The Nature of Prediction (and the Prediction of Nature)

The Nature of Prediction (and the Prediction of Nature)

Inaugural lecture

delivered upon accession to the office of
hoogleraar Numerieke analyse en dynamische systemen
at the University of Amsterdam
on 21 April 2011

by

prof. dr. ir. Jason E. Frank

This is inaugural lecture 412, published in this series of the University of Amsterdam.

Lay-out: JAPES, Amsterdam

Mevrouw de Rector Magnificus, Mijnheer de Decaan,
Geachte leden van het curatorium van de leerstoel Numerieke Analyse en Dynamische Systemen,
Geachte leden van het bestuur van de Stichting voor Hoger Onderwijs in de Toegepaste Wiskunde,
Beste collega's van het Centrum Wiskunde & Informatica en het Korteweg-de Vries Instituut voor de Wiskunde,
Zeer gewaardeerde toehoorders, Dear family and friends,

Introduction

'The IPCC climate simulations are far from being predictions.' That is the quote from this inaugural lecture that appeared on the University of Amsterdam web site along with the announcement of the lecture. I had been asked to provide a provocative quote for the announcement. The idea, that the Intergovernment Panel on Climate Change bases its findings on climate *projections*, and not climate *predictions*, is not original, but one that has been expressed by one of the world's foremost climate scientists and lead writers of the IPCC reports, to whom we will return later in the lecture.

The quote above suggests a tone of skepticism, and the sensitive nature of this subject was immediately confirmed when, prompted by the quote on the UvA website, a concerned citizen sent me an angry e-mail, accusing me of having no conscience, and suggesting that the statements I was planning to make could be used by policymakers as an excuse for inaction in the face of impending climate change. (To be honest, I had, at that moment, written no more of this speech than the first line, and hence found it ironic that my correspondent seemed to know what statements I was planning to make.)

The IPCC has recently come under fire in the Netherlands, among other sources in the book *De staat van het klimaat: een koele blik in een verhit debat* written by science journalist Marcel Crok[1]. At the end of his book, Crok argues that the proximity of science and politics in the climate issue is a detriment to objective science. Every scientific statement becomes politically loaded. My correspondent was concerned that any discussion of uncertainty in science would undermine science as a whole, increasing public mistrust of science. I wholeheartedly disagree. Misleading the public into thinking that science is free of uncertainty causes the public to mistrust science when its 'predictions' fail.

This lecture in no way calls into question the IPCC case on greenhouse gas forcing of climate. The IPCC case is based on broad evidence from a variety of sources, not just simulations. The simulations have a particular role, and the IPCC clearly communicates what that role is. Instead my goal here is to attempt to explain to you how mathematicians look at prediction, and to point out where challenges lie for scientists for improving climate prediction. A number of such challenges are already being taken up in the coming IPCC report.

How predictable is nature? On the NASA Eclipse Website[9] one can see a table listing all solar eclipses that will occur until the year 3000. For example, according to the catalog, on New Years Eve 2996, at 12:58:17pm a total eclipse will occur at 33°S latitude and 6°E longitude, having a path width of 86km. I expect it will be spectacular.

It may or may not surprise you that NASA is able to predict this eclipse so accurately. In this lecture I will explain how predictions of natural systems are made, and make some comments on the limitations of predictability. I will explain the *nature of prediction* in the context of the solar system (which is relatively simple), and then I will explain the *prediction of nature* in the context of the climate.

The Nature of Prediction…

Historically, the attempt to understand the motion of planets and other heavenly bodies was one of the driving forces behind the development of formal mathematics, along with commerce, surveying and architecture. Whereas the latter three were practical necessities, astronomy was a pure science in the sense that it was curiosity-driven and highly theoretical. It led to calculus and the branch of mathematics known as analysis. Johannes Kepler published the laws of planetary motion in 1609. By peering at the meticulous astronomical data of Brahe, Kepler had discovered that the planetary orbits were elliptical, and his second law, interpreted graphically in Figure 1, states that the line between a planet and the sun sweeps out an equal amount of area in equal periods of time: this implies that the planet speeds up when nearer the sun and slows down when farther away from it.

Kepler's observation that the planetary orbits were elliptical inspired Isaac Newton to devise the theory of gravitation. Newton proposed the laws of mechanics, which have three important consequences for the prediction of planetary motion: (1) a planet moves in a straight line unless acted upon by gravity, (2) the gravitational force effects a change in the velocity vector of a planet, (3) the gravitational force between any two bodies acts along the line between them and is inversely proportional to the square of their separation.

Differential equations and numerical integrators

In Figure 2 you see three computer simulations of the giant outer planets of the solar system – Jupiter, Saturn, Uranus and Neptune – and Pluto, which used to be a planet until astronomers demoted it in 2006. The simulations were computed using three different numerical methods, A, B, and C, which I will describe in a moment. Just like the porridge in the English story of Goldilocks and the Three Bears, Method A is 'too hot', Method B is 'too cold', and Method C is 'just right'. For the hot Method A the planetary orbits gradually grow in time, and the planets leave the solar system. If one looks closely one will see that Jupiter and Saturn nearly collide at the beginning of this simulation, which also throws the orbits off considerably. For the cold Method B the planetary orbits gradually converge upon the sun. When they get too close, they are slingshot off into space. Meanwhile, for the just-right Method C, the orbits are nicely periodic, corresponding to what we might expect after centuries of observations.

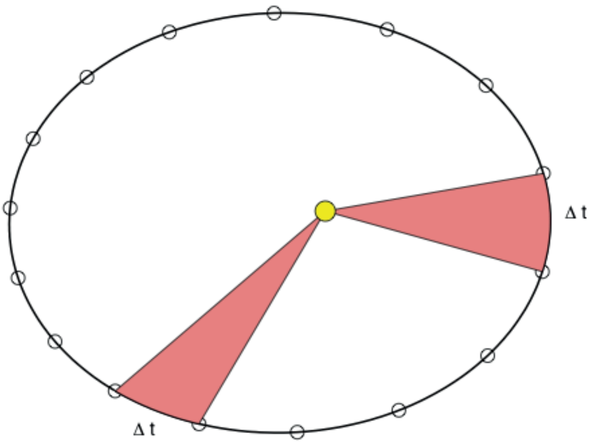

Figure 1 Kepler's second law of planetary motion – the line between the sun and an orbiting planet sweeps out equal areas in equal time.

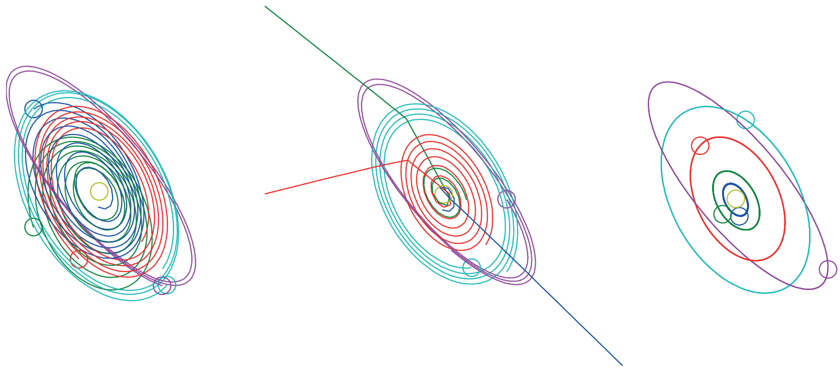

Figure 2 Simulations of the outer solar system: (left) Euler's method, (middle) 'backward' Euler, (right) Newton's method. The circles indicate the locations of the planets at the end of the simulations.

The *prediction problem* for the solar system is the following: Given the mathematical laws (in this case, Newton's equations) describing the motion of the planets, and sufficient information about their current state, determine their state at some future time T. For a single planet, Newton's equations amount to six equations – three for the position in three-dimensional space, and three to specify its velocity vector. For the solar system, including the sun, this corresponds to sixty equations. And depending on what we want to know about the planets, we may have to throw in a moon or an asteroid or two, at a rate of six equations each. The solution to such a problem would be sixty-plus functions of time, that specify the positions and velocities of all bodies in the solar system for all time. But we do not know how to solve that problem, nor does anyone believe it is possible. Fortunately mathematics tells us that we can solve the equations *approximately* if the time T is very small. Since T is generally not small, we divide up the period of time between 0 and T into a large number N of tiny time periods $0, t_1, t_2, \ldots, t_N = T$, and we denote by Δt the length of these tiny periods: $\Delta t = t_1 - t_0 = t_2 - t_1$, etc. Remember Δt, because we will mention it frequently in the discussion: Δt is the small amount of time for which we *can solve* the complicated mathematical equations, at least approximately. Now, all we have to do is solve Newton's equations on the tiny time interval Δt… One may think that this is not much easier than solving them on T, but the beauty of mathematics is that it makes a big difference if we can assume that Δt is small.

Let us examine how this may be done for a single planetary orbit – for example that of the Earth. When we speak of the *system*, in this case, we mean the Sun, which we assume to be fixed in space, and the Earth, which is moving. The goal is to determine the motion of the Earth over a small time step Δt. The nature of prediction is that we need Newton's equations, which tell how the system changes from one time to the next, *plus* a precise description of the state of the system at some initial time. This precise description is called the *initial condition*. Remember this too. Even though it seems innocuous enough, it plays an important role at the end of the lecture. For a planet, it turns out that its *state* is fully described by: its *location* in space (relative to the sun in this case) and its *velocity* vector. Recall that a velocity vector tells which direction something is moving, and how fast. For this lecture, all the mathematics that is needed is that of vector arithmetic: A vector is illustrated graphically by an arrow: the direction the arrow points represents its direction, and its length represents its magnitude (how fast, in the case of a velocity vector). An important property of a vector is that we are free to move it around in space, as long as we don't rotate it or magnify it. A second property is that to add two vectors, we just attach the tail of one to the head of the other (the order doesn't matter), and then draw a new vector from the free tail to the free head.

We denote the initial location of the planet by X_0 and its initial velocity by V_0. Together these constitute the initial condition for a planet. Stated another way, given the location and velocity of the planet at time t_0, we want to determine its new location and new velocity at time t_1, Δt units later. These states are like snapshots, or frames in a motion picture. In frame 0, the planet is located at X_0 and has velocity V_0. In frame 1, it is located at X_1 and has velocity V_1, and so forth. Our goal is to compute Frame 1 using the information in Frame 0.

There are hundreds if not thousands of methods for doing this. We will consider Methods A, B, and C above. Method A, the hot one, was first proposed by Leonhard Euler, a great Swiss mathematician. It proceeds as follows (Figure 3):

1. If there were no force acting on the planet, then it would move in a straight line in the direction of V_0 for a time Δt, and its new position would be $X_1 = X_0 + \Delta t\ V_0$. Euler just uses this value for the new position.
2. If the planet were standing still, on the other hand, the force acting on it would be constant, inversely proportional to the square of its distance from the sun, and acting along the line between the planet and the sun. Denote the force by F_0, as it is shown in the figure. The velocity is modified according to Newton's laws as follows: $V_1 = V_0 + \Delta t\ F_0$. Recall how to add vectors: match head to tail and draw an arrow. We take this as the new value of the velocity at time t_1.

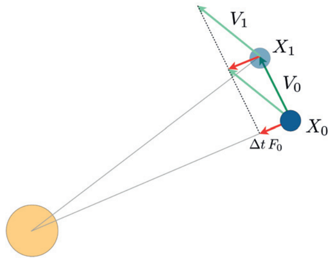

Figure 3 The method of Euler applied to the Earth orbit.

Method B is also attributed to Euler, but it is a bit more sophisticated (Figure 4). It is referred to as the 'backward' Euler method, for reasons I do not wish to go into. In this case we first *pretend* we know the position of the planet at time X_1. Knowing it, we can compute the force F_1 there, and given the force we can compute the change in velocity variable using the formula $V_1 = V_0 + \Delta t\ F_1$. In other words, the velocity vector is updated using the force at time t_1. Now, knowing the velocity we compute the position using the final velocity instead of the initial one to get $X_1 = X_0 + \Delta t\ V_1$. Of course, we didn't know X_1 to begin with, so these two equations have to be solved together. With a little luck one can proceed by guessing X_1, computing V_1, then computing a better estimate of X_1, then a better estimate of V_1 and so on, until one is satisfied that repeating this won't improve the solution any more. We say that X_1 and V_1 are defined *implicitly*, and we also refer to method B as the *implicit* Euler method. It was really popularized by people like John Butcher, a New Zealand mathematician who just last February was awarded the Van Wijngaarden prize in this very hall.

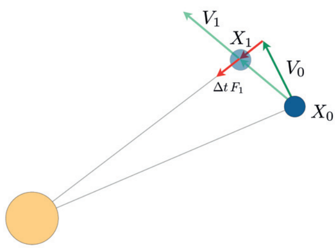

Figure 4 The 'backward' Euler method applied to the Earth orbit.

The just-right Method C was first used by Newton (Figure 5). In this case, the position is updated assuming the Sun is absent, $X_1 = X_0 + \Delta t\ V_0$, and *then* the velocity is updated assuming the planet is standing still *at its new location*: $V_1 = V_0 + \Delta t\ F_1$. Curiously, it turns out that this method satisfies Kepler's law that equal areas are swept out in equal times! And this fact is related to its just-right behavior. Newton's graphical proof of this fact is included in the *Principia*.

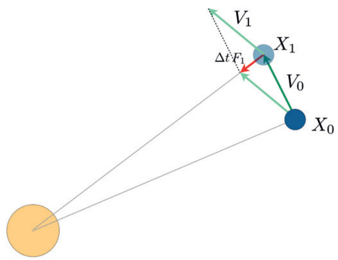

Figure 5 The method of Newton applied to the Earth orbit.

Once we know how to solve Newton's equations for a time step of size Δt, we can compute the locations of the planets at time t_1. At this point we are back to our original problem: the governing equations have not changed, and we have a new initial condition. We then solve the equations again for another time Δt to get the locations of the planets at time t_2, and so on, until we get to T. If the number of steps is very large (and we will see that it must be), this process could become rather tedious. Until the early 1950s, we paid a room full of people to do these computations; thereafter we developed the very first computers that put the very first people out of their jobs. The first computer in the Netherlands was built under the leadership of Adriaan van Wijngaarden who was one of my predecessors in the Professorial Chair of the *Stichting voor Hoger Onderwijs in de Toegepaste Wiskunde*[11]. Van Wijngaarden was the original head of the computing department at the *Mathematisch Centrum*, currently *Centrum Wiskunde & Informatica* and my employer. Later he served as director of the institute for many years.

Two limit cases: Numerical Analysis & Dynamical Systems

The astute listener may object, 'but the force and velocity are changing continuously during the time step Δt, so your answer is wrong!' This is true, we *have* made an error, and *error too is the nature of prediction*. However, we expect the error to be smaller if the time step is smaller. We can test this by computing one period of Earth's orbit, 365 days. If we take 12 time steps of size Δt equal to one month, the errors stack up and the orbit rapidly spirals away. This is successively improved using 52 time steps of one week and 365 time steps of one day. If we take time steps of size Δt equal to one hour, the orbit is nearly closed, even using Euler's method. However, the more steps we take, the more work for the computer (or the people), and the longer we have to wait for our answer. In practice Δt is chosen as a compromise between our desire for accuracy and our patience in waiting for the answer.

Mathematical analysis is often concerned with limits: what happens when some quantity becomes very large or very small? Two limits are of interest to prediction methods. The first limit we have just illustrated: it is the limit in which the length of the time interval T is kept fixed (one year), while taking Δt smaller and smaller (at the same time taking more and more steps). I refer to this as the *approximation limit*, because the prediction becomes ever more precise, closer and closer to the exact solution. The approximation limit belongs to the realm of *Numerical Analysis*, the first half of the name of the Chair of Numerical Analysis and Dynamical Systems. Numerical analysts try to show how rapidly the error decreases as we take two times as many steps, half as large. Convergence is important, but in practice computations are often done with large time steps and on time intervals much too long for the approximation limit to apply. A meteorologist-colleague once said, 'you can always recognize the mathematicians, because they explicitly state that Δt is a positive constant'.

But there is a second limit of interest to mathematicians, and that is the limit where Δt is kept fixed but the number of steps becomes large. For example, we think of repeating our calculation with Euler's method over and over to create an infinite sequence of snapshots of our solar system. We are interested in how the solution behaves in this limit, do the planets spiral away as with method A, crash into the sun as with method B, or follow nice ellipses as with Newton's method C? Questions of this nature refer to the *stability* of the method and belong more generally to the realm of mathematics called *Dynamical Systems*, the second half of the name of the Professorial Chair. For fixed Δt, the iterated numerical process defines a so-called *discrete dynamical system*. Questions pertaining to the stability of numerical prediction methods were studied in great detail by the previous two occupants of this Chair: Pieter van der Houwen, and Jan Verwer. I had the very great pleasure of working with both of them. In fact, at least two other occupants of the Chair also worked on discrete dynamical systems: Hans Lauwerier, who wrote a popular book on fractals[6], and Van Wijngaarden himself, who proposed a discrete computer calculus, which he suggested would be more appropriate for computational modeling than the continuum calculus now used[14].

… (and the Prediction of Nature)

At this point we have demonstrated the mechanical process of prediction. In most cases, however, there is a theoretical catch – and that is the question of predictability. To quote pioneering quantum physicist Niels Bohr, 'prediction is very difficult, especially if it is about the future'.

It has been postulated that our fascination for weather stems from its unpredictability: if one attempts to make use of the daily weather report, one may occasionally be disappointed. If one follows the multiple day forecasts, it is even more likely that the inaccuracy draws one's attention. There seems to be a problem with weather prediction. After our foregoing discussion, one may ask, do meteorologists who compute the weather need to use a smaller time step? Are the governing equations wrong? Or is the initial condition wrong? In fact, all of these are sources of error: the models do not account for all physical influences, the initial condition cannot be measured everywhere in the atmosphere, and undoubtedly the step size could be smaller. However, there is something else involved that causes the above effects to be grossly amplified: the governing equations exhibit 'chaos', a subject of mathematics that has been studied since the 1960s and which gained widespread popular attention in the late 1980s with the publication of several popular books, such as *Chaos: Making a New Science* by James Gleick[4].

The essential idea of chaotic behavior is that while the motion of the system remains bounded, two different solutions, no matter how close originally, grow apart at an exponential rate. This means also that errors made in computing the solution will grow exponentially. The property holds generically for most natural systems. It was studied in meteorology by the mathematician Edward Lorenz. Among the general public, the popular example of the 'butterfly effect' is familiar, whereby it is suggested that a butterfly flapping its wings in Brazil can trigger a series of growing instabilities that eventually result in a tornado in Texas. Now, while this is probably rather exaggerated, the salient idea is that small perturbations may lead to huge discrepancies. Lorenz first studied this phenomenon for the example of a system of equations describing circulating water in a heated box[7]. The warmed fluid rises, forcing the cooled fluid to descend, and a circular, overturning motion ensues. The state of the system is given by three variables – call them X, Y, and Z – where X represents the intensity of the overturning, Y represents the temperature difference between the ascending and descending fluids and Z represents the nonlinearity of the temperature profile.

Euler's method for the Lorenz system looks like this:

$$
\begin{aligned}
X_{n+1} &= X_n + \Delta t\,(sY_n - sX_n) \\
Y_{n+1} &= Y_n + \Delta t\,(rX_n - X_nZ_n - Y_n) \\
Z_{n+1} &= Z_n + \Delta t\,(X_nY_n - bZ_n)
\end{aligned}
$$

The numbers r, b and s are parameters: constant numbers chosen by Lorenz to be $r = 28$, $b = 8/3$, and $s = 10$. The variables X, Y and Z change in time. We can think of them as the coordinates of a point in three-dimensional space. In that case, our prediction for X, Y and Z is a sequence of such points, tracing out a curve, just like one of the planets but with a much more complex orbit. If we just examine how the variable Z varies in time, it seems unpredictable. Let us compare ten solutions of Lorenz's equations, each with a tiny error in the ini-

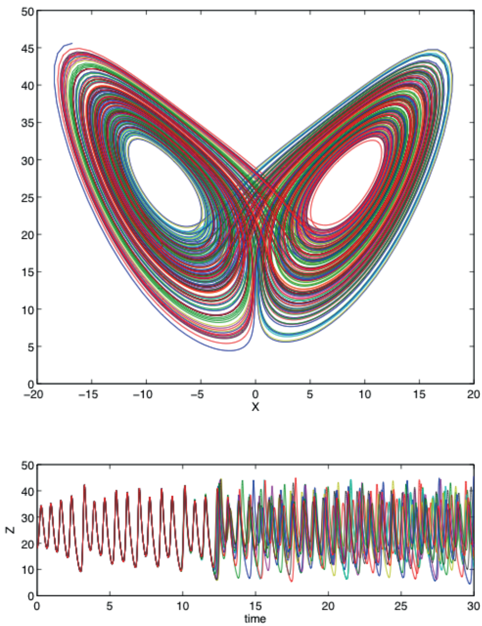

Figure 6 Chaotic divergence of trajectories of the Lorenz equations. Ten trajectories with small randomly perturbed initial condition: (top) trajectories in phase space (X-Z), (bottom) time series for the variable Z, showing predictability up to around time t = 12.

tial condition, say, less than one pro mille. In Figure 6 (bottom) we initially observe no difference in the ten solutions; they look like a single solution. Suddenly, after a certain time has passed, they all diverge completely. The predictability is lost. This divergence occurs at an *exponential* rate, just like the growth of bacteria populations, bank savings, or radioactive decay. We can speak of the half-life of a prediction. How long the half-life is, actually depends on the current conditions – a large high-pressure weather pattern has a much longer half-life than a low-pressure pattern. The implication of chaos is that there is a limit or horizon to prediction – errors are always present and may grow at an exponential rate.

Earlier we looked at the solar system; it is easy to think of the solar system as being periodic. The orbits of the planets seem to be stationary. One might think we can predict the state of the solar system forever. After all, eclipses can be predicted down to the second for thousands of years. In fact, the motion of a *single* planet around the Sun *would* be highly predictable. However, even the planetary system is chaotic, as soon as there are three bodies involved. This was noted by Henri Poincaré in 1890[10]. In a more recent paper appearing in the journal *Nature*, simulations of the solar system on a time interval of five billion years were carried out[5]. Small errors in the solar system grow by a factor of ten every ten million years. This means that the horizon for solar system simulations is around 200 million years. As a result, these five billion year simulations were not *predictions* in the sense we have been talking about. On such long time scales, the orbits of the planets look anything but periodic, the orbital ellipses rock back and forth, the orbital radii grow and decay, sometimes the order of the planets as we know them changes. For example, the orbit of Venus becomes larger than that of Earth. This will, among other things, be devastating for mnemonics (in Dutch, *ezelsbruggetjes*) for remembering the order of the planets, like 'My Very Educated Mother Just Served Us Nine Pizzas' (already obsolete since the demotion of Pluto). Children five billion years from now will have to think of new ones.

As another example of chaos, let us look at the climate simulations, such as those shown in Figure 7, which were carried out by the Dutch weather service KNMI in 2005 as part of the Dutch Challenge Project[3]. In this study, the global climate was simulated over 140 years from 1940–2080. In total, 31 simulations are shown, each with a miniscule disturbance of the temperature in the initial condition – less than one pro mille. Shown here are the results from the first month, January 1940, indicating the predicted temperatures in de Bilt. There is a 10°C temperature spread by the end of the month!

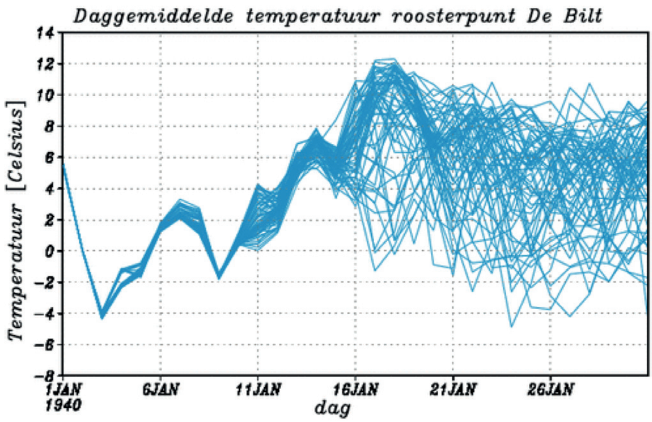

Figure 7 Ensemble climate simulation for the Dutch Challenge Project. (courtesy of KNMI)

Seeing this, one may wonder why anyone even bothers doing climate simulations in the presence of chaos. If we cannot trust the weather forecast two weeks ahead of time, what hope is there of predicting the whole climate 140 years in advance? The answer is, of course, that climate scientists are not interested in predicting the *weather*. That is, they are not interested in precisely predicting the temperature in Amsterdam on a Thursday in 2080, but in other quantities, such as for example the mean yearly temperature in Amsterdam in the period from 2070–2080, or the relative increase or decrease in rainfall for the summer months in the Netherlands between 2010 and 2080. Our premise is that such quantities *are* predictable, even if the precise state of the atmosphere on a given date cannot be specified.

Let us see how that might be. We return to the Lorenz system, and instead of showing the solution Z, let us just keep track of which values of Z are most likely to occur. We divide the interval from 0 to 60 up into 100 equal boxes, and with each time step of our Euler method we determine in which box the solution finds itself and count how many time steps fall within each of the 100 boxes. In this way we obtain a *statistical distribution* over all the values of Z, such as the one shown in Figure 8. The key point is that even though two solutions of Lorenz diverge exponentially, on a long interval and from a distance they all look more or less the same. In particular, *for any initial condition this same statistical distribution will result*. And now suppose the thing we want to know about the 'climate' of the Lorenz system depends only on the distribution of Z. For example, suppose we want to know the mean value and standard deviation of Z. Then even though the system is chaotic, this quantity is predictable. In this case, and for the Lorenz problem, it doesn't even depend on the initial condition!

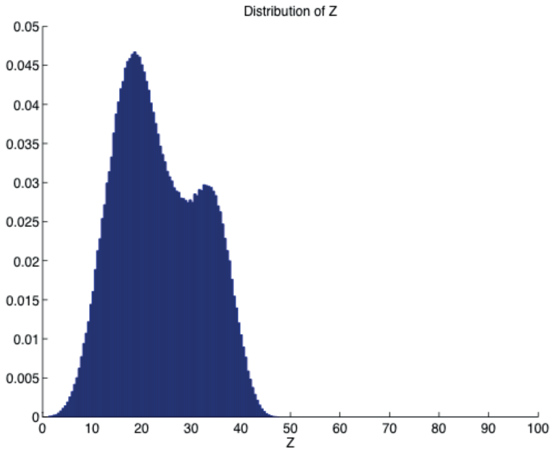

Figure 8 Despite the fact that the trajectories of the Lorenz equations (or the time series of Z) are unpredictable, the statistical distribution of Z over a long time simulation is independent of the initial condition.

Climate prediction of the First and Second Kinds

So we see that there exist quantities that can and those that cannot be predicted, even for the simple Lorenz system. In between are quantities that can be predicted somewhat accurately for longer times than the weather. What about something as complex as the climate? In other work, Lorenz proposes two concepts of predictability in climate, which he refers to as climate prediction of the first and second kinds[8].

Climate prediction of the first kind is similar to what we have already seen for the planets, except that one must determine which quantities are predictable on the time frame of interest, and then start from an initial condition that is consistent with the current climate. Probably multiple scenarios must be run, because the 'current climate' may correspond to many very different initial conditions.

Prediction of the second kind can be understood using the Lorenz example again. Suppose that instead of $r = 28$ we double this parameter and take $r = 56$ (for example, let us pretend that r represents the amount of CO_2 in the atmosphere – it doesn't, but just pretend), and we are interested in how a doubling of CO_2 will change the climate. Then we can repeat the simulation, using r = 56 this time, and compare the distributions of the variable Z. As shown in Figure 9, the distribution is changed. In this way we can study how the climate adapts to a change in some parameter such as CO_2 level. This is Lorenz's *prediction of the second kind*. With this approach we can predict how the statistics of climate – defined as the typical weather patterns – will change due to a change in parameters.

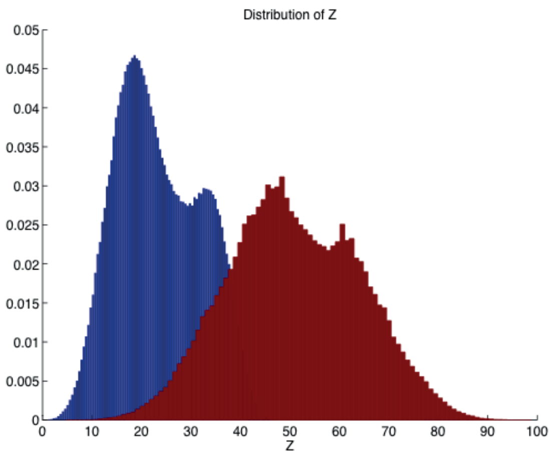

Figure 9 Comparison of the statistical distributions of the variable Z in the Lorenz equations, for r = 28 (blue) and r = 56 (red). This illustrates climate prediction of the second kind, specifically, how the 'climate' of the Lorenz attractor changes as the parameter r is doubled.

Climate sampling vs climate prediction

The Intergovernment Panel on Climate Change (the IPCC) has included in its reports climate simulations analogous to those just described in the simple situation of second kind prediction of Lorenz's equation. Roughly speaking, the IPCC fixes a value of CO_2 consistent with currently observed values and performs a long simulation, just as we did with the Lorenz model, to reach statistics that are stationary (i.e. unchanging with longer simulation). Subsequently they double the value of CO_2, and repeat the simulation to convergence of the statistics. Then these statistics are compared to make a prediction about the effects of CO_2 emissions. The differences between these statistical data are then used to extrapolate the present climate to a future one where CO_2 levels are doubled.

Kevin Trenberth is head of the climate analysis section of the National Center for Atmospheric Research in the USA and a lead author of the IPCC reports in 1995, 2001 and 2007. Trenberth submitted a letter to the weblog of the scientific journal *Nature* in 2007, which is very interesting in our context[12].

Just to make Trenberth's opinion on CO_2 emissions clear, I start with the conclusion of his letter. He writes,

> A consensus has emerged that 'warming of the climate system is unequivocal' and the science is convincing that humans are the cause. Hence mitigation of the problem: stopping or slowing greenhouse gas emissions into the atmosphere is essential. The science is clear in this respect.

and further,

> We will adapt to climate change. The question is whether it will be planned or not? How disruptive and how much loss of life will there be because we did not adequately plan for the climate changes that are already occurring?

Nonetheless, Trenberth's letter states

> In fact there are no predictions by IPCC at all. And there never have been. The IPCC instead proffers 'what if' projections of future climate that correspond to certain emissions scenarios… They are intended to cover a range of possible self consistent 'story lines' that then provide decision makers with information about which paths might be more desirable.
>
> Even if there were, the projections are based on model results that provide differences of the future climate relative to that today. None of the models used by IPCC are initialized to the observed state and none of the climate states in the models correspond even remotely to the current observed climate. In particular, the state of the oceans, sea ice, and soil moisture has no relationship to the observed state at any recent time in any of the IPCC models. There is neither an El Niño sequence nor any Pacific Decadal Oscillation that replicates the recent past; yet these are critical modes of variability that affect Pacific rim countries and beyond… I postulate that regional climate change is impossible to deal with properly unless the models are initialized.

> The current projection method works to the extent it does because it utilizes differences from one time to another and the main model bias and systematic errors are thereby subtracted out. This assumes linearity …

Hence, climate simulations, as employed by the IPCC, should not be confused with climate predictions – certainly not those of the first kind as defined by Lorenz. But in fact, there is also an important tacit assumption that goes into the *second kind* climate prediction, which almost certainly does not hold for the real climate: that is, that the results of a long simulation do not depend on the initial condition chosen. Let us demonstrate this, again using the Lorenz system.

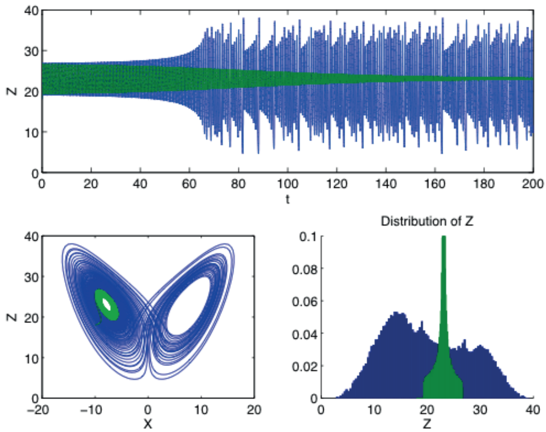

Figure 10 The statistics of trajectories of the Lorenz equations with r = 24.1 depends on the initial condition: (top) time series for Z, showing chaotic and asymptotically stable solutions; (under, left) the same trajectories in phase space; (under, right) probability distributions of Z.

We have seen the statistical distribution for the variable Z for the original choice $r = 28$ of Lorenz. This distribution is absolutely independent of the initial condition – mathematically speaking the system has a *global attractor*. Now let us choose $r = 24.1$, about 15% smaller. In this case, the character of the solution changes considerably. Two solutions are shown in Figure 10. For most initial conditions the chaotic behavior persists, yet for another class of initial conditions (such as the green one shown in the figure), the behavior is highly predictable. For this value of r, the statistical distributions also depend on the initial condition!

The IPCC takes care not to refer to its climate simulations as predictions. They speak of projections or scenarios, 'consistent and plausible' realizations of future climates. Nonetheless the simulation results are frequently misused by others to justify decisions in the face of the current warming; for example, to determine the need for higher dikes in the Netherlands. Trenberth warns that the IPCC simulations should not be used to predict regional change. In other words he says that whereas the simulations may give a plausible indication of the degree of *global* annual mean temperature increase, whether, say, the Netherlands *regionally* will be warmer or cooler, wetter or dryer, can not be deduced from the IPCC simulations. Trenberth continues:

> However, the science is not done because we do not have reliable or regional predictions of climate. But we need them. Indeed it is an imperative! So the science is just beginning. Beginning, that is, to face up to the challenge of building a climate information system that tracks the current climate and the agents of change, that initializes models and makes predictions, and that provides useful climate information on many time scales regionally and tailored to many sectoral needs.
> Of course one can initialize a climate model, but a biased model will immediately drift back to the model climate and the predicted trends will then be wrong. Therefore the problem of overcoming this shortcoming, and facing up to initializing climate models means not only obtaining sufficient reliable observations of all aspects of the climate system, but also overcoming model biases.

As he notes here, one challenge is to overcome model biases. This brings us back to the examples of the Goldilocks Methods A, B and C for the solar system at the beginning of the lecture, and the relation of all this with the research of my group at CWI. There we saw that different methods behaved differently in the dynamical systems limit of fixed Δt and T tending to infinity. Overcoming model bias means that the methods have to be designed such that their statistics agree with those of the real solar system.

In the same way that the just-right Method C retained the equal areas property of Kepler, for an atmospheric model one can construct numerical methods that respect other natural laws like energy or something exotic like *enstrophy* – the mean variance of the rotational component of the wind. With PhD student Svetlana Dubinkina we compared three methods that were identical besides conserving energy, enstrophy or both[2]. Using these we computed the average wind field from long simulations, in other words, the 'prevailing winds'. We found that the methods gave completely different results, as illustrated in Figure 11. The method that conserved energy predicted *no* prevailing wind at all – all fluctuations were equally likely. The method that obeyed only the enstrophy conservation law predicted a weaker prevailing wind with no mean rotational component, and the method that conserved both gave stronger winds that were presumably more realistic.

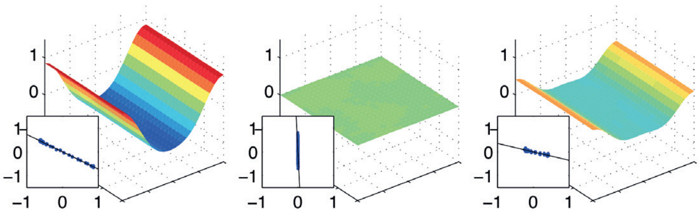

Figure 11 Mean stream functions and corresponding scatter plots of potential vorticity (inset) for long time simulations with discretizations due to Arakawa that conserve (left) energy and enstrophy, (middle) energy only, and (right) enstrophy only. The 'prevailing winds', which blow along the contours of the stream function surfaces, strongly depend on the method used.

Closing

From the first part of the lecture, there are three key elements to the nature of prediction: (1) given a mathematical rule that tells how the state of a system changes from one time to the next, and an initial condition describing the original state, we attempt to compute the state at a later time T; (2) the prediction is an approximation, by definition it is in error; (3) chaotic growth of error effectively places a horizon on predictability.

Nonetheless, certain statistical quantities that are insensitive to this error growth are predictable on long times, but only using numerical methods that accurately reproduce the climate statistical distribution. To conclude, I would like to outline where it appears to me, based on the discussion presented here, that progress can be made in climate prediction.

For effective second kind climate prediction, two ingredients are necessary:

1. For the reference simulation, the parameters must be consistent with the current climate, and the model able to reproduce the current cli-

mate, at least for some consistent class of initial conditions. This removes the assumption of linearity. In the words of Trenberth, the models must be *initialized*.
2. We must establish that the climate attractor is a global one such that initial conditions are irrelevant, or else explore and categorize the basins of attraction. Otherwise, the projected (future) climate cannot be initialized.

Alternatively, to predict climate in the first kind sense, which seems to me vastly preferable, a whole program of research must be carried out, including development of measures of accuracy of statistical quantities such as averages and time correlations, an analysis of which such quantities may be computed accurately on what time frames, an understanding of how that accuracy depends on the numerical discretization parameters, and the development of new computational techniques for statistically consistent parameterization of unresolved effects or other means of correction of statistical bias introduced by the numerics. There is much work to be done.

Ladies and gentlemen, the nature of prediction is uncertainty, but the prediction of nature may well succumb to the efforts of science. Lao Tzu's proverb 'Those who have knowledge, don't predict. Those who predict, don't have knowledge', holds only as a truism.

Acknowledgements

In closing, I would like to thank the *College van Bestuur* of the *Universiteit van Amsterdam* for approval of the Chair and its Occupant, similarly, the Dean of the *Faculteit Natuurwetenschappen, Wiskunde en Informatica*, and the board of the *Stichting voor Hoger Onderwijs in de Toegepaste Wiskunde*.

I would also like to thank the members of the curatorium, Hans van Duijn, Peter Sloot and Rob Stevenson for their input and their constructive remarks regarding the practical implementation of the Chair, and the director of the Korteweg-de Vries Institute, Jan Wiegerinck, for all of his efforts to help realize this appointment.

A number of people have served as advisors to me over the years. I would like to express my gratitude to the supervisors of my PhD research Piet Wesseling, Piet van der Houwen and Kees Vuik; and to Willem Hundsdorfer, who played a crucial role in my introduction to the mathematics community in the Netherlands. I had intended to stand here and give extensive acknowledgements to my predecessor in this chair, Jan Verwer, who took the initiative to

create a position for me at CWI, who supported my appointment as leader of the CWI research group MAC1, and who undoubtedly lobbied for my appointment in this Chair. I am saddened that Jan could not be here today.

In addition, I would like to thank two colleagues from abroad: Sebastian Reich of Potsdam University, with whom I have had a successful collaboration for many years, and Ben Leimkuhler of the University of Edinburgh, who introduced me to applied mathematics, and whose inexhaustible creativity is always an inspiration.

It is a pleasure to work with Daan Crommelin – together we are building a new MAC1. This cannot be done without the hard work of our PhD students and fellows.

Finally, I want to say some words to my family. My parents Jim and Sharon Frank were inspirational mathematics teachers from my earliest memories. They instilled both a respect for cleverness and learning, and a Christian perspective on life, for all of which I am very grateful. To Linda and Helen and Rosa: thank you for laughter and tears. You are everything to me.

Ik heb gezegd.

References

1. Crok, M., *De Staat van het Klimaat: een Koele Blik om een Verhit Debat*, FMB Uitgevers, Amsterdam, 2010
2. Dubinkina, S. & Frank, J., 'Statistical Mechanics of Arakawa's Discretizations', *Journal of Computational Physics*, 227, p. 1286–1305, 2007
3. Dutch Challenge Project, Koninklijk Nederlands Meteorologisch Instituut, http://www.knmi.nl/onderzk/CKO/Challenge_live/index.html, 2003
4. Gleick, J., *Chaos: Making a New Science*, Viking Penguin, 1987
5. Laskar J. & Gastineau, M., 'Existence of collisional trajectories of Mercury, Mars and Venus with the Earth', *Nature*, 459, 08096, 2009
6. Lauwerier, H., *Fractals: Endlessly Repeating Geometrical Figures*, Princeton University Press, 1991
7. Lorenz, E.N., 'Deterministic Nonperiodic Flow', *Journal of the Atmospheric Sciences* 20, p. 130–141, 1963
8. Lorenz, E.N., 'Climatic Predictability', *GARP Publication Series*, p. 132–136, 1975
9. NASA Eclipse Website, http://eclipse.gsfc.nasa.gov/solar.html, 2011
10. Poincaré, J.H., 'Sur le problème des trois corps et les équations de la dynamique. Divergence des séries de M. Lindstedt', *Acta Mathematica*, 13, p. 1–270, 1890
11. *Stichting voor Hoger Onderwijs in de Toegepaste Wiskunde*, http://staff.science.uva.nl/~thk/SHOTW/, 2011
12. Trenberth, K., 'Predictions of climate', Climate Feedback blog of *Nature*, http://blogs.nature.com/climatefeedback/2007/06/predictions_of_climate.html, 4 June, 2007
13. Universiteit van Amsterdam, *Nieuws en Agenda* website, http://www.uva.nl/actueel/promoties-oraties/oraties.cfm/80A6F298-546C-4A34-97BE45B6B0ED22C7, 21 April, 2011
14. Van Wijngaarden, A. 'Numerical Analysis as an Independent Science', *BIT Numerical Mathematics*, 6, p. 66–81, 1966