UNIVERSITY OF CAMBRIDGE

UNIVERSITY OF BATH

MEICON

Minimising Energy in Construction

Design Uncertainty

Report

Dr John Orr

EP/P033679/2

www.meicon.net

Uncertainty

"To err and err and err again, but less and less and less" [1]

1 Outline

This short report considers some key questions surrounding uncertainty in the design process. The present-day influence on structural design of two key figures in probability theory (Thomas Bayes and Ronald Fisher) is considered in the context of the Eurocodes. The report offers potential starting points from which actions taken in the design process might be critiqued with a view to improve design and reduce inefficiency in the use of structural materials. The report has been written as part of the wider project "Minimising Energy in Construction (MEICON)" (www.meicon.net) which has a long term vision for the built environment to be designed cost-effectively, based on whole life cycle energy consumption using minimum material resource for appropriate performance.

Throughout the report, questions raised in the first MEICON report "Survey of structural engineering practice" (www.meicon.net/survey2018) are highlighted when appropriate. These "Industry Questions (*IQ*)" and "Research Questions (*RQ*)" remain open for all readers to help the MEICON team to answer.

1.1 Recommended reading

The author recommend the following to give some additional background:

- Silver, N., 2012. *The signal and the noise. The art and science of prediction*. London: Penguin Books.
- Beal, A., 2011. A history of the safety factors. *The Structural Engineer*, vol. **89** (20).
- Bolton, M., 1993. *What are partial factors for?* International Symposium on Limit State Design in Geotechnical Engineering, Copenhagen, pp.565-583
- Bayes, T., 1763. An essay towards solving a problem in the doctrine of chances. Philosophical Transactions of the Royal Society, vol 53.
- Melchers, R., 2018 *Structural reliability and analysis*, 3rd Edition. Australia: John Wiley.

2 Probability

The mathematics behind probability was developed in the sixteenth century, and has its roots in the analysis of games of chance [1]. In 1763, in his posthumously published paper "An essay towards solving a problem in the doctrine of chances" Thomas Bayes marked a breakthrough in thinking about probability [2]. Applications of Bayesian probability are now found in a multitude of fields, from email filters to medicine [3]. Yet this was not always the case, and for much of the 250 years since his paper was published, the Bayesian approach had been side lined.

In his essay Bayes sets out how probabilistic beliefs about the world can be formulated as new data is encountered. In essence, it proposes that we learn about the universe through approximation, and with each gathering of evidence get closer to the truth [4].

Thomas Bayes was interested in finding the "inverse probability" of a cause given the results. We might use this to answer problems such as if a coin is tossed five times and gives five heads, what is the probability of a biased coin? Without some idea of the prevalence of biased coins, this question is unanswerable. Bayes set out a method that identifies the power of some prior

[1] From the poem "The Road to Wisdom", by Piet Hein

knowledge – if you can estimate a base line of how often heads-biased coins are minted, you can estimate how likely it is to have a heads-biased coin. And, crucially, as your information about the prevalence of heads-biased coins improves, so too can your probability estimates.

The crux of Bayes' paper is the logic of learning from evidence. You may begin your analysis assuming a certain chance of a biased coin, but the coin that yields five heads in a row should change that initial probability. Bayes' theorem shows that evidence supports a theory to the extent the theory makes that evidence likely (five heads supports having a biased coin, because a biased coin makes five heads more likely). This is known as conditional probability – the theorem tells us the probability that a hypothesis is true, *if* some event has happened. This is stated mathematically as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\neg A)P(\neg A)}$$

For proposition *A* and evidence *B*:

1. *P(A)* is the prior probability - the initial degree of belief in *A.*
2. P(¬A) is the initial degree of belief against *A*, and P(¬A) = 1-P(A).
3. P(B|A) is conditional probability, the degree of belief in *B* given that proposition *A* is true.
4. P(B|¬A) is conditional probability, the degree of belief in *B* given that proposition *A* is false.
5. P(A|B) is the posterior probability, the probability for *A* after taking into account for *B*, for and against *A*.

It is thus necessary to estimate (or know) three quantities – the *prior probability* (what was the probability of *A* before evidence *B* was known); the *conditional probability* of evidence *B* occurring if *A* is true, and the *conditional probability* of evidence *B* occurring if *A* is false.

Bayes' method requires prior-probabilities to be assigned. It was this "*problem of the priors*" that turned another statistical group against Bayes. This group, led by Ronald Fisher, objected to the idea that scientific reasoning could hinge (at least initially) on a personal hunch - subjectivity had (in their view) no place in the objectivity of science. Instead, the idea of "significance tests" was developed. Such tests require you to reject a hypothesis, if you observe results that would be very unlikely if the hypothesis was true. This "Fisherian" approach requires the analyst to pick some level of improbability that she will not tolerate, and reject the hypothesis if it implies the observed data are less likely than that level.

The normally recommended level for such a method is **5%**.

> *"Most people and even most scientists still don't know much about statistics except that there is something good about the number 0.05'' Efron [5].*

The Fisherian approach is based on the idea that uncertainty results from sampling – collecting data from a sample of the population, rather than the entire population. This will be familiar in the context of election polling, and the Fisherian approach directly allows a "margin of error" to be quantified. For example, we can quantify the error due to asking 200 people how they voted, out of a population of 2 million. Statistics students have, for some generations, been taught significance testing, and as noted by Efron [5], most of us assume there is something "good" about the 5% level but are not quite sure why. Yet the idea of drawing such a line is completely arbitrary, as Fisher himself notes:

> *"If one in twenty does not seem high enough odds, we may, if we prefer it, draw the line at one fifty (the 2 per cent. point), or one in hundred (the one per cent. point)." Fisher [6].*

The approach implies that if you collect all the data, the error will approach zero. Yet in many fields, including engineering, data is sparse. We must make decisions in conditions of real uncertainty, and an arbitrary cut off at 95% that ignores any real-world context is unhelpful – since it is not obvious what the real-world difference between 94.9% and 95.1% might be. The Fisherian approach also has many underlying assumptions – for example that uncertainty in the data follows a normal distribution, or that we can indeed define a sample population. The main critique against the Fisherian approach is thus that it results in findings that are effectively sealed off from the real world [4]: "statistically significant" results that are in all other senses manifestly ridiculous.

An unintended consequence of significance testing is perhaps found in the current replicability crisis within academia. Every so often a researcher obtains a result by luck, reports it as it passes a significance test, and other researchers rely on that. To demonstrate this problem Nature polled 1500 scientists in 2016 [7] – 70% of them had failed to reproduce another scientist's results. One source of this lack of repeatability is a reliance on significance testing. As noted by Papineau [3]:

> *Since the whole methodology of significance tests is based on the idea that we should tolerate a 5 per cent level of bogus findings, statistical traditionalists are not well placed to dodge responsibility when bogus results are exposed.*

In response to such criticism, Fisherians may argue that we should simply raise the significance level, perhaps to 0.1%. That would examine findings occurring once in a thousand, instead of once in twenty (5%). Yet this argument is also fatally flawed, since the problem does not appear to lie with the significance level, but with the test itself, and its inability to consider prior probability.

The Bayesian approach, with a reliance on prior beliefs, does have its own challenges. Yet no matter where the prior probability is placed, in the light of new evidence all predictions should converge towards each other (and towards a true answer). Engineering, therefore, should be able to learn from the data and in light of the data adjust our understanding of the probabilities of certain events occurring (loading, human response, failure). This process, of truly learning from real world data and updating our understanding as a result of new information, has not yet been embedded within structural design.

3 Structural Design

3.1 Principles

Limit state design is a philosophy under which structures are designed such that the probability that a number of performance criteria are exceeded is deemed to be acceptably small during the required functional lifetime of the structure. When a structure, or element within a structure, ceases to satisfy one or more of these performance criteria it is deemed to have exceeded a *limit state* and thus now fails to fulfil satisfactorily the design requirements.

BS EN 1990 [8] requires that "a structure shall be designed and executed in such a way that it will, during its intended life, with appropriate degrees of reliability and in an economical way sustain all actions and influences likely to occur during execution and use, and meet the specified serviceability requirements for a structure or structural element".

3.2 Limit states

The ultimate limit states are those which concern "the safety of people and/or the safety of the structure" [8]. The serviceability limit states are those which concern the "functioning of the structure or structural members under normal use, the comfort of people, the appearance the construction works" [8]. The requirements of limit state design may be met by design directly based on probabilistic methods (Annex C of EN1990 [8]), or by the partial factor method. The partial factor method is understood to be by far the dominant method used in practice.

Using the partial factor method, the designer must verify that limit states are not exceeded. This requirement is summarised in Eq.(1) and Eq.(2):

$$E_d \leq R_d \tag{1}$$

Where E_d is the design value of the effects of actions and R_d is the design value of the corresponding resistance

$$E_d \leq C_d \qquad (2)$$

Where E_d is the design value of the effects of actions specified in the serviceability criterion, and C_d is the limiting design value of the relevant serviceability criterion

Evidence from MEICON [9] shows that structural materials are often inefficiently utilised, and rarely is R_d set equal to E_d. This mitigation against uncertainty, or additional "sleep at night" factor, perhaps reflects a significant downside in getting it wrong, with little upside (currently) to be found in being materially efficient.

IQ15: Is there a perception that utilisation ratios of 1.00 (E_d = R_d) are dangerous? If so, why?

3.3 Actions

An action can be a *direct* force (load) applied to the structure (e.g. dead loads (self-weight) and live loads (i.e. variable loads such as traffic, pedestrian loading)) or an *indirect* action such as an imposed deformation or acceleration caused, for example, by temperature changes, settlement or earthquakes. They are normally classified according to their variation in time (permanent, variable, or accidental), which gives us some hint as to how they are modelled.

In the Eurocodes, a "characteristic" value of an action is the main representative value. For permanent actions, a single value may be used if the variability of the permanent action is small over the working life of the structure. In general, there is much greater certainty as to the value of permanent actions such as self-weight, since these are directly measurable on each structure and are normally defined by the designer.

For variable actions, the characteristic value according EN 1990 [10] to corresponds to either:

a) An upper value with an intended probability of not being exceeded, or a lower value with an intended probability of being achieved, during a specific reference period, or:
b) A nominal value, which may be specified in cases where a statistical distribution is not known.

Figure 1 illustrates the potential of option (a), where a probability density curve is drawn from some measured values of loading, and a characteristic value taken at the often assumed (see §2) 95% fractile. There are clear problems with such assumptions – for example the median floor live loading from people in offices is 0kN/m^2 (75% of the time, most office space is empty).

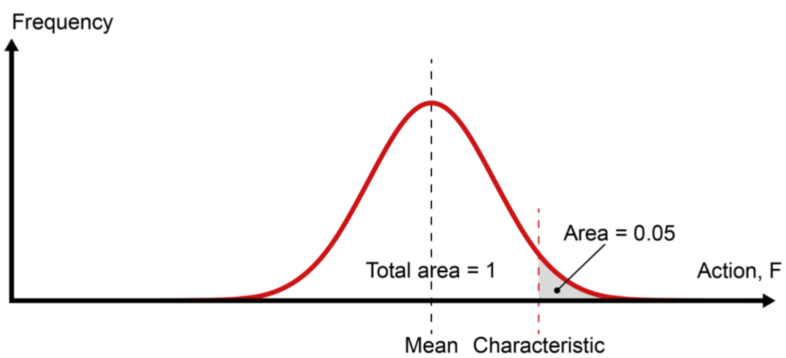

Figure 1: Theoretical actions (loading)

In practice it is very difficult to obtain reliable data on the actual loading on structures. As a result, loads specified in design codes are usually derived by consensus amongst experts at a level considered to represent conservatively the maximum load likely to be experienced by a structure during its design life with an acceptably low probability of being exceeded. For example BS EN 1991-1-1 [11] gives imposed loads for office floors as nominal values.

Thus, whilst it may be the understanding of the designer that the load chosen from a Eurocode table has some statistical basis, it may in reality simply be a value based on historic practice. This in a large part reflects the difficulty of measuring real loading on buildings. If such loading could be measured in a reliable, repeatable, low-cost manner across multiple typologies and locations, we might be able to adopt a more Bayesian approach in which our loading values could get less and less and less uncertain over time.

RQ3: What is the real envelope of floor loading for which most designs should be undertaken?

IQ4: What might the benefit be of design code floor loading values being based on data gathered from a systematic global survey of loading levels in buildings?

IQ7: How might real time building loading information be integrated into building management systems to provide "traffic light" load levels to aid facility management?

3.4 Material properties

Properties of materials are also represented by characteristic values, determined by standardised tests performed under specified conditions. For resistance properties characteristic values are usually taken at either a 5% fractile where a low value is unfavourable or a 95% fractile where a high value is unfavourable. Characteristic material densities are taken as mean values. For steel, a minimum yield strength is taken for design (see §4 for discussion on this point).

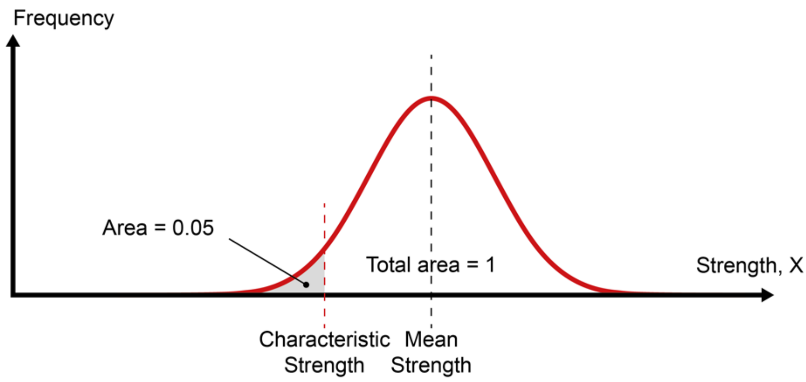

Figure 2: Theoretical characteristic property (e.g. material strength)

3.5 Probability of failure

Having defined design loading, and design resistance, one might be concerned with how likely a structure is to fail during its lifetime. In practically all designs this concern is dealt with by using recommended values for partial factors, see §3.6.

The basic reliability problem is to understand that failure occurs if the resistance (R) is less than the effect of actions (E), Figure 4. In the Eurocodes, a *reliability index* (β) is defined. If we assume that both R and E are normally distributed in one specific case, the safety margin Z (equal to R – E) is also normally distributed and has a mean μ_z and standard deviation σ_z (values can be computed from the R and E curves). The reliability index is defined by:

$$\beta = \frac{\mu_z}{\sigma_z}$$

In other words, β is a measure of the number of standard deviations from the mean of the failure function (Z) to zero. BS EN 1990 [10] gives values of target reliability indices for different time periods and limit states. As failure can occur in many different ways, such studies must examine all possible modes, and beyond simple examples of beams in isolation such calculations are unwieldy. The keen reader is directed to Melchers [12] for more expansion on this topic.

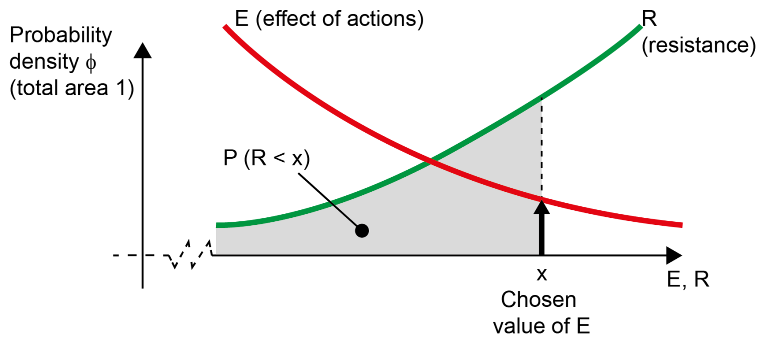

Figure 3: Probability of failure (simplified).

In addition to probability of failure, the consequences of failure should be considered more carefully. Quite clearly, a failure at the ultimate limit state must be avoided. Temporarily breaching a serviceability limit state might be palatable [9], yet in design philosophically the only difference is a single factor. Since serviceability limit states can govern material consumption for some structures, they should be examined very carefully in the light of cold data as to the real performance of structures.

RQ2: How can continuous measurement of floor loading in real buildings be used to provide certainty to the statistical basis for SLS loading, and how can this data be used to understand the extent to which loading conditions are "peaky" so that decisions about SLS requirements can be made?

IQ5: What might the unintended consequences be of changing live load values at ULS based on measured data?

IQ6: What might the unintended consequences be of changing live load values at SLS based on measured data?

3.6 Partial factors

In the Eurocodes, characteristic actions, and characteristic material resistance, are converted to design effects and design resistances as shown in Figure 4.

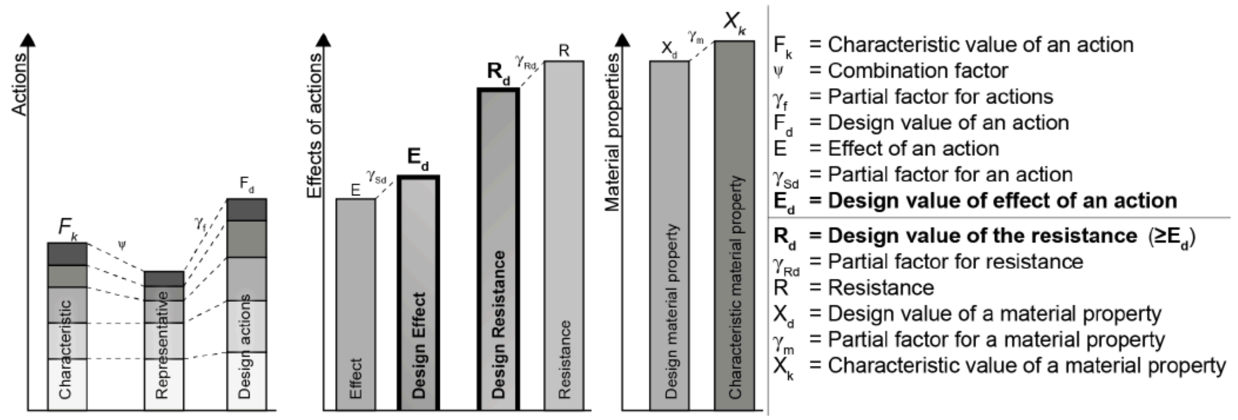

Figure 4: Partial safety factors – characteristic to design values

Values for the four "γ" values and the "ψ" value in Figure 4 may be defined in the Eurocode framework in one of two ways:

a) On the basis of calibration to a long experience of building tradition
b) On the basis of statistical evaluation of experimental data and field observations

Option (b) requires design to be undertaken within a probabilistic framework. Option (a) is identified in BS EN 1990 as the leading principle within currently available Eurocodes. Therefore, values for "γ" "ψ" are defined largely based on historical practice, rather than measured performance. We should remember that partial factors are relied on to satisfy requirements for structural reliability. The ideas behind Bayesian statistics of learning from experience might be of benefit in the iteration (rather than simply calibration) of partial safety factors.

The first MEICON report [13] identified that the unknown potential for construction error influences structural member sizing decisions. As true levels of error are unknown, these effects are hidden within partial safety factors, material properties, connections, and loadings. It may be more fruitful to measure the possibility of construction error, based perhaps on the level of workmanship, and to review this in light of changes in the construction industry as site based automation becomes more prevalent.

RQ4: What might the benefits and consequences be of reducing material and load partial safety factors?

RQ5: How should partial safety factors for workmanship change as construction becomes increasingly automated?

IQ14: How might we agree on a "ratchet" of increasingly stringent design requirements, allowing time to adjust design culture whilst recognising the imperative need to reduce carbon emissions?

3.6.1 A note on Eurocode 7

Eurocode 7, Clause 2.4.5.2(10) states:

> *If statistical methods are employed in the selection of characteristic values for ground properties, such methods should differentiate between local and regional sampling and should allow the use of a priori knowledge of comparable ground properties.*

The use of *a priori* knowledge appears to be a nod towards a Bayesian approach. The following clause, 2.4.5.2(11) then states:

> *If statistical methods are used, the characteristic value should be derived such that the calculated probability of a worse value governing the occurrence of the limit state under consideration is not greater than 5%*
>
> *NOTE: In this respect, a cautious estimate of the mean value is a selection of the mean value of the limited set of geotechnical parameter values, with a confidence level of 95%; where local failure is concerned a cautious estimate of the low value is a 5% fractile.*

Which perhaps demonstrates the degree to which significance levels, and the "5%" are embedded in the thinking behind characteristic values.

4 Calibration

In the development of new design codes, there is normally extensive calibration against previously published design codes. In 1987, Chabowski, Judge [14] undertook a calibration exercise following the European Economic Community decision to develop the Eurocodes. Their calibration was between BS 449, and the then new BS 5950. The calibration method is described [14] as follows:

> *"The method for calibrating a new code against an existing code described in this report is based on a comparison of the maximum design capacities for individual structural elements calculated in accordance with each code. For convenience, the ratio of the maximum capacity of an element calculated using the new code to that calculated using the existing code has been called the 'comparison indicator' (CI). The method allows the optional introduction of weighting factors to reflect the usage of different elements. It can also take account of comparisons between research results and design capacity calculated to a specific code."*

The analysis by Chabowski, Judge [14] compared designs for beams and columns across all the standard section sizes. The analysis was weighted based on percentage of section use, itself derived by questionnaire of designers. For example, 39% of sections used were assumed to be restrained beams under uniformly distributed loadings.

The material partial safety factor, γ_m, was derived from data on steel coupon tests - in which only 13 of 11,526 tests (0.1%) fell below the guaranteed yield strength. It was therefore determined to use the minimum yield strength as the design value, hence γ_m of 1.0 was chosen. Chabowski, Judge [14] do not report other data on the yield strength tests, and so we do not know anything about the distribution of the 11,526 test points, other than they are all greater than the minimum yield strength.

The discussion surrounding partial factors for loading was concerned with possible changes in mass of the elements between the two design codes. For example, the report [14] states:

> *"1) … use of [an overall partial safety factor of] γ_o = 1.6 would result in approximately the same number of section sizes being increased as are decreased, suggesting that economy would probably not be affected. This would not be reasonable because the introduction of new equations into the draft BS5950 was expected to result in an overall reduction in section sizes.*
>
> *"2) … use of … γ_o = 1.5 would theoretically result in approximately 70% or 80% of sections sizes being reduced … This would seem more acceptable than the case given in 1 above.*

The overall minimum factor of safety of 1.5 led then to partial factors for dead (γ_D) and live load (γ_L). These are based around two key assumptions: 1) that the values for γ_D and γ_L are independent of the ratio of dead to imposed loading, and 2) that values for γ_D and γ_L are independent of the ratio of γ_D/γ_L. These two assumptions were examined by sensitivity analysis, and the report uses a ratio of γ_D/γ_L = 0.875, "*as used in most limit-state design codes"* [14] to calculate that γ_D = 1.4 and γ_L = 1.6.

Calibration was also most recently in the development of the Eurocodes. Beal [15] outlines the history of the development of safety factors, and notes that the Eurocodes appear to offer safety factors of 5-10% lower than previous British Standards.

Given that there is very little data on the real loading on structures, and their real performance, there is a very real question as to how much these factors of safety should be reduced. Beal [15] likens this to "*tiptoeing towards the edge of a cliff*".

The small but incremental reductions in overall factors of safety between the two codes mentioned above are being undertaken without reference to what happens in the real world. Measurement of real performance is clearly an essential tool in ensuring that we are in fact converging on a "true" position, not a cliff drop.

RQ9: How might Failure Mode Effects Analysis (FMEA) be feasibly applied to the design of buildings to incorporate more detailed consideration of the consequences of failure and an appropriate level of risk?

RQ10: How can structural models be checked in an automated fashion? How can we reduce error rates in structural engineering design? Should there be a partial safety factor for analytical errors in all structural design, and how might this change over time as automation increases?

5 Measurement

Unlike jet engines, which are leased from their manufacturer, constantly monitored, maintained and updated, the design and construction of buildings is largely divorced from whole life performance, particularly when we consider embodied energy consumption. Short term views are taken by designers and contractors, who normally do not occupy or rent the structures they design. Evidence from MEICON [9] shows that structural materials are often inefficiently utilised, in part as mitigation against uncertainty – an additional "sleep at night" factor. It is not clear that this can be compatible with the use of reliability theory.

A key missing link in reducing uncertainty in design is data. How do structures really perform? What stresses are they subject to? What loading really occurs? Are computational models accurate, or not? The variability of structural modelling is an important but currently unknown design parameter, as is the potential for computer models to be unrepresentative of the physical structure – releases being the classic example. Yet these questions are difficult to answer, since we do not always know what data to collect to answer them – where are the sensors installed, how is data downloaded, and what do the numbers really mean? Unpicking these questions is a large and mostly unanswered research question.

To reach a position where certainty in the design process can enable minimum embodied energy structures requires a concerted interdisciplinary research program that will combine architecture, construction, computer science, artificial intelligence and machine learning, engineering and mathematics, Figure 5.

RQ19: What is the role of big data, computer science, and machine learning in changing the process of design?

RQ20: How do people interact with buildings? How does this change when they are lightweight? Are there any unintended consequences of lightweighting that change the user experience?

RQ21: What should be taught in Universities to prepare new engineers for the demands of design. What disciplines will be needed to work collaboratively in the future design office?

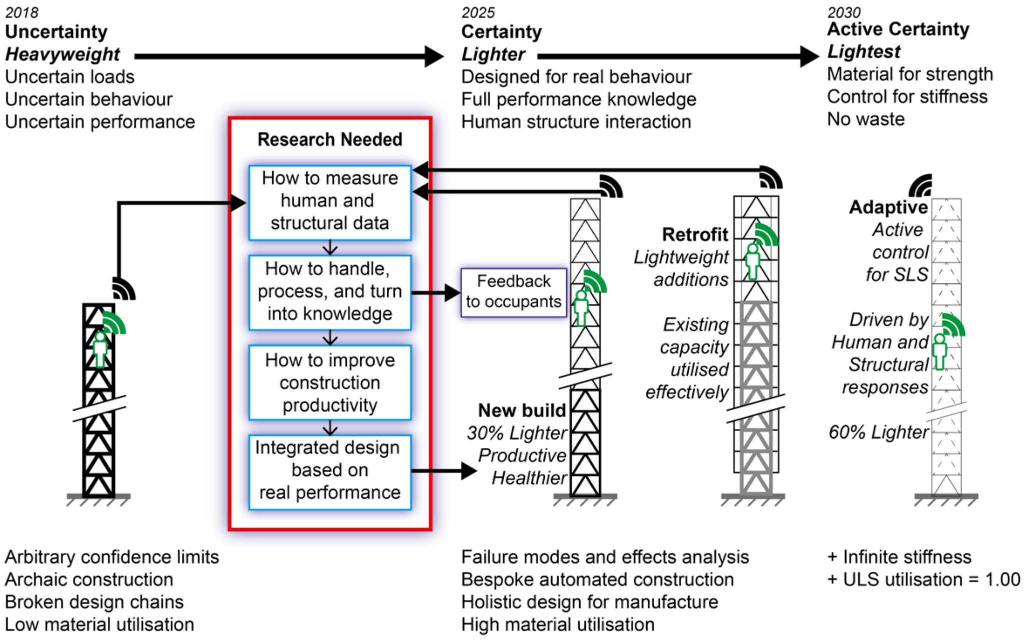

Figure 5: Research need to drive lightweighting

6 Conclusion

The measurement of real performance, correlated against loading and human response, offers a potential pathway to reduce design uncertainty and increase the efficiency with which we use structural materials. This has the potential to support targets in the UK for the construction sector associated with emissions, productivity, cost, and the trade balance [16].

If the process of learning from performance is to be embedded in the design, manufacture, assembly, use, reuse, and deconstruction of assets, then we must collectively be open to the potential that such measurement will reveal performance that can be improved.

The benefits of this research will be evident through efficient construction, minimal whole life energy use, better internal environments for occupants, and establishing, for the first time, design based on certainty.

7 References

1. Hájek, A. *Interpretations of Probability*. The Stanford Encyclopedia of Philosophy (Winter 2012 Edition) 2012 [cited 2019 04 March]; Available from: https://plato.stanford.edu/archives/win2012/entries/probability-interpret/.
2. Bayes, T., *LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S.* Philosophical Transaction, 1763. **53**: p. 49.
3. Papineau, D., *Thomas Bayes and the crisis in science*, in *TLS*. 2018, The Times Literary Supplement: Online.
4. Silver, N., *The signal and the noise. The art and science of prediction*. 2012, St Ives: Penguin Books.
5. Efron, B., *R. A. Fisher in the 21st Century.* Statistical Science, 1998. **13**(2): p. 95-114.
6. Fisher, R.A., *048: The Arrangement of Field Experiments.* 1926.
7. Baker, M., *1,500 scientists lift the lid on reproducibility.* Nature, 2016. **533**(7604).
8. BSI, *BS EN 1990: 2002+A1:2005*, in *Basis of structural design*. 2009, BSI.
9. Orr, J., et al., *Minimising energy in construction: Practitioners' views on material efficiency.* Resources, Conservation and Recycling, 2018. **140**.
10. *EN 1990: Eurocode: Basis of Structural Design*. 2002, EC.
11. BSI, *BS EN 1991-1-1*, in *Eurocode 1 - General actions — Densities, self-weight, imposed loads for buildings*. 2002, BSI: London.
12. Melchers, R., *Structural reliability and analysis, 3rd Edition*. 2018, Australia: John Wiley.
13. Orr, J., *Minimising Energy in Construction. Survey of Structural Engineering Practice*. 2018, University of Cambridge: Cambridge.
14. Chabowski, A., et al., *A deterministic calibration of the British Standard BS5950, 'The structural use of steelwork in buildings'*. 1987, BRE: Watford.
15. Beal, A.N., *A history of the safety favtors.* The Structural Engineer, 2011. **89**(20): p. 20-27.
16. BIS, *Construction 2025. Industrial Strategy: government and industry in partnership.* 2013, The Stationery Office: London.

Contact
Dr John Orr
Department of Engineering
University of Cambridge
Trumpington Street
Cambridge CB2 1PZ

+44 (0)1223 332 623
jjo33@cam.ac.uk
www.meicon.net

UNIVERSITY OF CAMBRIDGE

UNIVERSITY OF BATH

EPSRC
Engineering and Physical Sciences Research Council