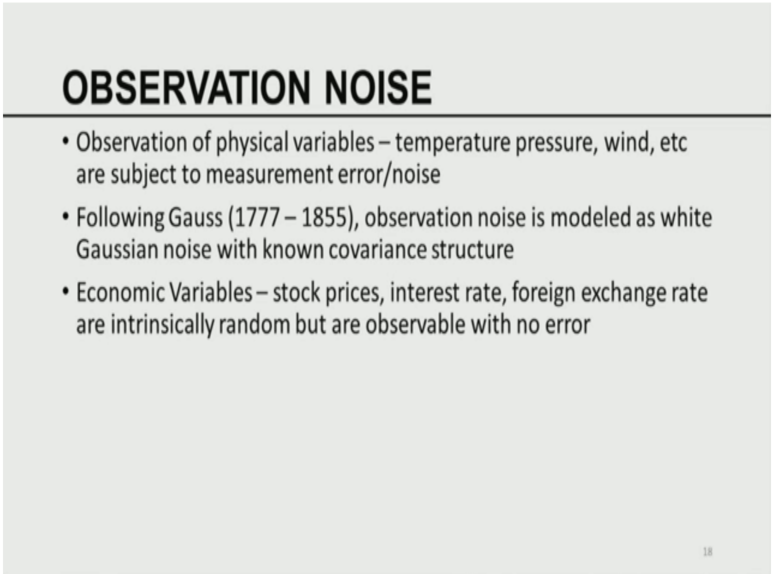
Dynamic Data Assimilation

Prof. S Lakshmivarahan
School of Computer Science
Indian Institute of Technology, Madras

Lecture -02
Data Mining, Data assimilation and prediction

In the last lecture, we have been talking about the holistic view of data assimilation, the role of models, the role of observation.

(Refer Slide Time: 00:28)



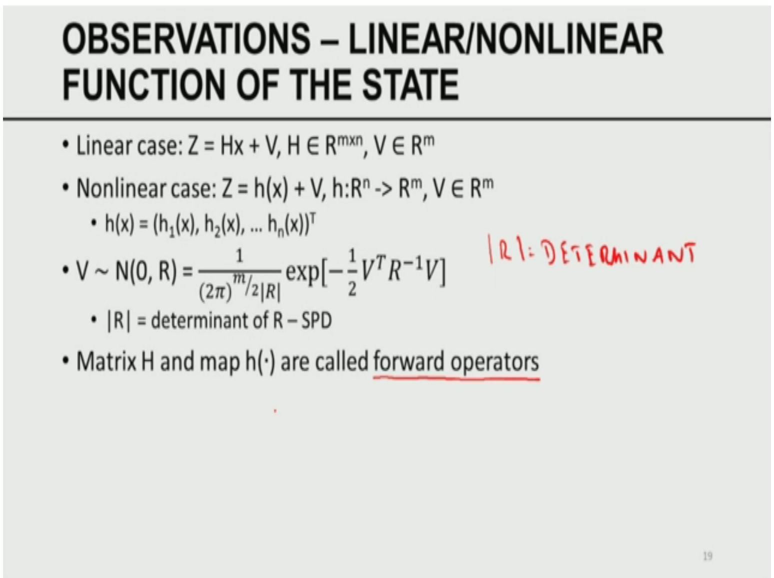
Within the context of observation, we started talking about observation noise. Observation noise depends on the instruments that are used to measure the observation. Satellite observations have certain kinds of errors structure. Radar observations have different kinds of error structures. All these physical quantities. Pressure, temperature, humidity; whatever you want to measure, we measure with instruments and they are always inherently associated with that observation noise. And this observation noise is modeled as a White Gaussian noise with a known covariance structure. That was the theme we saw towards the end of the last talk.

However, there is a fundamental difference between observations in other branches as supposed to engineering and physical sciences. For example, in economics we talk about stock prices we talk about interest rate we talk about foreign exchange rate. These are intrinsically random. But one distinguishing feature is that there are no observational errors, we cannot say I b m prices 22 dollars per stock plus or minus 5 percent. When you say IBM price, you precise there is no error. The observations with variant in the economic quantities are without error; however, observations of pressure temperature they are erroneous these are 2 fundamental differences between the variables in one domain against variable another domain.

Foreign exchange rate it changes from day to day there is a natural variation. There are lot of factors that go into the value of foreign exchange rupee versus dollar or dollar versus euro and so on. And so, there is a intrinsically natural random behavior, but there is no observational error. But observation error whenever is present is random. So, there is randomness associated both the observation, but the randomness come from different sources, one from the source of the error another from the sources of natural variability. In the case of temperature as I we already discussed the temperature a also varies seasonally. So, there is an intrinsic variation on the top of it we superimpose the observation errors.

So, understanding the kind of instruments that are used in observation. Understanding the properties of the errors that the instrument may be associated with. Understanding the natural variations of processes these are all fundamental under these are all very fundamental to how one is going to use observations in data assimilation systems.

(Refer Slide Time: 03:28)



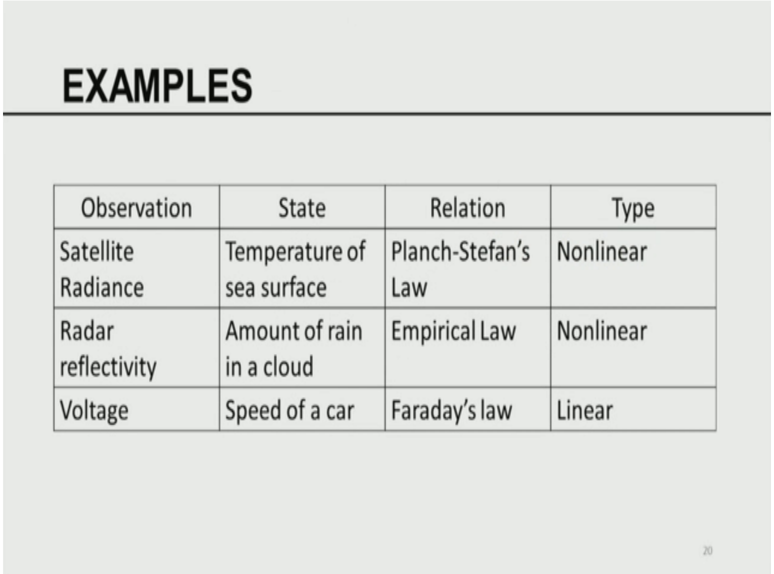
I would like to little a I will give a brief review of some of the models of observations, and how to relate observations with the model. In order to use the observations and to fit the model's observation there must be a bridge between observations and the model, and that is what the bridge we are going to now build that bridge could be a linear relation or a non-linear relation. These relations are functions they are functions of the state. The linear case let us take an example, x is a state of a system, state vector we already talked about. Z are the observation.

X is equal to H times x. V are the observational noise vector. If the observations are perfect V will be 0, if V is not 0 the observations are imperfect. They are noise imperfect in the sense, they are noisy. Here H is the matrix x is a n vector. Z is a m vector by our definition. H is a matrix of size m by n. V is a vector of size m. V represents the observation noise. In the non-linear case the same Z is given by H of x H is a function, and V is again a noise vector. H is a map from R n to R m we will discuss more about the definitions of these things in the mathematical preliminaries that we will talk about. H of x is called the vector valued function of a vector.

So, H of x is a vector consisting of H 1 of x H 2 of x this must be H M of x instead of n H M of x transpose. V is a vector normally distributed there is a Gaussian distribution or normal distribution. It is a 0 mean with a known covariance R. The mathematical expression that describes the Gaussian distribution is given by this. 1 over 2 pi to the power m by 2, the determinant of the matrix R in the in in in this the symbol R represents the determinant of this matrix. And V transpose R inverse V is called a quadratic structure. This essentially represents the bell-shaped curve. H is a matrix in this case linear, H is a map the case of non-linear both of these in the context of meteorology as well as geosciences they are called forward operators.

It is these forward operators that describe the relation between the model state and observation. Just to be able to talk about it little pictorially in terms of examples.

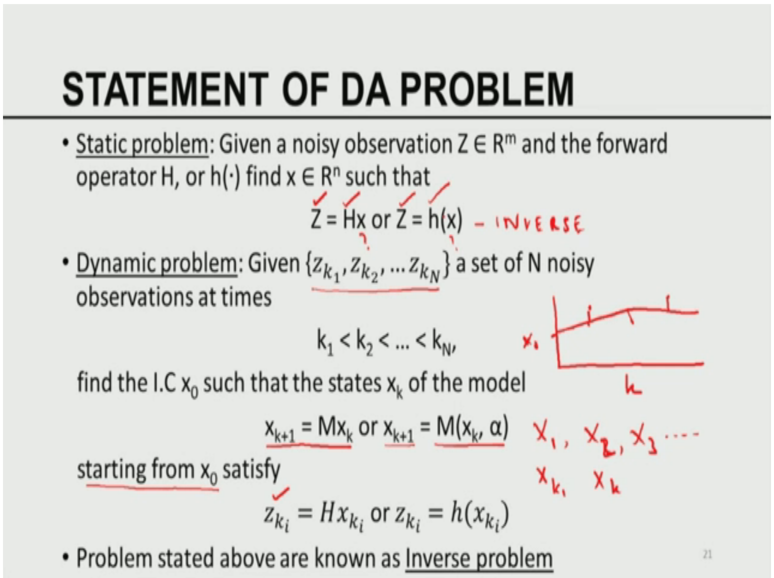
(Refer Slide Time: 06:36)



Observation	State	Relation	Type
Satellite Radiance	Temperature of sea surface	Planch-Stefan's Law	Nonlinear
Radar reflectivity	Amount of rain in a cloud	Empirical Law	Nonlinear
Voltage	Speed of a car	Faraday's law	Linear

For the state could be a sea surface temperature that is the actual thing that I want. Observations are satellite radians in the infrared domain. The relation that relates the radiance the temperature is called the plank Stefan's law. The relation is non-linear, I think it is plank p l c k. Radar reflectivity, the actual state of the amount of rain the law is empirical is non-linear. The observation could be voltage, the state of the car could be the speed of a car. The physical relation is called Faraday's law that is the linear relation. So, we could have non-linear models, linear models, we have the state of the system. We also have the nature of the observables.

(Refer Slide Time: 07:34)



So, what is the statement of a typical data assimilation problem? I would like to describe the problem either as a static problem or as a dynamic problem. Given a set of noisy observations Z belonging to R m, and the forward operator the matrix H or the function H of x. Find x such that Z is equal to H of x or Z is equal to H of x. So, this is the linear problem, this is a non-linear problem. Now you can see the following. We are given you are given Z, you know H. I had to find what x is. Or you are given Z you know the function, we have to find x a.

That is the problem that we have to solve. These kinds of problems as we will soon see are also called inverse problem. Why this called inverse? We will talk about it in a minute. Within the context of dynamic setup, I am now going to have observation taken at different times. So, let k 1 k 2 k n be the different time (Refer Time: 08:53) Z k 1 is the observation taken at time k 1, Z k 2 observation time k 2, Z k n is the observation taken a time k n. So, it could be o, k 1 k 2 k n could be known for the next case. Z k could be the maximum temperature at noon time in a given physical location. It could be downtown London, it could be downtown New York.

Or it could be the price of a stock. It could be foreign exchange between euro and yen Japanese yen on midday on Monday. So, these are the dates these are the observations. These observations are associated with noisy data. We are interested in physical sciences. In physical sciences almost all the observations come from instruments. So, they are inherently noisy, the model is a dynamic model. The model in this case is a discrete time linear model. X k is a state of the system at time k x k plus 1 is the state of the system at time k plus 1.

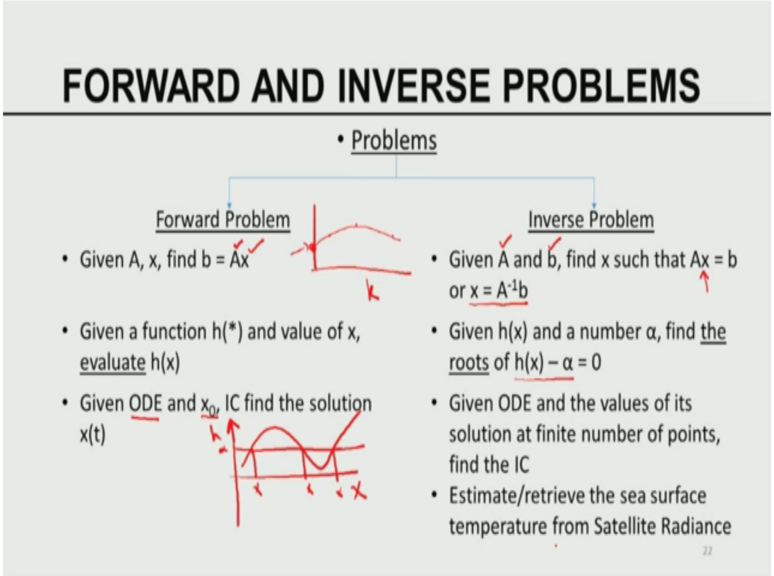
M is a transformation that transfers the state at time k to time k plus 1. And this transformation is a linear transformation. So, they are linear dynamics. Or M of x k alpha m again is a non-linear operator. X k is the state of the system at time k, alpha is the parameter, I know alpha, I know x k, I know the map m k. I can compute what is going to happen at time k plus 1. So, m could represents the price of a stock. X k is the price of a stock today x k plus 1 is the price of the stock tomorrow. M could be a model that predicts the temperature x k could be the maximum temperature today x k plus 1 could be a maximum temperature tomorrow.

So, time is discrete model can be linear or non-linear. So, the state of a system evolves in time. I have n observations of the state given by the Z's. So, what is that we would like to be able to find? The model states can be computed if I know the initial condition or the initial state. So, start the model at the initial state x naught. You compute the model state x naught is given x 1 can be found x 2 can be found x 3 can be found using either a linear model. Or a non-linear model starting from x naught x naught is given. So, x 1 x 2 x 3 these are called models evolution. I have observations at time k 1 observation time k 2 observation for time k n. So, the model has a state at time k 1.

The model has a state at time k 2 and so on. Our job is to be able to as to be able to make sure that my x k i when operated by H gives Z k i or x k i when operated by H equal to Z k i So, the problem stated above is called the inverse problem. Why this is the inverse problem? Look at this now. I have an initial, I have a time k. This is the initial state, I have observations at time k 1 observation at time k 2 observations are time 3 let us assume n is 3. Observations are noisy. I am I if I start at x naught my model is going to generate a trajectory which can be thought of like this. So, at time k 1 I have this this is the model predicted value that is observation this is the model predict the observation model predicted by the value observation.

So, these are called the errors. I want to use these errors to be able to estimate by initial condition. So, knowing the solution I want to find the best initial condition that fits the observation that is the inverse problem. Here knowing Z and H I would like to find x these the inverse problem. So, static problem give rise to inverse problem dynamic problem gives rise to inverse problem these inverse problems have different flavor with respect to solution process.

(Refer Slide Time: 16:21)



So, to be able to understand this inverse problem, now I am going to talk about an underlying fundamental classification of problems themselves at a very high level. Problems can divide into either the forward problem as an inverse problem. For example, what is a forward problem? Given a matrix A, A is given if the x is given I can multiply a matrix by a vector to get b. So, find b given a and x that is the called forward problem. Another example of a forward problem is; given a function H and the value x evaluate H of x for example, I have a polynomial I would like to be able to evaluate a polynomial at a 0.2.

I would like to be able to evaluate a polynomial at 0.75 polynomial is given the point x is given I simply want to be able to evaluate the value of the polynomial. Another example of our a problem, I am being given a differential equation. I am given x naught initial condition. So, given a differential equation I can find a general solution, if you tell me the given initial condition, I can find the specific solution that matches the initial condition. So, these are called examples of forward problems. In mathematics, they generally teach you to solve forward problems in matrix. Matrix multiply, evaluation of polynomials, solution of differential equations.

So much of the mathematical training in calculus and probability theory and other things a lot of effort is spent in in in in teaching how to solve forward problems. Why using the methods for solving forward problems? We have to we are called upon to solve inverse problems. In nature inverse problems are abundant. In order to be able to build expertise in solving inverse problem we need to be able to know how to solve forward problems. So, what is an example of an inverse problem in this context? Given A and the vector b, I would like to be able to find x such that A x is equal to b.

Now, you can see the difference. Given A given x finding b is the forward problem. And in other words, I am given all the quantities in the right-hand side I simply need to evaluate the expression to get the value the left-hand side. In the inverse problem, a part of left hand side is known a part of right hand side is known. There is certain things either in the left-hand side or the right-hand side are not known I have to be able to find the one that is missing. So, finding x given A and b x is equal to A inverse b inverse. A is called A inverse.

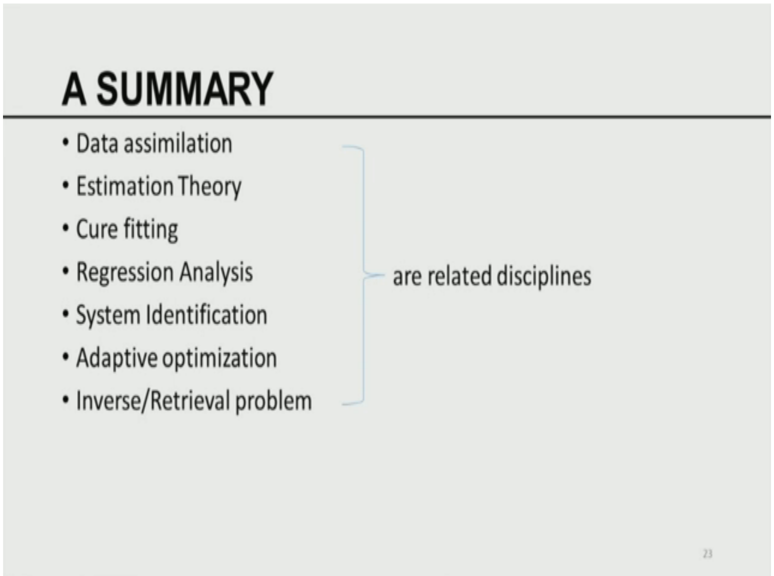
So, that is an example of inverse problem. What the other example of inverse problem? Given a function H of x and a number alpha, I would like to be able to find the roots of the equation H of x minus alpha is equal to is equal to 0 for example, if I have a function x this is H, this is the level alpha. The function may have this shape. So, I would like to be able to find the value of x here. The value of x here, the value of x here. This is alpha. I would like to be able to estimate the values of x at which the function takes the value of alpha. That is called finding the roots of equation if alpha is equal to 0 I am essentially finding the roots of the equation H of x is equal to 0.

So, root finding is an inverse problem, but evaluating a function is called the direct problem. I want you to look at the differences between direct problem and inverse problems. With respect to ordinary differential equation; if I know the initial condition I can compute the solution, but what is the inverse problem associated with the ordinary differential equation here is the inverse problem. This is the time k i. Know the differential equation I would like to be able to find the initial condition, such that the solution of the differential equation matches these values at different 3 different times in other words the solution may come like this I would like to be able to estimate this that is the initial condition I would like to estimate.

So, given an ordinary differential equation and the values of the solution at finite number of points find the initial condition when solved forward will match. The values at those given instances of in time. What is another inverse problem? Estimate are retrieve the sea surface temperature from the satellite radians. We already say said that example of a linear problem where the equatorial pacific is warmer by 3 to 4 degrees. So, somebody has made these measurements these measurements are given in terms of the energy received by the satellite. Energy depends on temperature.

So, knowing the energy I have to re-compute the temperature that led to the radiance. So, that is an inverse problem. So, in many of the applications, we have to solve the inverse problem, but before being able to solve the inverse problem I need to solve the forward problem. So, forward problem inverse problem data simulation intrinsically is an inverse problem.

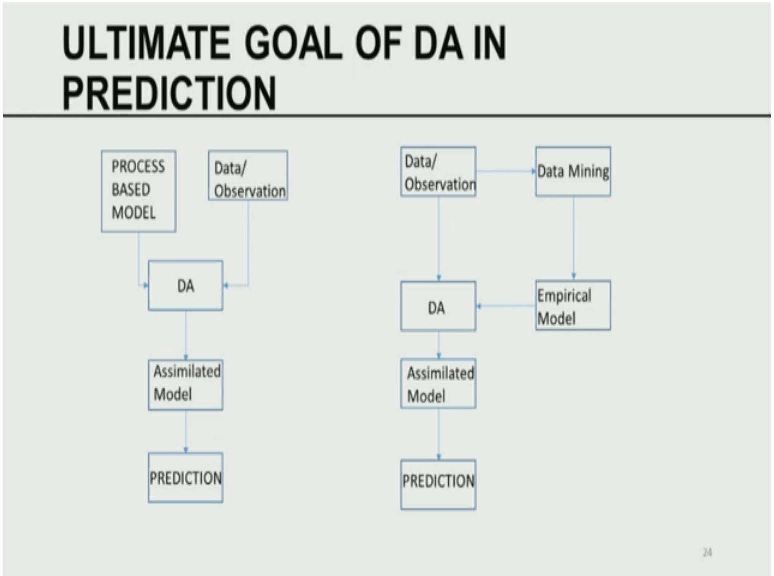
(Refer Slide Time: 18:26)



So, I am now going to summarize the discussions thus for. Data assimilation has been called by different names, data assimilation, estimation theory, curve fitting again there is a spelling error here I will correct. That it is also called regression analysis, it is called system identification is related to adaptive optimization. It is also called inverse or trivial problems, you can see all these are intrinsically related to underlying mathematics of it is called a data assimilation.

So, this is what I meant by 5 blind men looking in an elephant. You may be working in different areas, you may be looking at the elephant from different ways, but from an applied mathematical perspective a personal machine learning a person estimation theory a person numeric analysis. A person system identification, a person geology doing data assimilation. Mathematically they are all talking about the same underlying process, and that is the essence of this first module on point one. Now I would like to continue the discussion further by summarizing the discussion so far in in in a block diagram.

(Refer Slide Time: 19:41)



So, ultimate develop, what is why do we do all these things? Ultimate goal of data assimilation is prediction data assimilation is a part of the predictive process. So, data assimilation is a part of the predictive science if you wish to call. And prediction is fundamental to everything we do in life. So, in the process-based model setup, I have a model based on the process causality based, I have data observation, I bring the models in the observation in a data assimilation scheme. Once I have simulator I get what is an assimilated model. The course is largely con connected with the data assimilation part of it.

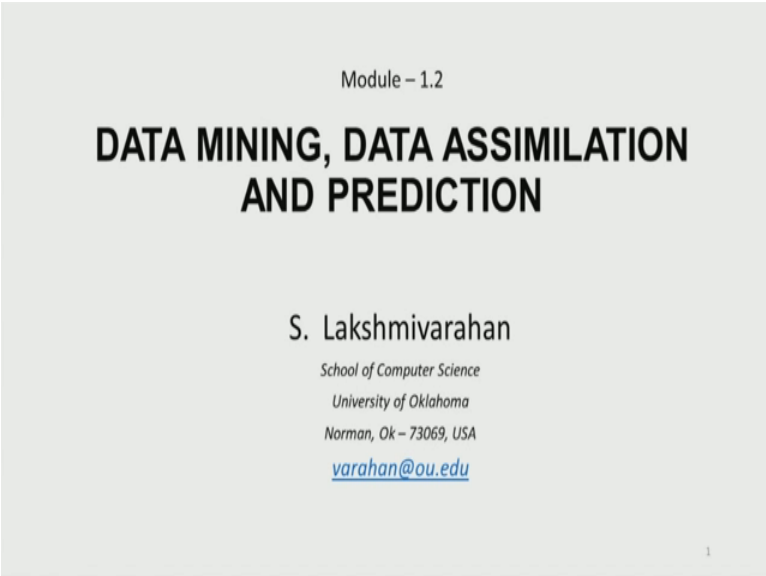
Once I generate an assimilated model, what do I do with it? I run the model forward in time and the model ran for a long time is the one that creates prediction what is going to there are unemployment a month from now. What is going to be the inflation year from now? What is going to be; what is the probability that there will be 30 inches of rain during summer monsoon in Bombay? And a given within 24 hours. What is the probability that the city of Chennai will again see 50 inches of rain within a week? So, all these kinds of predictions are needed to be able to predict we need to be able to do this process.

So, this is the line of argument that underlie the data assimilation within process based scientific engineering domain. As opposed to the empirical domain we have data observations, I need to do data mining. The data mining provides an empirical model. Please remember, model observation model observations, once model observations arise I do a data assimilation. I had an assimilated model I make predictions. So, that is the commonality the empirical model-based data mining is very popular in today's world data mining is A is a large and a vast area of science. The interest in data assimilation, and interest in data mining arise largely because of the need to predict prediction is fundamental to all the things that we do in life

So, that is the end of the first talk. First part of the module. Now I am going to start with. So, in one point one we talked about data assimilation in the widest possible perspective. The aim was even though each of us may be interested in our own sub domain. We would like to be able to see what kind of an animal it is, and how broad and how deep and who are other people doing it. So, that you people from different disciplines come together, if they can exchange ideas methods in one discipline can be transferred to methods another discipline all the disciplines can by this interaction benefit and grow.

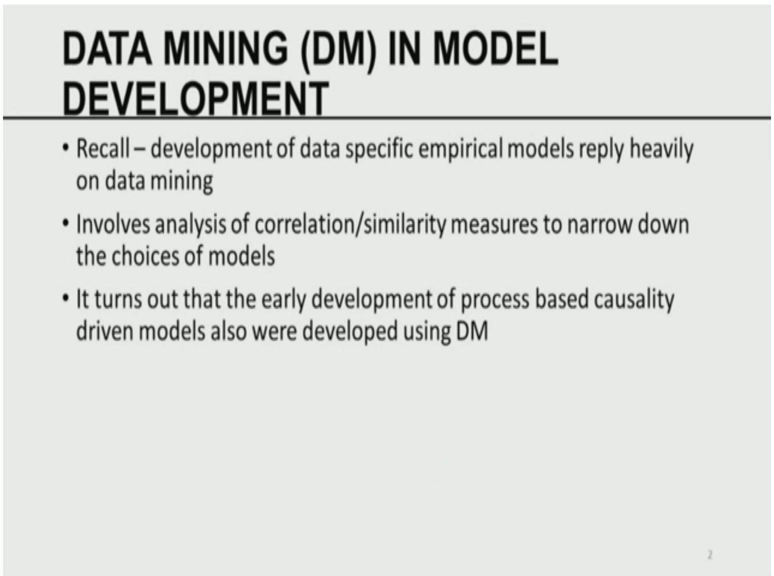
So, I would like to be able to promote that possibility that is why I provided such a broadest possible background, a canvas if we wish to call it that way. And to paint a as broad a picture of data assimilation as possible. Now I would like to go a little deeper into further relations between data mining data assimilation and prediction.

(Refer Slide Time: 23:29)



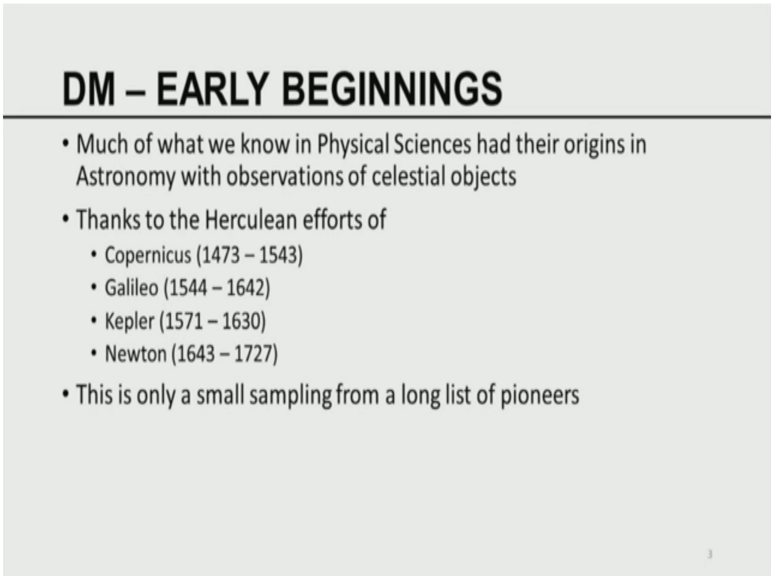
Please understand, data assimilation is the crux of what we do, prediction is the reason why we do what we do data mining is fundamental to everything we do in science. So, I am going to relate all these 3 from my perspective data assimilation data mining and prediction are parts of a continuum.

(Refer Slide Time: 23:53)



To understand this continuum, I would like to reemphasize data mining is fundamental to model development you. Please realize, data assimilation needs models and data. Where do models come from? Models always come from data mining. Models always come from data mining. So, I would like to be able to talk about the fundamental nature that data mining plays in the advancement of science. Recall the development of data specific empirical models rely heavily on data mining. Here what we did we compute computed correlogram similarity of data? So, data is used to develop models within the empirical context. This involves analysis of correlation, narrow downing narrowing down the choices of models, it turns out historically the early developments even in the causality-based models were developed primarily using data mining. So, data mining is not only used within the context of empirical model building, they are also fundamental to how scientists over ages have built models. I am now going to provide some instances to bring out the fundamental nature of data mining in all of model development.

(Refer Slide Time: 25:19)



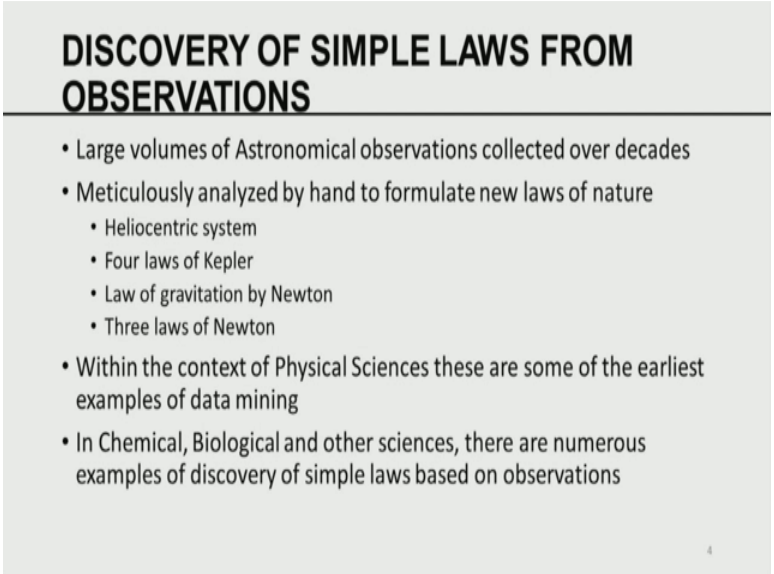
So, data mining it is early beginnings much of what we know in physical sciences today, either origins in astronomy with the observations by humans using simple celestial telescopes. We used these telescopes to observe various celestial objects at various time that bearings. Based on these astronomical observations early pioneers have developed various kinds of models, thanks to a sampling of the various efforts copernicus from geocentric to heliocentrics system, Galileo the notion of gravity, Kepler the notion of motion of planets around the sun in elliptical orbits that led to Kepler’s laws. In my view Kepler was probably the master data miner, he was a model data miner, what did he have?

He had 50 years of observations, he very meticulously analyzed all these observations by hand, and he had great ingenuity and imaginative power to be able to observe these models with the ability that he had to be able to condense all the observations into simple 4 basic laws that we have come to call Kepler’s laws. Every planet revolves around the sun in elliptical orbit, with the sun as it is focus. Each elliptic the in equal intervals of time they sweep equal areas, that gives an ability to compute the velocity. So, simply based on observations, he came up with 4 basic fundamental laws which in my view led to the birth of modern physics thus we have come to know.

Newton then further took those laws and developed the grain theory of gravitation, the 3 laws of Newton the Newtonian gravitational field. Now please look at this all these are empirical laws. There's no way to prove Newton's laws this, but Kepler had had had had beautiful vision he was able to fit a model to the data. So, Kepler’s laws were also empirically derived. Newton's laws are also empirically derived, but they had beautiful intuition that considered the basis for towards to the; to establishing the foundations of modern physics as we have come to know.

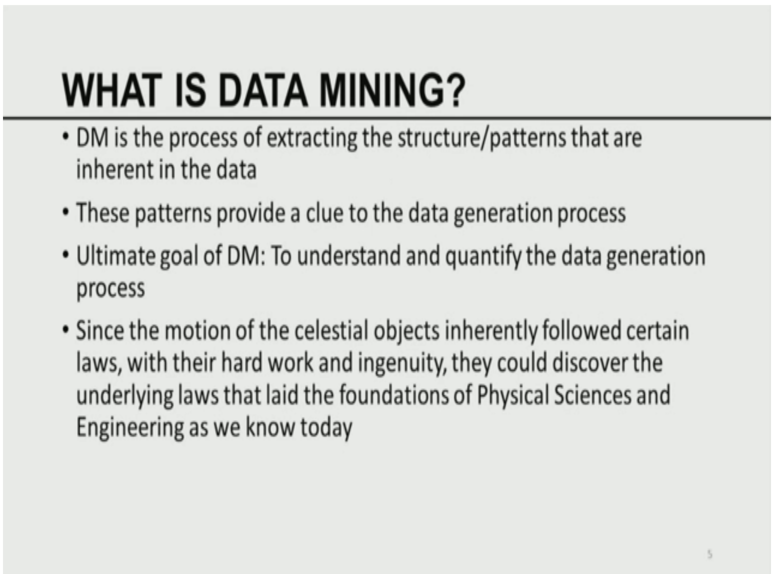
So, this is only a small sampling of a long list of pioneers, both in physics, I am not talking about biology, I am not talking about chemistry, I am not talking about eco sciences. In any and every branch of science that have been Kepler’s whose fundamental contribution began with a very careful very careful (Refer Time: 28:19) observations. So now, you can see data mining plays an absolutely fundamental role even in development of causality-based models. The causality the relation of causality came after understanding, Newton's laws be applied for everything if there is a force there is a equal and opposite force. If there is a force and a mass there's an acceleration so on and so forth. So, these are all some of the earliest examples of data mining within the context of physical sciences.

(Refer Slide Time: 28:49)



I have already talked about this I gained large volumes of astronomical observations are collected over indicates very meticulously analyzed by hand to formulate the following laws of nature. Heliocentric system loss of Kepler gravitation. So, within the context of physical sciences, data mining has played a very fundamental role, and still continues to apply. Almost all the laws in chemistry all the laws on biology the law of evolution by Charles Darwin is an empirical law that he enunciated. So, data mining is fundamental to advancement of any discipline not only in the development and empirical models, but also causality-based models as well.

(Refer Slide Time: 29:38)



So, what is data mining? In it is broader sense. Data mining is the process of extracting the structure of patterns that under that underlie or that are inherent to the observed data. For example, Kepler did not know that everything revolved around the sun in elliptical orbits. But the plants have been doing it, which we did not know by careful observations he was able to analyze hey this is how the plant is behaved, the ability to extract explain the phenomenon that underlies the astronomical objects and being able to summarize it in very simple laws. That is one of the most fundamental example of data mining. So, the patterns were already there. So, these patterns essentially generate, how Mother Nature generates this data.

So, the fundamental importance of data mining is to be able to understand that data generation process. If it can generate the process in the laboratory; that means, I have understood the system. So, the ability to extract the structure underlie that underlie the observation and summarize them in the form of simple laws, that is the ultimate goal of data mining to understand to quantify the data generation process, I would like to emphasize the data generation that is fundamental. Since the motional selected object inherently followed certain physical laws with their hard work and ingenuity they could discover.

Many of the underlying laws that constitute the fundamentals of physical sciences engineering as we know today. So, this is very fundamental to the advancement in science. So, we all have been doing data mining without calling it data mining analysis of data.

(Refer Slide Time: 31:36)

Now, I would like to come back and fast forward to the modern time. In today's technology thanks to this technology we are living in a world where data is much abundant. Let me go back, if you look at the technology of the time of Kepler he did not have anything. He did not have pencil as we have it he did not have the paper. As we have it he did not have the pen as we have it, but they have recorded observation in a particular way, but he know he was pretty good in being able to analyze very meticulously.

So, the amount of data from today's perspective may be small, but it is a large data set for him at that time given the technology. So, volumes of data doubles in every 3 years, thanks to technology. Computers, large scale storage devices, communication and sensor technologies. For example, in many parts of United States, we are prone to earthquakes, they have built tall buildings Los Angeles, basin, San Francisco, basin, they are they have been building on the top of falls. They know that there is a fault still they build, but what is the challenge they would like to be able to build these buildings that can withstand 8.5 Richter scale 7.5 Richter.

So, to be able to understand how this building behaves what is that they have done they have wired the bridges, the tall buildings with sensor devices, because if the earth shakes the building respond they would like to be able to record continuously the response of the buildings and bridges to tremor the tremors are happening all the time some 2.5, some 3.2, some 4.5.

So, by understanding these observations for very many different types of tremor, engineers can develop where the cracks develop. So, ability to observe has vastly improved thanks to the sense of technologies. Thanks to the ability to communicate. Thanks to the ability to store. In fact, United States there is data bank, where they have all the observations of all the radars ever since they were in implemented one database system that you can go back whatever that particular storm that affected by Miami in 1972.

What did the hurricane that hit Philippines in 1882 18 not 1882, 1985 and so on and so forth. So, data is available in abundance. Today's interesting data mining includes a very vast array of topics physical biological medical sciences. So, what is the one data mining project of great interest in medical sciences? We would like to be able to relate the structure of the genes to the occurrence of diseases. It is pretty much understood that defects in the genes caused the diseases. So, they would like to be able to identify what kind of defect causes pancreatic cancer. What kind of defect causes some other type of cancer associating the defect with the disease?

And then once I know this is the defect then I can try to compensate for the defect. So, in medicine there is a tremendous interest in in in in data mining, space exploration all branches of engineering, I would like to give an example of an engineering thing that is of great interest in civil engineering for example. As you know in United States we have highway system which is the envy of the world they have east west is about 3,000 miles north south is about 3,000 miles. You can think of a square 3,000 by 3,000 miles, they have roads everywhere. 1 mile of a 2-way highway cost 5 million dollars.

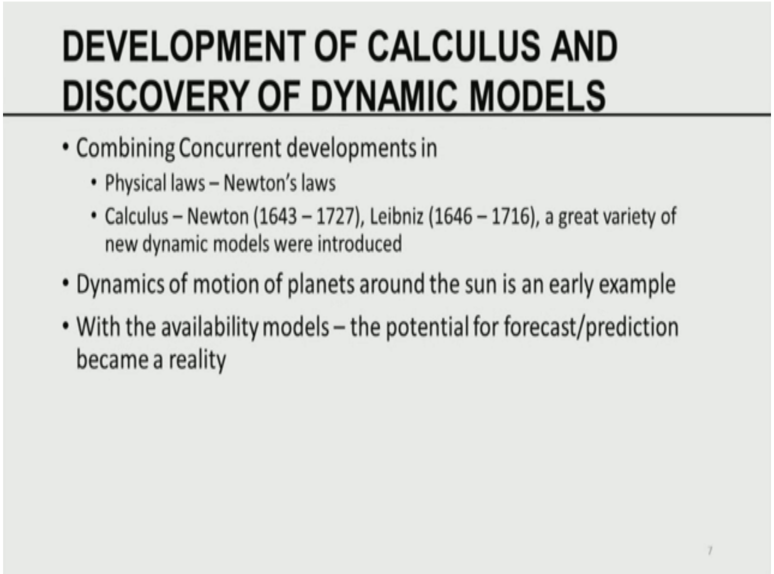
You can imagine there are 10 different parallel systems east west there are 10 different parallel systems north south each of the roads is well over 3,000 miles. You can see the amount of money that they have invested in developing infrastructure. 5, 3 to 5 million dollars per mile. So, what is that they would like to be able to do they would like to be able to optimize the performance of the concrete. The northern ranges of fringes of United States, the low degree the low temperature is minus 30. The high temperature could be 85. In southern fringes, the high temperature can be 100 and 10 the low temperatures can be minus 10. So, when you subject these road concrete to such a wide variation we need to understand the expansion contraction properties. So, in a laboratory they build concrete blocks of various compositions. They subject them to different kinds of other conditions.

They run an artificial wheel, they can adjust the pressure they can adjust the speed they can adjust the temperature surroundings, and they measure the wear and tear. So, they collect ton of empirical data, once they collect the empirical data they would like to be able to model the wear and tear of the pavement with respect to speed with this respect to temperature with respect to all kinds of quantities of interest. So, that is the science which is developing and it has tremendous implications in terms of transportation budget.

To maintain these roads is a very expensive. Environmental sciences, ecology, economic, social sciences, finance, banking, sports and creation in sports and creation, you are always want to predict which team will win the super bowl, which basketball team will be the national champion. So, there is a lot of betting goes on there's a lot of interesting predictions and sports. Of course, a government and private companies, what is an example of a data mining with respect to with respect to the to government a government wants to develop budget priorities. They would like to know what would be the net amount of revenue that will be available. So, that they can allocate this much for education that much for transportation that much for crime fighting so on and so far.

So, somebody has to predict what will be the budget available as for us to be able to spend in June first of 2016, and I have to predict that in December 2015. So, that the legislators can take 6 months to decide, how to divide the incoming budget into various process. Every government agency large and small, local and national, state level they all have to do this mechanism for prediction and that a data mining project, you look at the revenue stream or the past 15 years. You model the revenue stream and then predict. So, there is a lot of data mining projects involved. We will talk a little bit more about data mining later back to some of the early discoveries in astronomy.

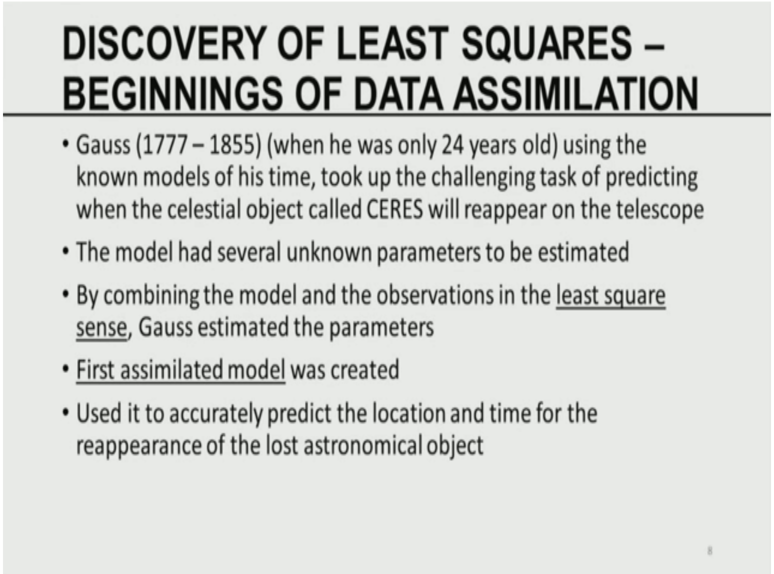
(Refer Slide Time: 38:46)



I would like to bring in data mining with respect to other developments. The development of calculus and the discovery of dynamic models, by Newton, Newton was essentially the first one to the Newton and Leibnitz are credited with the discovery of calculus. Once calculus was understood, the notion of dynamic models came into existence. The notion of a differential equation the first set of differential equation they developed is for the motion of the a pair of the sun motion of planets around the sun. So, the dynamics of motion of the planets from the sun was well understood in the early days of Newton it is himself. So, with the available look at this, now take us through Copernicus, Kepler, Newton discovery of calculus, they discovered dynamic models. When the dynamic models are known I need to know the initial condition I can make predictions.

So, the notion of dynamic model's model building data assimilation data mining prediction that as clear as it is today in the days of gauss in the days of Newton. So, you can see the interdependence between data mining data assimilation and prediction right from the early days. So, once we have models the potential for prediction our forecast forecasts and predictions are essentially same to synonymous words, and ability to predict became reality. So, we are able to predict many of the some of the aspects of the geophysical systems starting at as early as Kepler as early as gauss as early as Newton.

(Refer Slide Time: 40:30)



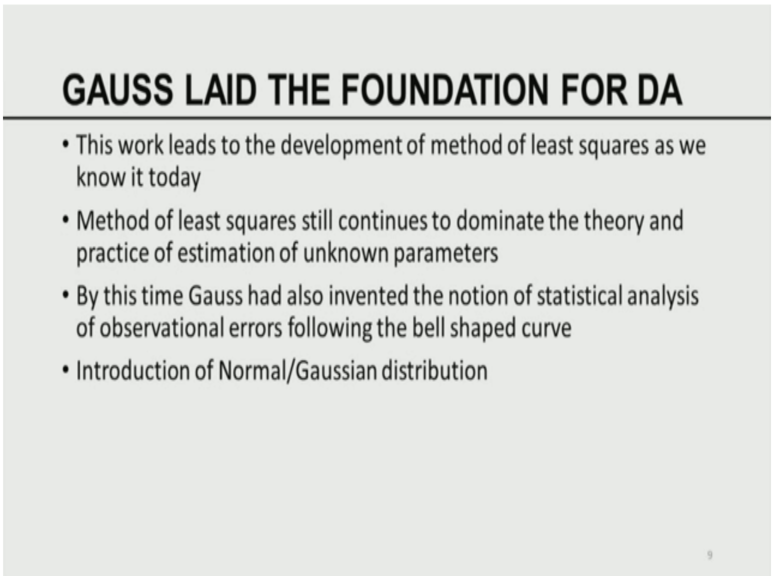
So, we talked about models, we talked about data mining we talked about process-based models, and empirical based models and so on, now I would like to talk about data assimilation, the birth of data assimilation. Discovery of least squares heralds the beginning of data assimilation. Historically gauss was only 24 years old. He was given a very challenging problem. What is that? Astronomers were observing an object called ceres, it is a moon. It went out of sight. So, they would like to know where it went, then it will reappear, what bearings did not reappear, what time it will reappear. All that he had was a bunch of satellite observations of the past track of the ceres take of it.

The problem now he was given only a data relating to the various positions taken from observations. So, what is that he has to do? He has to utilize that data to build a model. So, he built a static model to be with certain parameters. He combined that model with the observation he invented the least squares as we know today. So, gauss using the model he created from the data and use the data to estimate the unknown parameters. He invented the notion of least squares. He estimated the parameters of the model. He made a prediction when the objective what bearings it will appear sure enough a date. So, gauss is considered to be the father of modern data assimilation.

The very first assimilated model was created by gauss. Not only he created the he created the prediction, the method that he used to make the prediction is called the least squares method. The least squares method is the workhorse of the data assimilation industry. So, he used this model to be able to predict the location the time for the reappearance of the last of astronomical object, and that was the first success story about data mining, data assimilation, and prediction all rolled into 1.

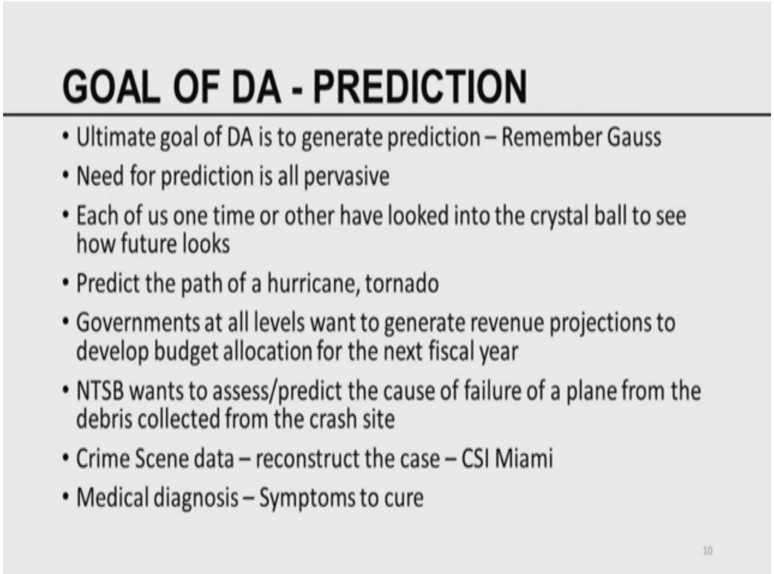
That was the first time. It is in this context he was trying to fit a model to the observation he found the observations were not consistent. He knew the observations had errors, it is at that time he also simultaneous discovered the Gaussian distribution as we now know the bell curve. So, you can see at that one moment he discovered the distribution of observation using bell curves. He invented the notion of least squares. He developed the first assimilated model ever known to mankind. He saw the first inverse problem ever known to mankind, and he by solving inverse problem made a prediction which is verified. So, that is a glorious time with the analysis of the data assimilation literature.

(Refer Slide Time: 43:38)



This work by gauss helped to laid the foundation for data assimilation, his work led to the development of method of least squares as we know today. The method of least squares tell continues to dominate the theory and practice of estimation of unknown parameters in all domain that we talked about stated ranging from curve fitting to restate a little regression to different problems that we already discussed by this time gauss also had co invented the notion of statistical analysis of observational errors the bell-shaped curve. That helped to introduce the notion of Gaussian or normal distribution.

(Refer Slide Time: 44:20)



Now what is the goal of data assimilation? Why do we do all these things? Why trouble ourselves with all this? Ultimately, we want to be a; to predict to ultimate goal of data assimilation to generate prediction. Remember gauss, he was interested in predicting where the object went and then it reappeared. Right from gausses days, the need for prediction has become all pervasive. Each of us at one time or the other have looked into the crystal ball to see how our own future looks like. I want to know some people go to us from astrologers, some people they do not believe in astrology, but they use other means how my life will evolve. We are all interested in our own future.

So, this need to be able to predict how we are going to evolve is a fundable is a fundamental interest to all of mankind. Here are some scientific examples. Predict the path of hurricane. I have seen a hurricane being just beginning to appear as a low-pressure system, just and the western fringes of Africa. The easterly carries them to the United States. We would not like to be able to make a prediction is going to kill Cuba and get into the gulf. I was going to go to Bermuda, and then turn right to the sea weather will cut into the United States. So, these kinds of predictions are very much needed especially from august through December is a very active season for hurricane.

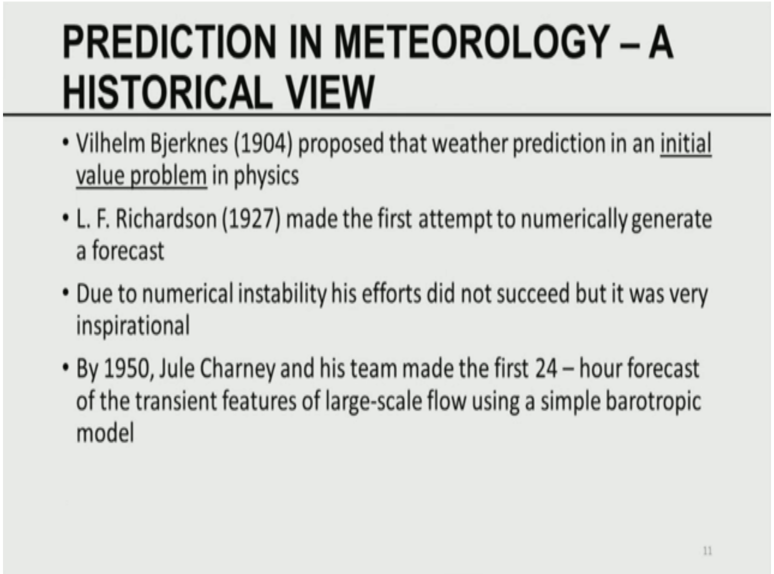
So, hurricane research center in Miami is extremely active in predicting paths and intensities of hurricane. Government at all levels want to generate revenue projections to develop budget allocations for the next fiscal year we already talked about that. Another example our prediction it essentially comes from national transportation safety board. Suppose a plane has crashed, planes do crash unfortunately. After cash happens they invite specialists. They collect the debris from the debris field they try to reconstruct and judge what must have brought the plane down. It must be a crack, it is a bomb.

So, it is a kind of a reanalysis that they have to do from the observations there to go to the cause. So, any kind of such analysis it is an inverse problem. If I had known I would have been able to predict, but I thought everything was very good the plane takes off, but something happens there's a failure. So, from the failure data I have to estimate the cost for the failure. Crime scene investigation I wanted to tell you who are all the other people who are in data mining. We are not the only one doing data mining the NTSB the national transportation safety board every time they are called upon where the train accident, or the bus accident or a plane accident.

They are called upon and they are experts in in in dynamics of various types looking at the crash structure. They will be able to rebuild the cause for the failure. Crime scene investigation. Police always comes to the scene after a crime was committed, if they were there the crime may not have been committed. So, the crime as I live in committed takes they take meticulously all the observations and they analyze these observations, and then they if there is a firing of a gun which direction could have come from what kind of bullet was used. They analyzed the reasons, you may remember a tv show called crime scene investigation C S I Miami.

C S I, New York is a very popular TV show in united states. A team of experts go analyze a case to be able to retrieve how who in what means committed. The crime medical diagnosis from symptom to the cure. I have symptoms diagnosis is an inverse problem from observations, I have to diagnose what the problem is. Once I know the problem I can I can then cure. So, symptoms to diagnosis to cure is an example of a data assimilation problem, a prediction problem a modelling problem and so on.

(Refer Slide Time: 49:02)

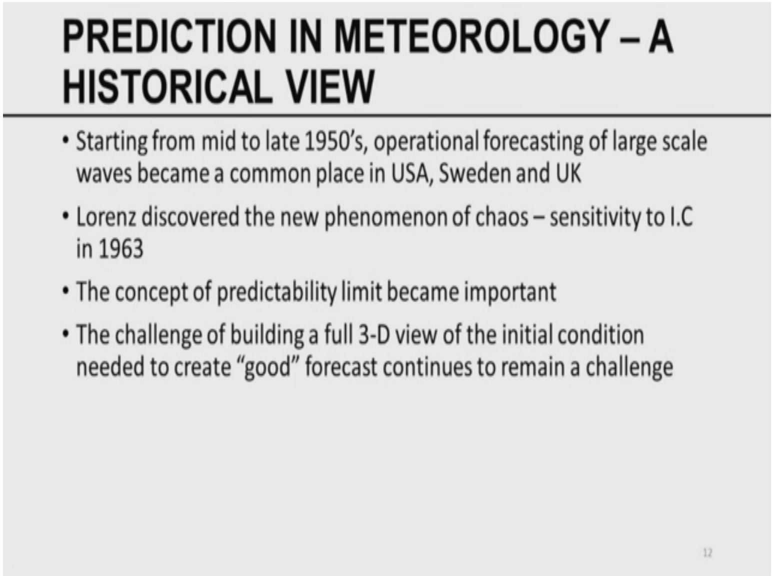


I am now to give some historical perspective. Because the course is going to be largely for scientific audience. Prediction within the context of meteorology historical view I am going to provide a quick review of some other things specific to atmospheric sciences. Vilhelm Bjerknes in 1904 was the first one to propose, the weather prediction problem as an initial value problem in physics. If the weather evolution can be described by a primitive equation forecasting, a solution in initial value problem mathematically. Richardson 1927 made the first attempt to numerically generate a forecast many of us would know that, but he used humans. So, he arranged people in a 6 by 6 array.

Each one of them represent a point of the grid. Each one of them did certain local calculations. They exchanged the calculations all around. And then the com computations evolved in time. Even though his experiment did not meet with success as we know today, but it is the model almost all the work that we do we can relate back to what Richardson did. So, he paved the fundamental foundations of modern predictive science, as it is used in geographical, I am sorry geological our geophysical domain. Due to numerical instability his efforts did not succeed, but it still continues to be very inspirational.

In 1950, Jule Chaney and his team using the very first computer in 1951, 52 made the first 24 hour forecast of the transient features of large scale flow using simple barotropic model. But that began the modern era of weather prediction. So, this is a very short history of predictions within the context of atmospheric sciences. You can see it all began in 1770 with gauss. So, within a span of 150 plus years, it has spread to other sciences and this is a very short history of what happened within the meteorological sciences.

(Refer Slide Time: 51:23)



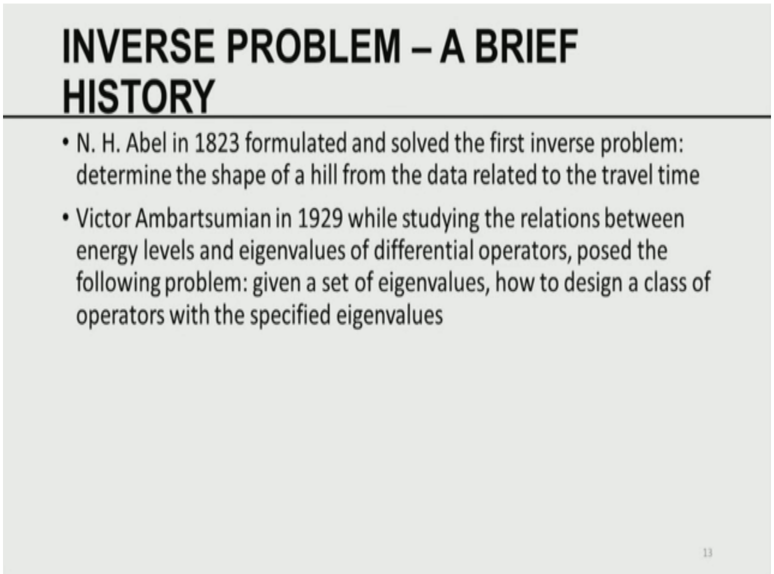
Starting from the late fifties operational forecasting centers became very much interested in in in predicting large scale waves. Sweden united states UK, they really led that efforts globally. In 1963 a great mathematician by name Edward Lorenz discovered the new phenomenon of chaos. Chaos relates to sensitivity of the model to initial conditions. So, if you're trying to do data assimilation, and you are trying to estimate the initial conditions. If the initially the modeled solution is very sensitive the initial condition, you have lot more challenges that troubles the headaches. So, the models chaotic is very difficult to do data assimilation that became very apparent with the fundamental work and the discovery of the phenomena of chaos in 1963.

With that also came the notion of what is called the predictability limit. So, what does it mean certain phenomena we can predict for 100 years lunar sooner eclipsed? In fact, there is a famous saying that goes somewhat like this and astronomers can predict, where the moons of Jupiter will be at midnight today, but he has no way of knowing where his teenage daughter will be at midnight today. The behavior of teenage daughter is very different from behavior of the moon of the Jupiter, one phenomena is predict perfectly predictable other phenomena is totally unpredictable.

So, that brings in the notion of what is called predictability limit. If there is a notion of a predictability limit, we need to be able to create good forecasts and we also need to be able to tell; what is that time span within which my forecast will hold. If you look at some TV predictions in another state they will give you a fixed prediction some TV will give 10 days prediction. If you carefully know 10-day prediction from today 9-day prediction, tomorrow eight-day prediction day after tomorrow you can see the variance of the prediction, for a given day made from 10 days, 8 days, 3 days, 5 days, 2 days, 1 day and the day.

So, this predict this this variance is very large. And that goes to tell that 10 days weather is not predictable. The good predictability for weather is hardly 3 to 5 days. We cannot we cannot predict better than that. So, the notion of predictability limit became very fundamental.

(Refer Slide Time: 53:39)

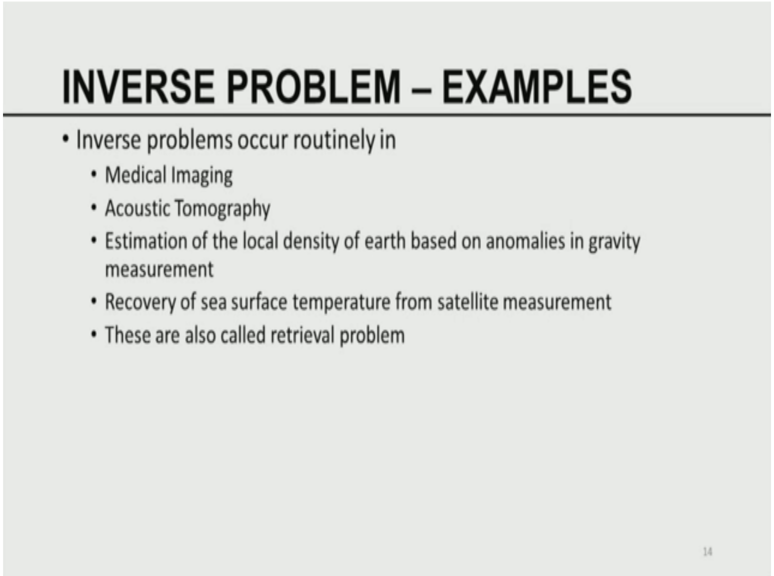


Importance of inverse problems in earlier parts of the lecture today. I would like to give a brief history of these inverse problems. Please remember, I've been giving historical backgrounds for prediction, because start with background for in inverse problems, data mining and so on. I think it is better to understand the history in a fundamental way. So, that we can appreciate the connections between what we do in our own trenches as opposed to the vastness. Inverse problem has a rich and a deep history.

The first example of inverse problem began in 1823 when Abel formulated and solved the very first inverse problem ever not mankind. And what the inverse problem? Determine the shape of a hill from the data related to the travel time. So, what the idea here pretend that there is a hill pretend that I have a ball, pretend that the ball is sliding down the hill, I have been able to make observations about the ball sliding the positions at various times. Looking at the positions of the object at various times my job is to be able to reconstruct the shape of the hill on which the ball has been rolling down. That is an inverse problem.

That is the very first problem known ever to mankind stated and solved. Another fundamental work within the analysis of inverse problem is by the Russian mathematician victor Ambartsumian in 1929. He was a physicist he was a mathematical physicist. He was trying to relate the energy levels in the atom with respect to the Eigen values of certain differential operators that relate to the states and the energy. And he said the following if I know the operator I can compute the Eigen values. That is the easy problem, it is like given a matrix. I can compute the Eigen values, but he for the first time ever asked the following question. If I specify a set of Eigen values, what kind of an operator will endow itself with this given Eigen structure? So, given a matrix I can compute Eigen value, but what is the matrix for which 3 and 5 are the Eigen values. So, from Eigen values to the matrix, matrix to the Eigen values. Matrix Eigen value is called the forward problem Eigen value to the matrixes inverse problem. So, these are considered to be the very first instances by mathematicians to consider solve important inverse problems.

(Refer Slide Time: 56:36)



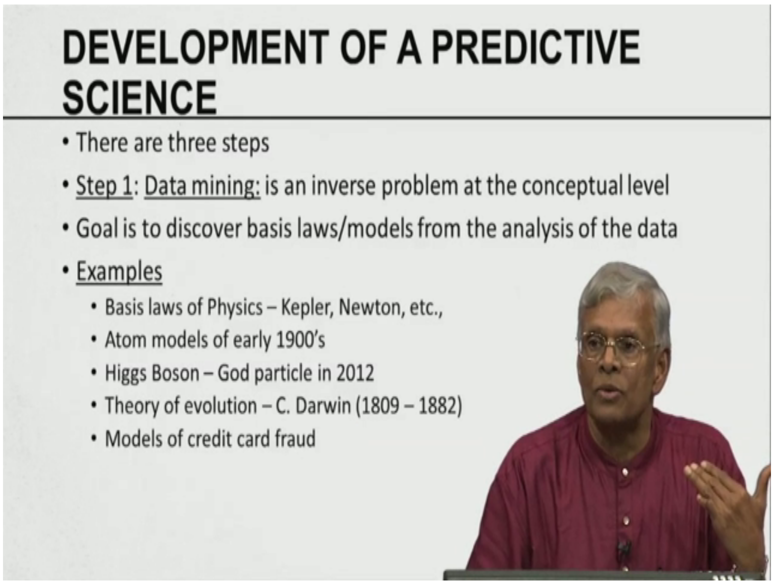
Inverse problems occur routinely here are some examples. In medical imaging how do doctors find there is a tumor? You go to through the cat scan, they take pictures, they analyze these pictures from the picture you have to tell whether you have a particular disease or not. That is an inverse problem. How do I know there's an air craft that has been flying for 30 years? I would like to be able to verify the air craft body does not have any cracks. I have to test it big, because if there is a crack it was not visible outside. It could be very dangerous to the function of the inner craft.

So, what do they do they send a very fine acoustic signal, and receive it another end and based on the properties of the received acoustic signal they can say whether there is a crack or not that is what is called acoustic tomography. Another place where inverse problem occurs. In physics in solving the regular problem we assumed acceleration due to gravity is 980 centimeter per second square, is that right? That is a standard valid, but that is an average value. But if you really take a gravity meter, I am take it to different places let us take a city called London.

You take the this gravity meter, the gravity changes from place to place to place to place to place. So, there is a fine anomaly in the value of the gravity. So, there are beautiful instruments that are designed to be able to measure these anomalies. So, what is that? What is the use of these anomaly measurements? The local anomaly depends on what is behind be below the earth. So, if there's a hard rock the gravity is more, if they are loose sand the gravity is less. So, geologists in their exploration they try to run this gravity anomaly meter and map the gravity field.

And once the gravity field is well mapped their geologists try to solve an inverse problem to discover oil, water, hard rock, granite and so on and so forth. So, that is an inverse problem. Another inverse problem we have already seen at the risk of repeating recovery of the surface temperature from satellite measurements. These are called inverse problem they are also called retrieval problems.

(Refer Slide Time: 59:10)



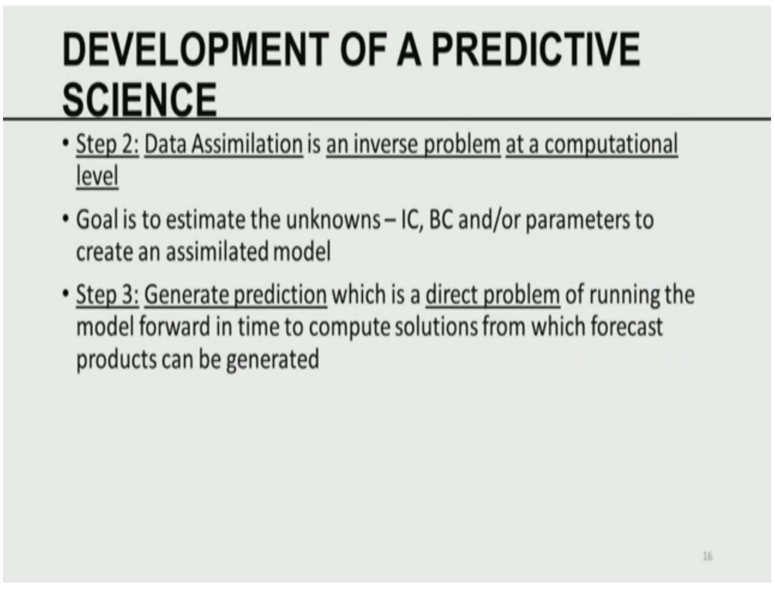
So, in trying to summarize our discussion of the development of predictive science. There are basically 3 steps the data mining step that is an inverse problem at the conceptual level at the highest level. The aim at this level is to be able to discover the basic laws. These are encapsulated as models. Which are the fundamental basis for analysis and they are discovered by analyzing data. Examples include Kepler's laws after models. In the early 1900's do we know the atom very well no, but we can do lots of things in physics with material science. But we will still do not understand the atom very well. Nobody has seen an atom yet data model still evolves you all know in 19 20 12 there is a big buzz about that the discovery of a new particle called Higgs Boson is just called Karte particle.

So, as one puzzle solve other puzzle comes into play at the accelerator center in Geneva, they continue to do the experiment. Almost all the atomic physicists from the world, they joined together in trying to uncover. What do they do? They smash atoms are very high speed, at the time when the splash happens there's a lot of energy is splashed. They take pictures of these energies flash. From the pictures and attacks of objects, they are going to identify this track must have been caused by this particle. This part track must have been caused by these kinds of particles.

So, looking at the tracks that happens at the kind of explosion is very fundamental to uncovering what must have exploded. So, that is an inverse problem. Even today in physics after thousand after nearly 200 years of interest in physics, we still tried to complete the picture of an atom. Theory of revolution by Darwin is an example of an inverse problem. And what are the models of interest in today? I would like to be able to identify credit card fraud. What is the kind of a profile of a person who will commit a fraud in not paying for the credit card?

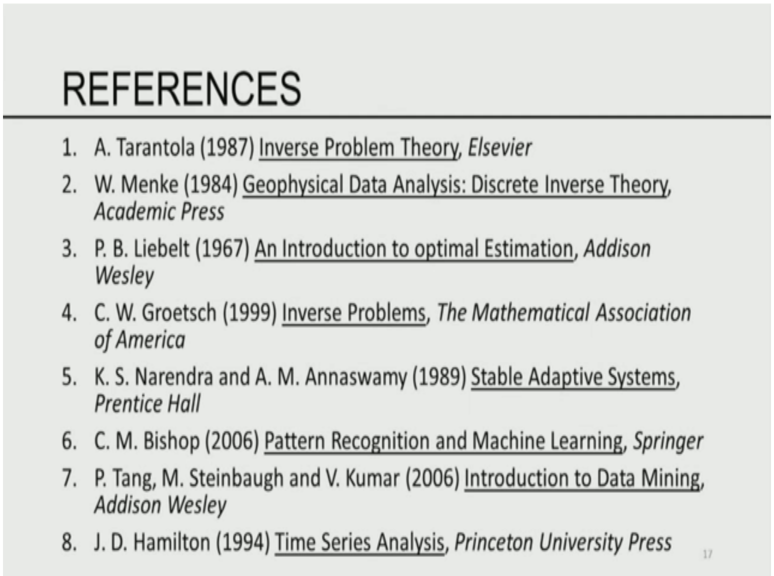
So, that is an example of a data mining project. You can see credit card information is the only thing we have, there is no law, there's nothing. I would like to be able to mine this data to be able to develop a profile of a person. And these profile ones developed are given to the bank, as you apply your indicators are compared with the profile. And then they you to predict yes, he seems to be his history tells me that he does not fit the profile. We can grant him a loan 100 lakhs to buy a house and so on and so forth.

(Refer Slide Time: 61:56)



Step 2, data assimilation is an inverse problem is the computational problem. So, data assimilation is largely an applied mathematical and a computational discipline. The goal is to estimate the unknown initial condition boundary condition parameters and to create an assimilated model. Step 3, once the assimilated model is created, we would like to be able to run the model forward in time to generate forecast. So, developing the forecast is a direct problem. Because you simply need to run the model forward in time which everybody knows? So, to be able to solve the direct problem of prediction to forward problem a prediction, I had to solve 2 inverse problems. One to develop the model another to do the data assimilation. So, inverse problems are 2 levels forward problem at the third level are intertwined that complete the picture of what underlies the predictive science.

(Refer Slide Time: 62:54)



Here are some references. Toronto 1987 inverse problem theory is largely for geophysical audience. The language is couched largely for geophysicist. Geophysical data analysis discrete inverse theory by Menke is a very nice little book. Am li built is the introduction to optimal estimation 1967. Inverse problems a mathematical associates America. This book for on inverse problem is for undergraduate students, he has is a very beautiful delightful book on simply simple inverse problem that you can introduce even to undergraduate students. Narendra and Anna Sami, they have written a very beautiful book on stable adaptive systems. You remember system identification, adaptive system that is one. Bishop a book on pattern recognition and machine learning.

Tang and Steinbaugh and Kumar introduction to data mining. Hamilton's book on time series analysis. These are some of the references taken from different disciplines. All have a common theme. The examples are different. The mathematics that underlies their always one of the same. So, I hope I have given you a broad overview of data assimilation, data mining and predictive science. This completes of a lecture 2. We will talk about various mathematical prerequisites in the next couple of 3 classes. Once the mathematical prerequisites are covered, then I will delve into different types of data assimilation problem in the coming lectures. Thank you for your patience. Bye.