Progress in CAUSAL digital twin

- a whitepaper

Dr. PG Madhavan

Seattle, USA
19 March 2021

About the author: PG launched his first IoT product at Rockwell Automation back in 2000 for predictive maintenance, an end-to-end solution including a display digital twin. Since then, he has been involved in the development of IoT technologies such as fault detection in jet engines at GE Aviation and causal digital twins to improve operational outcomes. His collected works in IoT is being published as a book, "Data Science for IoT Engineers" in June 2021. Rest of his career has been in industry spanning more major corporations (Microsoft, Lucent Bell Labs and NEC) and four startups (2 of which he founded and led as CEO). https://www.linkedin.com/in/pgmad/

Measurements by themselves are of limited use; there is nothing controversial in this statement – when we measure vibration signals from a bearing, it is NOT the vibration itself that is of interest; we want to know if the inner or outer race or the balls are failing! The real purpose of developing and deploying a digital twin is to understand what the "parameters of the underlying system" are so that we can *develop a CAUSAL understanding* of what the measurements are telling us about the system *("what is causing what?")*.

All of today's Machine Learning (ML) is correlation-based. And we know that "correlation is NOT causation"! Interest in causality began from time immemorial … in the last many decades, significant mathematical works have emerged. Most of the credit for the current "causality crusade" goes to Judea Pearl (2011 Turing award winner) who started his exposition of Causality Calculus from early 2000.

Before you take a medicine, randomized controlled trials have been conducted to prove the cause-and-effect of that medicine – this is the gold standard method of proving ***causality***. A physician will not ***prescribe*** it if it was only correlated with your illness! Taking a small leap, ***Causality* is required for *Prescriptive* analytics**.

You cannot easily do randomized controlled trials with industrial machinery! Digital twin (DT) is where cause-effect determination has to happen. And the time is now for CAUSAL digital twins

(CDTs); around late 2010s, there seems to be a "disillusionment" emerging around the value that expensive IoT deployments have created so far. In my opinion, it is the lack of **prescriptive analytics** which tell a business what to do so that operations and production can improve – which requires causal relationships to be quantified beyond just correlations.

Clearly, the first round of "display" digital twins provided a nice window into IoT data via a dashboard, etc. "Simulation" digital twins have already proved their value in machine design where Static and Structural aspects are important – one case is the design of a product using CAD/CAM techniques where we can visualize the stresses and thermal distribution useful in improving the design of a machine.

However, to improve operational aspects, we have to focus on the DYNAMICS of a system. In the specific case of "machine dynamics", we are interested in the kinetics and kinematics of the machine and NOT the Statics or Structural aspects. *In this article, we develop a Causal Digital Twin (CDT) solution that can lead to prescriptive analytics.*

Causality & IoT

In my previous articles on Causal Digital Twin (CDT), we have discussed the need for Causality (for example, see "What-if" digital twin); knowing what caused what beyond what is correlated with what ("Correlation is not Causation!") alone provides us with a model on which we can perform "what-if" analysis which is the basis of "set point" optimization for operations, say in an industrial plant (or any other system of connected assets). This activity is called Prescriptive Analytics. *In short, Causal digital twin makes prescriptive analytics possible that can lead to improvements in operations and hence generate business value from Internet of Things (IoT) investments.*

Most of the Causality work has been in the field of social sciences – epidemiology, econometrics, genetics, market research, etc. This background has made it difficult for Engineers to access this vast knowledgebase and translate them into IoT use cases; one of my main aims here is to simplify this process.

In the rest of this article, I rely on three research publications to formulate CDT. Clearly, there is a large volume of research and I apologize in advance to the authors of other key articles that I may have missed.

1. Causal discovery and inference: concepts and recent methodological advances (2016)
2. Estimation of a Structural Vector Autoregression Model Using Non-Gaussianity (2010)
3. Nonlinear Structural Vector Autoregressive Models with Application to Directed Brain Networks (2019)

The first two are of the "traditional" Causality type and the third one is by Signal Processing engineers. Another major source is Center for Causal Discovery for everything related to causality.

Causality Basics

Let us start from the traditional Causality school's approach. At the risk of over-simplifying, here is a way an IoT engineer can come to grips with the basic tenets of Causality.

Consider 4 entities (assets), B1, B2, B3, B4, as shown in figure 1.

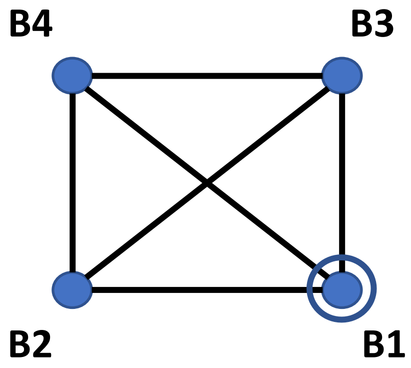


Figure 1. Fully connected graph

The focus now is on the STRUCTURAL aspects and not sensor data time series generated by each asset! This is a key point to keep in mind – we will bring time dependencies later.

The questions to ask are (1) Does all the links exist? (Some entities may not be connected); (2) If so, what is the direction of the link? And (3) What is the strength or weight of the links? The first 2 questions relate to Causal Discovery (of the structure) and the 3rd question relates to Causal Estimation (statistical/ signal processing methods when the data are sensor-measured time series from each of the 4 assets).

At the risk of oversimplifying, these are some of the main concepts in Causal structure discovery.

- A causal model is sufficient if it does not contain unobserved common causes or latent variables.
- Markov assumption: For causally sufficient sets of variables, all variables are independent of their non-descendants in the causal graph conditional on their direct causes (parents in the causal graph).
- Faithfulness: Causal influence is not hidden by coincidental cancelations.
- In directed acyclic graphs (DAGs), use "d-separation" ("d" for directional) to identify pair-wise independent nodes which means that there is no link connecting them. d-separation is a mechanical procedure to answer the equivalent conditional independence question.

These factors have to be satisfied for Causal discovery and estimation to be valid – which is very challenging to prove in traditional applications in Social Sciences. For example, if the nodes of the graph are Health, # cigarettes smoked, Age and Obesity, one can imagine how difficult it will be to draw the right directional links of the DAG and asserting Markov and Faithfulness assumptions!

Causality Insights

Classification and regression and causal inference are different. Peter Spites who is a long time leader of Causality at CMU presents the probabilistic reasoning for this in the 1st article cited earlier. There are important differences between the problem of predicting the value of a variable in an unmanipulated population from a sample (classification and regression) and the problem of predicting the post-manipulation value of a variable from a sample from an unmanipulated population (causality); the latter is called "counterfactual" analysis which reveals the difference.

- Causal factors in a DAG are pair-wise regression coefficient when Markov and Faithfulness conditions are satisfied and d-separation applied to the graph.
- The direction of the link, regress X on Y or Y on X, can be addressed by estimating the regression coefficient in each case, generating the residuals and applying statistical test for independence. If X truly causes Y, the residuals for the regression in this direction will be independent. Regression in reverse will generate residuals that are uncorrelated (by definition) but NOT independent.

Once given a DAG with the conditionally independent links between the nodes removed by d-separation and the direction of the link determined by checking the residuals, we have the Causal Graph. In practical use cases, we will have an outcome variable of interest; for example, in a market research study on a soap brand, the reported customer satisfaction or price may be the outcome variable and other nodes may be color, smell, size, etc. The market research team will pass around a questionnaire to many people and collect these data which forms the sample data at each node. Figure 2 is an example of a DAG with causal directions determined and outcome variable is B1 shown circled.

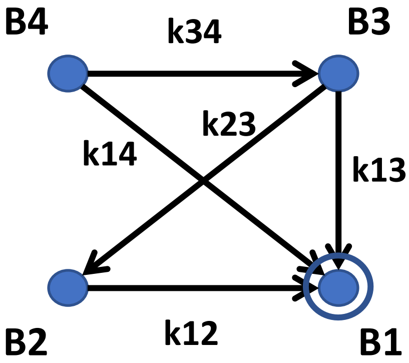


Figure 2. A causal graph with B1 as the outcome variable

From the market research study data collection example, it must be clear that TIME is not a variable in this causal analysis – this is called "structural" model or "instantaneous" causality. In a general case, the question arises of delayed or "lagged" measurements and its causal dependence. Note that in the soap brand study, lagged measurements make no sense – measurements are from different consumers which may have been done at different times but have no bearing on the analysis. Such Structural Causality is the realm of "traditional" causality.

In the IoT case, the factors influencing the outcome variable will have temporal dependencies in addition to the Structural ones in general. Since IoT measurements are typically time series measured from multiple assets interconnected on a plant floor say, instantaneous and lagged causality are important. If the locations of the asset lay out is significant, structural model is spatially distributed. *In the most general case, IoT data for DAG is a spatio-temporal multichannel time series.*

We are dealing with man-made systems in IoT. We will be working with a NASA Bearing Data case study later; a mechanic who has spent a lifetime repairing and rebuilding motors can tell you the key sources of vibration and how the sources are causally related. As a significant simplification for this expositional study, we will use the approach that domain experts have provided us the corresponding DAG and link directions already during an initial knowledge-discovery phase. Thus, our focus below will be on Causal Estimation.

Real Data Modeling: NASA Bearing data

We will develop Causal DT in a real-life setting using the popular NASA Prognostics Data Repository's bearing dataset. The data is from a run-to-failure test setup of bearings installed on a shaft. The rotation speed was kept constant at 2000 RPM by an AC motor coupled to the shaft via rub belts. A radial load of 6000 lbs is applied onto the shaft and bearing by a spring mechanism. All bearings are force lubricated. The arrangement is shown in figure 3.

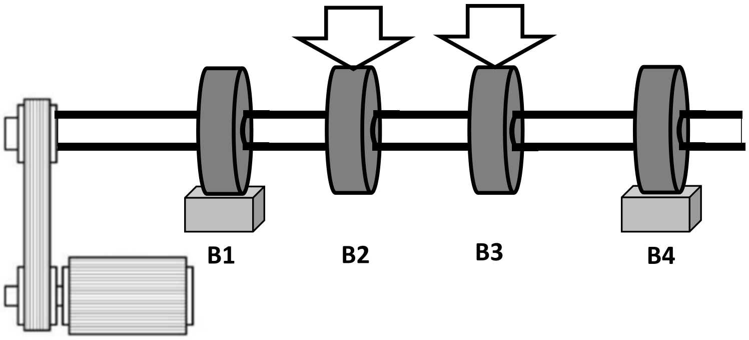


Figure 3. NASA Prognostic Data Repository Bearing data collection setup

There were 3 separate tests in the NASA dataset; we focus on Test 2. Data were collected from Feb 12, 2004 to Feb 19, 2004 when the tests were run continuously to failure in blocks of 10 minutes through the entire period. Bearing 1 will be the target of our study – its outer race failed on Feb 19.

From the structure of the setup and some knowledge of machine dynamics, bearing operation and vibration analysis, we can come up with the Causal Graph for what we will call “NASA bearing system”.

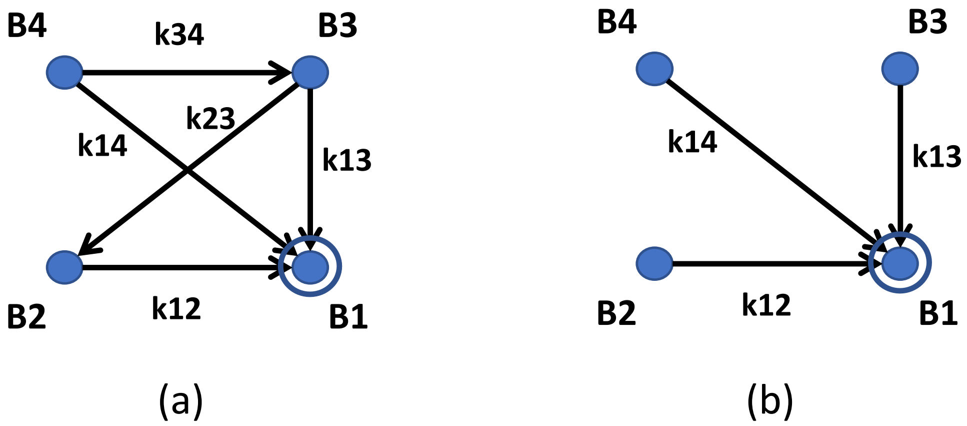


Figure 4. Causal graph structure possibilities

We simplify figure 4(a) as follows. Let us say that the domain expert has told us that Bearing 2 (B2) and Bearing 4 (B4) interaction will be negligible – hence link “k24” is zero (B2-B4 link is absent) and so on. For our purposes of demonstrating Causal Estimation for this IoT use case, we will use the further simplified DAG in figure 4(b) as well as assume that Markov and Faithfulness criteria for Causality have been met.

Instead of using the entire time series from Feb 12 to Feb 19 when Bearing 1 failed (measured continuously over 24-hour periods), we use a block of 10 minutes from each significant date. The time series of Bearing 1 vibration measurement is shown in figure 5.

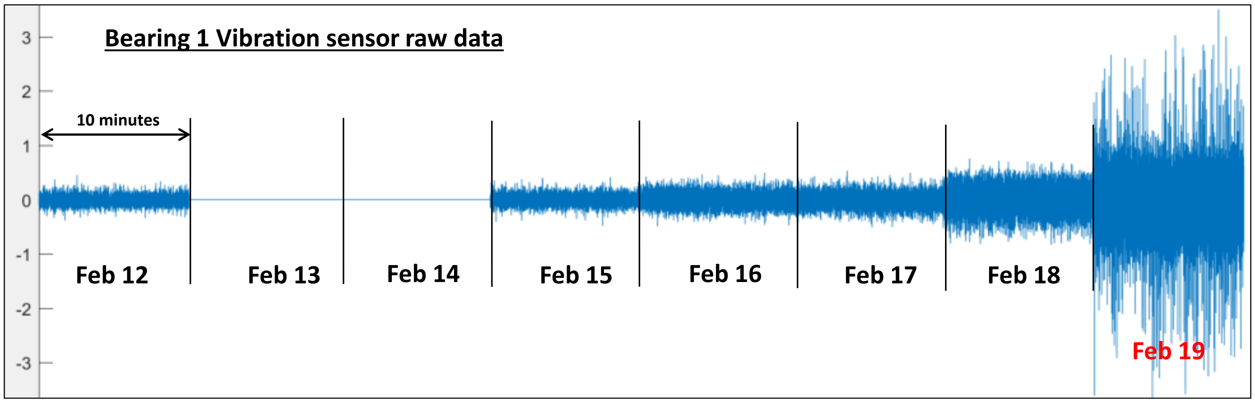


Figure 5. Time series of Bearing 1 vibration (NOT contiguous)

In figure 5, note that the data is NOT contiguous but 10-minute chunks from dates shown put together end to end. The causal factor estimation step in the next section will process each day separately. From figure 5, it is obvious that Bearing 1 vibration is excessive on Feb 19 (bearing failed) and on Feb 12, we see the normal vibration pattern. As one gets closer from Feb 16, 17 and 18, one can see clearly that something is going wrong with Bearing 1 with Feb 16 data being slightly indicative of an impending fault.

Recursive Estimation of Causal Parameters

References 2 and 3 in the initial section have addressed the estimation of instantaneous and lagged causal effects using Structural Vector Autoregressive (SVAR) models. While Reference 2 uses a linear model, Reference 3 extends the results to non-linear causal relationships. There are many more significant papers in social sciences on this topic which IoT engineers can draw from.

The general from of SVAR model:

$$\mathbf{y}_{\mathrm{t}} = \mathbf{A}^{0}\,\mathbf{y}_{\mathrm{t}} + \sum_{\mathrm{m}=1}^{\mathrm{M}} \mathbf{A}^{\mathrm{m}}\,\mathrm{y}_{\mathrm{t-m}} + \mathbf{e}_{\mathrm{t}}$$

All bolded quantities are vector or matrices. $\mathbf{A}^{0}$ has zeros for its diagonal entries since self-causality does not exist. $\mathbf{y}_{\mathrm{t}}$ are the measurements of the nodes of the DAG. First term is Structural causality and the second term is Lagged causality.

Consider the DAG in figure 4(b) as the model for NASA bearing system. The governing equations can be written as follows for our outcome variable B1:

B1 = $\mathcal{F}^{\mathrm{C}}$[B2, B3, B4] + $\mathcal{F}^{\mathrm{D}}$ [B1, B2, B3, B4]

$\mathcal{F}^{\mathrm{C}}$[.] – Current/ Causal/ Structural samples of [.]

$\mathcal{F}^{\mathrm{D}}$[.] – Delayed samples of [.]

Note that under $\mathcal{F}^{\mathrm{C}}$, B1 is missing – there is no concept of SELF-causal effect. From our causal graph in figure 4(b), we will ignore delayed B1 also and consider only 1 delay (for explanatory purposes) for the other bearings. We can write –

B1[n] = k12*B2[n] + k13*B3[n] + k14*B4[n] + k12d*B2[n-1] + k13d*B3[n-1] + k14d*B4[n]

In a vector form,

B1[n] = **h**[n]***k**[n] + e[n] where e[n] is the process noise. . . . (A)

$\mathbf{k}^{\mathrm{T}}$[n] = [k12 k13 k14 k12d k13d k14d]

h[n] = [B_2[n] B_3[n] B_4[n] B_2[n-1] B_3[n-1] B_4[n-1]]

Given equation (A), we can solve for **k** using the block version of Least Squares. We choose to use Recursive Least Squares (RLS) so that we can visualize **k**[n] since most likely, such solutions will be deployed as a real-time CDT. The actual RLS algorithm implemented is on page 54 of my Systems Analytics book.

Revisiting figure 5, on Feb 19, the abnormally high vibration shows the failure of Bearing 1. Feb 12 is an example of normally operating bearing. On Feb 16, 17 and 18, the vibration amplitude increases and is an indicator of impending failure. *If the ONLY objective is to predict failure, simply thresholding the vibration amplitude may be enough of a solution* (but will not be very robust due to the presence of noise, etc.).

Using SVAR model and RLS algorithm, we estimate **<u>k</u>**[n], the instantaneous and lagged causal factors as specified in figure 4(b). We will call **<u>k</u>**[n] "coupling factors" to be more consistent with machine dynamics terminology.

Each of the data blocks for each day (10 minutes long) were processed and **<u>k</u>**[n] estimated for each day. The results are in figure 6.

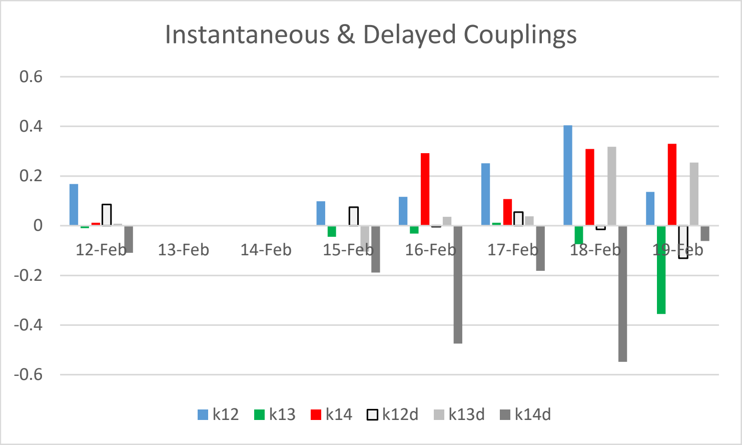


Figure 6. Instantaneous and deployed couplings (causal link weights)

On the day of failure (Feb 19), couplings actually diminish; certain couplings increase as Feb 19 approaches. This dynamic can be seen more clearly in figure 7.

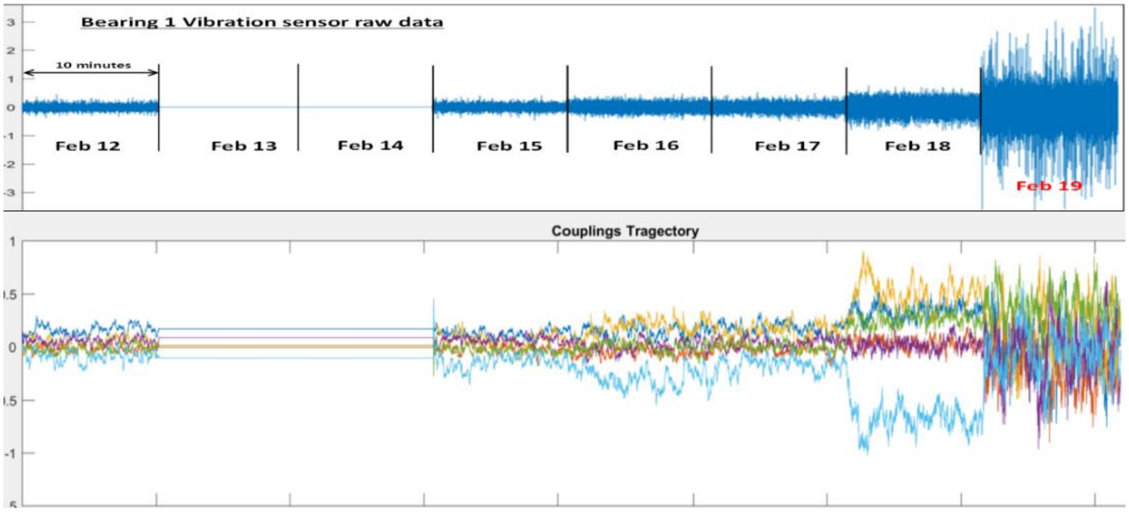
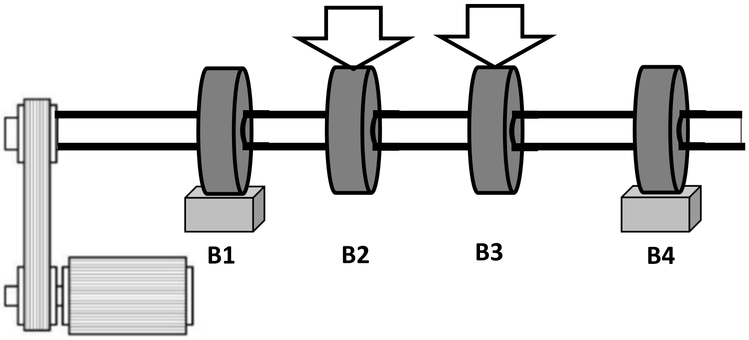


Figure 7. Bearing 1 time series (top) and all 6 couplings (bottom)

Key observations:

1. On Feb 19, failure occurred and the couplings reduced in value. This can be due the overall instability of the system of bearings.
2. Instantaneous couplings increase before failure; Lagged coupling between Bearing 1 & 2 shows very small variations but the other 2 delayed couplings show significant increases.



Findings:

- Compared to using just the vibration amplitude, couplings show a systematic increase as Bearing 1 failure approaches which opens up the possibility of a multi-factor prediction solution. How early a prediction could have been made is not visible in this study because only a 10-minute block of data per day was processed (Feb 17 data block chosen at random from 144 blocks on that day is suspect). If all continuous data were processed, it is very likely that the trajectories of **k**[n] will be smooth and thus be a robust basis for predicting failure point.
- Machine dynamics experts will take a special interest in delayed couplings; that is because the flexing of the shaft can be the root cause of failure and the delay accounts for a possible back and forth dynamic between Bearing 1 and the other bearings on the same shaft. This can lead to better design of the bearing arrangement or stiffer material for the manufacture of the shaft.

We will remark that we also studied simple factors such as variance ratios of pairwise bearing vibrations; no insightful information could be gathered.

In summary, it is very clear from figure 7 that couplings on Feb 18 are predictive of failure on Feb 19; it is noticeable that some of the couplings already show abnormal behavior as early as Feb 16. *It is also clear that if prediction of failure was the ONLY objective, vibration amplitudes (or variance) on Feb 16 compared to Feb 15 may be sufficient (though not a robust method in practice!).*

In addition to visualizing the IoT data from the bearings in a new way (as in figure 7 bottom panel), the real value of a causal model is to perform "What-If" analysis and do Counterfactual experiments. We demonstrate the power of SVAR model in this regard in the next section.

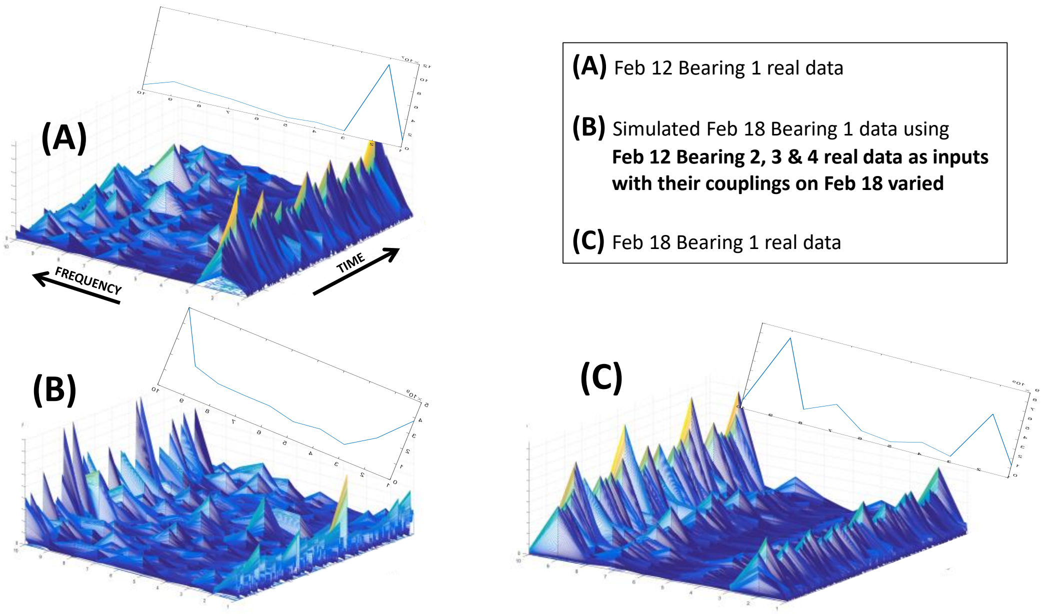
What-If analysis

SVAR digital twin provides a Causal Graph model with underlying parameters estimated from measured data. With a “converged” model in hand, one can perform off-line experiments – such as varying the parameters to assess the impact on Bearing 1. A simple study was undertaken to assess the “what-if” analysis capability of SVAR Causal digital twin.

Let us say that we are on Feb 12. We have all the vibration data on that day and the model in the last section. We varied 2 main delayed couplings, k13 and k14, such that Bearing 1 data output of this simulation were similar to Bearing 1 data on Feb 18, the day before failure. The value of this “what-if” analysis is that if lagged k13 and k14 reach those numbers on Feb 13, 14, . . , we can conclude that Bearing 1 will likely fail the next day.

Clearly, one cannot compare Feb 18 Bearing 1 real data time series to the one simulated on Feb 12 – being realizations of random processes, they will not have the same waveforms! We use a more sophisticated method called Time-Frequency Distribution (TFD) which is a type of time-varying power spectral density estimate. Once the time axis is expanded, power variations over frequency can be visualized more clearly and some qualitative assessments can be made.

Figure 8 shows the results of What-If analysis described above. A very low-resolution TFD was estimated for the purposes of this experiment. As the legend explains, all diagrams here are vibration data of Bearing 1. In figure 8 (A) and (C), the actual measured sensor data from the NASA dataset is displayed. Comparing them, you can see that Feb 12 TFD in (A) has very little high-frequency content (frequency axis increases from right to left in all diagrams in figure 8). Whereas in (C), one can see a row of high-frequency peaks all across the 10-minute time axis.



(A) Feb 12 Bearing 1 real data

(B) Simulated Feb 18 Bearing 1 data using **Feb 12 Bearing 2, 3 & 4 real data as inputs with their couplings on Feb 18 varied**

(C) Feb 18 Bearing 1 real data

Figure 8. What-If analysis using SVAR causal digital twin

This is made even more clear when you collapse all the time slices on to the frequency axis (thus yielding a traditional power spectrum estimate plot) which are shown floating on the right and top of each diagram. (C) has a clear high-frequency peak whereas (A) does not (note that frequency increases to the left).

(B) is the result of "what-if" analysis. What if we changed the coupling factors on Feb 12 data? Will the simulated Bearing 1 vibration look like the day before failure (Feb 18)? We see in figure 8(B) that for couplings from Feb 18, Feb 12th data produce Bearing 1 data that "looks like" Feb 18 data – (B) has more high-frequency peaks than (A) in the TFD and the power spectrum in (B) looks more similar to (C) than (A).

In essence, what we have done in this "what-if" analysis is to take Feb 12 data and see if the data will simulate a day before failure (Feb 18), which it does to some extent – *the value of this analysis is that we can simulate what may happen to the bearing system in the future*!

Another fascinating use is to run SVAR Causal digital twin in "fast forward" mode and explore what other effects it may exhibit in the future. In a more complex system than the relatively simple NASA bearing system, knowing the internal dynamics as specified by the Causal digital twin will have many more uses in assessing future state of the overall system and its individual components, remaining useful life (RUL) or simply predicting failure using causal graph parameters as a more robust (less false-positives and hence waste) method.

Counterfactual experiment

The discriminating power of causal models is the ability to perform counterfactual experiments AFTER the data has been collected. To provide a demonstration, we assess the effect of the absence of Bearing 3 on Bearing 1. To eliminate Bearing 3, we set the instantaneous and lagged versions of coupling, k13, to zero. The results for Bearing 1 data from Feb 18 (the day before failure) is shown in figure 9.

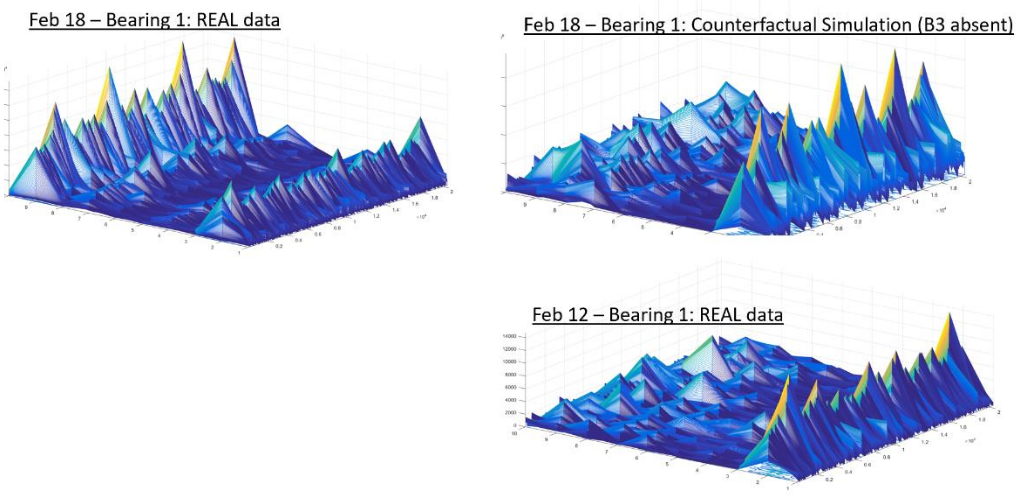


Figure 9. Counterfactual experiment using SVAR causal digital twin

In figure 9, Bearing 1 TFD for Feb 18 and Feb 12 are obtained the same way as in Figure 8 (A) and (C). If on Feb 18, if Bearing 3 was not in the system, what will the vibration of Bearing 1 look like? The top right plot in figure 9 shows that the Bearing 1 vibration will be more similar to Feb 12th Bearing 1 data! Notice the lack of high-frequency peaks compared to actual Feb 18th data on the top left of figure 9.

The practical use of this experiment for a machine dynamics expert is unclear but this counterfactual experiment is designed to show the power of Causal Graph models in the form of a Causal Digital Twin. *We can ask questions "after the fact" and seek answers to what could have happened* . . .

Significance of Causal digital twin

SVAR model allows the estimation of causal parameters of a dynamical system from measurements. Causal Graphs that satisfy causality assumptions make such analyses possible.

- Causal graphs can be elicited from domain experts or estimated directly from measured data – "causal discovery" (there are startups providing this service – for example, https://www.inguo.io/).
- SVAR model and recursive least squares provide a simple methodology for estimating instantaneous and lagged causal link strengths (or couplings).
- In the real data tests with NASA bearing data, we showed that SVAR Causal digital twin works in real-life cases and is able to unearth the physical system's unobservable parameters by learning them over time.
- One of the most unique features of SVAR Causal digital twin is that there is no hand-tuning of the digital twin! All steps are LEARNED iteratively as each data point arrives. Therefore, SVAR Causal digital twin is a true "real-time" digital twin. *Once a causal graph structure is chosen, all relevant parameters are learned "on the fly", there by instantiating each physical system automatically*.

Looking Ahead

We present this work as a foundational expository application (perhaps the first one) of the Causality framework to IoT which gives us Causal digital twins. There is a lot more to be borrowed from Causality theory – the references in the first section are great starting points. When it comes to estimation, SVAR model just scratches the surface.

Signal processing and Machine Learning experts have an extensive arsenal of recursive (real-time) methods based on conditional expectation estimation in a Kalman filter framework and other methods refined over the last 30 or so years. The extension to nonlinear causal factor estimation has already happened (3rd reference) and I foresee an explosion of Causal digital twins suitable for various IoT use cases.

The performance of "what-if" and counterfactual experiments using causal digital twins are the sources of actionable information for prescriptive analytics which can be used to make business decisions to improve operations and increase production.

Causal digital twin (CDT) solution presented here focuses on a system of connected assets and its dynamics; CDT is mostly unique in IoT methodology so far. Focusing on one asset at a time has its uses – in predicting that asset's potential failure and better structural design of the asset. However, when connected assets are treated as a system as used in a production environment, the results of ***Causal Digital Twin that learns in real-time become directly useful in enhancing business outcomes such as increased production volume, better quality and reduced waste – all contributing to increased gross margin for the business***.