

Mathematics in Atmospheric Sciences: An Overview

Pierre Gauthier

Department of Earth and Atmospheric Sciences
Université du Québec à Montréal
pierre.gauthier@uqam.ca

Abstract. Several sectors of human activities rely on weather forecasts to plan in preparation of high impact weather events like snow storms, hurricanes or heat waves. Climate studies are needed to make decisions about the long-term development in agriculture, transport and land development. Observing and modeling the evolution of the atmosphere is needed to provide key *reliable* information for both weather prediction and climate scenarios. This paper gives an overview of the scientific research underlying the development and validation of numerical models of the atmosphere and the monitoring of the quality of the observations collected from several types of instruments. A particular emphasis will be given to data assimilation which establishes the bridge between numerical models and observations. The mathematical problems arising in atmospheric research are diverse as the problem is one of stochastic prediction for which errors in both the model and the observations need to be considered and estimated. Atmospheric predictability is concerned with the *chaotic* nature of the nonlinear equations that govern the atmosphere. Ensemble prediction is one area that has expanded significantly in the last decade. The interest stems from the necessity to evaluate more than just a forecast: it aims at giving an estimate of its accuracy as well. This brings up more questions than answers.

1 Introduction

Modeling and observing the atmosphere provide the basic information through which we improve our understanding of atmospheric processes and their interactions. At heart, the atmosphere is a fluid governed by the Navier-Stokes equations, thermodynamic relationships, and numerous other processes that influence its dynamics. Modeling the atmosphere is necessary for numerical weather prediction (NWP) and climate modeling. Weather prediction is concerned with an initial value problem : knowing the current state of the atmosphere, the primitive equations (e.g., Navier-Stokes, thermodynamics) are numerically integrated in time to obtain forecasts up to 10 days ahead. In climate studies, similar integrations are carried out but over longer periods of time from decades, to centuries and even millenia (paleoclimatology) and one is interested in the evolution of the average of the atmospheric state, filtering out transient processes. One has

to be concerned with both the first moment (the mean) and the second moments (variability) of the time series of the evolving climate.

Both problems (NWP and climate) are faced with similar challenges. The primitive equations must be discretized which implies that approximations need to be introduced to include the effect of unresolved processes. Numerical schemes are only approximating the “exact” solution and may not conserve some basic physical quantities (e.g., the total mass of the atmosphere). Discretization of the atmosphere is done horizontally over the globe or limited areas. The spatial resolution of the model must be sufficient to resolve the key physical processes that have an impact on the evolution of the atmosphere. Even for weather forecasts for a few days ahead, global models are needed as small changes in the weather in distant regions may have an influence on what will happen here a few days later. This is often referred to as the *butterfly effect*. Local area models can only provide forecasts for shorter periods of time but can be run at higher spatial resolutions, which permit to resolve convection processes that are critical for accurate predictions of precipitations.

In both cases, computer efficiency of the numerical algorithms is also important and numerical models used for NWP is one of the main applications of high performance computing. NWP centres have some of the most powerful computers to be able to produce accurate weather forecasts within a limited time. More efficient algorithms and increasing the computing power enable to increase the complexity and accuracy of the models. In this presentation, an overview is presented of the current state of numerical modeling and some of the mathematical difficulties that are raised.

Numerical models include our knowledge about the physical laws that govern the motion of the atmosphere. They include also the limitations of our understanding and the results obtained with those models need to be validated by comparing their forecasts against observations. The global observing systems includes several types of instruments that provide measurements that are often only indirectly related to the atmospheric variables. Observations from satellite based instruments are measuring the electro-magnetic spectrum in the microwave, infrared and visible. This raises new problems associated with the propagation of electro-magnetic waves in highly heterogeneous media, particularly when clouds and aerosols are present. To make good use of the data collected from such instruments, it is important to be able to relate atmospheric variables to observations. Finally, acknowledging that both the numerical models and the observations have errors, the best estimate of the atmosphere is obtained by combining observations and models using the Bayesian framework of statistical estimation. The most important progress made over the last twenty years were in the development of data assimilation methods that are better suited for the assimilation of satellite data that are irregularly distributed in space and time.

This presentation aims at giving a sense of the mathematical problems that are raised when trying to build better numerical simulations of the evolving atmosphere.

2 Numerical Weather Prediction and Climate Modeling

The atmosphere being a fluid, it is governed by the Navier-Stokes equations. When expressed in a rotating referential, they can be written as

$$\frac{d\mathbf{v}}{dt} + 2\boldsymbol{\Omega} \times \mathbf{v} = -\frac{1}{\rho}\nabla p + \mathbf{D}_{dissipation} + \mathbf{F}_{External} \quad (1)$$

where \mathbf{v} is the wind vector, $\boldsymbol{\Omega} = \frac{2\pi}{T_e} \sin\varphi \mathbf{e}_r$ is the angular frequency associated with the rotation of the Earth, $T_e = 24$ hrs, φ is the latitude and \mathbf{e}_r is a unit vector oriented along the radius of the sphere. The other elements appearing here are the pressure, p , the air density, ρ , while \mathbf{D} and \mathbf{F} stand respectively for the dissipation and external forces acting on the air parcel. Pressure is related to density and temperature by the gas law ($p = R\rho T$) and thermodynamic exchanges imply that

$$\frac{d\ln\theta}{dt} = \frac{1}{c_p} \frac{dQ}{dt} \quad (2)$$

where $\theta = T(\frac{1000}{c_p})^{R/c_p}$ is the potential temperature and $\frac{dQ}{dt}$ represents the heating rate associated with external sources (e.g., radiative heating, release of latent heat due to condensation). Finally, mass conservation yields another important constraint which is

$$\frac{\partial\rho}{\partial t} + \nabla \bullet \rho\mathbf{v} = 0. \quad (3)$$

In a nutshell, numerical modeling of the atmosphere consists in solving this system of equations by using numerical schemes. One therefore must discretize in space and time and devise numerical schemes that will insure numerical accuracy and stability. Moreover, it is also important to impose some dynamical constraints that will insure that the numerical solutions preserve the known invariants of the atmosphere as much as possible. The spatial discretization imposes a limit of what can be resolved by a numerical model. Operational global models of the atmosphere now operates at horizontal resolutions as low as 20 km while in the vertical the domain extends all the way to 60 km to include the stratosphere. All in all, this means that the grid of those models comprises $2000 \times 1000 \times 80 \sim 10^8$ points and one such grid is needed for each of the 4 basic dynamical fields (horizontal wind components, temperature and humidity). Typical synoptic weather systems extend over a 1000 km and are well resolved at those resolutions but there are numerous atmospheric processes (e.g., convection) that require more than that.

The dynamics of the atmosphere leads to several types of waves over a wide range of frequencies. Waves can be classified with respect to their restoring forces. At the larger scale, one finds the planetary waves with very long wavelengths (5000-10000 km) governed primarily by the temperature gradient between the equator and the poles, and the variation of the Coriolis force. These waves evolve slowly (order of days) and govern the motion of weather systems in the extra-Tropics. Gravity waves are associated with convergence (divergence) in the

horizontal which creates vertical motion: those are external gravity waves. On the other hand, the atmosphere can be destabilized in the vertical due to heating at the surface: this is convection which is found within active cloud formations. The time scale associated with convection is of the order of hours. The Tropical regions are in a different regime as the Coriolis force is negligible there. Convection is very active and forced through different physical processes such as the heat exchange between the ocean and the atmosphere. Heat from the ocean is in a sense the fuel which permits to Tropical cyclones to intensify and become hurricanes (or Typhoons). Finally, the atmosphere being a gas, sound waves are also present and those are associated with very fast time scales.

Referring to [Durran, 1998], numerical schemes can become unstable if the time step is longer than the frequency of the waves. On the other hand, others are not unstable but will have the effect of damping any wave having a shorter wavelength. Different approaches can be taken: the problem can be cast in an Eulerian framework, a Lagrangian one or somewhere in between as in the semi-Lagrangian scheme [Robert, 1982]. Each has its own merits and weaknesses and must be looked at in the perspective of all the other elements that need to be included in the atmospheric models. The external forcing term of (1) includes many processes that are key to the way the atmosphere evolves. It includes the solar heating, the exchanges between the land and the atmosphere, and many other components. Finally, as stated earlier, an important aspect of this problem is its size. Algorithms must be able to take advantage of the computer architectures at hand. Nowadays, models are implemented on parallel architectures for which it is equally important to take into account the performance in terms of number of operations to do and the level of communication between the thousands of processors involved. The size of the problem demands such computing power and atmospheric numerical modeling is one of the important applications considered in high performance computing.

In numerical weather prediction, a model is run for a period up to ten days and the initial conditions play a key role: the weather that will occur over any specific region depends on the current conditions not only over that region but over the globe. Accurate forecasts in the 3 to 5-day range often critically depend of what is going in very distant regions on the globe. For example, weather forecasts over Canada depend of what is going on in the North Pacific, over Asia and in the Tropics near the South East of Asia.

Climate simulations are concerned with longer range predictions and only consider averages over long period of time. The objective is to describe the equilibrium state of the atmosphere due to the external forces exerted upon it. Climate models pay a lot of importance in having a model that, under certain conditions, preserves certain invariants of the basic equations. For instance, removing dissipation and external forces, total mass and total energy should be preserve. It is then important to test if the numerical schemes employed preserve such quantities. This is to make sure that if a climate simulation is indicating a warming of the atmosphere, that this is not the result of the numerical scheme.

A substantial part of the effort in model development is to try to validate the atmospheric model against known properties or directly against observations if available.

3 Observations and Model Validation

Numerical models are approximate and need to be validated by comparing forecasts against observations. The global observing system now includes a variety of observing systems from instruments on the ground, in ships and aircrafts. At the moment, most observations are collected from passive satellite instruments measuring the electro-magnetic spectrum in the microwave, infrared and visible bands. Active instruments like Lidars and radars are now onboard satellites and provide detailed information about the vertical structure of the atmosphere. This information is nevertheless only indirectly linked to the temperature, winds, humidity and other components of the atmosphere (e.g., chemical species, aerosols). A complex radiative transfer problem must be solved to be able to compare the observed radiances against those associated with the atmospheric state of a numerical model. Moreover, however abundant, observations do not provide neither the spatial or temporal coverage to uniquely characterize complete atmospheric fields, even when discretized. Some satellite instruments can cover the whole globe but it takes a number of hours. Moreover one must retrieve temperature out of radiances which only measure the integrated effect of temperature on the radiance emitted by the atmosphere. Methods based on observations alone compensate for the lack of spatial coverage by considering all observations available over a period of time (e.g., a month) to obtain an estimate of the mean temperature over that period. The *inversion problem* consists in reconstructing those atmospheric fields out of observations. Most inversion methods require a *background estimate* and one is trying to find how it can be changed to fit the observations. The background state (or *a priori* state) can be a reference standard atmospheric profile or a climatological mean. As will be discussed in the next section, a short-term forecast from a numerical weather prediction model is also a good *first-guess*.

Resolution is often lacking in observations to get quantitative measurements of atmospheric structures associated with particular physical processes. Convection in the Tropics is regulated by the diurnal cycle and numerical models still have difficulties to capture this. Measurements of precipitation are useful to validate the representation of convection. The Tropical Rainfall Measurement Mission (TRMM) had the objective of observing precipitations in Tropical regions to better understand physical processes associated with Tropical cyclones. Instruments provided observations about the distribution of precipitations and the vertical structure of the distribution of water within clouds associated with such cyclones. However, the satellite being on a low orbit (~ 300 km) only flew over a given region at irregular intervals. Nevertheless, an average of the precipitation budget can be obtained by considering data over the whole hurricane season. This has been found to be very useful to understand the role of hurricanes on climate for instance.

Comparing observations with a model prediction must take into account the error that can occur in observing the atmosphere and modeling it. Knowledge of a prediction and an observation must be complemented by knowledge of their associated errors. Conversely, comparing a model against observations provides information about its accuracy. Moreover, observations from satellite instruments are not independent and error from measurements in different wavelengths but from the same instrument can be correlated. Analysis of the error can be used to estimate how much independent information we get from a large volume of data. Finally, it is important to say that the link between the atmosphere and the observed radiation from space is complex. Electromagnetic waves are traveling through a highly heterogeneous atmosphere which includes clouds. Detailed modeling of the diffusion, transmission and reflection of radiation is complex and is addressed by Monte-Carlo methods in which thousands of photons are emitted and their paths calculated. Some information about the average behavior can be obtained [Barker *et al.*, 2003].

Finally, geostationary satellites provide images of nearly a full hemisphere at frequent time intervals. Information about atmospheric winds can be retrieved by an analysis of these images. Clouds are detected in the visible and infrared and their motion can be related to the underlying atmospheric winds. The techniques employed involve pattern recognition to detect identifiable structures from one moment to the next. One of the difficulties however is to assign correctly the height of the clouds observed on the satellite image.

Although observations do provide a large volume of data about the atmosphere, they come from different sources and it is not obvious to piece those together to get a comprehensive representation of the atmosphere and its evolution in time.

4 Data Assimilation and Analyses

Considering that both the observations and the forecasts from a numerical model have errors, the current state of the atmosphere can be obtained by blending the two considering their respective accuracies here represented by the probability density functions (pdf) of observation and background error. Given that a set of observations $\mathbf{y} \in \mathbb{R}^m$ has been observed and an *a priori* knowledge $\mathbf{x}_b \in \mathbb{R}^n$, both having error characteristics defined by their respective pdf. In a Bayesian sense, the problem can be stated as

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})P(\mathbf{x})}{P(\mathbf{y})} \quad (4)$$

with $P(\mathbf{x})$ being the *a priori* pdf of having \mathbf{x} as the true state, \mathbf{x}_t , $p(\mathbf{y}|\mathbf{x})$, the conditional probability of \mathbf{y} being the true value *given* that $\mathbf{x} = \mathbf{x}_t$, and $p(\mathbf{x}|\mathbf{y})$ is the probability of \mathbf{x} being the true value given that \mathbf{y} has been observed. Finally, $P(\mathbf{y})$ is the marginal probability of observing \mathbf{y} .

The mode of $p(\mathbf{x}|\mathbf{y})$ is the most likely state given the information we have and it can be obtained by minimizing the objective function

$$J(\mathbf{x}) = -\ln(p(\mathbf{x}|\mathbf{y})) \quad (5)$$

$$= -\ln(p(\mathbf{y}|\mathbf{x})) - \ln(P(\mathbf{x})) + C. \quad (6)$$

The complexity of the problem stems from its size as the number of observations m is of $O(10^6)$ while the dimension of the model space is $O(10^7)$. With such large dimensions, our knowledge of the probability distributions for the observation and background error can only be partial. Assuming these error distributions to be Gaussian, then

$$\begin{aligned} P(\mathbf{x}) &\propto \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b)\right) \\ p(\mathbf{y}|\mathbf{x}) &\propto \exp\left(-\frac{1}{2}(\mathbf{H}(\mathbf{x}) - \mathbf{y})^T \mathbf{R}^{-1}(\mathbf{H}(\mathbf{x}) - \mathbf{y})\right) \end{aligned}$$

in which case

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + \frac{1}{2}(\mathbf{H}(\mathbf{x}) - \mathbf{y})^T \mathbf{R}^{-1}(\mathbf{H}(\mathbf{x}) - \mathbf{y}). \quad (7)$$

This is the variational form of the data assimilation problem. When formulated at a given time, this corresponds to the 3D case (3D-Var) while if $\mathbf{x} = \mathbf{x}(t)$ is time dependent, this is extended to the 4D case (4D-Var). Over the last 10 years or so, 4D-Var has been implemented in several operational numerical weather prediction centres including Europe (ECMWF, [Rabier *et al.*, 2000]), Météo-France [Gauthier and Thépaut, 2001], the United Kingdom [Rawlins *et al.*, 2007] and Canada [Gauthier *et al.*, 2007]. It is used to assimilate more than a million data daily to produce meteorological analyses that define the initial conditions used to produce weather forecasts up to 10 days.

As they are our best estimate of the current state of the atmosphere, analyses are also useful to validate numerical weather forecasts by running historical cases. If available over long periods of time, they have proven to be extremely useful to validate climate simulations. However, observations of the atmosphere become scarce as we go back in time. The modern era includes large volume of data that became available only since the end of the 20th century (~ 1980). Reanalyses are obtained by assimilating past observations using the most recent data assimilation systems that benefited from the latest advances made in data assimilation methods and numerical modeling. As climate studies are concerned with periods of decades to centuries and longer, reanalyses over the complete 20th century would be needed to validate climate models. Because of the scarcity of observations, the extension of reanalyses to the beginning of the 20th century presents a challenge. From the experience gained in using 4D-Var, it is known that because it embeds the dynamics of the atmosphere within the assimilation process, it has the ability to “propagate” information in a dynamically consistent way to regions that are poorly observed. In a recent study, [Uppala *et al.*, 2008] showed that with 4D-Var, it is possible to reconstruct the complete state of the atmosphere using only surface observations. Their experiments was to redo a portion of the reanalysis using only available surface observations. The analysis based on surface observations compared remarkably well to the reference analysis that used all observations.

A data assimilation system requires that the error statistics for both the background and the observations be correct. The background error statistics are the results of the error growth in a short term weather forecast associated with the initial uncertainty in the initial conditions. Dynamical instabilities depend on the prevailing conditions and it is those that define the background error statistics. These uncertainties can be captured by realizing Monte-Carlo simulations which produce an ensemble of forecasts out of which partial information is obtained about the probability distribution of the background error. Due to the limited size of the ensemble and the large dimension of the phase space, this gives rise to a rank deficiency problem associated with the estimation of the background-error covariances. Observations on the other hand can be contaminated by systematic error (biases) associated with the instrument or the observation operator used to link the raw observations to atmospheric variables. When extended to the 4D case, the observation operator includes also several integrations of the numerical atmospheric model to either iteratively minimize the objective function of 4D-Var or to perform the Monte-Carlo simulations needed of ensemble-based assimilation methods.

5 Mathematical Challenges

Numerical modeling has been applied in atmospheric sciences, oceanography, hydrology and atmospheric chemistry. Although significant advances have been made in the different areas, their individual behavior is influenced by the other components. Oceans act as a huge heat reservoir that regulates the temperature of the atmosphere and it plays a key role in climate studies. The atmosphere transport chemical constituents and aerosols which in turn can influence atmospheric heating by absorbing or reflecting atmospheric radiation. It is now necessary to have coupled systems to represent the Earth system as a whole. This raises a number of new problems that need to be addressed if such an endeavour is to be successful. The individual components often represent phenomena that act on very different time scales. The oceans evolve on longer time scales than the atmosphere and atmospheric chemistry includes fast and slow reacting species. To model a system with such a wide variety of time scales is not obvious. One could use the lowest time step that would be acceptable for all components but this would lead to very expensive models in terms of computing time. The same can be said about the spatial resolution that is needed. Several approaches have been considered involving adaptive grids, multi-grid models, etc., each with its own advantages and disadvantages. Numerous model integrations need to be carried out over long periods of time to study different climate scenarios and their sensitivity to changes in the model inputs. Consequently, the algorithms need to be very efficient on parallel computer architecture.

Another modeling problem is the radiative transfer in presence of clouds. Incident radiation on heterogeneous surfaces is partly reflected in different directions,

partly absorbed and then propagated through a highly heterogeneous medium [Barker *et al.*, 2003, Schertzer and Lovejoy, 1987]. In [Lovejoy *et al.*, 2001], fractals are used to represent the heterogeneous nature of the distribution of particles in the ocean and used for the assimilation of satellite measurements of ocean color, which relates to the presence of phytoplankton in oceans.

Climate studies and numerical weather prediction use models which involve a lot of unknowns. Some of these are empirical parameters that have been *tuned* to bring the model to agree as much as possible with observations. Error bars can be associated with those parameters and this immediately raises the question: how will this influence a weather forecast or a climate simulation? Ensemble prediction is used to examine the sensitivity of numerical simulations to perturbations of these parameters within the bounds of what is known about the uncertainty of the input parameters [Palmer and Hagedorn, 2006]. Finite ensembles of forecasts are generated by integrating the model with random perturbations of the parameters and the objective is to evaluate the time evolution of the probability density function (p.d.f.) of the atmospheric state. Given that the size of the sample is very small with respect to the number of degrees of freedom involved, this p.d.f. is only partially known. In [Candille and Talagrand, 2005], it is argued that the size of the ensemble puts a limit on what we can know about the p.d.f. in terms of both its reliability and resolution. These considerations are important for the design of ensemble prediction systems.

The use of observations in data assimilation is based on statistical estimation and requires a good knowledge of the forecast and observation error. Forecast error can be estimated from ensemble of forecasts produced by perturbing the initial conditions according to the analysis error. Since the forecast error covariances embed the error growth processes of associated with atmospheric instabilities, they are *flow-dependent*. These background-error statistics are used in the assimilation and should therefore reflect the changing nature of the meteorological situations. Casting the problem in the framework of the Kalman filter, the Ensemble Kalman filter (EnKF) estimates the analysis and forecast error covariances which are used in the assimilation of observations. This aims at correctly weighting the information gained from observations with respect to that of the background state [Evensen, 1994, Houtekamer *et al.*, 2005]. Forecast error can be attributed to error in the input to any given model but it can also be due to the error in the model itself. Somehow this must be taken into account but then, model error must be estimated through comparison with observations and it is not difficult to argue that there is simply not enough of data to properly estimate fully the characteristics of model error. All in all the difficulties associated with data assimilation methods are related to model error and nonlinearities of the dynamics which is not fully taken into account in ensemble methods. This is now investigated with particle filters [van Leeuwen, 2009] and Bayesian estimation. As for model errors, [Trémolet, 2006] presents the *weak-constraint* 4D-Var, which permits to use longer assimilation windows than is possible with the strong constraint 4D-Var.

6 Conclusion

Complex mathematical and physical problems are associated with understanding the Earth system, and the atmosphere in particular. Given the numerous approximations that must be made, it is important to validate those models through comparison to observations. This paper presented an overview of different areas in atmospheric sciences that are faced with some mathematical challenges. These need to be translated as efficient algorithms that can tap into the computing power now available. There are vast amount of resources invested to build high performance computing centres to experiment with models at higher spatial resolutions, to study the impact of including complex physical processes, to see how the ocean, the atmosphere and land interact, and investigate how these processes impact the evolution of the future climate or the weather forecasts that are critical in many sectors of human activities.

These issues are important enough to justify also the investment made in deploying satellite instruments to observe the Earth system. In recent years, good progress has been made in the development of new data assimilation methods but the fact remains that less than 10% of the total volume of incoming satellite data can be assimilated. The presence of clouds is one important obstacle and the effort is now focusing on representing and including clouds in models of the atmospheric radiative transfer in the assimilation. There are questions associated with the estimation of the observations error statistics, taking into account that, for satellite instruments, data cannot be assumed to be independent as their error is spatially correlated. Another difficulty is that even though there is information in satellite imagery that the eye can see, the current assimilation methods have difficulties in following recognizable patterns in clouds and other constituents advected by the winds. Finally, one should not forget that the new instruments onboard satellite create extremely large volume of data that must be transmitted from the satellite to the ground, and then have to be processed and disseminated to operational numerical weather prediction centres all around the world. If this is not dealt with properly, this can be the bottle neck and the data would not be received in time to be assimilated.

The scientific community is constantly consulted for guidance on how to make the best use of these computing capabilities or the huge amounts of data collected by the global observing system. The research that needs to be done is often the result of exchange of ideas between different disciplines. The advent of quadri-dimensional data assimilation (4D-Var) is a nice example. It was the result from discussions between atmospheric dynamicists and mathematicians working on control theory. Dynamicists were interested to understand the genesis of weather systems that develop in just a few days. The problem was boiled down to minimizing a functional that had the initial conditions as the control variable. Results had been obtained with low dimensional systems (just a few hundreds of degrees of freedom) but the approach taken was to compute the gradient of this functional by finite differences which required a number of model integrations equal to the number of degrees of freedom. This was to compute *one* gradient and this had to be repeated for each iteration required by the descent

algorithm. This was clearly a dead end unless an insane amount of computing power was available. The breakthrough came from control theory which stated that the same gradient could be obtained from a single *backward* integration of the adjoint model. This happened in the mid-1980's and now, 4D-Var is part of the operational suite of several NWP centres.

Some examples were given in the preceding section in which observing and modeling the atmosphere (and the Earth system) give rise to complex problems. Following the work of [Lorenz, 1963], it was found that atmospheric processes being inherently nonlinear can lead to what is now known as *chaos* meaning that a small perturbation to the system can lead to dramatic changes in a finite time. This idea has sparked a lot of interest in mathematics which resulted in significant advances in research on dynamical systems [Guckenheimer and Holmes, 1983]. The chaotic nature of the atmosphere is reflected in studies on predictability which led to the development of ensemble prediction and probabilistic forecasting. These applications are currently treated with severe approximations that leads to limitations in the information they provide. The question is to know what reliable information can we gain from the vast amount of observations and numerical results generated by these experiments. Analysing that much data is also a challenge.

References

- [Barker *et al.*, 2003] Barker, H.W., Pincus, R., Morcrette, J.-J.: The Monte-Carlo Independent Column Approximation: Application within large-scale models. In: Proceedings GCSS/ARM Workshop on the Representation of Cloud System in Large-Scale Models, Kananaskis, Al, Canada, 10 pp. (May 2002), <http://www.met.utah.edu/skrueger/gcss-2002/Extended-Abstracts.pdf>
- [Candille and Talagrand, 2005] Candille, G., Talagrand, O.: Evaluation of probabilistic prediction systems for a scalar variable. *Quart. J.R. Meteor. Soc.* 131, 2131–2150 (2005)
- [Durran, 1998] Durran, D.R.: Numerical methods for wave equations in geophysical fluids. *Texts in Applied Mathematics*, vol. 32, 463 pages. Springer, Heidelberg (1998)
- [Evensen, 1994] Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *em J. Geophys. Res.* 99(C5), 10143–10162 (1994)
- [Gauthier and Thépaut, 2001] Gauthier, P., Thépaut, J.-N.: Impact of the digital filter as a weak constraint in the preoperational 4DVAR assimilation system of Météo-France. *Mon. Wea. Rev.* 129, 2089–2102 (2001)
- [Gauthier *et al.*, 2007] Gauthier, P., Tanguay, M., Laroche, S., Pellerin, S., Morneau, J.: Extension of 3D-Var to 4D-Var: implementation of 4D-Var at the Meteorological Service of Canada. *Mon. Wea. Rev.* 135, 2339–2354 (2007)
- [Guckenheimer and Holmes, 1983] Guckenheimer, J., Holmes, P.: *Nonlinear Oscillations, Dynamical Systems and Bifurcations of vector fields*, 453 pages. Springer, Heidelberg (1983)

- [Houtekamer *et al.*, 2005] Houtekamer, P.L., Mitchell, H.L., Pellerin, G., Buehner, M., Charron, M., Spacek, L., Hansen, B.: Atmospheric Data Assimilation with an Ensemble Kalman Filter: Results with Real Observations. *Mon. Wea. Rev.* 133, 604–620 (2005)
- [Lorenz, 1963] Lorenz, E.N.: Deterministic non-periodic flow. *J. Atmos. Sci.* 20, 130–141 (1963)
- [Lovejoy *et al.*, 2001] Lovejoy, S., Schertzer, D., Tessier, Y., Gaonach, H.: Multifractals and Resolution independent remote sensing algorithms: the example of ocean colour. *Inter. J. Remote Sensing* 22, 1191–1234 (2001)
- [Palmer and Hagedorn, 2006] Palmer, T., Hagedorn, R.: *Predictability of Weather and Climate*, 695 pages. Cambridge University Press, Cambridge (2006)
- [Rabier *et al.*, 2000] Rabier, F., Jrvinen, H., Klinker, E., Mahfouf, J.-F., Simmons, A.: The ECMWF operational implementation of four dimensional variational assimilation. Part I: experimental results with simplified physics. *Quart. J.R. Meteor. Soc.* 126, 1143–1170 (2000)
- [Rawlins *et al.*, 2007] Rawlins, F., Ballard, S.P., Bovis, K.J., Clayton, A.M., Li, D., Inverarity, G.W., Lorenc, A.C., Payne, T.J.: The Met Office global four-dimensional variational data assimilation scheme. *Quart. J.R. Meteor. Soc.* 623(623), 347–362 (2007)
- [Robert, 1982] Robert, A.: A semi-implicit and semi-Lagrangian numerical integration scheme for the primitive meteorological equations. *J. Meteor. Soc. Japan* 60, 319–325 (1982)
- [Schertzer and Lovejoy, 1987] Schertzer, D., Lovejoy, S.: Physical modeling and Analysis of Rain and Clouds by Anisotropic Scaling of Multiplicative Processes. *Journal of Geophysical Research* D8(8), 9693–9714 (1987)
- [Trémolet, 2006] Trémolet, Y.: Accounting for an imperfect model in 4D-Var. *Quart. J.R. Meteorol. Soc.* 132, 2483–2504 (2006)
- [Uppala *et al.*, 2008] Ippala, S., Simmons, A., Dee, D., Källberg, P., Thépaut, J.N.: Atmospheric reanalyses and climate variations. In: *Climate variability and Extremes during the past 100 years*, 364 pages, pp. 103–118. Springer, Heidelberg (2008)
- [van Leeuwen, 2009] van Leeuwen, P.J.: Particle filters in geophysical systems. To appear in *Mon Wea. Rev.* (2009)