Dealing with Materials Data: Collection, Analysis and Interpretation

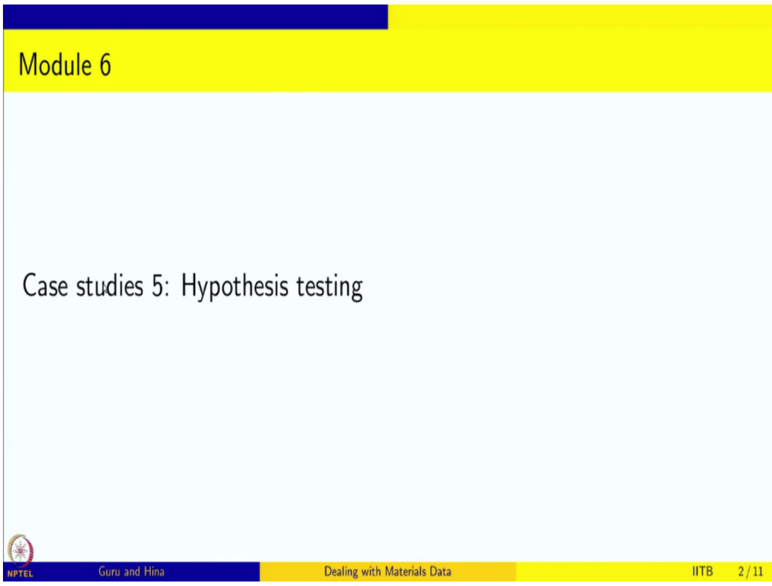
Professor M P Gururajan
Professor Hina A Gokhale
Department of Metallurgical Engineering and Materials Science
Indian Institute of Technology, Bombay

Lecture No. 98

Case study 5: Hypothesis testing

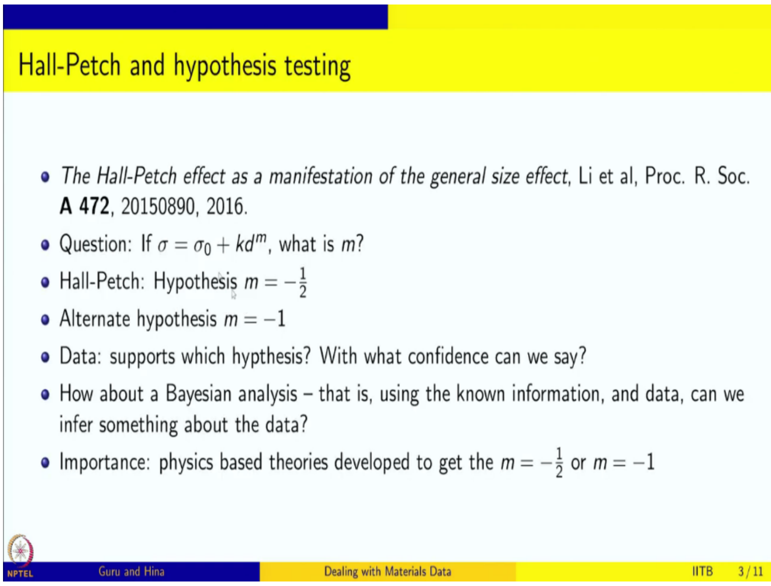
Welcome to Dealing with Materials Data. We are looking at the Collection, Analysis and Interpretation of Data from Materials Science and Engineering.

(Refer Slide Time: 0:25)



We are in module 6 which is on case studies and we are at the last case study which is on Hypothesis Testing. And as you will see this also is a nice conclusion not only for the case studies but for entire course that we have been going through.

(Refer Slide Time: 0:43)



So let us start with the Hypothesis testing with specific reference to Hall-Petch relationship. This is based on a paper the Hall-Petch effect as a manifestation of a general size effect by Li et al, published and proceedings of Royal Society A in 2016. I strongly recommend that you read this paper and try to reproduce some of the results and this entire discussion of this session is based on this paper.

So we are not going to do any R coding because most of what is there you should be able to do on your own but we will just discuss this paper with specific reference to hypothesis testing and see how this actually brings together everything that we have learned so far. The question that this paper tried to answer is the following: If sigma which is a flow stress is some sigma 0 plus kd to the power m Hall-Petch assumes that this is minus half and then fits for these two quantities k and sigma 0.

That is Hall-Petch hypothesis, m is minus half. But there are also other hypothesis, one of them, for example, says m is minus 1. Should, I choose minus 1 or minus half is the kind of question that we are trying to understand. We did do Hall-Petch, we took upper data and we tried to fit it for d to the power minus half and we did get sigma 0 and k that is reported in the literature, this is form MIMS data, the NIST data.

But is it true that it is minus half? Is the question that is being asked here. So this is hypothesis that we are trying to test to see if it is true. So they have collected large number of data from literature and trying to see if the data supports either of this hypothesis and let us say it supports this hypothesis, then with what confidence can we say that we are sure that this is true, right?

How about a Bayesian analysis – that is, using the known information and if you get a new data on some strength versus grain size today, can be improve our confidence level, for example. So these are the type of questions that this paper tried to answer. This also, the paper also tries to look at some physics-based theories which are developed to get m as minus half or m as minus 1.

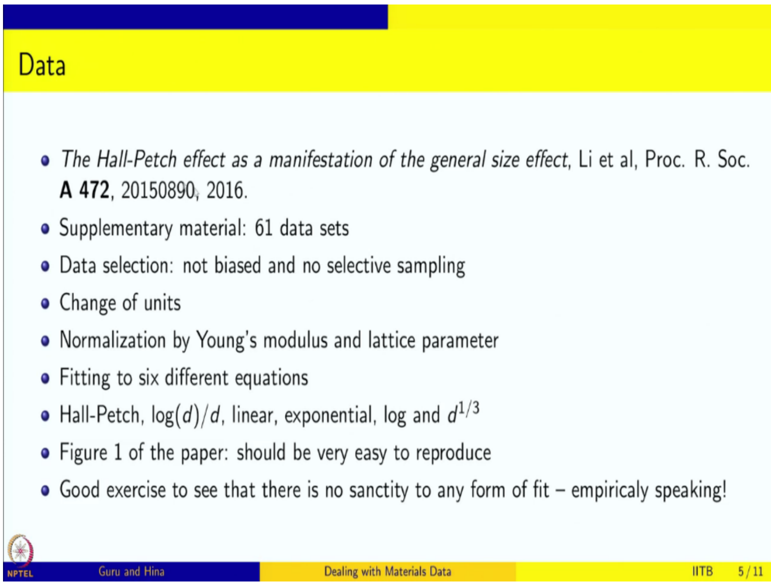
For example, you can say, you can think of it as a purely empirical thing, so we take data grain size and we take data on floor stress and just fit them but you can also say why this exponent should be minus half or 1 minus 1 and then you can try to analyse.

(Refer Slide Time: 3:33)

In fact, the historically speaking, that is also described in this paper, so Eshelby, Frank, Nabarro published a dislocation pile-up model and apparently Hall and Petch based on Eshelby theory came up with this fitting model sigma and sigma naught plus kd to the power minus half.

And Frank who is also involved here has other paper with van der Merwe and Mathews and co-workers, so all of them are referred to in this paper. They have come up with some critical thickness theories which said that it should go as log(d)/d , so there are two different models are hypothesis to fit the data.

(Refer Slide Time: 4:15)



So this paper also has a supplementary material in which some 61 data sets are shared and the data selection is not biased and there is no selective sampling, so they have actually described a way they came up with this data sets and to show that there is no selective sampling. So as you can see even though this is a case study in hypothesis testing, we are talking about collection, because collection is very important and then the analysis and then the inference.

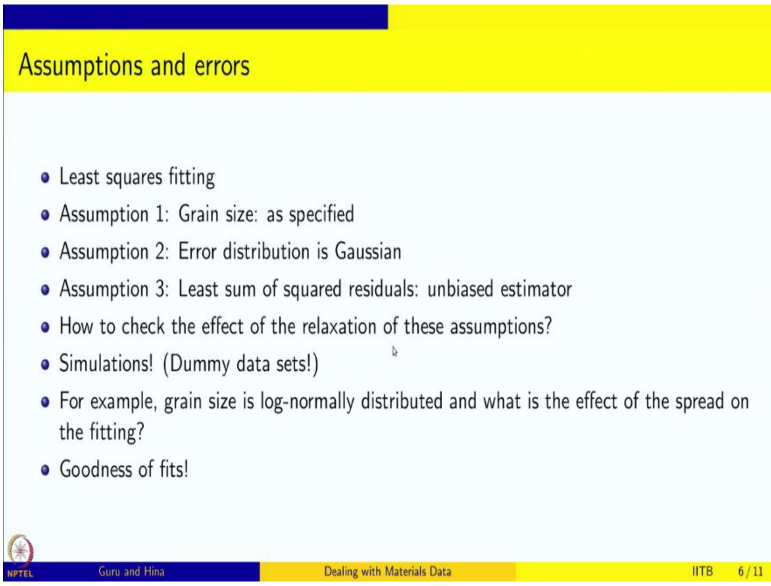
So all the parts that we are discussing or we were interested in this course so far, they are all involved in this case study and that is why it is a fitting case study to do at the end. So the data selection they described that it is not biased, there is no selective sampling. Of course different data sets are in different units, so first is to convert all of them to a single units, set of units, that is the SI units and then they also normalise by the stress and the grain size by Young's modulus and lattice parameter.

This is arbitrary just to make sure that the numbers are not in much broader range, they are doing this, so this is not essential. After they take the data and do all this change of units and normalisation, they fit it to 6 different equations; Hall-Petch is one, log d by d is another and then linear exponential log and d to the power one third, so these are the other models that they tried to fit.

And in figure 1 of the paper, they actually show you the data and the different fits that they have done. So this should be very easy for you to produce because the data is given in supplementary, so you can take those files convert them into CSV, load the data, plot the data, do the fitting for different forms that is given in the paper and draw those fitted lines and you should be able to see that the same figures that they have produced in figure 1 of the paper.

So this is a good exercise because we have already done this sort of exercise, but the exercise also shows you when you do that there is no sanctity to any form of it, right. You empirically speaking all fits seem to fit well for different data sets, right.

(Refer Slide Time: 6:42)



So why is this so? So in order to understand this, they go back to the fitting process which is a least square fitting and they list out the assumptions that are made in making this fitting. First assumption is grain size is a number that is as given, right. And the error is Gaussian which is random, it is a noise, right. But if the grain size is not a number or it had a distribution which is not the grain size distribution is not normal, right.

So taking mean and assuming that any error is because of random distribution could even be wrong. So error distribution could also have systematic deviations and so it might not be Gaussian. And so the third assumption, least some of squared residuals which is what we are trying to do when we are fitting this might not be an unbiased estimators, so there are lots of assumptions that are implicit and that go in.

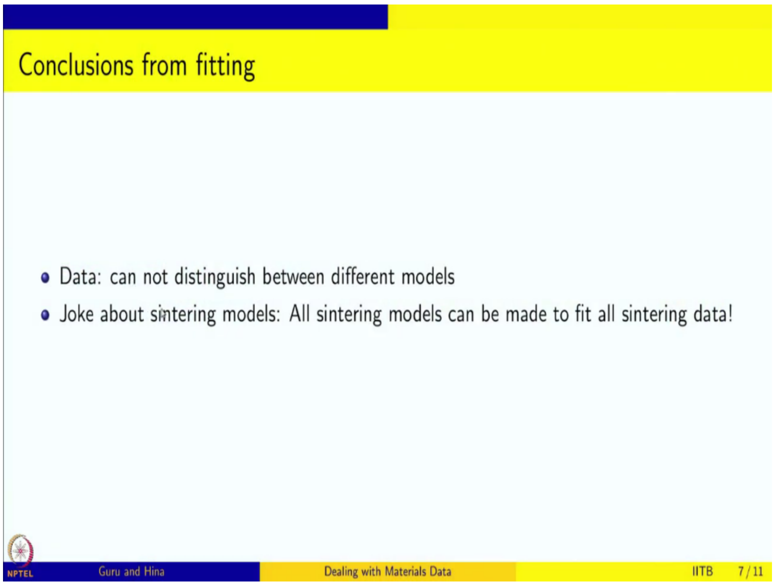
And now you might say that oh, we have to look at all this. For example, if the grain size is not just a number, it is a distribution what happens? And if that distribution is not the normal distribution but it is log normal or as we have seen, we have seen one grain size distribution for example which was fitting very well to some beta or something like that. So if you have distribution, what happens?

So in order to check the effect of relaxation of all this assumptions, they do simulations and generate what they call as dummy data sets. One of the simulations, for example, is that grain size is log normally distributed, you assume and with the same grain size that is the average being given grain size, but if it is log normally distributed, what is the effect of the spread on the grain size on the fitting process, on the fitting parameter.

So this is something that they tried to do. Of course, there is also lots of discussion in terms of which are the fits which are good statistically speaking. So this is very-very essential, we do the fitting, we get the parameters but how good is the fit is something that we should look at, so all this is described in the paper and the dummy data sets are also the results from the dummy data sets are also reproduced in the paper.

Of course we have done several such simulations ourselves as part of this course, so you should also be able to generate some data sets on your own and try to do this analysis or repeat this analysis.

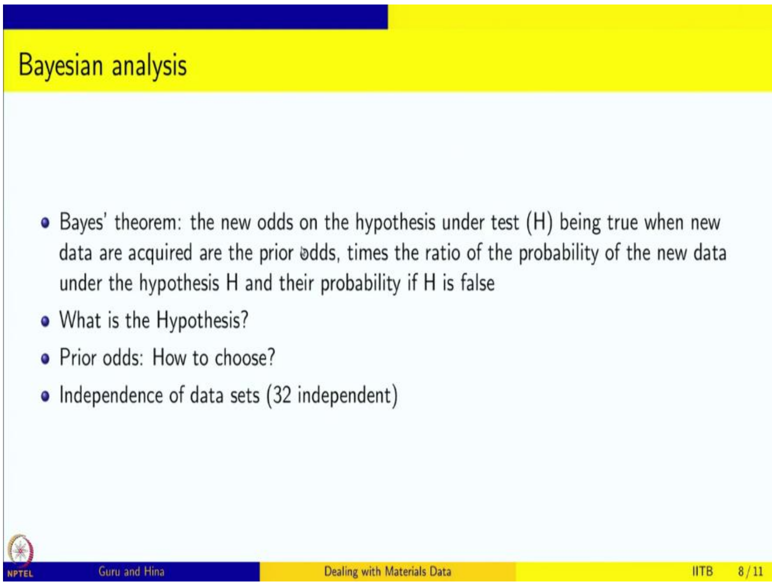
(Refer Slide Time: 9:22)



So the conclusion from the fitting exercise is that the data cannot distinguish between different models, so this reminds me of a joke about sintering models which is there in one of the ceramics textbooks. It says that all sintering models are can be made to fit to all sintering data, right. The models are of no use if they cannot distinguish between different data sets or if suppose you build a model which is all encompassing, it makes every data look right, then right data, wrong data it does not matter.

So the model loses its usefulness unless it can distinguish between different possibilities, right. The most or the broadest model which will fit all data is really of no great use. So the conclusion based on all these different data sets that they have collected from the literature is that the data cannot distinguish between different models, which is sort of disappointing because we have been living with Hall-Petch for probably about 70 years now.

(Refer Slide Time: 10:25)



So then they do the Bayesian analysis, so you have heard about Bayes' theorem and Bayesian analysis. In base theorem, basically we want to say what is the probability that this is true or the hypothesis is true, to say that we actually look at prior information we have and we also use the information on the probability that new data that we are having, so it is a ratio of the probability of the new data under the hypothesis and the probability if the hypothesis is false.

So assuming that it is true assuming, it is false look at the ratio of this probabilities multiplied by the prior odd and you get the new odd for the hypothesis being true. Now, of course, one can do the Bayesian analysis, we have seen an example in the other part of the course but to do the Bayesian analysis you have to state what your hypothesis is. Once you have stated the hypothesis, you have to choose prior odds for that hypothesis.

So the simplest example, for example, is that okay so if you have a coin and if you think that it is a biased coin, what is the prior odd that you should choose if you want to be unbiased, you will say, okay, I will assume that it is a 50 percent chance that it is biased and 50 percent chance that it is not biased and then toss few times and see if it consistently gives one head or tail over the other, more number of times then you can decide and you can improve your prediction.

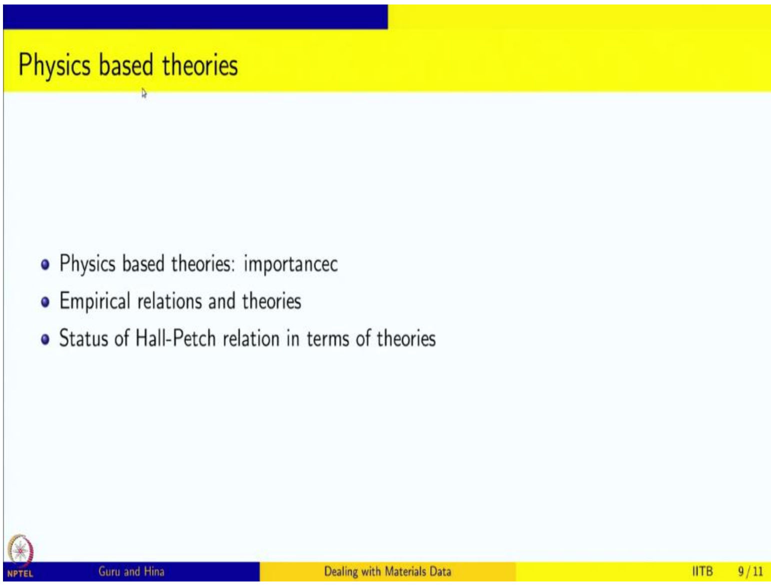
You have done it 10 times and it gives 8 times head, probably it is, the probability that it is biased is much higher than 0.5. But suppose it gives some 4 or 6 then it is still difficult to design whether it is biased or not. So we are using some prior information and prior you can also assume that okay, uniformly it might be any probability, right of giving heads, not just 0.5, it would be 0.3, it would be 0.7, 0.8.

So any probability between 0 and 1 is allowed, you can assume because if it is biased, it will be away from 0.5, so it could be anything and then you keep looking at the experiments and finding out what the probability, what is the number of times the head or tails is turning up and from that you keep improving, what is the probability that this coin will give a head? Is it 0.5 or it is away from 0.5, is it much less than 0.5 or much greater than 0.5.

So to do Bayesian analysis you have to have a hypothesis, so in this case, what is the hypothesis we choose and of what is a prior odd we choose for that given hypothesis? So there is a nice discussion on these two parts and this is an example that is coming from metallurgy material science, so that should be easier for you to follow and of course there is also a discussion about independence of data sets.

Even though we had more than 60 data sets, only 32 can be considered as independent and then using this kind of analysis is carried out. So this is also described in the paper, so it is a nice idea to go take a look at the paper and understand the arguments.

(Refer Slide Time: 13:47)



So that takes us to the next part of the paper which we are not going to discuss in detail but physics-based theories are important. The experiments are done and empirical data is generated and we can just fit the data to some functional form and using data we will be able to predict, we will be able to bill the things and we will be able to do well, but making theories which are physics based in order to understand why the empirical relationship the way it is actually helps improve and make progress much faster.

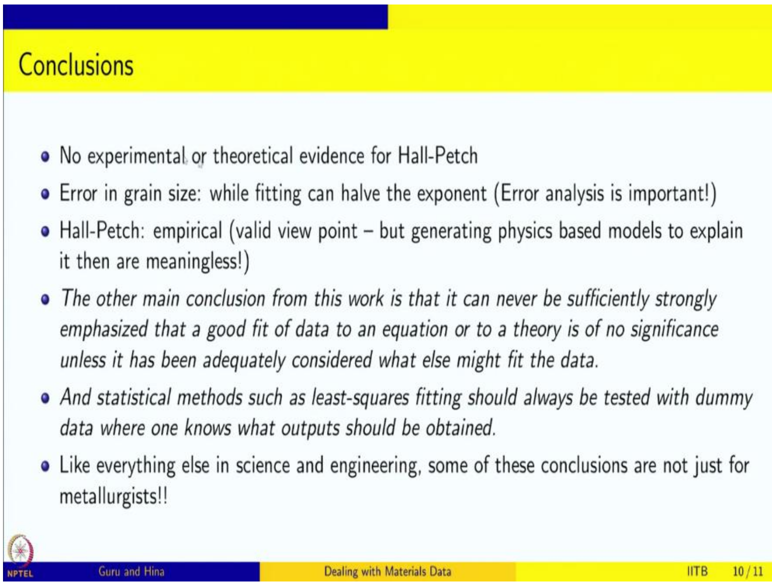
So there is an interesting piece by professor Roddam Narasimha called on the Needham's question which was as to why India and China which were doing great suddenly lost out in terms of science and technology, mathematic, etc. And the answer involves all these things. Do we just take empirical relations and live with them or do we build theories? And when we build theories, how do we test that they are true or not true?

Do we do computations or not? So, all these are actually much broader questions then just deciding on one. We are taking the example of Hall-Petch, but this has a much broader implication. For example, if it is m is minus half is what empirical relationship is giving, can we come up with a physics-based theory as to why it should be minus half. If it is minus 1 then can we come up with the theory?

Now if you get that you know m minus half, minus 1, everything can fit your data, then you cannot really come up with a physics based theory to explain why things are the way they are, which means we are going to lack some sort of deeper understanding into what is happening in the material that we see a particular type of relationship. So the status of Hall-Petch relation in term of these theories is discussed in the paper and the conclusion is that Hall-Petch is not at all supported.

In fact, none of the theories have consistently explained all the observations is the basic conclusion, they discuss some three or four different theories, so those of you who are from metallurgy background will also actually enjoy reading this part and try to understand what the arguments are.

(Refer Slide Time: 16:08)



So that comes brings us to the conclusion of the paper. The paper concludes that there is no experimental or theoretical evidence for Hall-Petch. Experimental evidence because no data does not support and theoretical evidence because none of the models consistently give conclusions that are consistent with what you see.

They also show that the error in grain size while fitting can half the experiment because grain size is not a number, it is a distribution and that it can have very number can actually also affect the fitting. So this shows that error analysis is very important, so it is a good idea to have this practice of doing error analysis all the time.

Because we have even made mistake once when we were not paying attention to error then we are doing the fitting for the cyanide and the hydrogen reaction for example, we were just looking at the numbers and then we realised that the way we used error to weight, this give statistical weight to the data points was not correct.

But later in the case study we have corrected it. So error analysis is important and it should become part of our practice to always do the error analysis. You can consider Hall-Petch to be an empirical relationship which is valid view point but if you do that, then generating physics based theories to explain Hall-Petch coefficient is meaningless.

If you want to generate physics based modules to explain the relationship that you see then we have to discard Hall-Petch because that does not seem to be the right way to go. So that is the conclusion of this paper. And this is a very strong conclusion, anybody who is from metallurgy and most people are from metallurgy and most people from material science know about Hall-Petch and to say that there is no experimental or theoretical evidence for Hall-Petch is really strong statement.

But what is more important and why I have chosen this paper to be a case study and the final case study in this course is that it also has other conclusions which are equally important and little bit philosophical and also about the way we look at things, so it is very important to understand these other viewpoints also when we are doing data analysis because data analysis we have discussed with specific reference to materials science and engineering or metallurgy.

But it need not be only material science and engineering, we are bombarded with data on all fronts, it might be data, it might be a statistics about employment, about car sales or about the demography or pretty much everything. I mean you want to buy insurance, you want to put money in a pension scheme, so everything you have to look at the data and you have to analyse and then you have to come to your conclusions.

So collecting, analysing and interpreting data is much broader skill that one has to develop, just that we are getting to do it using data from material science and engineering in this course, but it is much broader than that and any data for that matter. So, in fact we believe that after going through this course, if you can change your attitude to other types of data and realise that same ideas apply elsewhere, then that would be a big success for this course.

So they say the other main conclusion from this work is that it can never be sufficiently strongly emphasized. That a good fit of data to an equation or to a theory is of no significance unless it has been adequately considered what else might fit the data. So you take a data, you have a theory, you fit it and you see that oh, it is fitting very well. You cannot stop there, you should also think of other models if they can fit the same data.

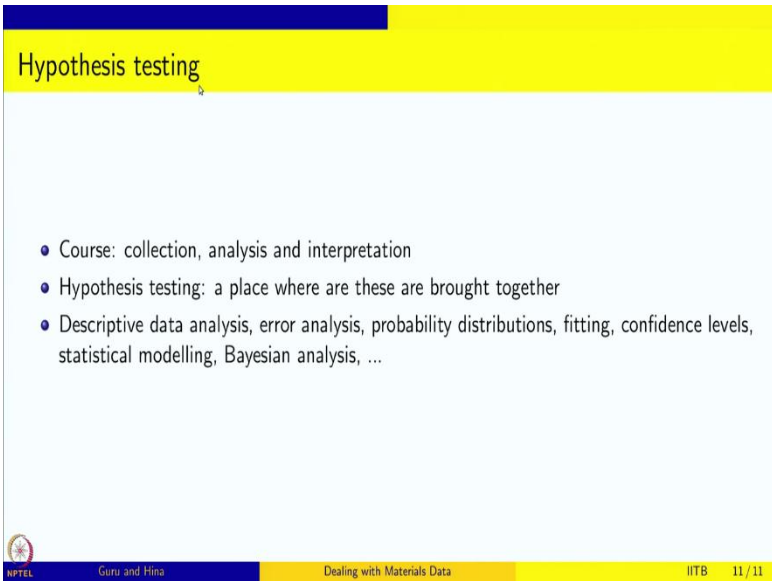
If there are more than one model which can fit the same data, then both the fits have only an empirical status and a physics-based theories then have to step in and say why one should be preferred over the others, so this is very-very important. Many of times we make this mistake of collecting data, having a prior idea or model in mind and just fitting it and see that it fits well and then saying okay, everything is done.

But it is a much better idea to also think about alternate explanations, are there any? If so, can we rule them out, right? And more importantly this is another thing that is not at all emphasized anywhere but it is very important especially if you are a beginning researcher, is that statistical method, such as least square fitting should always be tested with dummy data where one knows what output should be obtained, right.

So these methods and doing this statistical simulation and looking at models is a very-very important thing and we should get into the habits of several things, one is looking at the error analysis, another one is looking at alternate explanations and ruling them out and third one is doing such statistical models to get the maximum information out of the data and more confident about our interpretations.

Those things can happen only if we do all these things. So these conclusions lost to at least for example, like everything else in science engineering, they are not just for metallurgists, materials scientist, I mean this is generally the way in which we are doing the science and engineering.

(Refer Slide Time: 21:55)



So hypothesis testing is a case study and we have kept it as a lost case study because it is a nice case study. This course is about collection, analysis and interpretation of data and this case study brings all of it together, we have seen information about how the data was collected, we have seen several different ways in which analysis can be carried out and we have seen how the data should be interpreted or the analysis should be understood.

And all this are brought together very nicely, for example, whatever we have done as part of this course descriptive data analysis, error analysis, probability distribution, fitting, confidence level, statistical modelling, Bayesian analysis and so on and so forth. So everything has been bought together in this one problem which is very-very well-known problem and it leads to some surprising conclusions.

And so this paper is not very old as you can see it was published sometime in 2016, so its just three years ago, so that brings us to the end of this module on hypothesis testing as well as this course, whatever we have done so far you can see that those tools and techniques are of great use to understand the new data that you will generate as part of your study and your research.

It will also help you analyse existing data and methods much more rigorously and come to better conclusion and some of those conclusions could be surprising like the Hall-Petch conclusion that there is no experimental of theoretical evidence supporting Hall-Petch which is very very strong statement to make. So we will conclude this module as well as this course at this point, but this is only conclusion for this course and we hope that you will continue working with data.

You will use all this knowledge and analysis for your other problems and as part of the course, we will also share some more data sets that has been generated by our research students and we welcome you to play with them, understand them or share your data with us and share your experience of working with the data with us. So welcome to Materials Data Analysis, Collection and Interpretation. Thank you.