data is smoothed with a moving average of 10 time steps.

B-08 & B-15

In Figure B.3, the pressure response between the stimulating injection well B08/BHP and the target injection well B15/THP is shown. The time delay of the response in B15/THP is approximately 6 hours. Interestingly, the pressure response in B15/THP appears to be weaker when the pressure in B08/BHP is increasing, as opposed to a decrease in the pressure of B08/BHP.

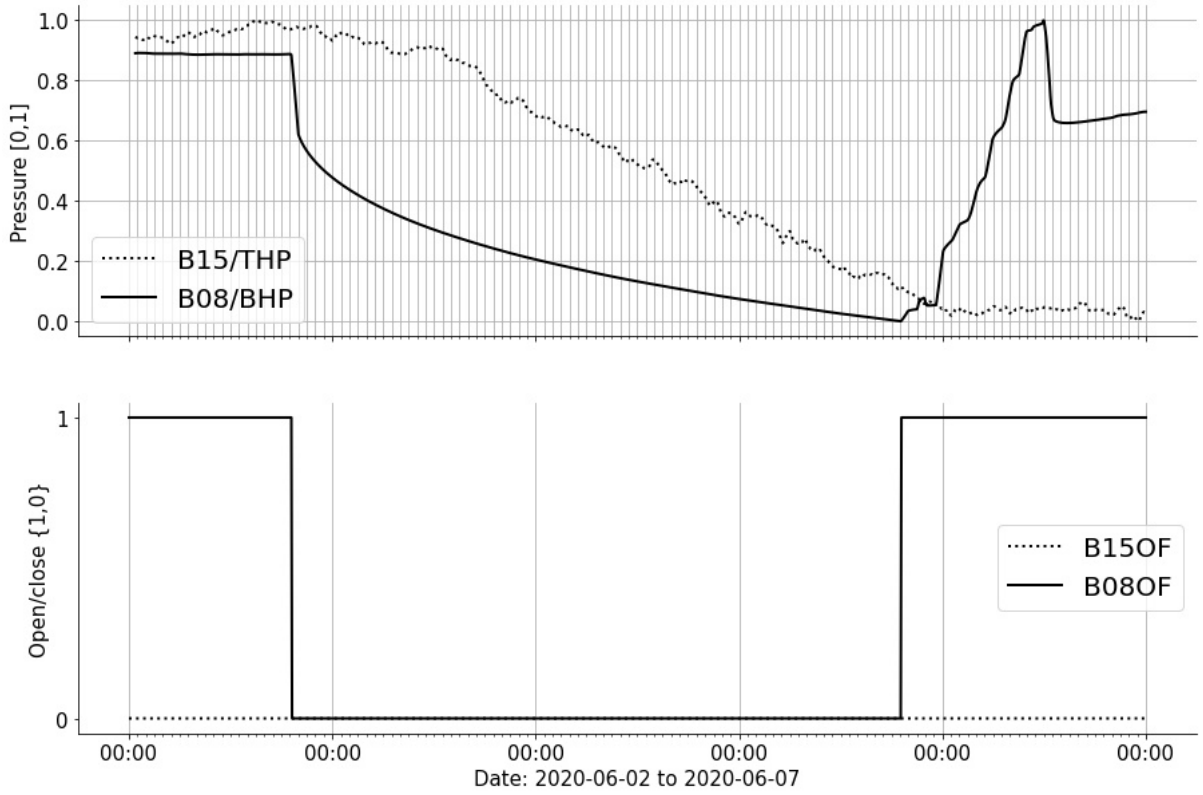


Figure B.3: Zoomed in view of a pressure response between B08/BHP and B15/THP. Pressure is normalized on a [0,1]-range within this period. Data is smoothed with a moving average of 10 time steps.

In Figure B.4, the pressure response between the stimulating injection well B15/THP and the target injection well B08/BHP is shown, along with the injection well B20/BHP. The wells B15/THP and B20/BHP have a strong coupling in that they are often closed and opened in the same period. This makes it difficult to separate these as stimulating wells. The pressure response in B08/BHP is near instantaneous for the coupled stimulus from B15/THP and B20/BHP.

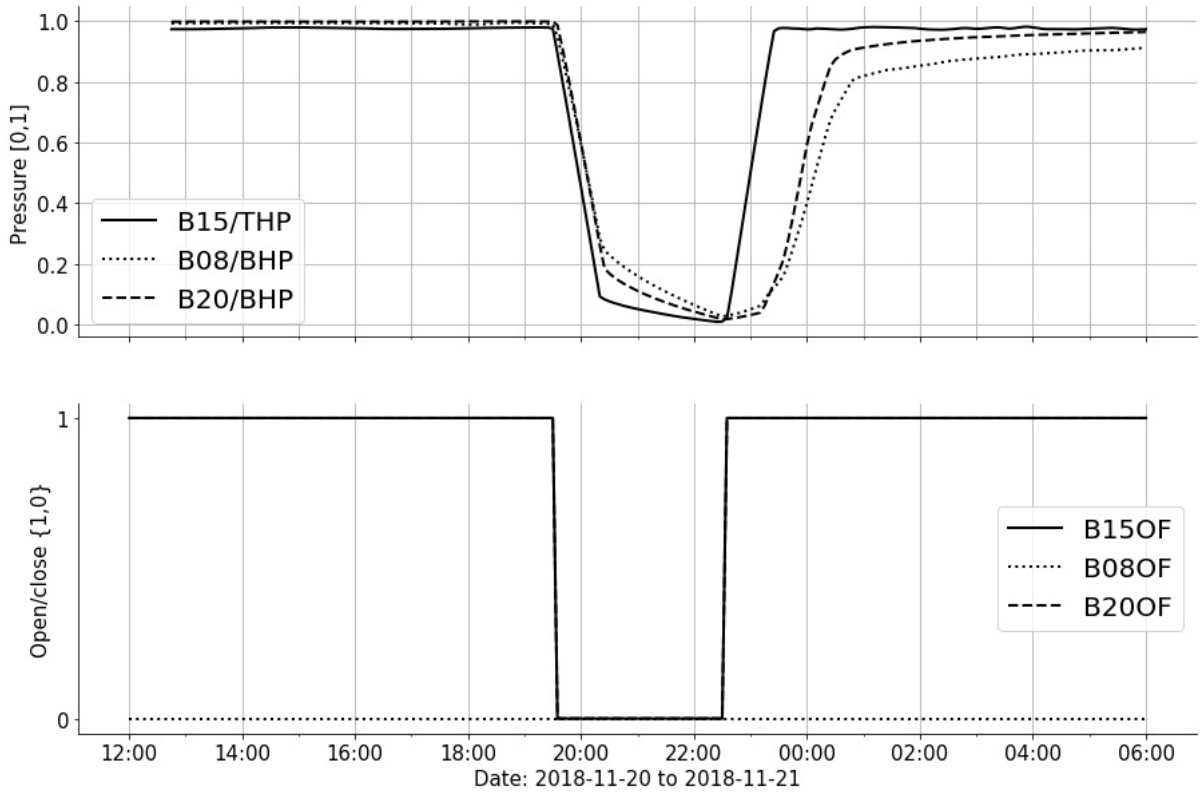


Figure B.4: Zoomed in view of a pressure response between B15/THP and B08/BHP. Pressure is normalized on a [0,1]-range within this period. Data is smoothed with a moving average of 10 time steps.

B-20 & B-15

In Figure B.5, the reciprocal pressure response between the stimulating injection well B20/BHP and the target injection well B15/THP can be seen. This includes the injection well S18/BHP, which closes in the same period as B20/BHP. Similarly to the B15/THP and B20/BHP coupling, there is a strong coupling between B20/BHP and B08/BHP. Since there is a known connectivity between B08/BHP and B15/THP, we selected one of very few periods when B08/BHP did not close at the same time as B20/BHP. The maximum time delay is approximately 2 to 3 hours.

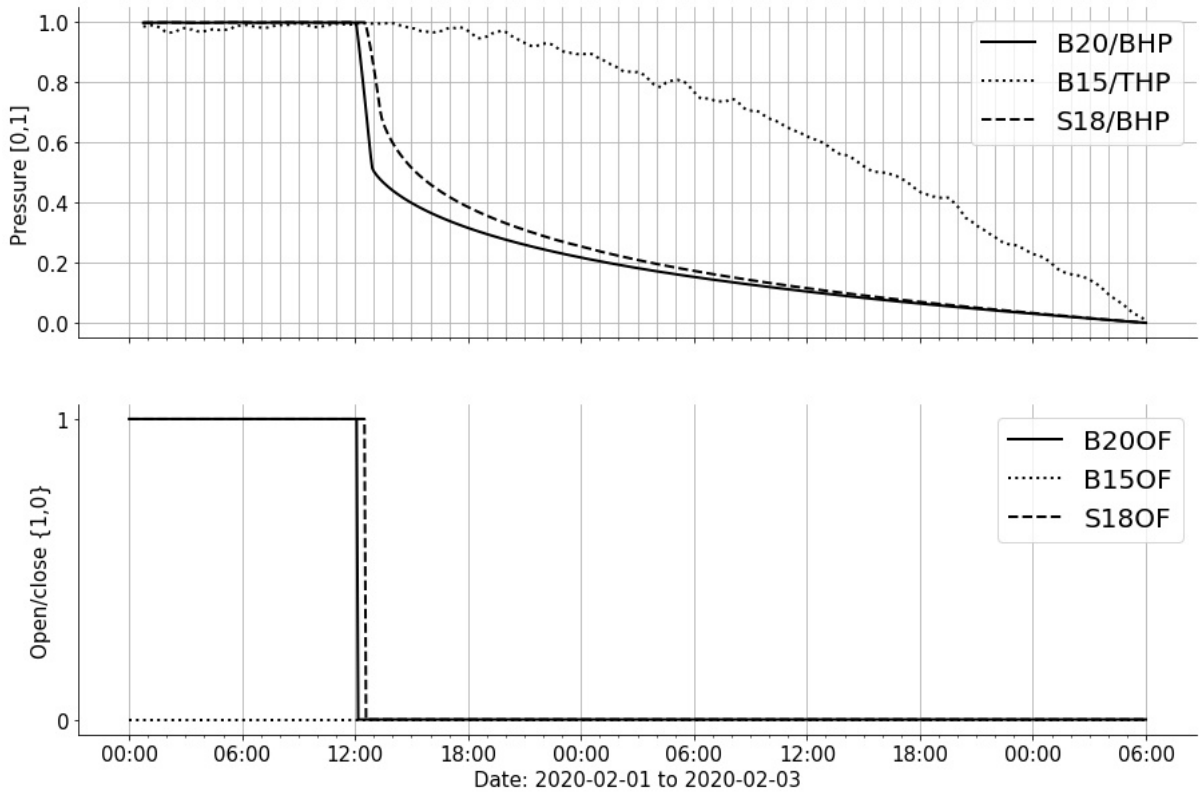


Figure B.5: Zoomed in view of a pressure response between B20/BHP and B15/THP. Pressure is normalized on a [0,1]-range within this period. Data is smoothed with a moving average of 10 time steps.

S-17 & A-12

In Figure B.6, a very weak pressure response between the stimulating injection well S17/BHP and the target production well A12/THP can be seen. The maximum time delay is approximately 2.5 hours. We were not able to identify a reciprocal link with A12/THP as the stimulating well.

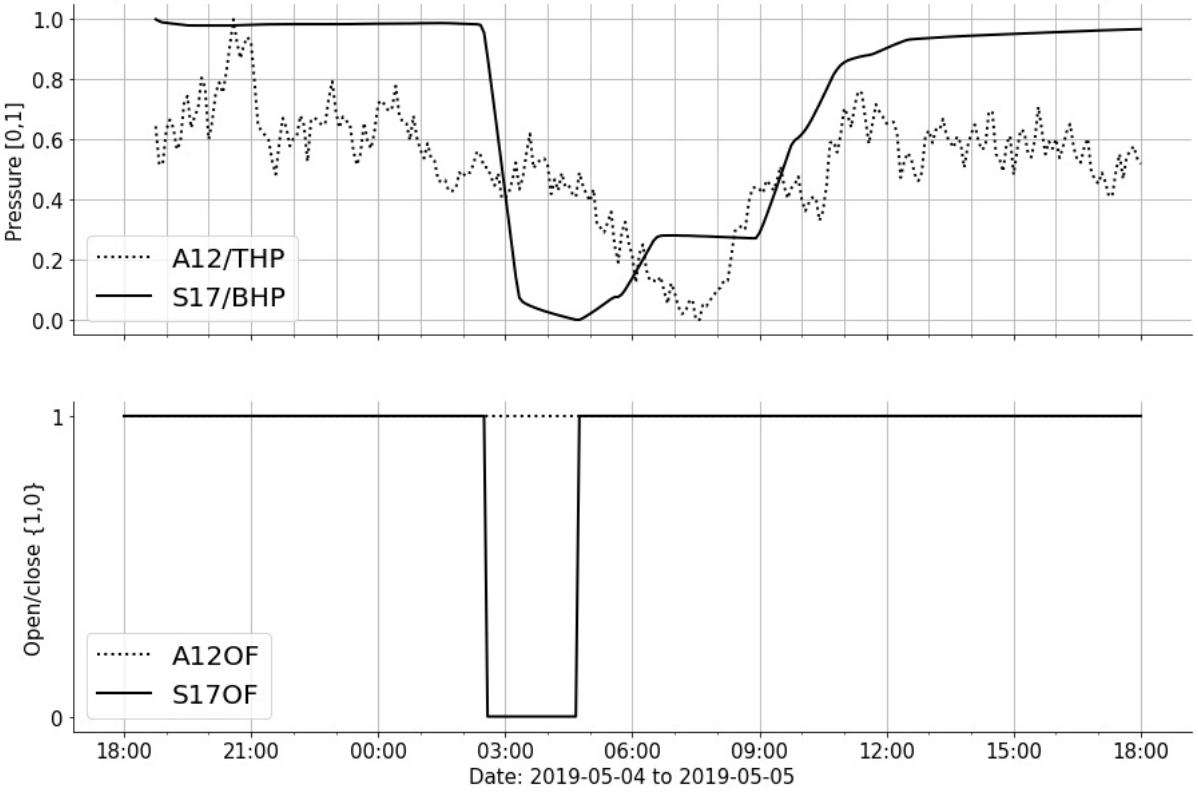


Figure B.6: Zoomed in view of a pressure response between S17/BHP and A12/THP. Pressure is normalized on a [0,1]-range within this period. Data is smoothed with a moving average of 10 time steps.

S-18 & B-20

In Figure B.7, a pressure response between the stimulating injection well S18/BHP and the target injection well B20/BHP can be seen. The maximum time delay is approximately 1.5-2 hours. During the opening of S18/BHP and the subsequent pressure increase, the pressure response in B20/BHP seem to be triggered at a certain pressure threshold, with the response occurring approximately 0.5 to 1 hour later.

In Figure B.8, the reciprocal pressure response of the stimulating injection well B20/BHP and the target injection well S18/BHP can be seen. The maximum time delay is approximately 0.5 hours. Similarly to S18/BHP acting on B20/BHP, there is a pressure threshold that needs to be hit before the pressure responds after 0.5 to 1 hour in S18/BHP.

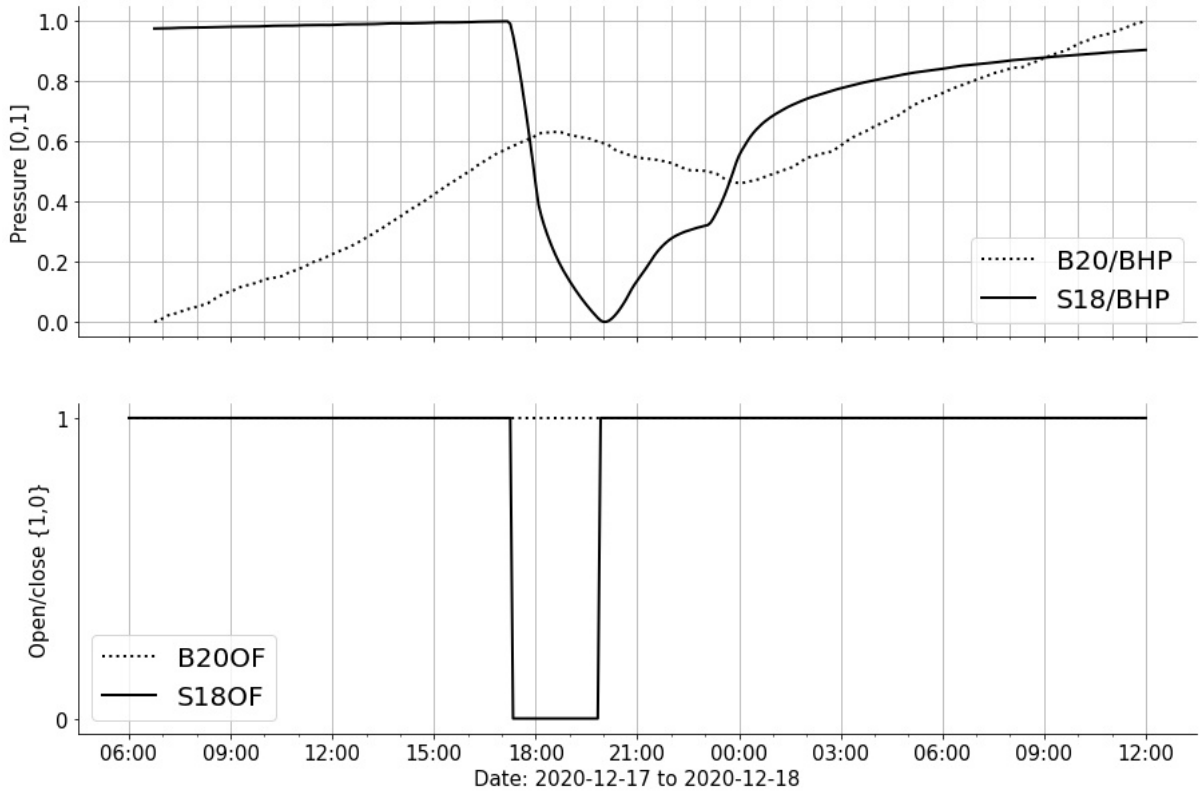


Figure B.7: Zoomed in view of a pressure response between S18/BHP and B20/BHP. Pressure is normalized on a [0,1]-range within this period. Data is smoothed with a moving average of 10 time steps.

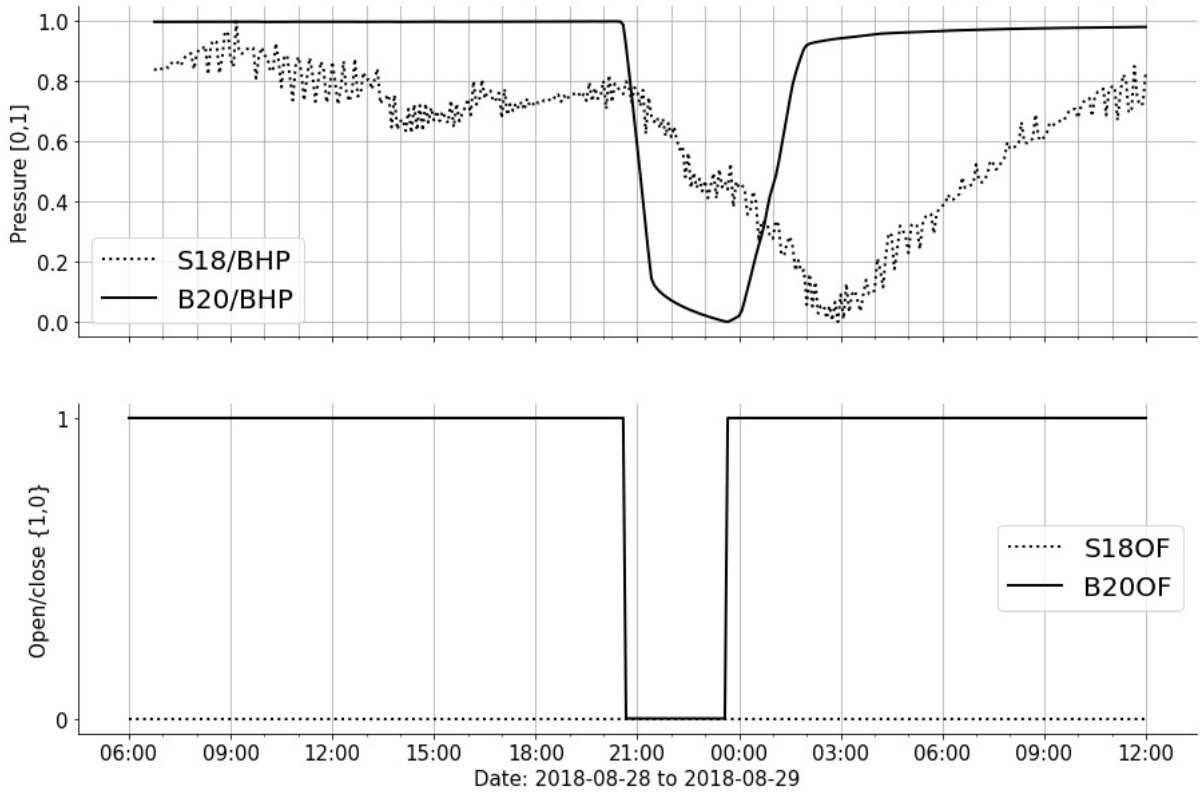


Figure B.8: Zoomed in view of a pressure response between B20/BHP and S18/BHP. Pressure is normalized on a [0,1]-range within this period. Data is smoothed with a moving average of 10 time steps.

Appendix C

Method demonstration raw output

C.1 PCMCI+ no hidden confounder

```
##
## Step 1: PC1 algorithm with lagged conditions
##

Parameters:
independence test = par_corr
tau_min = 1
tau_max = 10
pc_alpha = [0.01]
max_conds_dim = None
max_combinations = 1



## Resulting lagged parent (super)sets:

    Variable $X^1$ has 2 parent(s):
        ($X^1$ -1): max_pval = 0.00000, min_val = 0.601
        ($X^3$ -5): max_pval = 0.00885, min_val = -0.156

    Variable $X^2$ has 2 parent(s):
        ($X^2$ -1): max_pval = 0.00000, min_val = 0.734
        ($X^1$ -3): max_pval = 0.00000, min_val = 0.365

    Variable $X^3$ has 2 parent(s):
        ($X^3$ -2): max_pval = 0.00000, min_val = 0.614
        ($X^1$ -5): max_pval = 0.00000, min_val = 0.335

##
## Step 2: PC algorithm with contemp. conditions and MCI tests
##

Parameters:
```

```
independence test = par_corr
tau_min = 0
tau_max = 10
pc_alpha = 0.01
contemp_collider_rule = majority
conflict_resolution = True
reset_lagged_links = False
max_conds_dim = None
max_conds_py = None
max_conds_px = None
max_conds_px_lagged = None
fdr_method = none

## Significant links at alpha = 0.01:

    Variable $X^1$ has 1 link(s):
        ($X^1$ -1): pval = 0.00000 | val = 0.613

    Variable $X^2$ has 2 link(s):
        ($X^2$ -1): pval = 0.00000 | val = 0.614
        ($X^1$ -3): pval = 0.00000 | val = 0.375

    Variable $X^3$ has 2 link(s):
        ($X^3$ -2): pval = 0.00000 | val = 0.508
        ($X^1$ -5): pval = 0.00000 | val = 0.329
```

C.2 PCMCI+ hidden confounder

```
##
## Step 1: PC1 algorithm with lagged conditions
##

Parameters:
independence test = par_corr
tau_min = 1
tau_max = 10
pc_alpha = [0.01]
max_conds_dim = None
max_combinations = 1



## Resulting lagged parent (super)sets:

    Variable $X^2$ has 1 parent(s):
        ($X^2$ -1): max_pval = 0.00000, min_val = 0.693

    Variable $X^3$ has 1 parent(s):
        ($X^3$ -2): max_pval = 0.00000, min_val = 0.635
```

```
##
## Step 2: PC algorithm with contemp. conditions and MCI tests
##

Parameters:

independence test = par_corr
tau_min = 0
tau_max = 10
pc_alpha = 0.01
contemp_collider_rule = majority
conflict_resolution = True
reset_lagged_links = False
max_conds_dim = None
max_conds_py = None
max_conds_px = None
max_conds_px_lagged = None
fdr_method = none

## Significant links at alpha = 0.01:

    Variable $X^2$ has 2 link(s):
        ($X^2$ -1): pval = 0.00000 | val = 0.696
        ($X^3$ 0): pval = 0.00186 | val = 0.186 | unoriented link

    Variable $X^3$ has 2 link(s):
        ($X^3$ -2): pval = 0.00000 | val = 0.560
        ($X^2$ 0): pval = 0.00186 | val = 0.186 | unoriented link

## Ambiguous triples:

    ($X^3$ -2) --> $X^3$ o-o $X^2$
    ($X^2$ -1) --> $X^2$ o-o $X^3$
```

C.3 TCDF no hidden confounder

```
Arguments: Namespace(cuda=False, data=['data/tcdf_demo_1.csv'],
    dilation_coefficient=10, epochs=100, ground_truth=None, hidden_layers=0,
    kernel_size=10, learning_rate=0.01, log_interval=500, optimizer='Adam',
    plot=True, seed=1111, significance=1.0)

 Dataset: tcdf_demo_1.csv

 Analysis started for target: X^1
Epoch: 1 [1%]   Loss: 2.625633
Epoch: 100 [100%]  Loss: 0.935598
Potential causes: [0]
Validated causes: [0]
```

```
 Analysis started for target: X^2
Epoch: 1 [1%]   Loss: 5.410668
Epoch: 100 [100%]  Loss: 1.234993
Potential causes: [1, 0]
Validated causes: [1, 0]

 Analysis started for target: X^3
Epoch: 1 [1%]   Loss: 3.954392
Epoch: 100 [100%]  Loss: 1.153275
Potential causes: [2, 0]
Validated causes: [2, 0]

==================Results for tcdf_demo_1.csv
   ==================================
X^1 causes X^1 with a delay of 1 time steps.
X^2 causes X^2 with a delay of 1 time steps.
X^1 causes X^2 with a delay of 3 time steps.
X^3 causes X^3 with a delay of 2 time steps.
X^1 causes X^3 with a delay of 5 time steps.
================================================================================
```

C.4 TCDF hidden confounder

```
Arguments: Namespace(cuda=False, data=['data/tcdf_demo_2.csv'],
   dilation_coefficient=10, epochs=100, ground_truth=None, hidden_layers=0,
   kernel_size=10, learning_rate=0.01, log_interval=500, optimizer='Adam',
   plot=True, seed=1111, significance=1.0)

 Dataset: tcdf_demo_2.csv

 Analysis started for target: X^2
Epoch: 1 [1%]   Loss: 4.860878
Epoch: 100 [100%]  Loss: 1.420413
Potential causes: [0]
Validated causes: [0]

 Analysis started for target: X^3
Epoch: 1 [1%]   Loss: 3.277262
Epoch: 100 [100%]  Loss: 1.148035
Potential causes: [1]
Validated causes: [1]

==================Results for tcdf_demo_2.csv
   ==================================
X^2 causes X^2 with a delay of 1 time steps.
X^3 causes X^3 with a delay of 2 time steps.
================================================================================
```