This is a preprint draft. The published article can be found at: https://doi.org/10.1007/s10694-023-01405-8.

Please cite this paper as:
Naser M.Z., Çiftçioğlu A, (2023). "Revisiting Forgotten Fire Tests: Causal Inference and Counterfactuals for Learning Idealized Fire-induced Response of RC Columns". *Fire Technology*. https://doi.org/10.1007/s10694-023-01405-8.

Revisiting Forgotten Fire Tests: Causal Inference and Counterfactuals for Learning Idealized Fire-induced Response of RC Columns

M.Z. Naser[1], Aybike Özyüksel Çiftçioğlu[2]
[1]School of Civil & Environmental Engineering and Earth Sciences (SCEEES), Clemson University, USA
[1]Artificial Intelligence Research Institute for Science and Engineering (AIRISE), Clemson University, USA
E-mail: mznaser@clemson.edu, Website: www.mznaser.com

[2]Department of Civil Engineering, Manisa Celal Bayar University, Turkey, E-mail: aybike.ozyuksel@cbu.edu.tr

Abstract
The expensive nature and unique facilities required for fire testing make it difficult to conduct comprehensive experimental campaigns. As such, engineers can often afford to test a small number of specimens. This complicates attaining a proper inference, especially when addressing questions in the form of what would have been the fire response of a particular specimen had it been twice as large? Or had it been made from a different material grade? In hindsight, answering causal and hypothetical (counterfactual) questions goes beyond the capacity of statistical and machine learning methods which were built to address observational data. To overcome the above challenges, this paper presents a causal approach to answering such questions. In this approach, principles of causal inference are adopted to reconstruct the deformation-time history of reinforced concrete (RC) columns and propose an idealized fire response for these columns. The findings of this study indicate the significant influence of the loading level, aggregate type, and longitudinal steel ratio on the deformation history of fire-exposed RC columns.

Keywords: Causal inference; Fire response; Fire tests; Reinforced concrete columns.

Introduction
Fire tests are complex and expensive and hence, are likely to be relatively small in size. Given the need for unique facilities and expertise, testing full scale specimens under fire conditions become of limited nature, and such experimental campaigns become effectively rare. While most publicly available works often contain 2-6 specimens, only a few of these programs have examined and reported a significantly large number of tests [1–5].

These programs exist in part as they were funded by governmental efforts to establish or modernize fire building codes and standards. Some of the publicly available campaigns include those sponsored by the National Bureau of Standards (NBS, and now the National Institute of Standards and Technology (NIST) [6], National Research Council (NRC) of Canada [7], and Eurocode [8]).

The results of such tests turn valuable on a number of fronts. First, examining a large number of specimens within one program implies a greater degree of consistency than examining those to be collected from a collection of studies. Such consistency arises from maintaining many of the commonly difficult-to-control latent variables as consistently as possible [9]. Such variables may include the fabrication process, use of the same fire testing facilities and equipment, and quality and synergy of the reported results, all of which allow us to better infer the outcome of fire testing through meaningful comparisons.

Another front pertains to having a coherent direction of research that leads to the development of a detailed and wide-ranging test matrix. In a typical matrix, one specimen is left as a benchmark, and all other specimens are altered. The most common means of designing a test matrix is only to alter one variable for each specimen. This allows a direct comparison between the benchmark and each altered specimen, as well as between the specimens themselves [10]. Following this approach, any change in an observed response from the benchmark specimen can be traced back to the altered variable.

The one-at-time approach, when statistically meaningful by providing a good sample size in the test matrix, also allows us to develop predictive tools [11]. The most commonly used tools in the structural fire engineering domain are charts, tables, and formulas. The goodness of such tools stems from the goodness of the data (i.e., the results of fire tests) used in creating such tools [12]. Thus, the uniformity provided in large-sized fire campaigns becomes elemental to the success and predictability of the resulting estimation or prediction tools [13].

Generally, even the most notable fire testing programs do not examine every possible variation and/or combination of factors. This is true in the sense that practical limitations persist with regard to the time, financial resources, and vision of stakeholders. Hence, we often revert to extending the experimental findings via validated numerical (e.g., finite element) models. This practice has been well accepted and remains the primary means to complement fire tests or evaluate fire response as permitted by building codes and standards [14]. For example, Section 4.3 of Eurocode 2 defines the above under "*advanced calculation methods… [they] shall be based on fundamental physical behavior leading to a reliable approximation of the expected behavior of the relevant structural component under fire conditions*."

At the moment, we continue to lack a robust definition and procedure for building and validating these calculation methods. Similarly, we lack an understanding of the established standardized inputs, solution procedures, and outputs used in such methods. These items remain an ongoing scene of debate that require serious progress [15].

This work stems inspiration from the parallel fields of statistics, medicine, and social and computer sciences related to establishing an approach to causal inference [16]. Causal inference draws conclusions pertaining to the existence of a causal connection between the variables. Such a relationship is often mistaken for a correlational relationship, an elemental means to analyze experimental results. However, the difference between the two is quite substantial. For instance, the latter is defined as a general trend where two variables increase or decrease together (i.e., on average, smaller specimens have a lower fire resistance time than larger specimens). On the other hand, the former is defined when a causing variable is partly responsible for generating the effect variable, and this variable is partly dependent on the first [17].

Just like we adopt experimental and finite element principles to carry out tests or create advanced models, to identify causal relations and answer causal questions, we must employ causal principles. At this point, the domain of structural fire engineering lacks the front of causality and causal inference. As a matter of fact, a search with the key terms of "*causality*", "*causal inference*" and "*structural fire engineering*" returns very little to no work on this front [18]. Fortunately, the rise of modern machine learning (ML) now makes it possible to arrive at causal estimations of

This is a preprint draft. The published article can be found at: https://doi.org/10.1007/s10694-023-01405-8.

Please cite this paper as:
Naser M.Z., Çiftçioğlu A, (2023). "Revisiting Forgotten Fire Tests: Causal Inference and Counterfactuals for Learning Idealized Fire-induced Response of RC Columns". *Fire Technology*. https://doi.org/10.1007/s10694-023-01405-8.

various phenomena with ease [19–21]. This presents an exciting opportunity to explore within the realm of this domain. This is also a key motivation behind this work.

From this perspective, the current study showcases the merit of causal inference in one of the most fundamental problems in structural fire engineering by reconstructing the deformation-time history of various RC columns. Our analysis will leverage a compiled dataset from one of the largest physical fire tests in recent years carried out by Prof. TT Lie and coauthors from the National Research Council (NRC) of Canada. Our analysis demonstrates that it is not only possible to infer new findings from forgotten fire tests causally but that these findings can lead to the development of idealized models that can extend beyond the original tests. In addition, the results of this analysis indicate that the fire response history of RC columns is heavily influenced by the present loading level, aggregate type, and longitudinal steel ratio of the fire-exposed RC columns. This study also compares the causal approach to that obtained from statistical and traditional data-driven ML to highlight the importance and merit of adopting causality.

Description of TT Lie's fire testing programs at NRC

Of interest to this study is the testing program conducted by TT Lie, which is considered one of the most systematic and comprehensive fire campaigns in the last three decades [1–5]. The testing program was conducted at the National Research Council (NRC) of Canada with joint capacity from the Portland Cement Association (PCA).

Overall, 41 full-scale RC columns were tested under three phases wherein the following parameters were investigated:1) cross-sectional area, 2) cross-sectional shape (square, rectangular, circular), 3) thickness of concrete cover, 4) percentage of longitudinal reinforcing steel, 5) lateral reinforcement (tied or spiral), 6) concrete mixture (type of aggregate), 7) concrete strength, 8) moisture content of concrete (relative humidity), 9) end conditions, 10) axial or rotational restraint, 11) load intensity, 12) load eccentricity, and 13) fire exposure intensity. Each test was documented by providing a complete temperature-time and deformation-time history, and time to failure. In addition, the residual strength of a few columns was also measured. This testing program is informally known as Internal Report No. 569.

The overall goal of this testing program was twofold:1) to generate measured fire resistance data on RC columns designed in accordance with the American Concorde Institute (ACI) and the Canadian Building Codes (CBC), and 2) to develop general methods for the calculation of the fire resistance of concrete columns. It is worth noting that this testing campaign builds upon two earlier and smaller fire tests by Lie et al. [2] and [3] (published in 1972 and 1974, respectively).

Most of the 41 tested RC columns were tested under fixed-fixed restraints, except five, which were tested under various restrained conditions. In addition, seven columns were eccentrically loaded. Two specimens were made of high-strength concrete. Two of circular shape, two of a rectangular shape, and two were made with lightweight aggregate. One column was tested at ambient conditions and two were tested under residual conditions. Finally, one specimen was tested under an intense fire that exceeded the standard fire. Please note that the first fire-tested column was unloaded and hence the deformation-time for this curve was not provided.

This is a preprint draft. The published article can be found at: https://doi.org/10.1007/s10694-023-01405-8.

Please cite this paper as:
Naser M.Z., Çiftçioğlu A, (2023). "Revisiting Forgotten Fire Tests: Causal Inference and Counterfactuals for Learning Idealized Fire-induced Response of RC Columns". *Fire Technology*. https://doi.org/10.1007/s10694-023-01405-8.

Of the above RC columns, 14 square columns (13 of 305×305 mm and one 406×406 mm) and one circular column are of comparable features and hence only these columns from [1] were considered in this study. All columns were cast from normal strength concrete, were reinforced with Grade 414 MPa steel, and most of the columns had humidity measured to be within the range of 70%. The features of the columns are listed in Table 1.

To further complement the above columns, two additional rectangular columns studied by TT Lie [4] and another six more columns (of identical size) but different properties were also added from another test by TT Lie [5]. Hence, there were twenty three 3810 mm long RC columns examined herein.

Table 1 Features of examined columns.

No.	No. in Ref.	Size (mm)	f_c (MPa)	f_y (MPa)	ρ	P (%)	Aggregate type	Humidity$^+$ (%)	Failure time (min)	Ref.
C01	2a	305×305	36.9	414.0	0.022	0.69	Silicate	15.0	170	[1]
C02	3a	305×305	34.2	414.0	0.022	0.44	Silicate	70.0	218	[1]
C03	4a	305×305	35.1	414.0	0.022	0.38	Silicate	63.0	220	[1]
C04	7a	305×305	36.1	414.0	0.022	0.57	Silicate	74.0	208	[1]
C05	8a	305×305	34.8	414.0	0.022	0.97	Silicate	74.0	146	[1]
C06	9a	305×305	38.3	414.0	0.022	0.67	Silicate	75.0	187	[1]
C07	8f	305×305	42.6	414.0	0.044	0.38	Silicate	61.0	252	[1]
C08	9f	305×305	37.1	414.0	0.044	0.57	Silicate	61.0*	225	[1]
C09	10b	305×305	40.9	414.0	0.022	0.38	Carbonate	75.0	510	[1]
C10	11b	305×305	36.9	414.0	0.022	0.56	Carbonate	75.0	366	[1]
C11	12b	305×305	39.9	414.0	0.022	0.87	Carbonate	76.0	216	[1]
C12	6c	305×305	46.6	414.0	0.022	0.47	Lightweight	79.0	188	[1]
C13	7c	305×305	42.5	414.0	0.022	0.44	Lightweight	80.0	259	[1]
C14	10g	406×406	38.8	414.0	0.025	0.66	Silicate	80.0	262	[1]
C15	11h	D355	41.6	414.0	0.022	0.51	Silicate	65.0	240	[1]
C16	5h**	305×457	42.5	414.0	0.017	0.46	Silicate	65.0	396	[1]
C17	6h***	203×914	42.1	414.0	0.012	0.19	Silicate	58.0	330	[1]
C18	1	305×305	36.0	340.0	0.017	0.70	Silicate	63.2	97	[5]
C19	2	305×305	29.0	340.0	0.017	0.84	Carbonate	91.8	164	[5]
C20	3	305×305	28.0	340.0	0.017	0.86	Silicate	98.0	109	[5]
C21	4	305×305	31.8	340.0	0.017	0.77	Carbonate	80.0	175	[5]
C22	5	305×457	32.5	340.0	0.018	0.67	Carbonate	69.3	232	[5]
C23	6	305×305	26.4	340.0	0.014	0.86	Carbonate	66.7	175	[5]

*Assumed based on C07. **Also appears in [4] as column no. 2. **Also appears in [4] as column no. 3. $^+$defined by TT Lie as the moisture content of concrete.

The following discussion presents a description of the series of comparisons between the RC columns listed above. We begin by showcasing the results pertaining to the thermal response and then move to the deformation response.

Thermal response

It should be noted that the discussion on the thermal performance of these columns was kept to a minimum, as further details and explanations can be found in the cited reports as well as in the open literature. Some changes in such a rise were observed, especially in larger columns, owing to the size effect or those made from different aggregates. Table 2 lists the columns selected from Table 2 for comparison purposes. All columns had the same concrete cover of 48 mm, and their temperatures were measured at the steel rebar level. For completion, a brief discussion on the effects of shape, aggregate type, size, and humidity is presented. It is worth noting that the mechanical features (load level, steel ratio) have little to no influence on the thermal response of columns.

Table 2 Features of examined columns.

No.	No. in Ref.	Size (mm)	f_c (MPa)	f_y (MPa)	ρ	P (%)	Aggregate type	Humidity (%)	Failure time (min)	Ref.
C01	2a	305×305	36.9	414.0	0.022	0.69	Silicate	15.0	170	[1]
C06	9a	305×305	38.3	414.0	0.022	0.67	Silicate	75.0	187	[1]
C12	6c	305×305	46.6	414.0	0.022	0.47	Lightweight	79.0	188	[1]
C14	10g	406×406	38.8	414.0	0.025	0.66	Silicate	80.0	262	[1]
C15	11h	D355	41.6	414.0	0.022	0.51	Silicate	65.0	240	[1]
C16	2	305×457	42.5	414.0	0.017	0.46	Silicate	65.0	396	[4]
C17	3	203×914	42.1	414.0	0.012	0.19	Silicate	58.0	330	[4]
C19	2	305×305	29.0	340.0	0.017	0.84	Carbonate	91.8	164	[5]
C22	5	305×457	32.5	340.0	0.018	0.67	Carbonate	69.3	232	[5]

Effect of shape

C06 and C15 are the two closest columns with different shapes and comparable cross-sectional areas within 6%. C06 is a square column and C15 is a circular column. As one can see in Fig. 1, the circular columns experience a lower temperature rise as compared to the squared column.

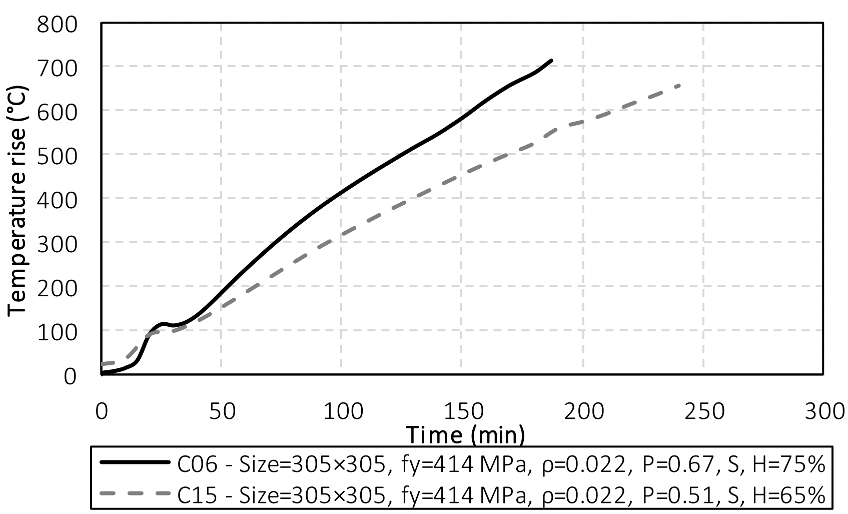


Fig. 1 Effect of shape [Note that C15 is a circular column]

Effect of size

It is clear that C01, C14, C16, and C17 have identical temperature rises during the first 60 min of fire exposure (see Fig. 2). Then, the temperature rise slowly differs and reaches a maximum of

This is a preprint draft. The published article can be found at: https://doi.org/10.1007/s10694-023-01405-8.

Please cite this paper as:
Naser M.Z., Çiftçioğlu A, (2023). "Revisiting Forgotten Fire Tests: Causal Inference and Counterfactuals for Learning Idealized Fire-induced Response of RC Columns". *Fire Technology*. https://doi.org/10.1007/s10694-023-01405-8.

100°C beyond 180 min. Overall, the size effect in these columns do not have a large influence on temperature rise[1] – given that the concrete cover was kept constant at 48 mm. However, the failure time is drastically different between these columns, which is due, as will be shown in an upcoming section, to the level of loading and steel reinforcement ratio.

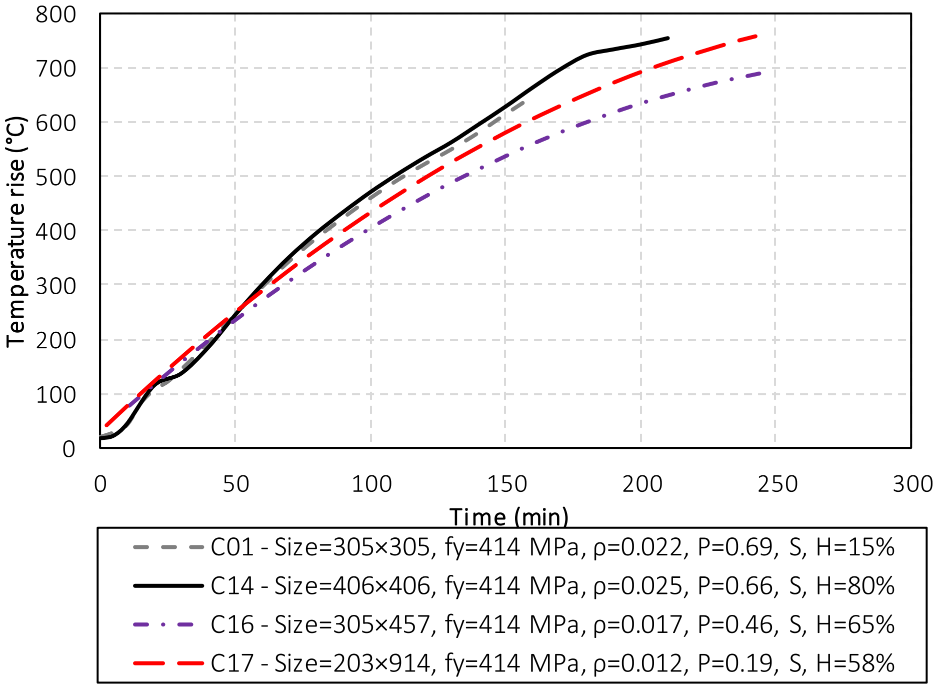


Fig. 2 Effect of size

<u>Effect of aggregate type</u>

Figure 3 shows that the effect of the aggregate was apparent after the 60 minutes mark. As expected, C06, which was made from silicate concrete, exhibited the highest temperature rise. Both C12 and C19 experienced a slightly lower temperature increase of approximately 100°C at 120 min and 150°C at 180 min. The same observation can also be seen by comparing the response of C16 and C22.

[1] It should be noted that the size effect is more likely to influence the cross sectional temperature distribution as well as core temperature of columns. The disucssion of this section is limited to the temperature rise in steel rebars which happen to be at 48 mm away from the surface of the concrete for all columns.

This is a preprint draft. The published article can be found at: https://doi.org/10.1007/s10694-023-01405-8.

Please cite this paper as:
Naser M.Z., Çiftçioğlu A, (2023). "Revisiting Forgotten Fire Tests: Causal Inference and Counterfactuals for Learning Idealized Fire-induced Response of RC Columns". *Fire Technology*. https://doi.org/10.1007/s10694-023-01405-8.

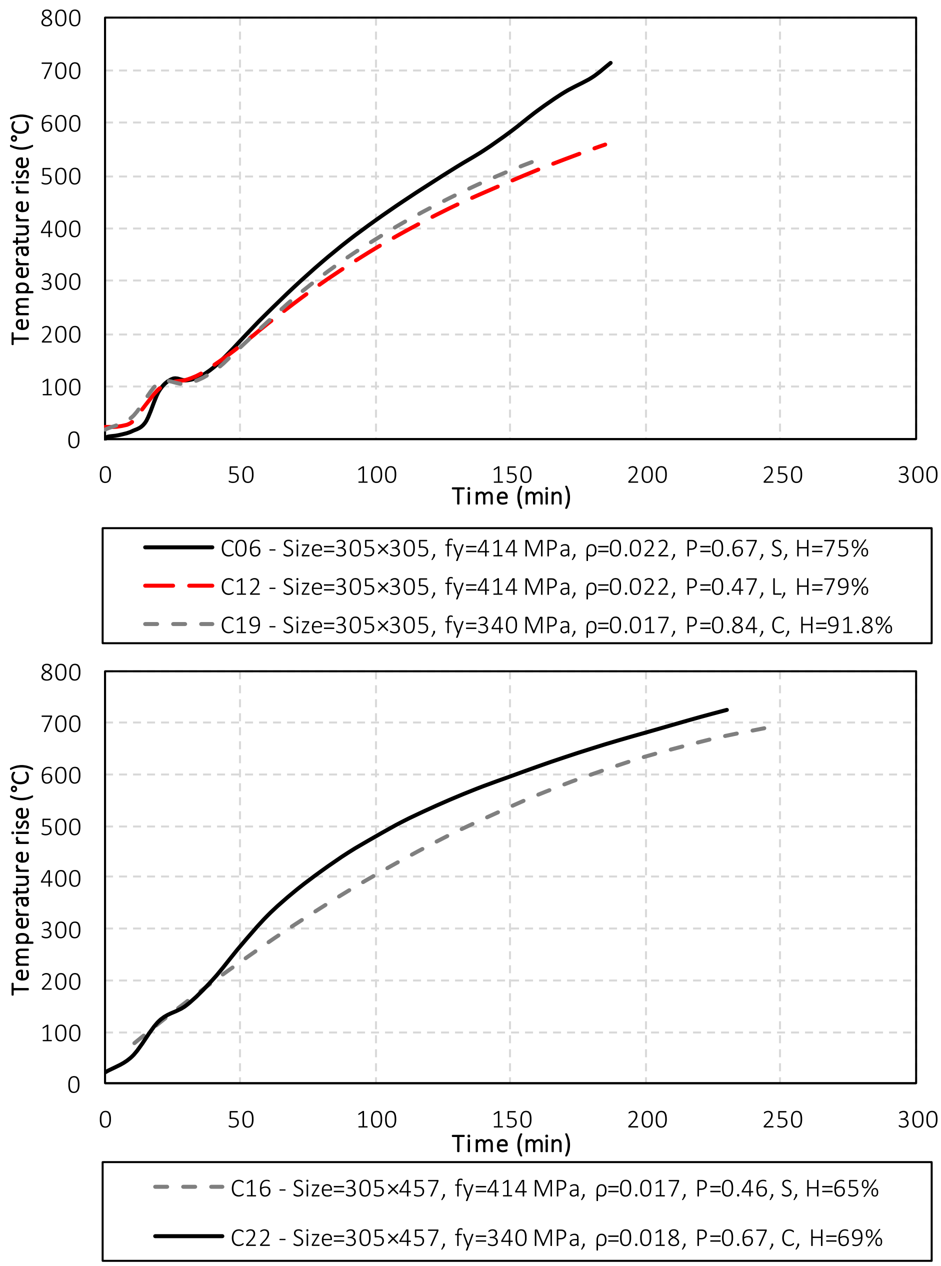


Fig. 3 Effect of aggregate type

<u>Effect of humidity (moisture content)</u>

The effect of humidity seems to be minor, as shown in Fig. 4. However, at this point, we cannot clearly identify the magnitude of this effect because both columns share almost identical features with regard to size, yield strength, and steel ratio. Both columns were loaded with loads that are 2% apart. However, C01 (15% humidity) failed at 170 min, whereas C06 (75% humidity) failed at 187 min.

This is a preprint draft. The published article can be found at: https://doi.org/10.1007/s10694-023-01405-8.

Please cite this paper as:
Naser M.Z., Çiftçioğlu A, (2023). "Revisiting Forgotten Fire Tests: Causal Inference and Counterfactuals for Learning Idealized Fire-induced Response of RC Columns". *Fire Technology*. https://doi.org/10.1007/s10694-023-01405-8.

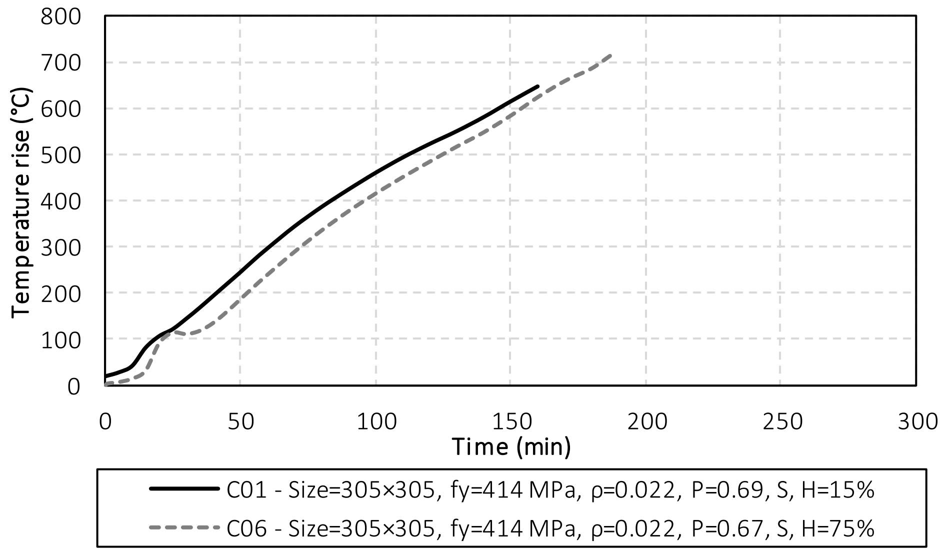


Fig. 4 Effect of humidity

Deformation response

The effects of column size, longitudinal steel ratio, level of loading, aggregate type, and degree of humidity were compared. In this comparison, the RC columns were matched based on how close they were to each other in terms of the variables presented in Table 1.

In one particular comparison concerning the effect of column size, identical matching was not possible due to the lack of a tested identical column(s). However, this comparison was maintained for illustration purposes. It should be noted that all the columns had ties spaced at 305 mm, except for C17 at 203 mm. Finally, the range of the horizontal axis was kept constant across all figures, while the vertical axis was not kept constant to allow for maximum legibility.

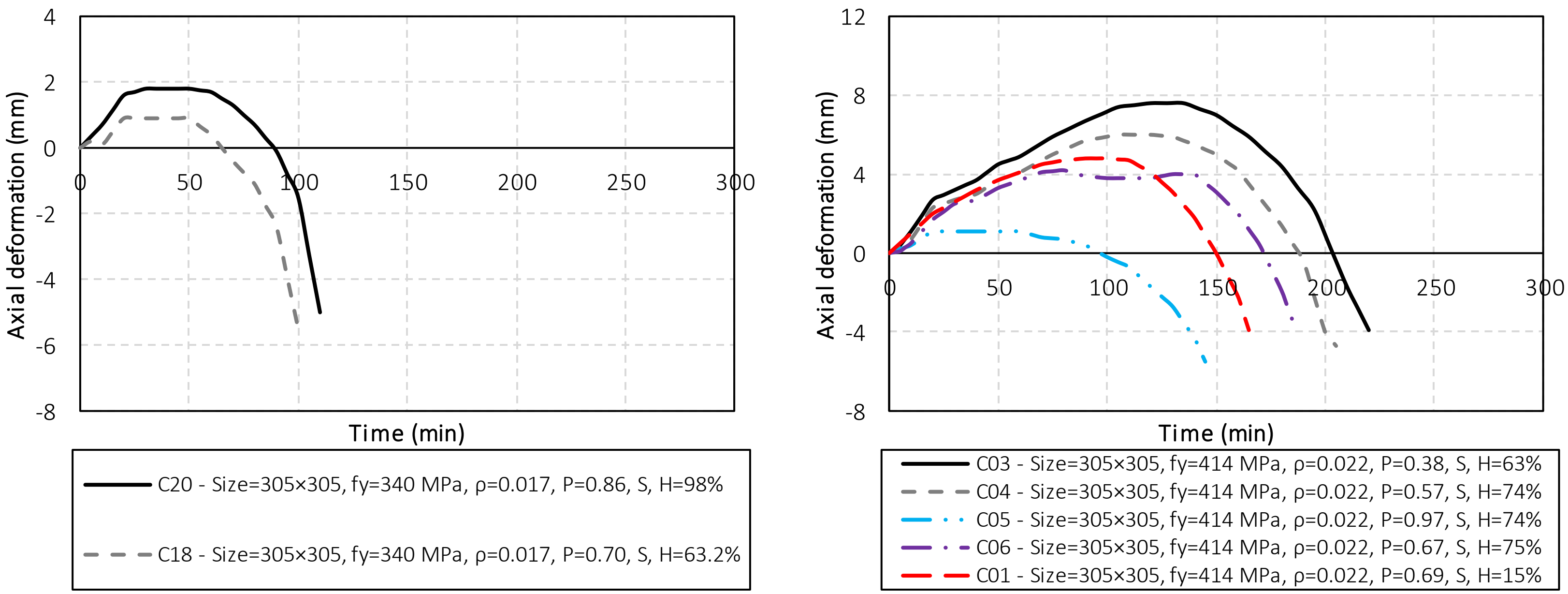
<u>Effect of loading</u>

Figure 5 shows the effect of loading on the deformation history of RC columns with the most resemblance. This figure shows such effect for columns made of silicate and carbonate in two series (steel grade 340 MPa and 414 MPa). Overall, the deformation history is short for heavily loaded columns. In contrast, lightly loaded columns experienced a larger expansion on average. It is worth pointing out that columns of silicate aggregates have a steeper and sharper decline when approaching failure than columns made from carbonate aggregates. Furthermore, heavily loaded columns did not exhibit much elongation under fire when compared to lightly loaded columns (<50%). In all cases, the thermal elongation of heavily loaded columns appears to be within 1-2 mm.

As all the depicted columns are of the same size, the effect of loading can be described with a rotation that takes place within the first 30 min of fire exposure, as the deformation history becomes heavily dependent on the level of loading at that time. Interestingly, C19 (84%) and C21 (77%) showed very small deviations, which could be due to the high level of loading. This observation could not be verified by columns from the silicate group because of the lack of two columns with such load levels.

This is a preprint draft. The published article can be found at: https://doi.org/10.1007/s10694-023-01405-8.

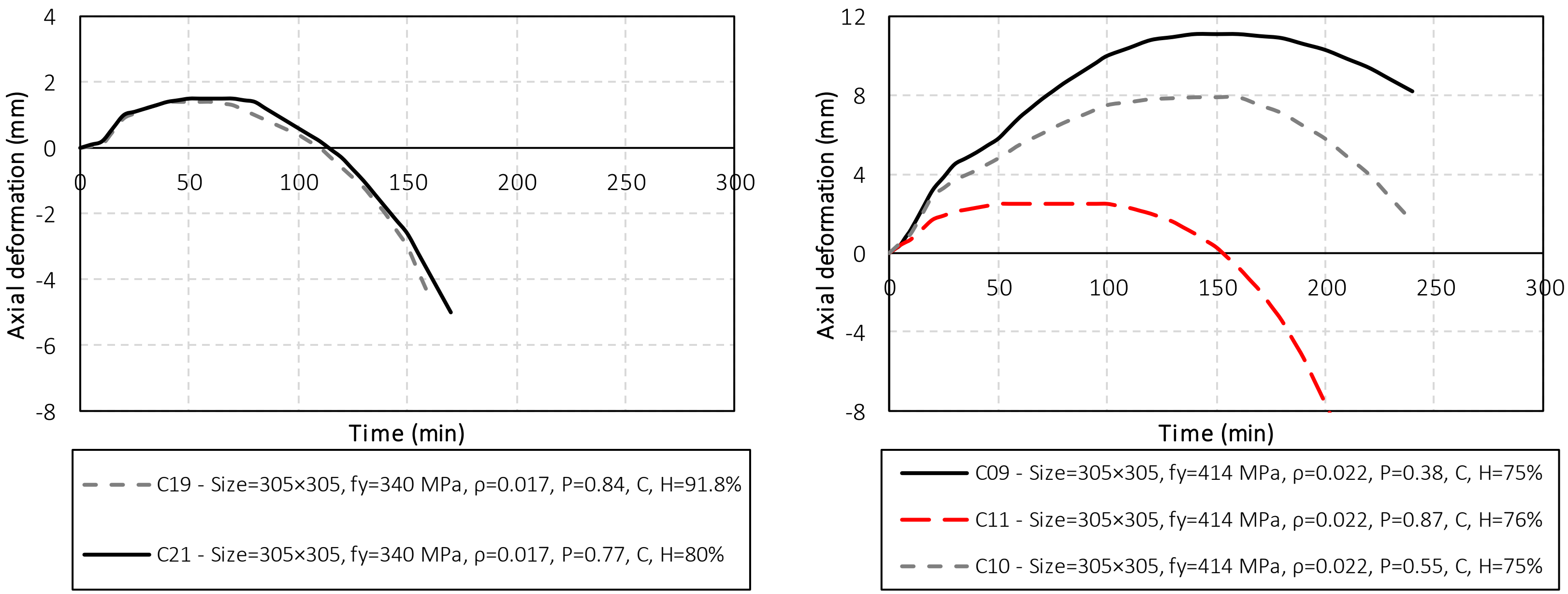
Please cite this paper as:
Naser M.Z., Çiftçioğlu A, (2023). "Revisiting Forgotten Fire Tests: Causal Inference and Counterfactuals for Learning Idealized Fire-induced Response of RC Columns". *Fire Technology*. https://doi.org/10.1007/s10694-023-01405-8.



(a) Silicate aggregate

This is a preprint draft. The published article can be found at: https://doi.org/10.1007/s10694-023-01405-8.

Please cite this paper as:
Naser M.Z., Çiftçioğlu A, (2023). "Revisiting Forgotten Fire Tests: Causal Inference and Counterfactuals for Learning Idealized Fire-induced Response of RC Columns". *Fire Technology*. https://doi.org/10.1007/s10694-023-01405-8.



(b) Carbonate aggregate

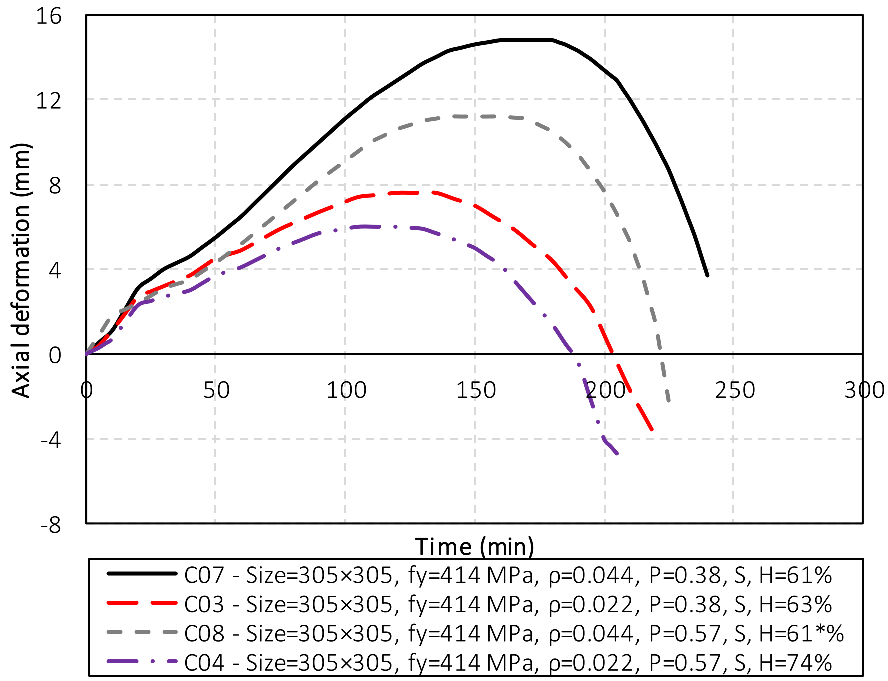
Fig. 5 Effect of loading

This is a preprint draft. The published article can be found at: https://doi.org/10.1007/s10694-023-01405-8.

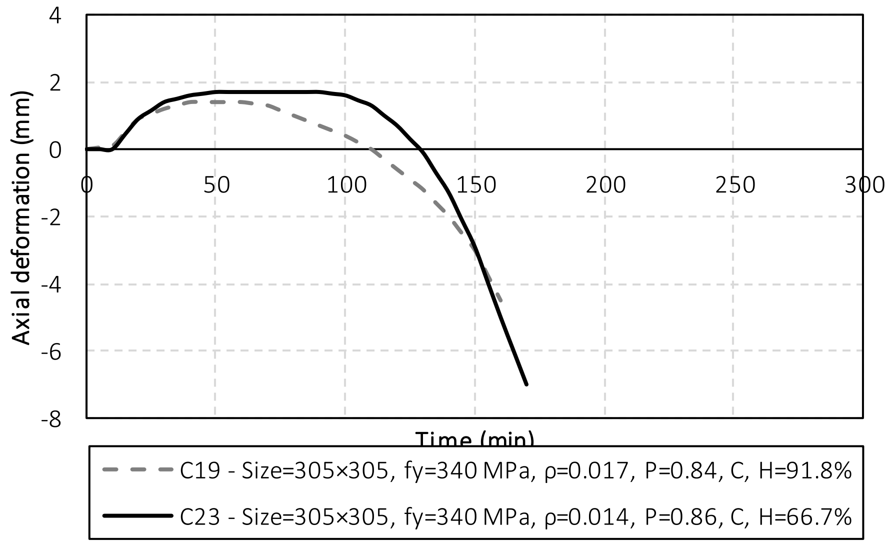
Please cite this paper as:
Naser M.Z., Çiftçioğlu A, (2023). "Revisiting Forgotten Fire Tests: Causal Inference and Counterfactuals for Learning Idealized Fire-induced Response of RC Columns". *Fire Technology*. https://doi.org/10.1007/s10694-023-01405-8.

<u>Effect of longitudinal steel ratio</u>

Figure 6 presents the effect of varying the longitudinal steel reinforcement on silicate and carbonate RC columns. It is evident that this effect is more pronounced in columns made from silicate aggregates. More specifically, columns with higher steel ratios tended to exhibit larger axial deformation. A clear kink appears around 30 min of fire, which also matches that which takes place as noted by increasing the load level. For lightly loaded columns (C03 and C07), this kink is quite large. The same also appeared to a lesser extent in C04 and C08. At the moment, a conclusive observation in the case of columns made from carbonate aggregate is not possible, given that the two most similar columns (C19 and C23) are heavily loaded at 84 and 86%, respectively.



(a) Silicate aggregates



(b) Carbonate aggregates

Fig. 6 Effect of steel ratio

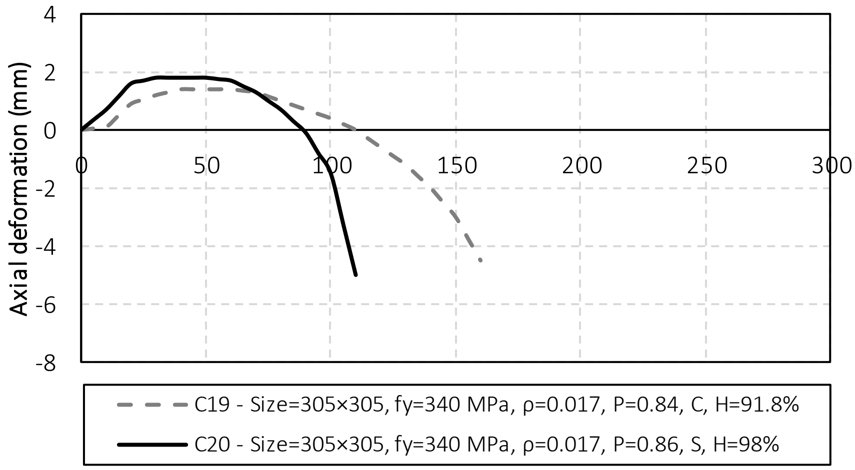
<u>Effect of aggregate type</u>

In general, columns made from carbonate and lightweight aggregates tend to have a fuller curvilinear deformation history than those made from silicate aggregates. This observation was valid for columns with low and medium loads. At higher loading levels, all columns, regardless of the type of aggregate, tended to have a small and short deformation history, indicating an accelerated failure.

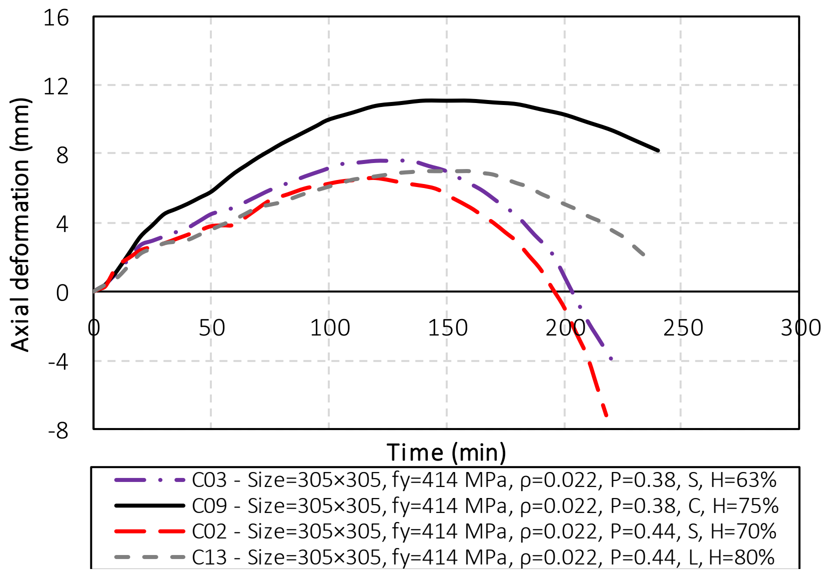
A look into Fig. 7a shows that C19 (silicate) and C20 (carbonate) share the most resemblance in the columns of grade 340 MPa. These two columns have a similar deformation history of up to 60 min, after which C20 starts to show signs of failure. Evidently, C19 and C20 failed at 164 min and 109 min, respectively. This shows the significant impact of carbonate aggregate on the response of columns, which marks an increase of about 1 h rating (60 min to 120 min). The same observation is also made in Fig. 7b by comparing C09 and C03 (but with a much larger variance at failure). Lightweight aggregates (C13) also outperform silicate aggregates (C02). A direct comparison between the columns of carbonate and silicate aggregates was not possible.

This is a preprint draft. The published article can be found at: https://doi.org/10.1007/s10694-023-01405-8.

Please cite this paper as:
Naser M.Z., Çiftçioğlu A, (2023). "Revisiting Forgotten Fire Tests: Causal Inference and Counterfactuals for Learning Idealized Fire-induced Response of RC Columns". *Fire Technology*. https://doi.org/10.1007/s10694-023-01405-8.



(a) Grade 340 MPa



(b) Grade 414 MPa

Fig. 7 Effect of aggregate type

Effect of shape

The effect of the column shape can be examined by comparing the deformation responses of the C15 (circular) and C10 (square) columns – see Fig. 8. It is clear that the circular column tends to undergo smaller deformation than the square column. However, both columns eventually seem to share the same response to failure. The reader is to remember that C10 is made from carbonate aggregate, whereas C15 is made from silicate aggregate. As such, the shown comparison is to be examined keeping the differences from the previous section in mind as the influence of aggregate type can be substantial.

This is a preprint draft. The published article can be found at: https://doi.org/10.1007/s10694-023-01405-8.

Please cite this paper as:
Naser M.Z., Çiftçioğlu A, (2023). “Revisiting Forgotten Fire Tests: Causal Inference and Counterfactuals for Learning Idealized Fire-induced Response of RC Columns”. *Fire Technology*. https://doi.org/10.1007/s10694-023-01405-8.

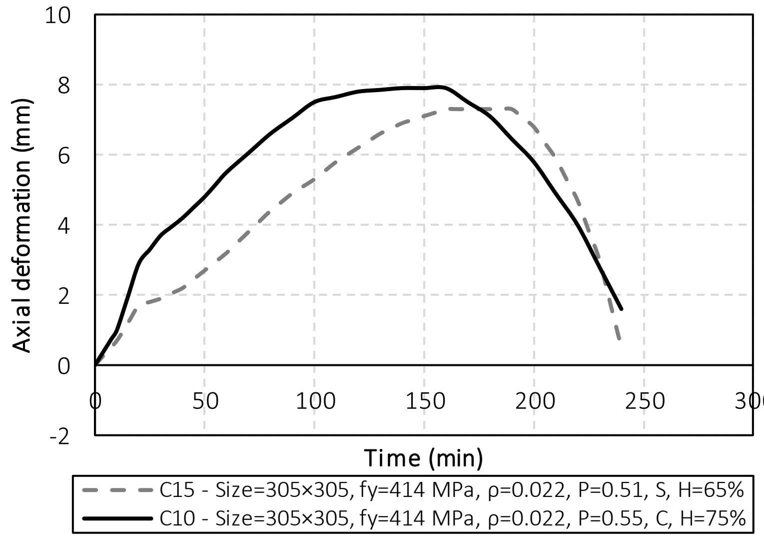


Fig. 8 Effect of shape [Note that C15 is a circular column]

<u>Effect of size</u>

Figure 9 shows the deformation history of four RC columns that share the most resemblance of all other columns. The closets of resemblance can be seen in C06 (305×305 mm) and C14 (406×406 mm). It can be seen that while both columns initially seem to have a similar deformation history, C15 continues to have a longer survivability under fire. In fact, C15 failed at 262 min versus 187 min, as shown in the case of C06. It should be noted that C16 and C17 are presented as the loading level, and the steel ratio significantly differs from those of C15 and C06.

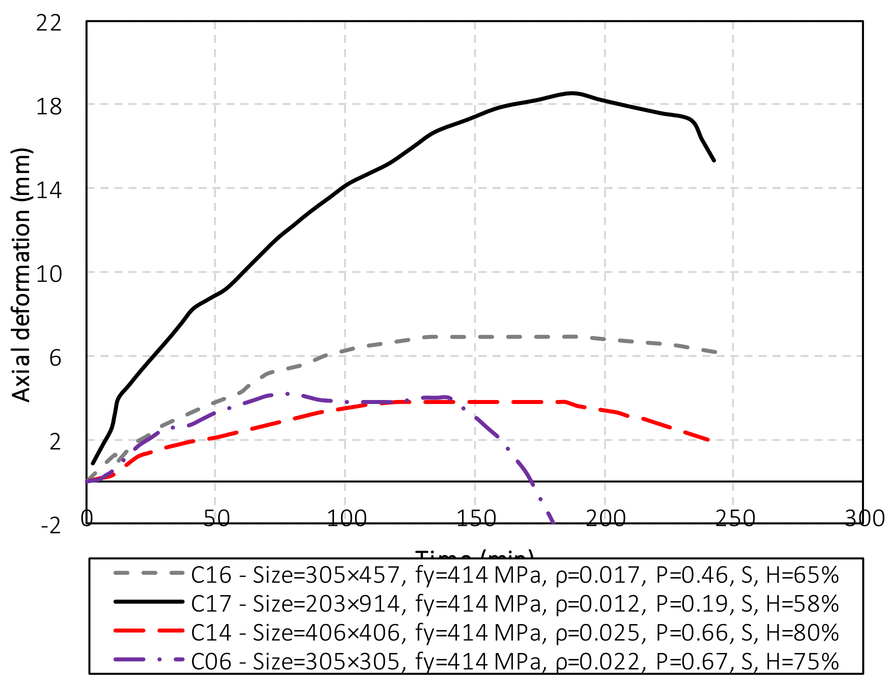


Fig. 9 Effect of cross section size

Effect of humidity (moisture content)

Unfortunately, the effect of humidity was unclear given the large variation between all tested columns with respect to this factor.

Comparison between all effects

This section presents a general comparison between all examined columns based on four distinct items: column size, load level, humidity, and yield strength. Figure 10 presents this comparison as a function of the aggregate type. It is worth noting that the vertical axis, horizontal axis, and size of the data points were fixed as the fire resistance time, steel reinforcement ratio, and loading levels in all sub-figures shown, respectively.

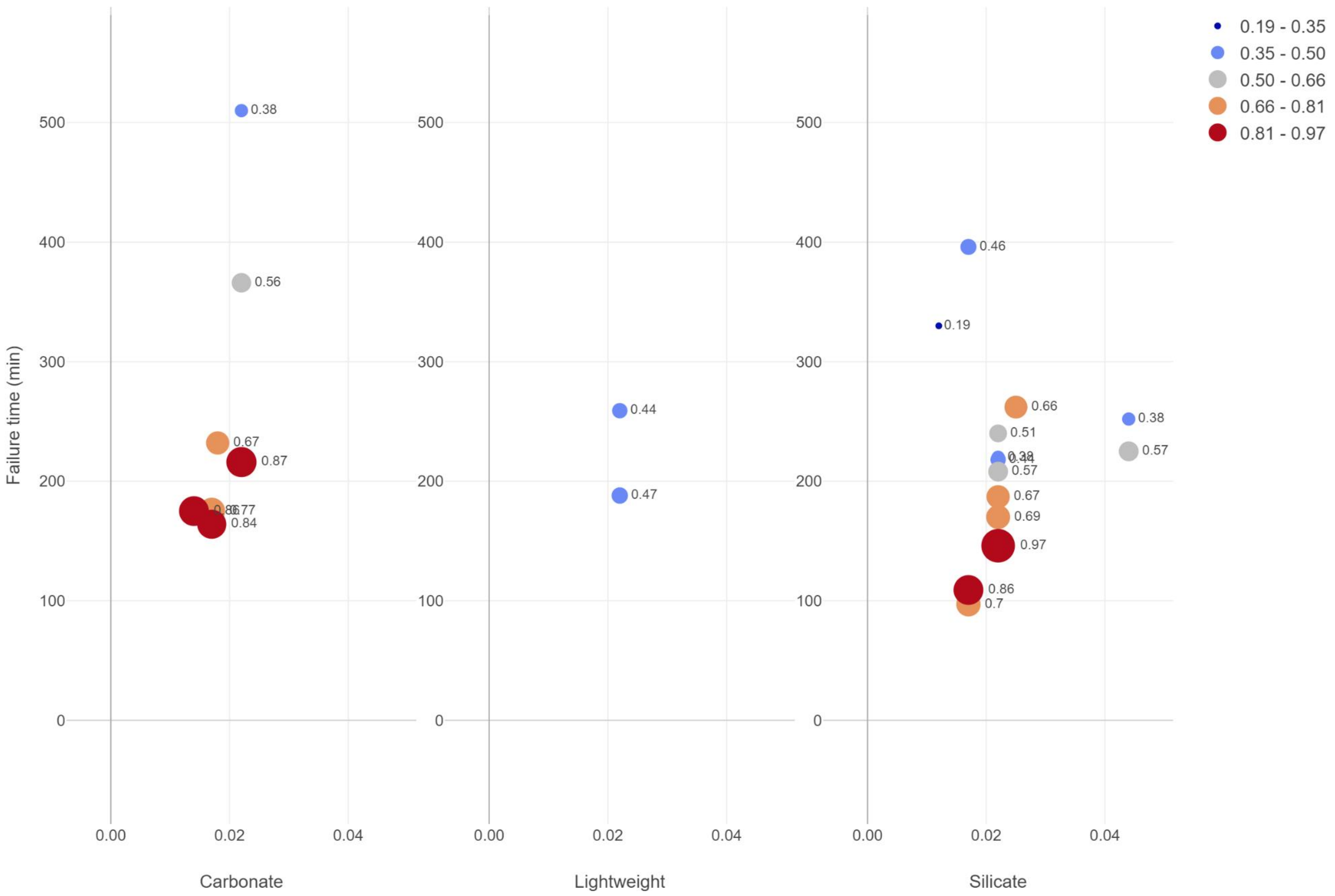
As we can see, Fig. 10a indicates that larger columns are associated with longer failure times. There is also a clear indication that the failure time is strongly associated with the level of loading. Further, Fig. 10b noted that heavily loaded RC columns tend to naturally have low fire resistance. Of these columns, all columns made from carbonate and lightweight aggregates passed the 2 hour mark and the majority exceed the three hour mark. The failure times of the columns made from silicate aggregates had a much wider range of failure times.

Given the large range of reported humidity values compared to the available columns, it is quite difficult to draw clear conclusions. However, the columns with the highest humidity failed in a relatively short time. As expected, columns made with reinforcement from a low steel grade failed at shorter failure times.

This is a preprint draft. The published article can be found at: https://doi.org/10.1007/s10694-023-01405-8.

Please cite this paper as:
Naser M.Z., Çiftçioğlu A, (2023). "Revisiting Forgotten Fire Tests: Causal Inference and Counterfactuals for Learning Idealized Fire-induced Response of RC Columns". *Fire Technology*. https://doi.org/10.1007/s10694-023-01405-8.

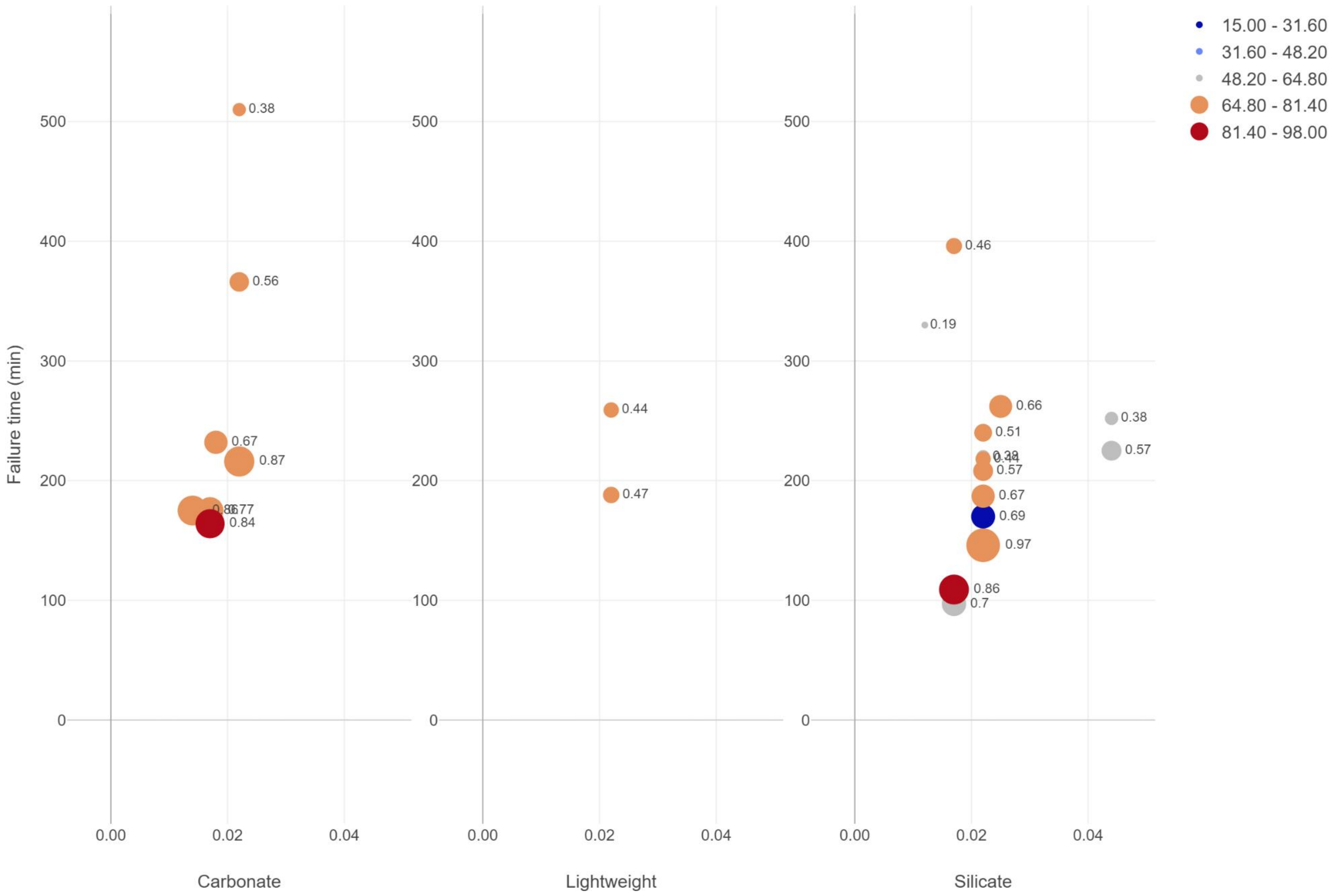
(a) Effect of section size



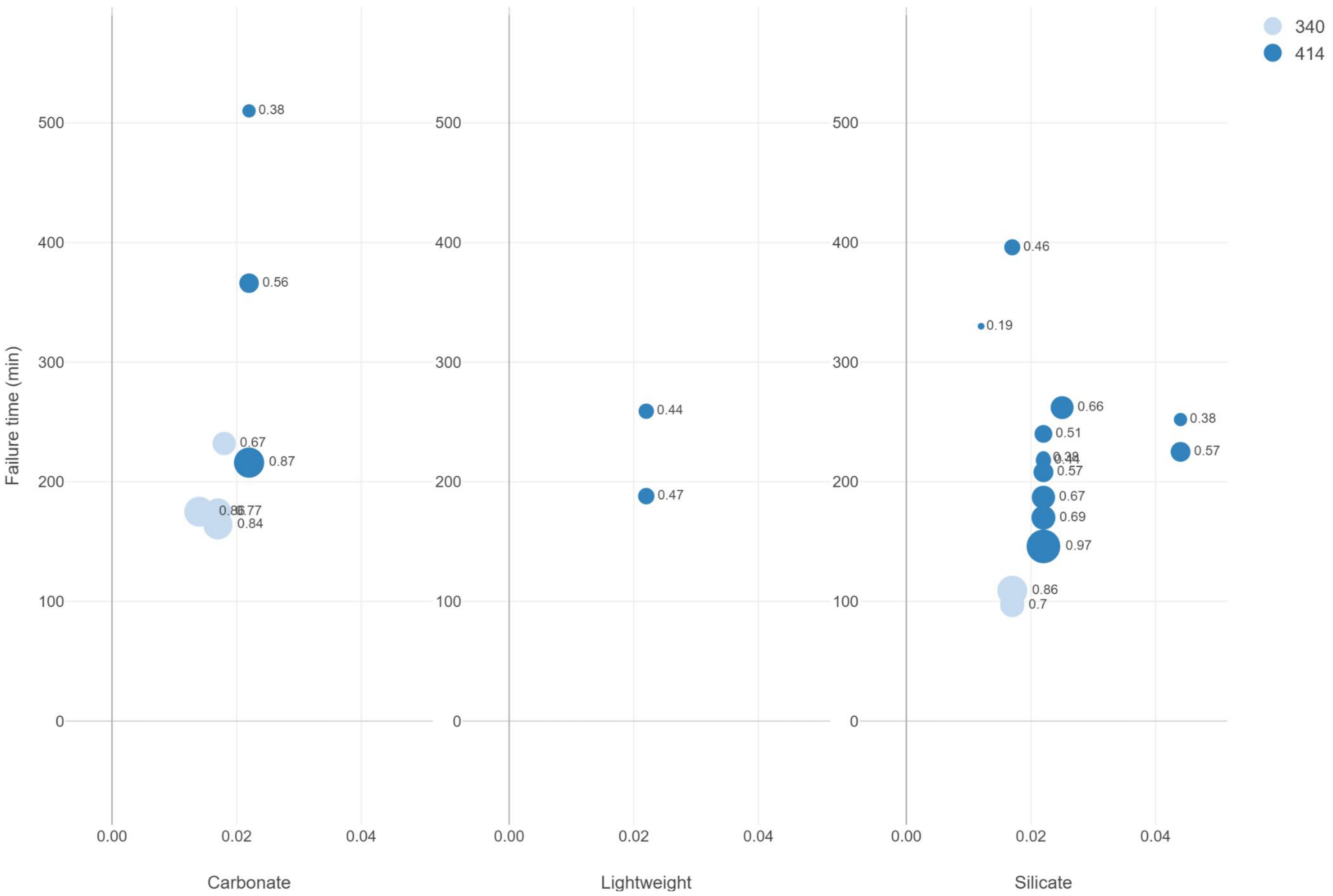
(b) Effect of loading level

This is a preprint draft. The published article can be found at: https://doi.org/10.1007/s10694-023-01405-8.

Please cite this paper as:
Naser M.Z., Çiftçioğlu A, (2023). "Revisiting Forgotten Fire Tests: Causal Inference and Counterfactuals for Learning Idealized Fire-induced Response of RC Columns". *Fire Technology*. https://doi.org/10.1007/s10694-023-01405-8.



(c) Effect of humidity



(d) Effect of yield strength

Fig. 10 Comparison between examined RC columns [Note: the value next to each data points represents the load level, and the horizontal axis represents the ratio of steel reinforcement].

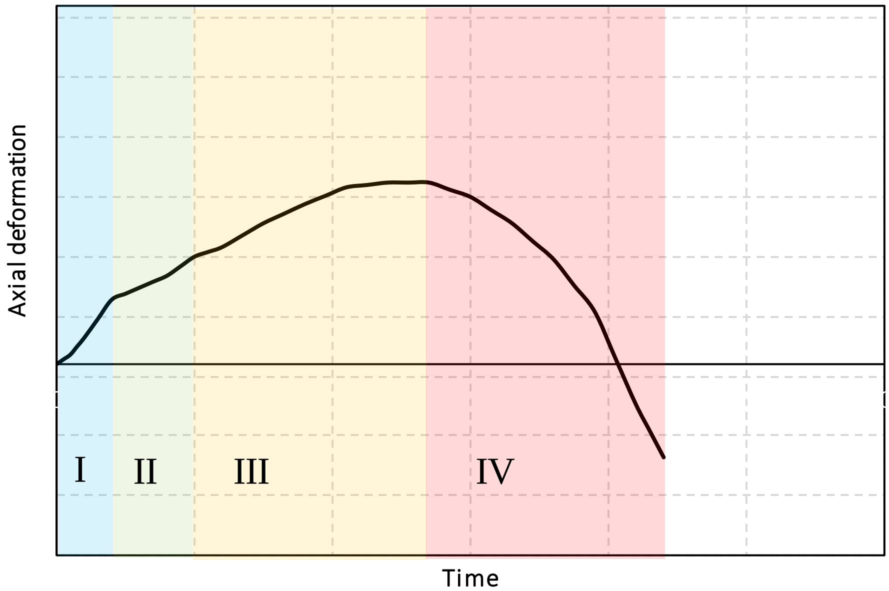
Idealized deformation response under fire conditions

Based on the above comparisons, the deformation history of the fixed-fixed RC columns made from silicate aggregates can be simplified into four stages (see Fig. 11). The first stage of deformation is marked with a rise that takes place at–30-45 degrees and continues for approximately 15-20 min irrespective of the features and loading of the column. The magnitude of this deformation was small. This initial slope reduces by approximately half for columns with a steel ratio of approximately 2% and approximately a third for columns with a higher steel ratio (i.e., 4%).

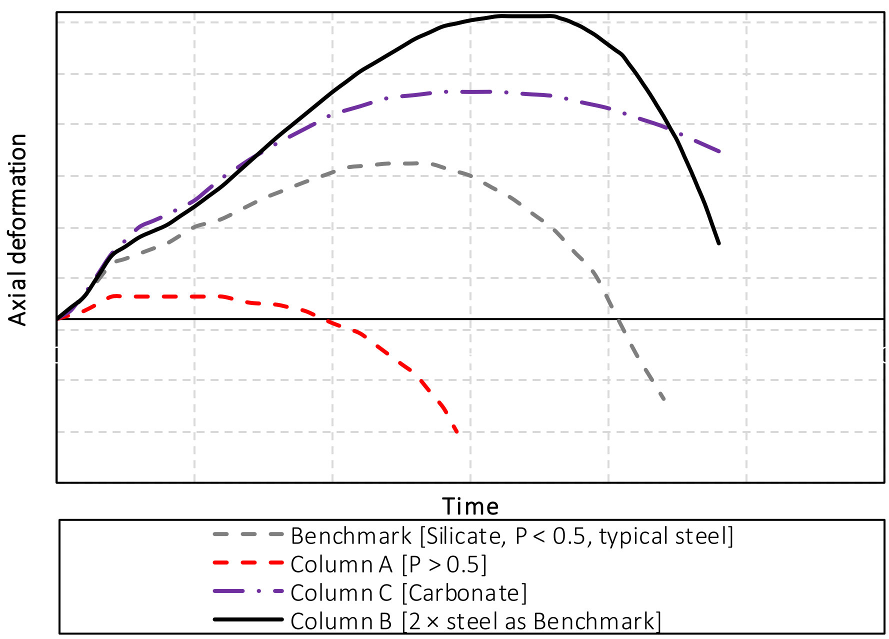
This rotation marks the second stage of deformation, and its magnitude and slope are tied to the available longitudinal steel and level of applied loading, wherein lighter loads lead to a slight reduction in the slope, and higher loads rotate this slope more towards the horizontal.

Beyond this stage, the deformation continues to rise at a slow rate until it peaks, which could occur within a few hours. The third stage was often the longest. Such a peak marks the end of the third and the start of the final stage, after which the column shifts from an expansion mode into a contraction mode. At this stage, the rate of deformation increases until it is almost parallel to the vertical axis. This stage often lasts for 10-40% of the total fire exposure duration. In other words, once the column shifts its mode, it is likely that such a column is a near failure. For example, if this occurs at 120 min, the column is very likely to fail within the next 20-80 min. The analysis in the previous section clearly shows that the duration of each stage, as well as the associated deformations, is highly dependent on the loading level.

A description similar to that outlined above can also be seen in the case of RC columns made from carbonate aggregate (despite the fact that the influence of loading is much more pronounced because all but one column were heavily loaded). The key difference between the two types of columns is that the transitions of a column made from carbonate aggregate are much smoother, implying higher endurance (longer time to failure with a minimum of an additional 75 min) under fire.



(a) Demonstration of the four stages



(b) Comparison between idealized columns

Fig. 11 A look into idealized response under fire

The above idealization sets the foundation for establishing an approach to estimate the deformation curve of RC columns. For example, looking at all the curves presented so far, we can deduce that the deformation response of RC columns under standard fire conditions is likely to follow a curvilinear trend that can best fit via a polynomial form. This idealization breaks free from the thermo-mechanical coupling often used in fire analysis and allows an engineer to predict the deformation response without the need to perform thermal analysis, given that such a response is expected under standard fire conditions.

It is likely for such an idealization to transform the fire response of RC columns, and possibly other members, into a scaling problem; wherein if a benchmark behavior is selected, then future responses of variants of such behavior can be deduced with moderate to a reasonable accuracy. Such a practice already exists and is often titled as *rules of thumb*. In this case, these rules of thumb were arrived at by comparing columns of identical or similar features in Figs. 1-9. The same will also be examined via casual assumptions in a later section of this paper.

Thus, a standard regression analysis was conducted to derive two empirical formulas that can be used to plot the deformation history of RC columns, taking into account the loading level, reinforcement yield strength and ratio, and fire exposure time. Figure 12 shows a visual comparison of the predictivities of these formulas. As can be observed, these expressions achieved good performance metrics. With this accuracy in mind, these formulas may underestimate the deformation in the final stage of a fire exposure of 1-3 mm.

Deformation history of the silicate RC columns

$$Deformation\ history = P + 0.0003T \times S + 7.056T^2 \times G^3 - 0.138T \times P - 8.111 \times 10^{-5} \times G \times T^3 \quad (1)$$[2]

MAE = 1.0 mm, $R^2 = 0.84$.

Deformation history of the carbonate and lightweight RC columns

$$Deformation\ history = 71945.75 + 0.0002T \times S - 8.46 \times 10^{-5} \times T^2 - 71946.23 \times tanh(14.024P) - 0.00056P \times T^2 \quad (2)$$

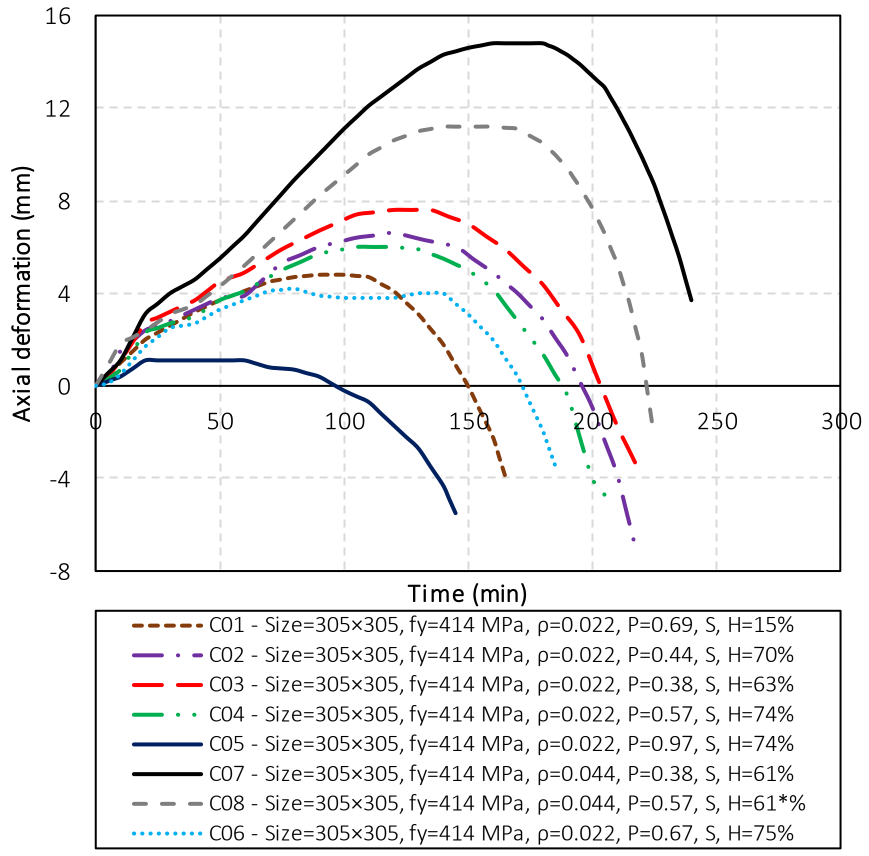
MAE = 0.40 mm, $R^2 = 0.96$.

Please note that: *S*: yield strength of steel (MPa), *T*: time under standard fire (min), *P*: loading level (%), and *G*: steel ratio (%). These expressions are verified for the columns of 305×305 sections int his study.

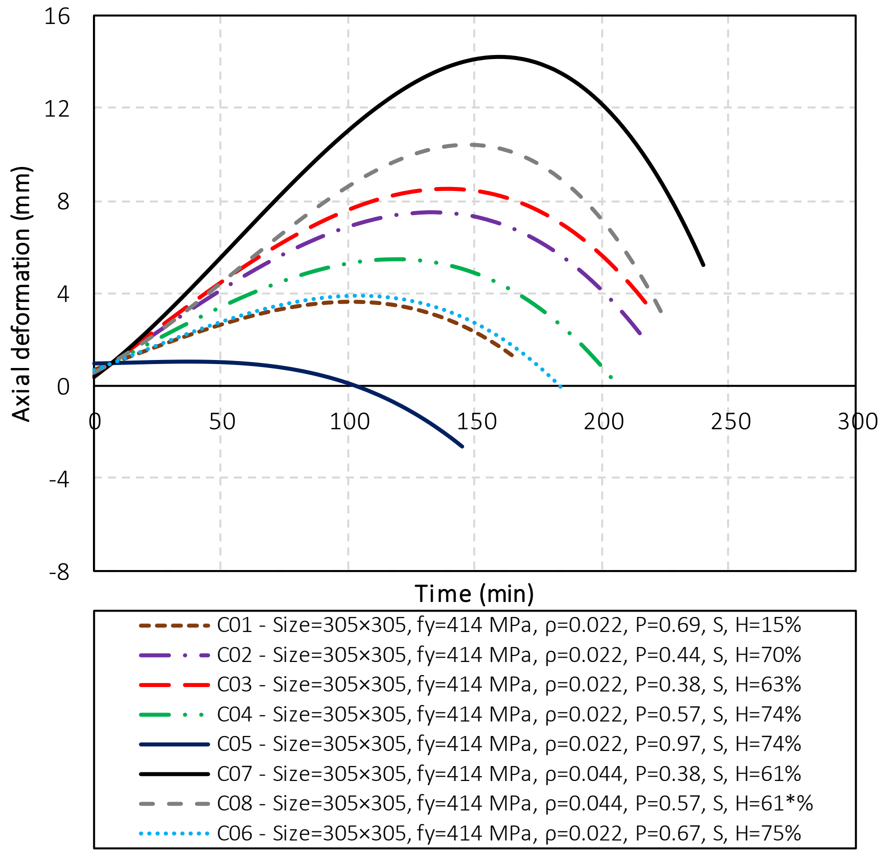
[2] Please note that, Mean Absolute Error (MAE), and Coefficient of Determination (R^2).

This is a preprint draft. The published article can be found at: https://doi.org/10.1007/s10694-023-01405-8.

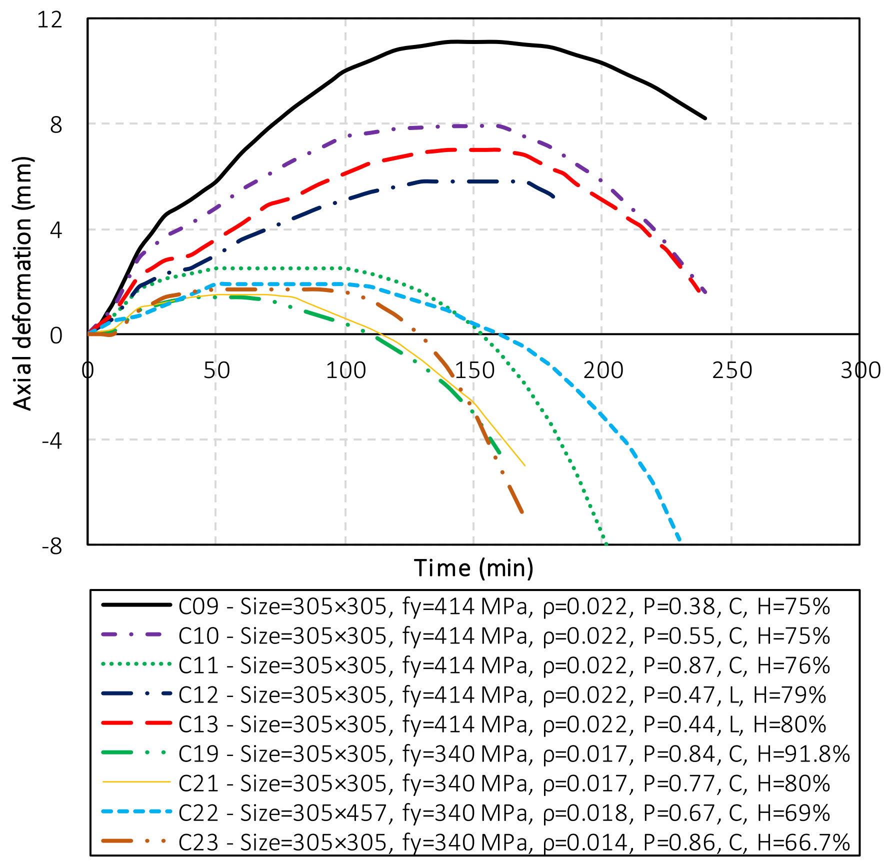
Please cite this paper as:
Naser M.Z., Çiftçioğlu A, (2023). "Revisiting Forgotten Fire Tests: Causal Inference and Counterfactuals for Learning Idealized Fire-induced Response of RC Columns". *Fire Technology*. https://doi.org/10.1007/s10694-023-01405-8.



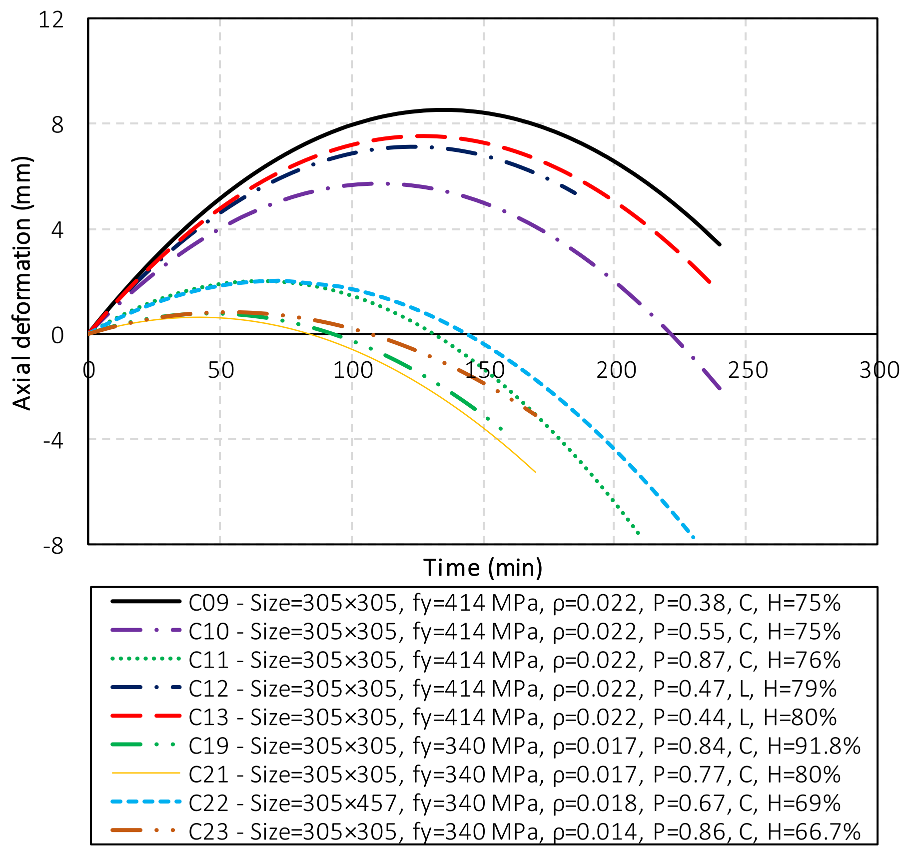
(a) Measured response of silicate RC columns



(b) Predicted response of silicate RC columns



(c) Measured response of carbonate and lightweight RC columns



(d) Predicted response of carbonate and lightweight RC columns

Fig. 12 Predictivity of newly derived formulas

Data-driven ML model and analysis

In an effort to maintain the theme of this paper, the results of one ML algorithm are shown herein. This algorithm is a light gradient boosted tree (LGBT) and was selected in the aftermath of a sensitivity analysis that included two more algorithms (XGBoost and Random Forest). For brevity, the results of this sensitivity analysis are not shown herein.

The LGBM is a tree-based algorithm built upon the success of the original AdaBoost algorithm [22]. Unlike the Random Forest algorithm, LGBM fits the trees in a successive manner and then fits their residual errors in each iteration and focuses on those errors to improve its predictivity. The used algorithm can be found online at [23] with the following default settings: learning rate = 0.05, maximum depth = "none," number of boosting stages = 1000, etc.

In addition, our dataset is healthy as it contains 9081 data points and satisfies the conditions set by:

- Van Smeden et al. [24] – having a minimum set of 10 observations per feature.
- Riley et al. [25] – having a minimum of 23 observations per feature.
- Frank and Todeschini [26] – maintaining a ratio of 3 and 5 between the number of observations to the number of features.

The LGBM was trained using collected data. First, the data were randomly shuffled and split into training (T), validation (V), and testing (S) sets. The model was trained and validated against the T and V sets and then examined on the S set. The LGBM was trained following a *k*-fold cross-validation procedure, wherein the collected dataset was randomly split into test and training sets of $k = 10$ groups. The model was trained using *nine* sets and validated on the tenth set. This training was repeated *ten* times until each unique set was used as the validation set.

The performance of the model was then quantified using three metrics: the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R^2). These metrics are commonly accepted in structural fire engineering publications [27,28] and are listed in Table 3. MAE represents the mean average error of all observations. Thus, the low MAE values were favorable. R^2 is the square of the coefficient of correlation (r) and measures the degree of association between the observed and predicted values. Higher positive R^2 values indicate a strong and positive prediction capability. The RMSE describes the model errors in a scale-independent fashion, with lower values representing a high prediction capability. Finally, the behaviour of the model was visually examined and is deemed suitable as seen in Fig. 13.

Table 3 List of common performance metrics.

Metric	***Formula***	***T***	***V***	***S***
MAE	$MAE = \frac{\sum_{i=1}^{n}\lvert E_i\rvert}{n}$	0.371	0.322	0.282
RMSE	$RMSE = \sqrt{\frac{\sum_{i=1}^{n} {E_i}^2}{n}}$	0.604	0.533	0.461
R^2	$R^2 = 1 - \sum_{i=1}^{n}(P_i - A_i)^2 \,/ \sum_{i=1}^{n}(A_i - A_{mean})^2$	0.981	0.982	0.982

This is a preprint draft. The published article can be found at: https://doi.org/10.1007/s10694-023-01405-8.

Please cite this paper as:
Naser M.Z., Çiftçioğlu A, (2023). "Revisiting Forgotten Fire Tests: Causal Inference and Counterfactuals for Learning Idealized Fire-induced Response of RC Columns". *Fire Technology*. https://doi.org/10.1007/s10694-023-01405-8.

E: error, A: actual measurements, P: predictions, and n: number of data points.

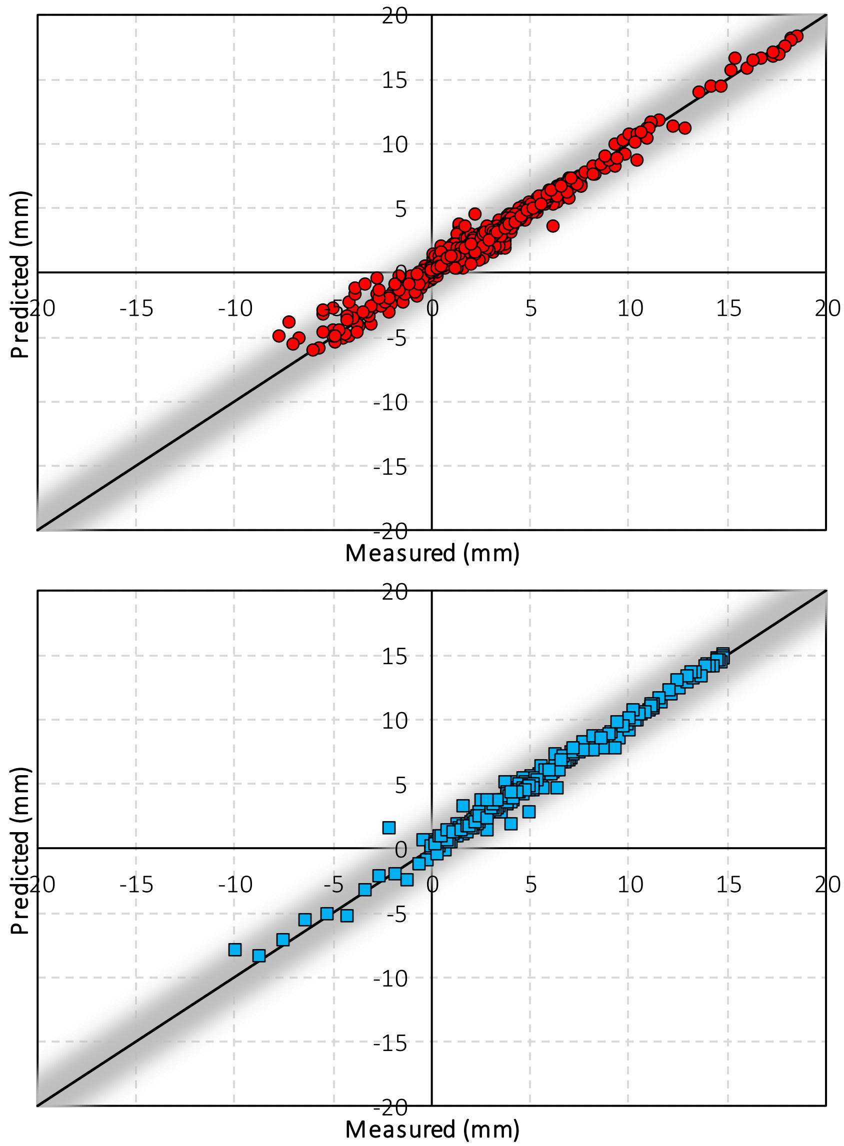


Fig. 13 Comparison against LGBM predictions [Note: training data in red circles and validation and testing data in blue squares]

Causal inference analysis

Causal inference aims to identify causal relationships between variables. This inference process can be broken down into three stages: identification, estimation, and refutation. In the identification stage, a list of potential causal variables is created. In the estimation stage, these variables are constructed, and their effects on the outcome are estimated. Finally, in the refutation stage, the causal conclusion is tested by creating a list of potential confounding variables and checking whether their effects are significant.

A standard causal analysis has three steps. In the first, a causal discovery process is used to uncover the underlying structure between the selected features (which can be identified from domain knowledge or physical tests). The underlying structure is then built by satisfying three causal principles, namely the Markov causal assumption, the causal faithfulness assumption, and the causal sufficiency assumption [29–31]. The Markov causal assumption states that a variable is independent of all other variables (except its *own* effects) conditional on its direct causes. This assumption is checked via the *d-separation* criterion [29], which entails whether a variable is independent of another given a third by associating independence. The *casual faithfulness* assumption states that a causal graph has independent relations through the d-separation criterion. The c*ausal sufficiency* assumption refers to the absence of hidden or latent parameters that we do not know nor are aware of. The readers are invited to review the following work for a detailed discussion on each of the aforenoted stages [32–34]. The readers are also to note how these three assumptions are not present in commonly adopted statistical methods, which also serves to contrast these two methods.

In this paper, we carry out our causal inference analysis using the Python-based DoWhy and EconML packages. The DoWhy library [35], a Bayesian graphical model for causal inference, provides three key contributions to causal inference models. First, it provides a principled way of modeling problems as causal graphs by explicitly expressing all underlying assumptions so that they can be used later in calculations and predictions. Second, it unifies many popular methods of causal inference that use the graphical approach and potential outcomes approach to causality. Third, the model automatically checks if the estimates are valid or not (if possible) and assesses their robustness [36].

The graphical causal model (GCM) in DoWhy is a probabilistic linear graphical model that has been developed to provide a framework for representing and reasoning causal relationships. GCM-based inference generates counterfactuals for future scenarios by considering what would happen if a variable changes or stays unchanged [37]. Unlike predictions via regression, which assumes the world is constant, in counterfactual prediction, specific aspects of the world are predicted using data as if the world were different. Counterfactual explanations can be used to justify forecasts of specific instances in interpretable ML. The event is what the machine has predicted to happen as a result of input values, and causes are its particular inputs that predicted this outcome [38].

On the other hand, EconML [39] estimates individualized causal responses from different types of data, such as observational or experimental using a nonlinear causal model. This package is designed to allow users to easily explore the effects of various models and features on causal estimates and to provide tools for estimating average treatment effects with small samples. It provides an interface for estimating individualized causal responses from observational data, with a focus on the interpretability of estimates. It includes the estimation of the parameters of linear and nonlinear models using maximum likelihood methods, as well as an inference based on those parameter estimates [40]. A complete discussion of both of these packages can be found in their original sources cited above.

This is a preprint draft. The published article can be found at: https://doi.org/10.1007/s10694-023-01405-8.

Please cite this paper as:
Naser M.Z., Çiftçioğlu A, (2023). "Revisiting Forgotten Fire Tests: Causal Inference and Counterfactuals for Learning Idealized Fire-induced Response of RC Columns". *Fire Technology*. https://doi.org/10.1007/s10694-023-01405-8.

Analysis

We analyze causal relationships in three different ways that all contribute to causal interpretation: linear estimation (DoWhy), EconML estimation (nonlinear), and counterfactual estimation.

In order to run the causal analysis with linear estimation, we initially need to import the relevant libraries (DoWhy or EconML). We provide an estimation using the linear estimation model after identifying the causal model (see Fig. 14). Then, the treatment values are determined, and a causal model is established between the input variables, output variables, and treatments. The treatment values used the average values in each of the selected variables as obtained from our dataset of the columns examined by Lie. Finally, a refutation process that allows us to evaluate the accuracy of model predictions is carried out. This process includes three tests:

- Random Common Cause: Adds randomly drawn variables to the database and re-runs the analysis to see if the causal estimate changes or not. The causal estimate shouldn't change by much due to a random variable.
- Data Subset Refuter: Creates subsets of the data and checks whether the causal estimates vary across subsets. In order to effectively measure causation, there should not be large variances in the estimates.
- Placebo Treatment Refuter: Randomly assigns a variable as a treatment and re-runs the analysis. If a causal relationship exists, then the causal estimate will move toward zero.

This observed data and the new value of the input in it to be changed are defined. This provides us with counterfactual values of what would happen if we changed our specified input (namely, Humidity, H, Aggregate type, A, loading level, P, yield strength of steel, S, steel ratio, G, and exposure time, T) in the observed data, with no other changes.

Causal structure

Our causal model (i.e., directed acyclic graph (DAG)) disregards the effects of all variables on each other and assumes that they only have an influence on the deformation history, as shown in Fig. 14. In this DAG, we assumed that all variables only have a direct causal link with the deformation history (i.e., without any inter-relation to other variables).

This is a preprint draft. The published article can be found at: https://doi.org/10.1007/s10694-023-01405-8.

Please cite this paper as:
Naser M.Z., Çiftçioğlu A, (2023). "Revisiting Forgotten Fire Tests: Causal Inference and Counterfactuals for Learning Idealized Fire-induced Response of RC Columns". *Fire Technology*. https://doi.org/10.1007/s10694-023-01405-8.

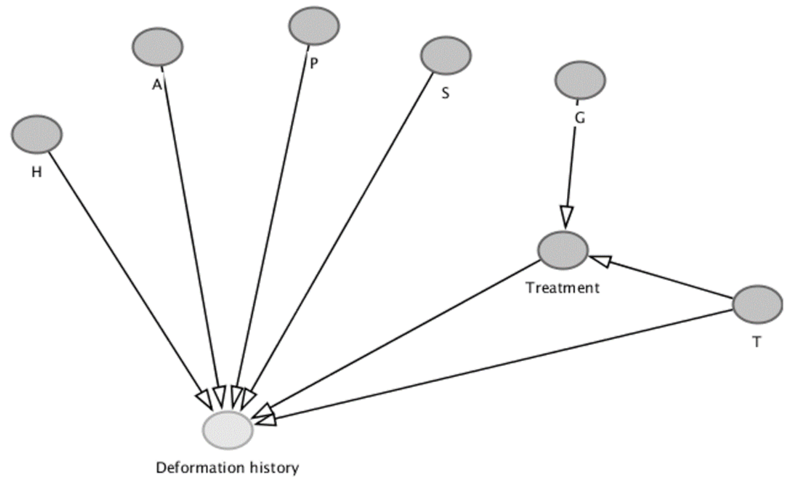


Fig. 14 Hypothetical model [Note: *T*: intervention/treatment]

Tables 4 and 5 list the impacts of each feature on the axial deformation (Y) from the linear and nonlinear models. It is worth noting that there is good agreement between the linear and nonlinear causal models. The analysis from the selected DAG seems to satisfy all refuting models and hence can be deemed successful. It is interesting to note that the impact of loading level and steel grade is the largest on the axial deformation of RC columns exposed to fire.

Table 4 Results of the *DoWhy* analysis in terms of output (linear model)

Treatment variable	Estimate		Refute					
	Mean value	p-value	Random Common Cause	p-value	Data Subset Refuter	p-value	Placebo Treatment	p-value
S	5.52	-	5.12	0.06	5.65	0.56	0.023	1.24
G	3.70	-	3.75	0.84	3.91	0.62	-0.03	1.38
P	5.92	-	5.38	0.00	5.83	0.64	-0.01	1.46
A	1.31	-	0.08	0.18	1.62	0.30	-0.02	1.46
H	3.14	-	3.19	0.78	3.25	0.76	-0.0089	1.28

Table 5 Results of the *DoWhy* analysis in terms of output (nonlinear model)

Treatment variable	Estimate		Refute					
	Mean value	p-value	Random Common Cause	p-value	Data Subset Refuter	p-value	Placebo Treatment	p-value
S	4.69	0.26	4.69	0.88	4.69	0.89	-0.03	0.94
G	4.92	4.12e-05	4.92	0.92	4.93	0.96	0.038	0.92
P	4.64	0.03	4.64	0.86	4.63	0.89	4.64	0.94

This is a preprint draft. The published article can be found at: https://doi.org/10.1007/s10694-023-01405-8.

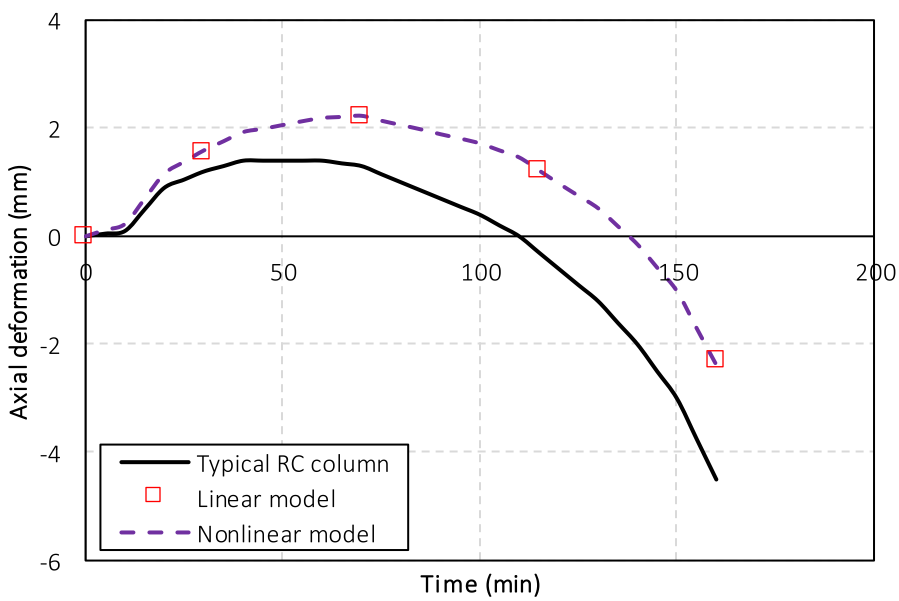
Please cite this paper as:
Naser M.Z., Çiftçioğlu A, (2023). "Revisiting Forgotten Fire Tests: Causal Inference and Counterfactuals for Learning Idealized Fire-induced Response of RC Columns". *Fire Technology*. https://doi.org/10.1007/s10694-023-01405-8.

A	1.59	0.77	1.59	0.94	1.60	0.98	-0.008	0.89
H	3.24	0.33	3.24	0.86	3.24	0.94	0.037	0.94

Causal inference

Now that the causal model (DAG) is verified, we can use this model to infer the deformation history under fire conditions for a variety of interventions. For the sake of this discussion, we will showcase the *average* change in the deformation history in columns. To simplify this process, we selected column C19 as a representative column. The *average* change in the deformation history of this column was then inferred when aggregates changed from carbonate to silicate when the steel reinforcement ratio is increased to 2.4%, when humidity and loading were reduced to 60%, and when the yield strength of steel increased to 414 MPa (from 340 MPa).

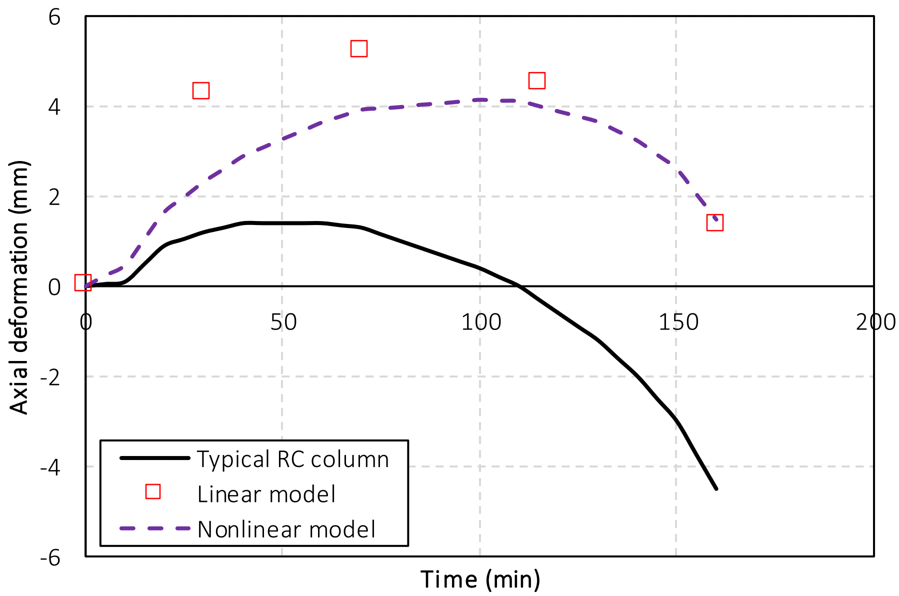
As one can see in Fig. 15, the inferred axial responses match those identified by the rules of thumb obtained from examining the fire tests conducted by Lie et al. Overall, both the linear and nonlinear models have similar trends but can vary in the magnitude of the obtained deformation. Noting how the deformation varies in terms of a few millimeters, this difference can be neglected, and we can focus our comparison on the obtained responses. We advocate for the use of a nonlinear model given that it can provide the full deformation response as opposed to the linear model (which can only generate a discrete number of points). A future work will extend the created causal model to infer the failure and fire resistance of RC columns.



(a) When the aggregate type changes from carbonate to silicate

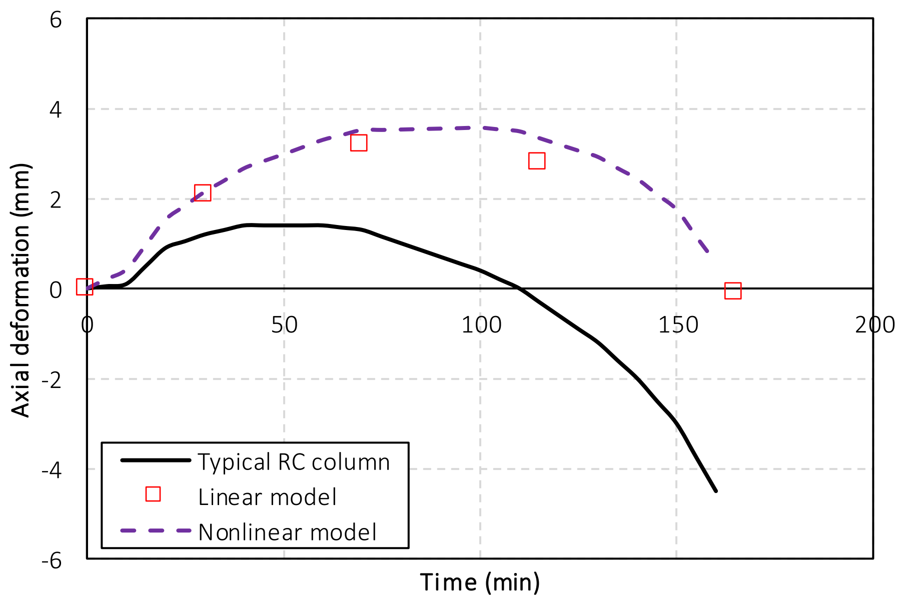
This is a preprint draft. The published article can be found at: https://doi.org/10.1007/s10694-023-01405-8.

Please cite this paper as:
Naser M.Z., Çiftçioğlu A, (2023). "Revisiting Forgotten Fire Tests: Causal Inference and Counterfactuals for Learning Idealized Fire-induced Response of RC Columns". *Fire Technology*. https://doi.org/10.1007/s10694-023-01405-8.



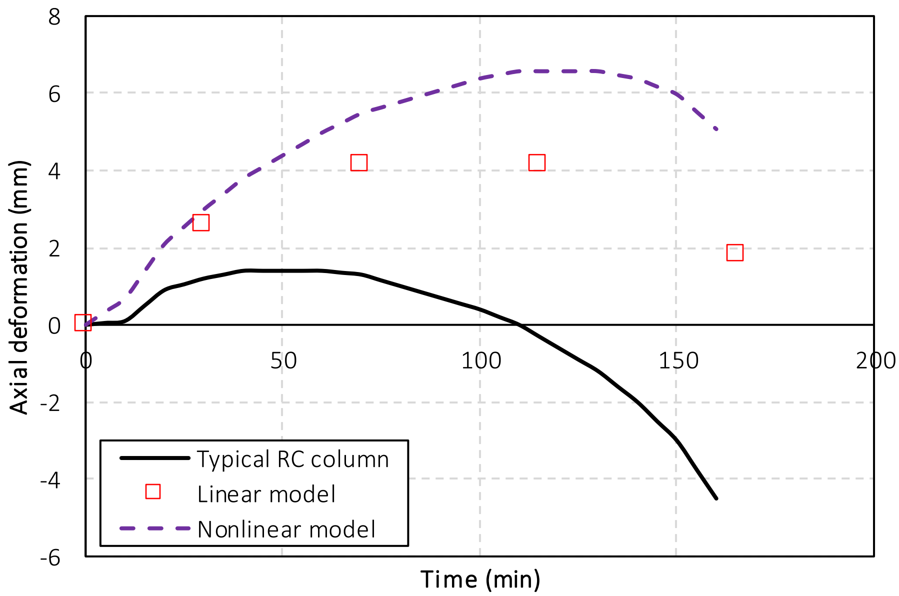
494

495 (b) When the steel reinfrcment ratio is larger than 2.4%

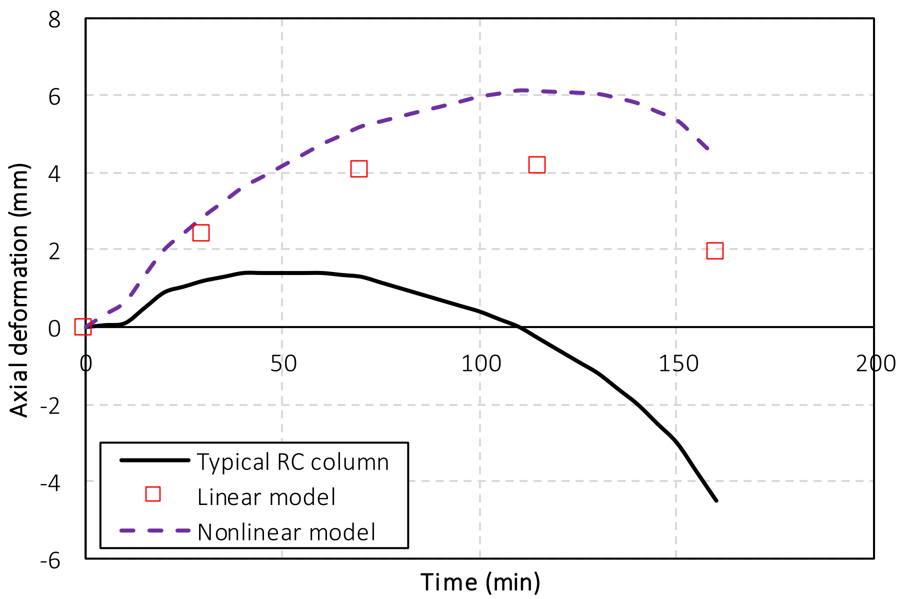


496

497 (e) When humidity is about 60%



(f) When the load level is less than 60%



(g) When the yield strength of steel increases to 414 MPa

Fig. 15 Effect of possible interventions upon a typical RC column (C019) [Note: the vertical axis is different for different cases]

Counterfactuals

A look into Fig. 13 and Table 3 shows that the proposed formulas and the LGBM performs well and can predict the deformation history of the examined RC columns. So, formula no. 2 and the ML model were used to predict the deformation response of a randomly selected column (C19 to continue our comparison, which had carbonate aggregate, f_y = 340 MPa, and P = 84.1%) if it was to be made from silicate aggregates, if it had steel reinforcement of Grade 414 MPa, and if it was subjected to a loading level of 50%.

As the reader can see, all three if-based questions are hypothetical and were not observed in the fire tests. The reader is also to note that here, we are inferring what would the deformation history of C19 be had it been made with the above features, as opposed to what is the *average* change in the history deformation in this column as presented in the previous section.

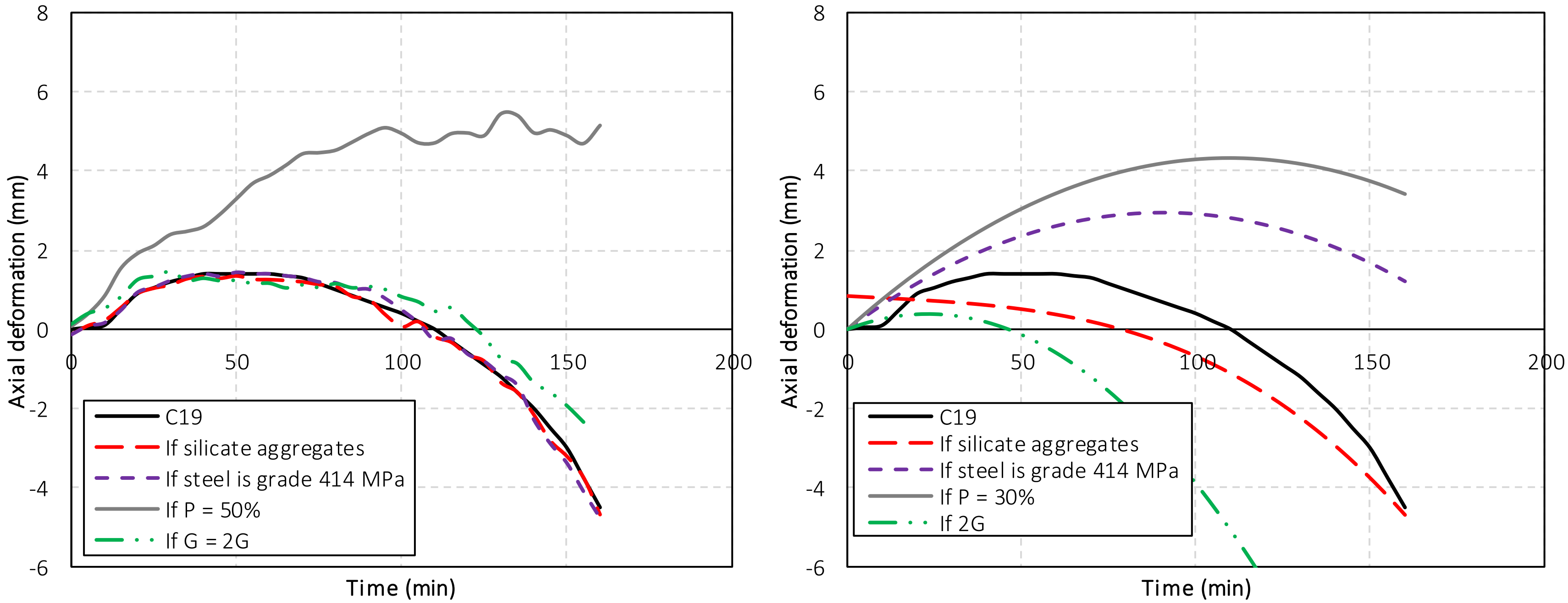
Figure 16 shows the outcomes of the data-driven, formula, and counterfactual inference predictions. It is quite clear that the ML predictions are unlikely to be correct as they 1) do not conform to the findings noted in the previous section, 2) do not show that changing the yield strength, nor aggregate type, seem to affect the deformation response, and 3) reducing the load level from 84% to 50% while leads to more deformation; however, this deformation continues to rise awkwardly.

On the other hand, the majority of formula predictions match the same rules of thumb identified earlier with the exception of the prediction for the increase in steel ratio. This is likely due to the fact that the derived formula for the RC columns of carbonate aggregate did not have as many columns with larger steel ratios (the reader is to note that the two columns with the largest steel ratios were made from silicate aggregate).

Finally, the results of the causal model seem to be the most realistic of all the other presented approaches (especially throughout the full history). These results match that from the fire tests as well as inferences made in the previous section.

This is a preprint draft. The published article can be found at: https://doi.org/10.1007/s10694-023-01405-8.

Please cite this paper as:
Naser M.Z., Çiftçioğlu A, (2023). "Revisiting Forgotten Fire Tests: Causal Inference and Counterfactuals for Learning Idealized Fire-induced Response of RC Columns". *Fire Technology*. https://doi.org/10.1007/s10694-023-01405-8.

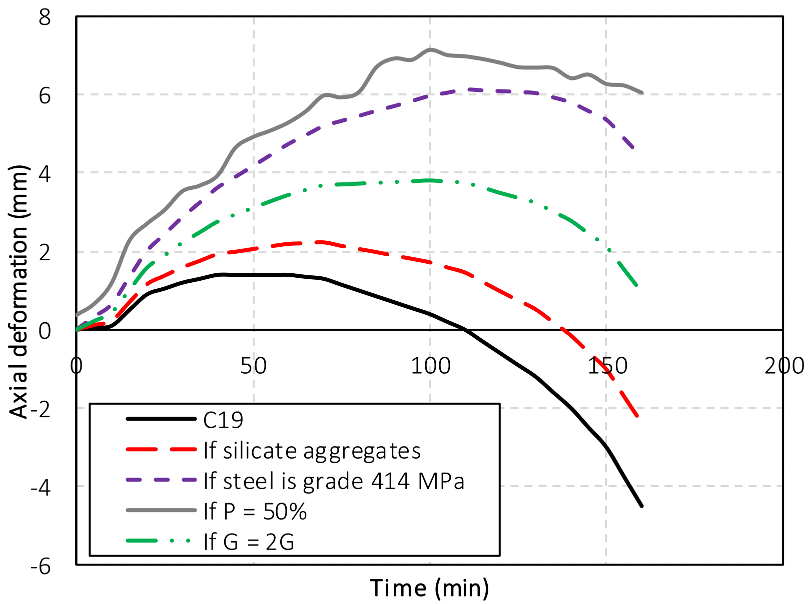


(a) ML predictions

(b) Predictions from formulas

This is a preprint draft. The published article can be found at: https://doi.org/10.1007/s10694-023-01405-8.

Please cite this paper as:
Naser M.Z., Çiftçioğlu A, (2023). "Revisiting Forgotten Fire Tests: Causal Inference and Counterfactuals for Learning Idealized Fire-induced Response of RC Columns". *Fire Technology*. https://doi.org/10.1007/s10694-023-01405-8.



(c) Inferences from causality [using the nonlinear model]

529 Fig. 16 Comparison between predictions

Conclusions

This paper adopts causal inference principles to address causal and counterfactual questions pertaining to the fire response of RC columns in order to overcome the limitations of statistical and machine learning methods. The results of the conducted analysis show that the deformation history of fire-exposed columns is heavily influenced by the amount of loading level, type of used aggregates, and magnitude of longitudinal steel ratio. The following inferences can also be drawn from the findings of this study.

- The thermal history of RC columns is influenced by the shape, aggregate type, and humidity, particularly after 60 min of fire exposure. The size effect seems to be minor for columns of the same concrete cover.
- In general, circular columns, columns with large cross sections, and those made from lightweight and carbonate aggregate tend to have a slow temperature rise compared to their counterparts. This slow rise in temperature can be quantified at 100-200°C and on a case-per-case basis.
- The deformation history of the fixed-fixed RC columns can be considered to have four stages. The profile and duration of each stage are dependent on the loading level, type of aggregates used, and magnitude of the longitudinal steel ratio.
- The point in time at which a column shift from expansion to contraction marks the initiation of failure.
- RC columns made from silicate aggregates are likely to fail within 20-80 min of reaching the failure stage. However, fire-exposed columns made from carbonate aggregates are more likely to outperform silicate columns and fail at later times.
- Data driven ML is very much likely to fail to address causal and counterfactual questions in the context of the examined phenomenon herein.

Data Availability

Some or all the data, models, or codes that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgment

We would like to thank the Editor and Reviewers for their support in this work and for constructive comments that enhanced the quality of this manuscript.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] T. Lie, J. Woollerton, Fire Resistance of Reinforced Concrete Columns - NRC Publications Archive - National Research Council Canada, 1988. https://doi.org/http://doi.org/10.4224/20386656.

[2] T. Liet, D. Allen, Calculation of the fire resistance of reinforced concrete columns - NRC Publications Archive - Canada.ca, n.d. https://doi.org/10.4224/40001205 (accessed November 10, 2022).

[3] D. Allen, T. Lie, Further studies of the fire resistance of reinforced concrete columns - NRC Publications Archive - Canada.ca, 1974. https://doi.org/10.4224/40001183 (accessed November 10, 2022).

[4] T.T. Lie, R.J. Irwin, Method to calculate the fire resistance of reinforced concrete columns with rectangular cross section, ACI Struct. J. 90 (1993) 52–60. https://doi.org/10.14359/4210.

[5] H. Wu, T. Lie, Fire resistance of reinforced concrete columns: experimental studies - NRC Publications Archive - Canada.ca, 1992. https://doi.org/10.4224/20375195 (accessed November 11, 2022).

[6] J.R. Lawson, A History of Fire Testing, NIST Tech. Note 1628. (2009). https://doi.org/10.1038/130562a0.

[7] M.A. Sultan, Review of the NRC Canada Studies on Fire Resistance of Walls: Results, Research Gaps and Design Guidelines, Fire Technol. (2022) 1–33. https://doi.org/10.1007/S10694-022-01312-4/FIGURES/9.

[8] J. Janss, Statistical analysis of fire tests on steel beams and columns to Eurocode 3, Part 1.2, J. Constr. Steel Res. (1995). https://doi.org/10.1016/0143-974X(94)00017-C.

[9] V. Kodur, M.Z. Naser, Structural Fire Engineering, 1st ed., McGraw Hill Professional, 2020.

[10] M.Z. Naser, Causality in Structural Engineering: Discovering New Knowledge by Tying Induction and Deduction via Mapping Functions and eXplainable Artificial Intelligence, AI Civ. Eng. (2022). https://doi.org/10.1007/s43503-022-00005-9.

[11] G.E.P. Box, J.S. Hunter, W.G. Hunter, Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building, John Wiley & Sons, 1978.

[12] C. Wade, G. Cowles, R. Potter, P. Sanders, Concrete Blade Columns in Fire wade - Google Scholar, in: Concr. 97 Conf., Adelaide, Australia, 1997. https://scholar.google.com/scholar?hl=en&as_sdt=0%2C41&q=Concrete+Blade+Columns+in+Fire+wade&btnG= (accessed November 6, 2022).

[13] A. Jain, H. Patel, L. Nagalapatti, N. Gupta, S. Mehta, S. Guttula, S. Mujumdar, S. Afzal, R. Sharma Mittal, V. Munigala, Overview and Importance of Data Quality for Machine Learning Tasks, in: Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., 2020. https://doi.org/10.1145/3394486.3406477.

[14] ECS, EN 1993-1-2: Eurocode 3: Design of steel structures - Part 1-2: General rules - Structural fire design : European Committee for Standardisation : Free Download, Borrow, and Streaming : Internet Archive, 2005.

[15] J. Ferreira, T. Gernay, J. Franssen, O. Vassant, Discussion on a systematic approach to validation of software for structures in fire - Romeiro Ferreira Joao Daniel, in: SiF 2018 10th Int. Conf. Struct. Fire, Belfast, 2018. http://hdl.handle.net/2268/223208 (accessed April 1, 2020).

[16] L. Yao, Z. Chu, S. Li, Y. Li, J. Gao, A. Zhang, A Survey on Causal Inference, ACM Trans. Knowl. Discov. Data. 15 (2021) 1–46. https://doi.org/10.1145/3444944.

[17] J. Pearl, D. Makenzie, The Book of Why: The New Science of Cause and Effect-Basic Books, Basic Books, 2018.

[18] M.Z. Naser, A.O. Ciftcioglu, Causal Discovery and Causal Learning for Fire Resistance Evaluation: Incorporating Domain Knowledge, (2022). https://doi.org/10.48550/arxiv.2204.05311.

[19] A.R. Nogueira, A. Pugnana, S. Ruggieri, D. Pedreschi, J. Gama, Methods and tools for causal discovery and causal inference, Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 12 (2022) e1449. https://doi.org/10.1002/WIDM.1449.

[20] S. Khalilpourazari, M. Mohammadi, A new exact algorithm for solving single machine scheduling problems with learning effects and deteriorating jobs, (2018). https://doi.org/10.48550/ARXIV.1809.03795.

[21] S. Khalilpourazari, H. Hashemi Doulabi, Designing a hybrid reinforcement learning based algorithm with application in prediction of the COVID-19 pandemic in Quebec, Ann. Oper. Res. 312 (2022) 1261–1305. https://doi.org/10.1007/s10479-020-03871-7.

[22] Y. Freund, R.E. Schapire, Experiments with a New Boosting Algorithm, Proc. 13th Int. Conf. Mach. Learn. (1996). https://doi.org/10.1.1.133.1040.

[23] LightGBM, Welcome to LightGBM's documentation! — LightGBM 3.1.1.99 documentation, (2020). https://lightgbm.readthedocs.io/en/latest/ (accessed February 9, 2021).

[24] M. van Smeden, K.G. Moons, J.A. de Groot, G.S. Collins, D.G. Altman, M.J. Eijkemans, J.B. Reitsma, Sample size for binary logistic prediction models: Beyond events per variable criteria:, Https://Doi.Org/10.1177/0962280218784726. 28 (2018) 2455–2474. https://doi.org/10.1177/0962280218784726.

[25] R.D. Riley, K.I.E. Snell, J. Ensor, D.L. Burke, F.E. Harrell, K.G.M. Moons, G.S. Collins, Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes, Stat. Med. (2019). https://doi.org/10.1002/sim.7992.

[26] I. Frank, R. Todeschini, The data analysis handbook, 1994. https://books.google.com/books?hl=en&lr=&id=SXEpB0H6L3YC&oi=fnd&pg=PP1&ots=zfmIRO_XO5&sig=dSX6KJdkuav5zRNxaUdcftGSn2k (accessed June 21, 2019).

[27] M.Z. Naser, · Amir, H. Alavi, Error Metrics and Performance Fitness Indicators for Artificial Intelligence and Machine Learning in Engineering and Sciences, Archit. Struct. Constr. 2021. 1 (2021) 1–19. https://doi.org/10.1007/S44150-021-00015-8.

[28] A. Tapeh, M.Z. Naser, Artificial Intelligence, Machine Learning, and Deep Learning in Structural Engineering: A Scientometrics Review of Trends and Best Practices, Arch. Comput. Methods Eng. (2022). https://doi.org/10.1007/s11831-022-09793-w.

[29] J. Pearl, Causal inference in statistics: An overview, Stat. Surv. (2009). https://doi.org/10.1214/09-SS057.

[30] C. Heinze-Deml, M.H. Maathuis, N. Meinshausen, Causal Structure Learning, Annu. Rev. Stat. Its Appl. (2018). https://doi.org/10.1146/annurev-statistics-031017-100630.

[31] R. Scheines, An Introduction to Causal Inference, (1999).

[32] K. Kuang, L. Li, Z. Geng, L. Xu, K. Zhang, B. Liao, H. Huang, P. Ding, W. Miao, Z. Jiang, Causal Inference, Engineering. 6 (2020) 253–263. https://doi.org/https://doi.org/10.1016/j.eng.2019.08.016.

[33] P.J.E.A. Javier, M.P. Liponhay, C.V.G. Dajac, C.P. Monterola, Causal network inference in a dam system and its implications on feature selection for machine learning forecasting, Phys. A Stat. Mech. Its Appl. 604 (2022) 127893. https://doi.org/https://doi.org/10.1016/j.physa.2022.127893.

[34] S. Kleinberg, G. Hripcsak, A review of causal inference for biomedical informatics, J. Biomed. Inform. 44 (2011) 1102–1112. https://doi.org/https://doi.org/10.1016/j.jbi.2011.07.001.

[35] A. Sharma, E. Kiciman, DoWhy: A Python package for causal inference, (2019).

[36] A. Sharma, E. Kiciman, DoWhy: An End-to-End Library for Causal Inference, (2020). https://doi.org/10.48550/arxiv.2011.04216.

[37] P. Blöbaum, P. Götz, K. Budhathoki, A.A. Mastakouri, D. Janzing, DoWhy-GCM: An extension of DoWhy for causal inference in graphical causal models, Arxiv.Org/Pdf/2206.06821.Pdf. (2022) 1–7.

[38] B.A. Dickerman, M.A. Hernán, Counterfactual prediction is not only for causal inference., Eur. J. Epidemiol. 35 (2020) 615–617. https://doi.org/10.1007/s10654-020-00659-8.

[39] K. Battocchi, E. Dillon, M. Hei, G. Lewis, P. Oka, M. Oprescu, V. Syrgkanis, EconML: A Python Package for ML-Based Heterogeneous Treatment Effects Estimation, GitHub. (2019).

This is a preprint draft. The published article can be found at: https://doi.org/10.1007/s10694-023-01405-8.

Please cite this paper as:
Naser M.Z., Çiftçioğlu A, (2023). "Revisiting Forgotten Fire Tests: Causal Inference and Counterfactuals for Learning Idealized Fire-induced Response of RC Columns". *Fire Technology*. https://doi.org/10.1007/s10694-023-01405-8.

[40] V. Syrgkanis, G. Lewis, M. Oprescu, M. Hei, K. Battocchi, E. Dillon, J. Pan, Y. Wu, P. Lo, H. Chen, T. Harinen, J.-Y. Lee, Causal Inference and Machine Learning in Practice with EconML and CausalML: Industrial Use Cases at Microsoft, TripAdvisor, Uber, in: 2021 Knowl. Discov. Data Min., 2021.

[41] J. Pearl, Causal Diagrams and the Identification of Causal Effects, in: Causality, 2013. https://doi.org/10.1017/cbo9780511803161.005.

[42] G.W. Imbens, Potential outcome and directed acyclic graph approaches to causality: Relevance for empirical practice in economics, J. Econ. Lit. (2020). https://doi.org/10.1257/JEL.20191597.