Scientific challenges of convective-scale numerical weather prediction

Article

Accepted Version

AMERICAN
METEOROLOGICAL
SOCIETY

Bulletin of the American Meteorological Society

EARLY ONLINE RELEASE

This is a preliminary PDF of the author-produced manuscript that has been peer-reviewed and accepted for publication. Since it is being posted so soon after acceptance, it has not yet been copyedited, formatted, or processed by AMS Publications. This preliminary version of the manuscript may be downloaded, distributed, and cited, but please be aware that there will be visual differences and possibly some content differences between this version and the final published version.

The DOI for this manuscript is doi: 10.1175/BAMS-D-17-0125.1

The final published version of this manuscript will replace the preliminary version at the above DOI once it is available.

If you would like to cite this EOR in a separate work, please use the following full citation:

Yano, J., M. Ziemianski, M. Cullen, P. Termonia, J. Onvlee, L. Bengtsson, A. Carrassi, R. Davy, A. Deluca, S. Gray, V. Homar, M. Koehler, S. Krichak, S. Michaelides, V. Phillips, P. Soares, and A. Wyszogrodzki, 2017: SCIENTIFIC CHALLENGES OF CONVECTIVE-SCALE NUMERICAL WEATHER PREDICTION. Bull. Amer. Meteor. Soc. doi:10.1175/BAMS-D-17-0125.1, in press.

version:DOC/TWG-NWP/bams/ms.tex : 19 October 2017

SCIENTIFIC CHALLENGES OF CONVECTIVE–SCALE NUMERICAL WEATHER PREDICTION

Jun-Ichi Yano, Michał Z. Ziemiański, Mike Cullen, Piet Termonia, Jeanette Onvlee, Lisa Bengtsson, Alberto Carrassi, Richard Davy, Anna Deluca, Suzanne L. Gray, Víctor Homar, Martin Köhler, Simon Krichak, Silas Michaelides, Vaughan T. J. Phillips, Pedro M. M. Soares, Andrzej A. Wyszogrodzki

Submitted to BAMS

Affiliations:

Yano– CNRM, CNRS&Météo-France, Toulouse, France; Ziemiański, Wyszogrodzki – Institute of Meteorology and Water Management, National Research Institute, Warsaw, Poland; Cullen – Met Office, Exeter, UK; Termonia – Royal Meteorological Institute, Belgium; Onvlee – KNMI, De Bilt, Netherlands; Bengtsson – CIRES-NOAA, Boulder, USA; Carrassi, Davy – NERSC, Bergen, Norway; Deluca – Institut Catal de Cincies del Clima, Barcelona, Catalonia, Spain; Gray – University of Reading, Reading, UK; Homar – Universidad de les Islas Baleares, Palma, Spain; Köhler –DWD, Offenbach, Germany; Krichak – Tel Aviv University, Tel Aviv, Israel; Michaelides – The Cyprus Institute, Nicosia, Cyprus & Cyprus University of Technology, Limassol, Cyprus; Phillips – University of Lund, Lund, Sweden; Soares – Instituto Dom Luiz, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisbon, Portugal

Correspondence Author:

Jun-Ichi Yano, CNRM, Météo-France, 42 Av. Coriolis, Toulouse, France. E-mail: jun-ichi.yano@meteo.fr

Capsule:

Numerical weather prediction (NWP) models are increasing in resolution and becoming capable of explicitly representing individual convective storms. Is this increase in resolution leading to better forecasts? Unfortunately, we do not have sufficient theoretical understanding about this weather regime to make full use of these NWPs.

Abstract:

After extensive efforts over the course of a decade, convective–scale weather forecasts with horizontal grid spacings of 1–5 km are now operational at national weather services around the world, accompanied by ensemble prediction systems (EPSs). However, though already operational, the capacity of forecasts for this scale is still to be fully exploited by overcoming the fundamental difficulty in prediction: the fully three–dimensional and turbulent nature of the atmosphere. The prediction of this scale is totally different from that of the synoptic scale (10^3 km) with slowly–evolving semi–geostrophic dynamics and relatively long predictability on the order of a few days.

Even theoretically, very little is understood about the convective scale compared to our extensive knowledge of the synoptic-scale weather regime as a partial–differential equation system, as well as in terms of the fluid mechanics, predictability, uncertainties, and stochasticity. Furthermore, there is a requirement for a drastic modification of data assimilation methodologies, physics (*e.g.*, microphysics), parameterizations, as well as the numerics for use at the convective scale. We need to focus on more fundamental theoretical issues: the Liouville principle and Bayesian probability for probabilistic forecasts; and more fundamental turbulence research to provide robust numerics for the full variety of turbulent flows.

The present essay reviews those basic theoretical challenges as comprehensibly as possible. The breadth of the problems that we face is a challenge in itself: an attempt to reduce these into a single critical agenda should be avoided.

Background

The improvements in numerical weather prediction (NWP) over the last half century may overall be considered as an outcome of a *straightforward extrapolation* of technologies: increase of model resolution; relaxations of the dynamical approximations, from the quasi-geostrophic to the primitive equation system, and with the removal of the hydrostatic balance approximation; introduction of more complex physics as well as parameterizations[1]; and a more careful procedure for preparation of the forecast initial conditions. These model upgrades have been rather dramatic, thanks to an exponential growth in computer capabilities. These upgrades have been, in turn, contributing to the steady improvements of NWP forecast performance to date (*cf.*, Bauer *et al.* 2015).

The effort to straightforwardly–extrapolate technological capability has reached such a level that operational regional forecast models are now running with horizontal mesh sizes of 1–5 km worldwide. For example, in Europe, the French AROME (Applications de la Recherche à l'Opérationnel à Méso–Echelle) forecasts over France are run operationally with a grid spacing of 1.3 km, the Met Office in the UK uses a grid spacing of 1.5 km, and MeteoSwiss runs the COSMO (Consortium for Small-scale Modelling) model with a grid spacing of 1.1 km.

NWP capacity has reached a critical threshold: NWP models now begin to resolve individual convective elements within multicell, mesoscale, and synoptic–scale storms (*i.e.*, they are "convection–permitting" models). This tendency to higher resolution will continue: it is planned that the COSMO model will be run with a horizontal grid spacing of 500 m by 2020, thus convection will be more resolved. A goal of convective-scale NWP is to accurately forecast high-impact storms, including their locations and intensities, which has the potential to bring a wide range of benefits to society. Forecast guidance from convective-scale

[1] Note that unlike the common custom in atmospheric modeling, the present essay strictly distinguishes between physics and parameterizations: physics always refers to explicit physical processes, whereas parameterization always refers to subgrid–scale processes.

NWP is already operationally available today. At the same time, this threshold also marks an end of straightforward extrapolation of technologies for NWP, even in the crudest sense: the convective–scale regime is very different from the well–studied synoptic weather regime, calling for a qualitatively different approach. The transition to forecasting at the convective–scale is hardly a matter of straightforward extrapolation. There are several important gaps in our understanding: our basic and overall theoretical understanding of this regime is much weaker than for the synoptic–scale regime. The convective–scale regime is far more complex, even more so than as suggested by existing theoretical studies on convective dynamics (*e.g.*, Moncrieff and Green 1972; Thorpe *et al.* 1982; Rotunno *et al.* 1988; Yano and Plant 2012).

Though specific issues for convective–scale NWP may be found discussed in the literature, the big-picture view is missing: we can properly tackle the convective–scale NWP problems only by taking into account the full breadth of all the issues. Some of these challenges are particularly problematic: the "convection–permitting" regime is sometimes called the "grey zone", referring to a transition from a regime in which convection is fully parameterized to a regime in which convection is fully resolved, especially in the convection community. However, we should not reduce the problems of this regime just to that of convection parameterization. The extent of the challenge at the convective scale becomes apparent only when seeing all of the challenges together.

The practical issues faced by European weather services may be understood by the fact that, for example, a typical public user requirement in Switzerland would be a prediction of precipitation in a specific valley. A more specific example is a thunderstorm event at the Belgian music festival Pukkelpop in August 2011 (de Meutter *et al.* 2015). During the music festival, at which about 60,000 people were present, a short-lived downburst occurred. Five people were killed and at least 140 were injured. An operational failure to predict this downburst event was something to be criticized from a public perspective, although the downburst had a width of only 100 m and so was far too small to be resolved by current operational NWP models.[2] Weather services naturally need to follow those public expectations. In responding to such expectations from the public, we also need to shift the focus to the finer scales and more fully exploit the information from convective–scale NWPs.

The present essay has emerged from a sense of an urgent need for action within European NWP consortia — ALADIN (Aire Limitée Adaptation dynamique Développement InterNational), COSMO, and HIRLAM (High Resolution Limited Area Model) — in responding to these challenges. This essay complements previous BAMS articles, including Mass *et al.* (2002), Fritsch and Carbone (2004), Mass (2006), Stensrud *et al.* (2009), and Sun *et al.* (2014). As discussed therein, we clearly acknowledge that currently there are extensive research efforts at the operational level to improve convective–scale NWP by exploiting various existing observations as well as modeling techniques. The main emphasis put forward in the present essay is an urgent need to properly address more fundamental theoretical issues. With our lack of basic understanding of this regime, current efforts will sooner or later otherwise become deadlocked. A good awareness of these more fundamental issues and of the limits of the current operational efforts is crucial just for good continuation of the current progress, even though those fundamental problems may not be immediately solvable.

To keep a reasonable focus, so that we can discuss the issues in depth, this essay addresses only the most basic theoretical issues. We recognize that other issues could be equally important, such as observation-related issues, but here we limit ourselves to only discussing these in the theoretical context. As we clearly acknowledge the current operational efforts are of crucial importance, but for the sake of keeping focus they are not covered herein.

In the next section, these fundamental issues are examined one by one. Discussions begin with the most basic issues of partial differential equations (PDEs), then turn to the issues of fluid mechanics, and then gradually move to more operational issues. Though the argument

[2] See further discussions on the parameterization problems in the subsection *Parameterizaton.*

as a whole evolves over the section, since the issues to be discussed are so extensive each subsection on an issue is written in an almost stand–alone manner for ease of reading. In this manner, this essay provides a full breadth of the most fundamental problems of the convective–scale NWP.

Scientific Challenges

Partial–differential equation problem

The synoptic weather system of the 10^3–km scale can be described by the primitive equation system under hydrostatic balance. The basic mathematical structure of this system is relatively well understood (Petcu *et al.* 2008). This is in stark contrast to the nonhydrostatic anelastic system, a standard formulation adopted for convective–scale modeling.[3] This system is far more difficult to analyze mathematically, hence it is much less well known.

The synoptic–scale weather system can, furthermore, be approximated by quasi–geostrophy or, alternatively and better, by semi–geostrophy, based on the fact that the system exhibits a close balance between the Coriolis and the pressure–gradient forces. This basic feature enables us to understand, to a large extent, synoptic–scale weather in terms of balanced dynamics (*cf.*, Leith 1980).

Unfortunately, under the convective–scale regime, we lose these basic balances of the system, making it much harder to understand the fundamental characteristics of the system. Even a basic proof of nonsingularity associated with latent heating has only recently been established for the simplest case (Temam and Tribbia 2014). Understanding of these flows may partially be accomplished by identifying a wide variety of subsystems defined as asymptotic limits. However, such an understanding requires a much broader knowledge of fluid dynamics and thermodynamics, even without considering full microphysics, than for the traditional synoptic–scale system. However, these subsystems under various asymptotic limits

[3]Strictly speaking, many operational models do not follow the anelastic formulation, but adopt the fully–compressible formulation. However, these models are still designed *not to* fully resolve the sound waves by adopting semi–implicit methods for the time integration.

occupy only a small fraction of the vast parameter space in the convective–scale regime. No asymptotic representation is likely to be identified in a bulk part of this regime.

Though all these aspects may sound purely mathematical, our lack of understanding at this most basic level hinders crucial progress at more practical levels (*cf.*, *Numerics*).

Dynamical System

Synoptic–scale flows may be understood as a type of dynamical system because mathematically they reside on a *slow stable manifold* (Leith 1980), which is only weakly coupled to the much more complex dynamics of smaller–scale convection. Thus dynamics on these scales can be described with a relatively limited number of effective degrees of freedom, *i.e.*, low-dimensional dynamics like Lorenz's (1963) strange attractor. Furthermore, such an effective low–dimensionality of the system guarantees relatively stable, reliable, long-term model forecasts, even though the evolution may be somehow *chaotic*.

In the convective–scale regime on the other hand, although a wide variety of asymptotic regimes emerge, nothing equivalent to *geostrophic balance* is found: the effective dimension of the system is suddenly increased. As a result, the dynamical–system approach mostly developed for low–dimensional systems no longer works effectively. Furthermore, this transition severely restricts predictability (*cf.*, *Probability*).

Turbulence

Atmospheric flows are turbulent at almost all the scales of practical interest according to a standard definition of turbulence in fluid mechanics based on the Reynolds number, which measures the importance of nonlinearity relative to viscous dissipation (*e.g.*, Fritsch 1995). Unfortunately, this feature is often neglected due to a custom of calling planetary–boundary layer (PBL) turbulence "atmospheric turbulence", leaving an impression that turbulence is only found in the PBL of the atmosphere. It is also typical that a distinction is made between turbulence and convection, which further adds to the impression that atmospheric convection is not turbulent. While the nature of turbulence within convective cells is non-Kolmogorov, and so has different properties to that typically found in the PBL, it is fundamentally a turbulent process.

At the synoptic scale, the turbulent nature of the flow is limited by the stratification and rotation of the atmosphere and so tends to be quasi two–dimensional. An important feature of two–dimensional turbulence is that the energy is overall transferred from smaller scales to larger scales (an "inverse cascade"). As a result, atmospheric flows tend to be organized at larger scales which maintains a relative smoothness of the flow (*cf.*, Tennekes 1978). This property of two–dimensional turbulence allows us to treat synoptic–scale flows as a low–dimensional dynamical system.

On the other hand, once the horizontal scale of the system reaches below $O(10\text{ km})$, the aspect ratio of the flow becomes unity,[4] hydrostatic balance is no longer satisfied, there is no longer constraint from rotation, and the flow becomes fully three–dimensional: this is the essence of the convective scale. These flows are far more complex than two–dimensional turbulence, more transient and intermittent (*i.e.*, they lack balance) and they are associated with a much larger degree of freedom. Thus, three–dimensional turbulent flows are much harder to predict than the chaotic system found in low–dimensional dynamical systems: in the fully–turbulent regime, the number of active modes keeps increasing with increasing resolution and prediction becomes increasingly harder with no sign of convergence.

To understand fully three-dimensional convective atmospheric turbulence, the basic nature of the energy interactions between these many active modes in the system should first be properly understood. In fully three–dimensional turbulence, energy is predicted to be

[4]Observation (*cf.*, Nastrom and Gage 1985) shows that the slope of the kinetic energy spectrum as a function of the wavenumber, k, turns from k^{-3}, as expected for the two–dimensional turbulence, to $k^{-5/3}$ at about the few–hundred kilometer scale (roughly corresponding to the radius of the deformation) in a virtual contradiction to this aspect ratio argument. This regime with a $k^{-5/3}$ spectrum above the 10–km scale (often called "stratified turbulence") is still quasi–two dimensional, arising from a strong influence of the stratification on this scale range (*cf.*, Lindborg 2006).

transferred overall to the smaller scales, but some of the energy at smaller scales is also transferred to the larger scales leading to a tendency for organized convection. Although the basic mechanism of organized atmospheric convection is classically attributed to vertical wind shear (*cf.*, Moncrieff and Green 1972; Thorpe *et al.* 1982; Rotunno *et al.* 1988), its full mechanism from a point of view of full turbulence dynamics is still to be established (*cf.*, Yano *et al.* 2012). Here, we also need to move beyond a conventional framework of interactions between convection and the large scale towards a true multi–scale framework.

Our current understanding of turbulent flows is essentially based on a straightforward extrapolation of Kolmogorov's theory for homogeneous, three–dimensional turbulence (*cf.*, Zilitinkevich *et al.* 2013). Existence of the stratification and an active role of buoyancy are likely to qualitatively change the basic nature of the flow. Such an investigation into the fundamental nature of self-organized turbulence has not yet been accomplished.

Predictability

The predictability of atmospheric flows is fundamentally limited because the errors in prediction exceed the typical amplitude of a signal of a given scale at a certain point in time. Once the error exceeds this amplitude, the prediction loses any practical value, although it is always possible to run an NWP model beyond this limit.

The fully turbulent nature of the convective–scale regime limits the predictability more severely than for low–dimensional chaotic flows (*cf.*, Palmer *et al.* 2014). In a chaotic system, an error of the initial condition limits the predictability. In principle, the predictability can always be extended by defining the initial condition more accurately. However, in a fully–turbulent regime, the accuracy of the initial condition no longer ultimately limits the predictability (Sun and Zhang 2016), although a denser observational network may extend the predictability to some extent. Rather, the intrinsic nature of the flow itself (notably its intermittency) becomes the ultimate limiting factor. More observations by, *e.g.*, a denser network, do not overcome this intrinsic predictability limit.

On the other hand, one may wish that the predictability of synoptic scale would be improved by explicitly resolved convection rather than an unreliable parameterized convection. However, even this is not obvious considering the complex multiple–scale interactions of the turbulent flows associated with convection (*cf.*, *Turbulence*).

Probability

The predictability of convective systems is about a few hours (*e.g.*, Hoheneger and Schär 2007), but this is not a fixed number. In some situations, the convective system is strongly controlled by a synoptic–scale process, giving a longer predictability. It is also spatially dependent. Detailed surface data (vegetation, soil types, topography) may further help to extend the predictability. Identifying situations with enhanced predictability is an important forecast issue in convective–scale NWP.

However, regardless of its precise value, there always exists a limit beyond which a forecast becomes so uncertain that it loses any deterministic usefulness. As a result, when an NWP model is run for a few days, as is the basic strategy of the NWP community (*e.g.*, ALADIN, COSMO, HIRLAM, Met Office), the resulting forecast can only be interpreted in terms of probabilities: we cannot say precisely when and where an afternoon shower should be expected on the next day, but only give a probability distribution in time and space. In this manner, convective–scale NWP must be inherently based on probability.

Unfortunately, probability is not an easy concept to understand.[5] It is true that there are already many methodologies for predicting the probability of weather events (*e.g.*, Schwartz

[5]Note that the probability is even not a measurable quantity. For example, if a 30% probability of rain is verified by actual rain by 30% of the time, this probability forecast is *statistically consistent* with the observation. However, this is not a *sufficient condition* to verify it. The true verification must be performed on the probability forecast for each event (or non–event) individually. Of course, this is not possible, because the actual realization is rain or no–rain without an intermediate state. In other words, we can never measure a probability observationally for an individual event, but only in a statistical sense. However, the latter is not sufficient for the verification.

et al. 2010). A typically–adopted approach is to estimate a probability by creating a large sample or ensemble. However, the frequency of an event within a certain sample is not equivalent to a probability of a single unique event of particular interest. Such frequency–based thinking may be helpful for analyzing a homogeneous sequence of tries (or events), such as the tossing of a coin or dice. In contrast, a sequence of rainfall events is hardly "homogeneous": each event happens under unique circumstances. In this case, a different probability must be assigned for each rainfall event, without creating a sample.

The current standard methodology for estimating weather probabilities, the ensemble prediction system (EPS), is also based on this sample–space based thinking (*cf.*, Leith 1974). Although the EPS is indeed a useful approach, it does not predict *by itself* a probability in any obvious manner: three rain forecasts out of ten ensemble members does not automatically mean a 30% chance of rain, unless the sample is defined in a homogeneous manner. Generating such a homogeneous sample with a reasonable, finite ensemble size is not a simple matter, and it becomes more difficult for a system with an increasing number of unstable modes (*cf.*, Uboldi and Trevisan 2015).

Frequency and probability must carefully be distinguished from each other, as Bayesian probability teaches us (*cf.*, Jaynes 2003). Furthermore, any probabilistic prediction system should be derived, ideally, from the basic physical principle for predicting probability, *i.e.*, the Liouville equation (Yano and Ouchtar 2017), although its practical use may appear difficult (*cf.*, *Data Assimilation*).

Stochasticity

Prediction of individual convective events is so difficult that it is tempting to deal with them as random events arising from stochasticities. Such a formulation also more naturally leads to a probabilistic description. However, we have to be cautious in proceeding in this manner.

Some of the physical processes may be intrinsically stochastic: Brownian motion is a classical example. Many complex microphysical processes that do not provide simple closed analytical expressions, *e.g.*, generation rate of the secondary ice crystals by a collision of two ice particles (Yano and Phillips 2016), may also be best considered to be stochastic. Following this line of reasoning, one may wish to consider any noisiness in a system as a consequence of stochasticity. However, such reasoning is not necessarily justified. For example, although turbulent flows are extremely noisy, their physics is completely deterministic and presented in a closed form by the Navier–Stokes equations: a relatively simple nonlinearity can easily produce a noisy time series. The choice between using a stochastic or nonlinear representation of a given process must therefore be made carefully.

We should also realize that noisiness in short–time and small–spatial scales does not necessarily lead to a stochastic influence at larger scales: the two levels of the processes must be carefully distinguished from each other. The method of homogenization developed under multi–scale asymptotic expansions (Pavliotis and Stuart 2007) provides a rigorous procedure for assessing whether the large–scale influences of those noise-like features are actually stochastic.

Generally speaking, we should not assume that all the difficulties in predicting the convective-scale regime arise from randomness: adding more stochasticity is not necessarily a solution. We should also carefully distinguish between the intrinsic stochasticity in physics and the stochasticity introduced as an artificial device in parameterizations. The latter must be addressed with more mathematical rigor (*cf.*, Berner *et al.* 2017).

Data Assimilation

As the horizontal resolution of NWP models increases, a denser observational network is also required. However, simply increasing the number of observations is not enough. NWP models require more information than is being measured: observations generally do not cover the entire model domain, and more importantly, observed quantities are often only indirectly related to model variables. Methodologies for estimating the model state from observations come from nonlinear filtering and optimal control theory (Jazwinski 1970; Crisan and Rozovskii 2011), also referred to as data assimilation (DA: *cf.*, Kalnay 2002) in geosciences.

The full problem of DA consists of estimating the so–called posterior probability: *i.e.*, the probability of the model–system state based on the observations as well as on our general knowledge of the system (prior information). This problem can be formally solved by invoking the Bayesian theorem (*cf.*, Jaynes 2003). The Liouville equation (or its generalization including stochastic forcing) predicts the time evolution of the probability. However, such a formal approach has so far been seen as unsuitable for NWP applications: the vast dimension of the systems involved renders impractical even just estimating the probabilities, let alone computing their time evolution.

To simplify the problem, Gaussian approximation has often been introduced so that only the mean and covariance of the uncertainty probability must be computed. The two most widely–adopted DA methods for operational NWP, four-dimensional variational assimilation (4DVar: Talagrand and Courtier 1987) and the ensemble Kalman filter (EnKF: Evensen 2009), adopt this simplification. To be even more practical, operational DA is further simplified by tuning the DA to just a single dominant scale, usually the synoptic scale.

On the other hand, as model resolution increases, new phenomena are resolved on a broader range of scales including convection, and so DA must also be designed to simultaneously keep control on all resolved scales. Studies suggest that this problem may, in principle, be solved by 4DVar (Lorenc and Payne 2007) and EnKF (Snyder and Zhang 2003). However, even more changes in DAs are required to efficiently deal with two main features inherent at the convective scale: (i) a much faster and intermittent error growth rate (*cf.*, *Predictability*) and (ii) the nonlinear and non-Gaussian characters of the underlying dynamics and error statistics.

The first issue is intimately related to the concept of *observability* (*cf.*, Jazwinski 1970) that may be defined as the problem of identifying the minimum spatio-temporal observational density to efficiently counteract error growth (Quinn and Abarbanel 2010). Observability is a necessary condition for the stability of a DA solution, which is in turn a necessary condition to reduce the state-estimation (and prediction) error (Carrassi *et al.* 2008). Observability can be achieved through development of the observational network itself as well as of the DA procedure. The former includes, for example, the development of a C–band dual–polarization Doppler–radar network under the European Operational Program for Exchange of Weather Radar Information (OPERA: Huuskonen *et al.* 2014). Surface measurement (*e.g.*, soil moisture) networks with sufficient spatio-temporal resolution also contribute, although they are still to be strengthened over Europe.

There are several approaches for dealing with the second issue, including the rank histogram filter applied to Kalman–filter methods (Anderson 2010). However, the most fundamental approach for dealing with this issue is to turn to a more basic principle based on fully Bayesian Monte Carlo methods (particle filters, PFs: Doucet *et al.* 2000). A problem with PFs is that the number of particles required for accurate performance grows exponentially as the system's dimension increases (Bocquet *et al.* 2010). Choosing the importance–proposal densities that give a larger overlap with the conditional density may delay the filter collapse, or even prevent it (Slivinski and Snyder 2016). Hybrid EnKF–PF methods are promising alternative approaches to this problem (Chustagulprom *et al.* 2016). The development of advanced PFs for DA in convection–permitting NWP models will be an important priority for the coming years (*cf.*, Poterjoy *et al.* 2017).

Cloud Microphysics

Increasing model resolution also demands more sophisticated physics. Unfortunately, the issues of physics are vast. Here, we deliberately limit our discussions to the cloud microphysics, due to its unique status.

Our knowledge of microphysical processes coming both from laboratory and theoretical studies is quite extensive (*cf.*, Pruppacher and Klett 1997), although our knowledge is hardly perfect and the existing bin–microphysics parameterizations certainly do not make full use of this knowledge. At the same time, even the current bin microphysical schemes are still too expensive to use for convective-scale NWPs. In short, we know the microphysics too well and we have to somehow simplify it for it to be included in operational NWP models while maintaining a reasonable model run speed. The main problem with current microphysical modeling is that these simplifications are made in a rather arbitrary manner without performing any *systematic* "investment–gain" analysis. For example, one can find many articles in the literature claiming an improvement of a model by upgrading, for example, from a single-moment to a double-moment scheme. However, a carefully balanced judgment is often missing on relative gain against a given investment. Here, Bayesian decision theory (Berger 1985) may be called for. A solid first step towards this direction is taken by *e.g.*, van Lier–Walqui *et al.* (2014).

The benefits of implementing more realistic, and more complex, descriptions of cloud microphysics may appear enormous: hail damage could be better estimated by fully considering the hail size and hardness (Phillips *et al.* 2014), and winter precipitation (due to ice, liquid, or a mixture of both) may be better predicted by using a more detailed description of the melting process (*e.g.*, Phillips *et al.* 2007). However, in the convective–scale regime, the expected improvements may not be attainable: with convective–scale turbulence intrinsically interacting with the enhanced cloud microphysics, an increase in the complexity of the microphysics may not automatically lead to a more reliable forecast, but may lead merely to higher forecast uncertainties as if adding white noise. A suitable level of sophistication in *deterministic* physics (not only microphysics, but surface processes, radiation, *etc*) must be objectively and quantitatively assessed, with this aspect being fully taken into account.

Parameterization

The role of subgrid-scale parameterizations becomes more subtle as convection starts to become explicitly resolved. In traditional NWP models, individual convective storms are key elements to be parameterized. Under the "convection–permitting" regime, these parameterizations become *almost* unnecessary. In fact, most operational "convection–permitting" NWP models turn off the deep–convection parameterization. However, the threshold resolution for turning it off is not well established.

It is more likely that the transition towards a situation where it is no longer necessary to parameterize deep convection should be more gradual, and certain intermediate procedures are required in this transition regime (*e.g.*, Gerard *et al.* 2009). These procedures should be performed without traditional parameterization assumptions such as scale separation and quasi–equilibrium. Some studies propose a stochastic formulation (*e.g.*, Plant and Craig 2008), although a formal formulation analysis shows that the system remains deterministic even without these traditional assumptions (Yano 2014).

The focus is likely to shift to the PBL (Ching *et al.* 2014). However, many new parameterization issues also arise there, including those for sub–cloud scales of deep convection: it is very likely that the turbulent mixing between the clouds and the immediate environment must be described more carefully than traditional entrainment–detrainment descriptions (*cf.*, de Rooy *et al.* 2013).

Overall, we face challenges for subgrid–scale parameterizations from two sides. On the one side, we need to further elaborate existing parameterizations (*e.g.*, deep and shallow convection, PBL). On the other side, we also need to introduce new parameterizations, *e.g.*, for the sub–cloud scale processes. It naturally follows that the consistencies between the existing and the new parameterizations must also be carefully established. The interactions between various subgrid–scale processes, *e.g.*, between the PBL and convection, also become more critically important.

To effectively tackle all these problems together, we face issues of *consistency and uni-*

fication. Here, we propose that the best solution would be to develop a single consistent unit of subgrid–scale parameterizations by returning to the first principles of explicit physics (*e.g.*, a large–eddy simulation PDE system), and re–construct everything from there. For specific procedures, see Yano *et al.* (2015), Yano (2016). Rebuilding everything from scratch is often much faster, in the end, than trying to *unify* something already in place, but developed without much regard for mutual consistencies. These more robust parameterizations are, furthermore, expected to make the subgrid–scale information more practically useful in forecasts (*cf.*, Kain *et al.* 2010, de Meutter *et al.* 2015).

Numerics

In the traditional synoptic–scale regime, which in essence resides on a low–dimensional dynamical system, increases in spatial resolution have, overall, contributed to a better convergence of the forecast quality. On the other hand, in the convective–scale regime, with so many modes actively involved in the dynamics, solutions of the governing equations are computable with much smaller accuracy at any practical resolution, and the solutions do not converge with increasing resolution. For example, the Met Office Unified Model finds no tendency towards forecast convergence when increasing horizontal grid spacing from 1.5 km to 100 m (Stein *et al.* 2015), since the increase of horizontal resolution gradually resolves more turbulent processes. As a conventional wisdom, grid spacings at least as fine as $O(10-10^2\ \mathrm{m})$ are required for large–eddy simulations (LESs) to be meaningful. The typical "convection–permitting" grid spacing is only just comparable to the size of the largest eddies within the PBL.

Prominent flow features are often realized right at the limit of the model resolution in "convection–permitting" scale simulations, making the simulations sensitive to details of subgrid-scale parameterizations as well as to the properties of the numerical algorithms. As a result, some artifacts in outputs may result. For example, investigating the flow over a heated plane, Piotrowski *et al.* (2009) find that anisotropic viscosity can artificially produce realisticlooking regular structures that mimic naturally–generated Rayleigh-Bernard cells. Clearly, verification of these numerical results critically depends on the availability of theoretically and mathematically correct solutions of the PDEs, which can help provide a more rigorously-defined testing and selection of the numerical algorithms suitable for convection-resolving computations.

Among the numerical algorithms, advection is common to every physical variable and therefore of particular importance. A good advection scheme must conserve the sign and the shape of a variable to be advected, when the system is purely advective, by suppressing artificial oscillations and numerical diffusion. Some advection schemes suppress numerical diffusion by introducing an anti-diffusion term ("limiter"). For example, the "flux corrected transport" method, as adopted by *e.g.*, Smolarkiewicz (2006), constructs advective fluxes as weighted averages of a flux computed by a monotonic, but diffusive, low order scheme and a flux computed by a high order scheme so as to suppress unphysical behaviour.

Semi-Lagrangian schemes (Staniforth and Côté 1991) are popular among NWP models because they permit a relatively large time step while still allowing the model to run smoothly. However, we must be cautious with their application to the turbulent convective–scale regime (*cf.*, Lauritzen *et al.* 2011). Although some successful turbulent applications may be found in the literature, semi-Lagrangian schemes work most efficiently for a relatively laminar flow.

In convective–scale turbulent calculations, the numerics must be robust.[6] Particular attention is required for the dynamical core, including the treatment of advection. Though no explict discussion is provided herein, attention must also be equally paid to the numerical solver for the physics and the subgrid–scale parameterization (Dubal *et al.* 2006, Termonia

[6]In certain situations, "robust" only narrowly refers to whether a given scheme is conditionally stable. On the other hand, here we use this notion in the more general sense that given numerics are not only stable, and insensitive to a change of the resolution, *etc.*, but also preserve the basic numerical properties predicted by theory.

and Hamdi 2007)

Conclusions

We have identified the following fundamental theoretical challenges in convective–scale NWPs:

- *PDE*: A lack of proper understanding both of the dynamics and the partial differential equations describing this regime poses serious difficulty, especially for the verification of numerical model results.

- *Turbulence*: A theory of turbulence must be developed going beyond the traditional approaches based on relatively straightforward extrapolation of Kolmogorov's theory for homogeneous turbulence, to the buoyancy–driven stratified case.

- *Probability*: Probability becomes a key variable to be predicted, because NWP models are run for much longer time–scales (a few days) than the predictability limit (a few hours). The intrinsic probability, as defined by the Bayesian probability theory, should be evaluated rather than the oft-used estimation of probability by frequency counting. The Liouville equation, as a basic physical principle of probability prediction, should be further exploited to accomplish this.

- *Data Assimilation*: New assimilation approaches such as the particle filters (PFs) must be pursued because the traditional assumptions of quasi–linearity and Gaussian-distributions are no longer valid.

- *Observational Network*: Although the development of a denser observational network may be crucial, it is meaningful only under the constraints of *observability*. Moreover, the intrinsic limit of predictability (a few hours) due to the fully turbulent nature of the convective–scale regime ultimately prevents us from extending predictability through the inclusion of more observations.

- *Stochasticity*: Stochasticity must be introduced into forecast models in a more robust and solid manner, for example, based on the method of homogenization under multi–scale asymptotic expansions. It is important to keep in mind that more than a mere existence of fluctuations is required to justify the introduction of stochasticity into physics.

- *Physics*: The degree of sophistication of the model physics, notably of the cloud micro-physics, must be decided by investment–gain analysis, *e.g.*, based on Bayesian decision theory. Some of the physical processes may be better represented simply as a stochasticity.

- *Parameterizations*: Subgrid–scale parameterizations should be re–developed from scratch in a unified manner, starting from a basic set of equations for the physics and dynamics, as given by *e.g.*, LES models, so that universality and consistency are ensured.

- *Numerics*: The fully turbulent nature of the convective–scale regime demands that the numerical algorithms be much more robust than in traditional NWP models, especially to avoid generation of artificially–organized structures at the scale of the model resolution.

Each research direction requires its own substantial investments, augmenting current efforts and being subject to development of more detailed research strategies. We do not even pretend that these investigations are easy. For example, at this stage, it would be impossible to make any progress with the convective–scale regime as a PDE problem if the traditional, rigorous methodologies are to be applied; a completely different approach would be required here. On the other hand, the assimilation problem can be addressed more easily as a continuation of the current efforts. Intensive investments into the currently–existing top–end methodologies are likely to lead to breakthroughs in the relatively short term.

It is also crucial to extensively exploit existing knowledge from non–atmospheric science literature, for example, from turbulence research. These fundamental scientific issues require our re–thinking and re–structuring, but also re–directing of some non–atmospheric science research to more fundamental problems. For example, non–Kolmogorov turbulence is not solely an atmospheric problem, but it has much wider applications. A well–organized research network, as well as supporting funding, would be required so that highly multi–disciplinary research may be formed to address these problems in full.

Acknowledgments

JIY appreciates extensive discussions with European consortia members as well as a number of mathematicians over the years, which led to this essay. The authors gratefully acknowledge all reviewers, including Andrew Lorenc, Lou Wicker, for their insightful comments and suggestions, which improved the clarity and presentation of this essay. The opinions expressed in the present essay do not necessarily reflect the official positions of the consortia, the agencies, or the institutions that the authors are affiliated with. AC has been funded by the project *REDDA* of the Norwegian Research Council.

References:

Anderson, J. L., 2010: A non-Gaussian ensemble filter update for data assimilation. *Mon. Wea. Rev.*, **138**, 4186–4198.

Bauer, P., A. Thorpe, and G. Brunet, 2015: The quiet revolution of numerical weather prediction, *Nature*, **525**, 47–55, doi:10.1038/nature14956.

Berger, J. O., 1985: *Statistical Decision Theory and Bayesian Analysis*, 2nd Ed., Springer–Verlag, 617 pp.

Berner, J., and Co-authors, 2017: Stochastic parameterization: Towards a new view of weather and climate models. *Bull. Amer. Meteor. Soc.*, **98**, 565–588, doi:

http://dx.doi.org/10.1175/BAMS-D-15-00268.1.

Bocquet, M., C. A. Pires, and L. Wu, 2010: Beyond Gaussian statistical modeling in geophysical data assimilation. *Mon. Wea. Rev.*, **138**, 2997–3023.

Carrassi, A., M. Ghil, A. Trevisan, and F. Uboldi, 2008: Data assimilation as a nonlinear dynamical systems problem: Stability and convergence of the prediction-assimilation system. *Chaos*, **18**, 023112.

Ching, J., R. Rotunno, M. LeMone, A. Martilli, B. Kosovic, P. A. Jimenez, and J. Dudhia, 2014: Convectively induced secondary circulations in fine-grid mesoscale numerical weather prediction models. *Mon. Wea. Rev.*, **142**, 3284–3302, doi:10.1175/MWR-D-13-00318.1.

Chustagulprom, N., S. Reich, and M. Reinhardt, 2016: A hybrid ensemble transform particle filter for nonlinear and spatially extended dynamical systems. *SIAM/ASA J. Uncertainty Quantification*, **4**, 592–608.

Crisan, D., and B. Rozovskii, 2011: *The Oxford Handbook of Nonlinear Filtering.* Oxford University Press, Oxford, 1080pp.

de Meutter, P., and Co-authors, 2015: Predicting small–scale, short–lived downbursts: Case study with the NWP limited–area ALARO model for the Pukkelpop thunderstorm. *Mon. Wea. Rev.*, **143**, 742–756.

de Rooy, W. C., P. Bechtold, K. Frohlich, C. Hoheneger, H. Jonker D. Mironov, P. A. Siebesma, J. Teixeira, and J.-I. Yano, 2013: Entrainment and detrainment in cumulus convection: an overview. *Quart. J. Roy. Meteor. Soc.*, **139**, 1–19.

Doucet, A., S. Godskill, and C. Anrieu, 2000: On sequential Monte Carlo sampling methods for Bayesian filtering. *Stat. Comput.*, **10**, 197–208.

Dubal, M., N. Wood, and A. Staniforth, 2006: Some numerical properties of approaches to physics–dynamics coupling for NWP. *Quart. J. Roy. Meteor. Soc.*, **132**, 27–42.

Evensen G. 2009: *Data Assimilation: The Ensemble Kalman Filter*, 2nd Ed. Springer–Verlag, Berlin Heidelberg, 306 pp.

Fritsch, U., 1995: *Turbulence: The Legacy of A. N. Kolmogorov.* Cambridge University Press, 296pp.

Fritsch, J. M., and R. E. Carbone, 2004: Improving quantitative precipitation forecasts in the warm season: A USWRP research and development strategy. *Bull. Amer. Meteor. Soc.*, **85**, 955–965.

Gerard, L., J.-M. Piriou, R. Brožková, J.-F. Geleyn, and D. Banciu, 2009: Cloud and precipitation parameterization in a meso-gamma-scale operational weather prediction model, *Mon. Weather Rev.*, **137**, 3960–3977.

Hoheneger, C., and C. Schär, 2007: Atmospheric predictability at synoptic versus cloud–resolving scales. *Bull. Amer. Meteor. Soc.*, **88**, 1783–1793, doc:10.1175BAMS-88-11-1783.

Huuskonen, A., E. Saltikoff, and I. Holleman, 2014: The operational weather radar network in Europe. *Bull. Amer. Meteor. Soc.*, **95**, 897–907.

Jaynes, E. T., 2003: *Probability Theory, The Logic of Science.* Cambridge University Press, Cambridge, UK, 727 pp.

Jazwinski A.H. 1970: *Stochastic Processes and Filtering Theory.* Academic Press, New York, 376pp.

Kalnay, E., 2002: *Atmospheric Modeling, Data Assimilation and Predictability.* Cambridge University Press, Cambridge, 368 pp.

Kain, J. S., S. R. Dembek, S. J. Weiss, J. L. Case, J. J. Levit, and R. A. Sobash, 2010: Extracting unique information from high resolution forecast models: Monitoring selected fields and phenomena every time step. *Wea. Forecasting*, **25**, 1536–1542.

Lauritzen, P. H., P. A. Ullrich, and R. D. Nair, 2011: Atmospheric transport schemes: Desirable properties and a semi–Lagrangian view on finite–volume discretizations. *Numerical Techniques for Global Atmospheric Models*, P. H. Lauritzen, C. Jablonowski, M. A. Taylor, and R. D. Nair, Eds., *Lecture Notes in Computational Science and Engineering*, **80**, Springer, 185–250.

Leith, C. E., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.*, **102**, 409–418.

Leith, C. E., 1980: Nonlinear normal mode initialization and quasi–geostrophic theory. *J. Atmos. Sci.*, **37**, 958–968.

Lindborg, E., 2006: The energy cascade in a strongly stratified fluid. *J. Fluid Mech.*, **550**, 207–242.

Lorenc, A. C., and T. Payne, 2007: 4D–Var and the butterfly effect: Statistical four–dimensional data assimilation for a wide range of scales. *Quart. J. Roy. Meteor. Soc.*, **133**, 607–614.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141.

Mass, C., 2006: The uncoordinated giant. *Bull. Amer. Meteor. Soc.*, **87**, 573–584.

Mass, C. F., D. Ovens, K. Westrick, and B. A. Colle, 2002: Does increasing horizontal resolution produce more skillful forecasts? *Bull. Amer. Meteor. Soc.*, **83**, 407–430.

Moncrieff, M. W. and J. S. A. Green, 1972: The propagation and transfer properties of steady convective overturning in shear, *Quart. J. Roy. Meteor. Soc.*, **98**, 336–352.

Nastrom, G. D., and K. S. Gage, 1985: A climatology of atmospheric wavenumber spectra of wind and temperature observed by commerical airicraft. *J. Atmos. Sci.*, **42**, 950–960.

Palmer, T. N., A. Döring, and G. Seregin, 2014: The real butterfly effect. *Nonlinearity*, **27**, R123–R141

Pavliotis, G. A., and A. M. Stuart, 2007: *Multiscale Methods: Averaging and Homogenization*. Springer, Berlin, 307pp.

Petcu, M. R. M Temam, and M. Ziane, 2008: Some mathematical problems in geophysical fluid dynamics. *Computational Method for the Atmosphere and the Oceans*, Special Volume, R. M. Temam, J. Tribbia, Guest Eds., *Handbook of Numerical Analysis*, **14**, 567–740.

Phillips, V. T. J., A. Pokrovsky, and A. Khain, 2007: The influence of time-dependent melting on the dynamics and precipitation production in maritime and continental storm-clouds. *J. Atmos. Sci.*, **64**, 338–359.

Phillips, V. T. J., A. Khain, N. Benmoshe, and E. Ilotovich, 2014: Theory of time-dependent freezing and its application in a cloud model with spectral bin microphysics. I: Wet growth of hail. *J. Atmos. Sci*, **71**, 4527–4557.

Piotrowski, Z. P., P. K. Smolarkiewicz, S. P. Malinowski, and A. A. Wyszogrodzki, 2009: On numerical realizability of thermal convection, *J. Comp. Phys.*, **228**, 6268–6290.

Plant, R. S., and G. C. Craig, 2008: A stochastic parameterization for deep convection based on equilibrium statistics. *J. Atmos. Sci.*, **65**, 87–105.

Poterjoy, J., R. A. Sobash, and J. L. Anderson, 2017: Convective–scale data assimilation for the weather research and forecasting model using the local particle filter. *Mon. Wea. Rev.*,

145, 1897–1918.

Pruppacher, H. R. and Klett, J. D., 1997: *Microphysics of clouds and precipitation*. Kluwer Academic Publishers, Dordrecht, 712 pp.

Quinn, J. C., and H. D. Abarbanel, 2010: State and parameter estimation using Monte Carlo evaluation of path integrals. *Quart. J. Roy. Meteor. Soc.*, **136**, 1855–1867.

Rotunno, R., J. B. Klemp, and M. L. Weisman, 1988: A theory for strong, long-lived squall lines. *J. Atmos. Sci.*, **45**, 463–485.

Schwartz, C. S. , and Co-authors, 2010: Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Wea. Forecasting*, **25**, 263–280.

Slivinski, L., and C. Snyder, 2016: Exploring practical estimates of the ensemble size necessary for particle filters. *Mon. Wea. Rev.*, **144**, 861–875.

Smolarkiewicz, P., 2006: Multidimensional positive definite advection transport algorithm: An overview. *Int. J. Num. Meth. Fluids*, **50**, 1123–1144.

Snyder, C., and F. Zhang, 2003: Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.*, **131**, 1663–1677.

Staniforth, A., and J. Côté, 1991: Semi–Lagrangian integration schemes for atmospheric models — A review. *Mon. Wea. Rev.*, **119**, 431–440.

Stein, T. H. M., and Co-authors, 2015 The DYMECS Project: A statistical approach for the evaluation of convective storms in high–resolution NWP models. *Bull. Amer. Meteor. Soc.*, **96**, 939–951.

Stensrud, D. J., and Co-authors, 2009: Convective–scale warn–on–forecast system. *Bull.*

Amer. Meteor. Soc., **90**, 1487–1499.

Sun, Y., and F. Zhang, 2016: Intrinsic versus practical limits of atmospheric predictability and the significance of the butterfly effect. *J. Atmos. Sci.*, **73**, 1419–1438.

Sun, J., X. Ming, J. W. Wilson, I. Zawadzki, S. P. Ballard, J. Onvlee-Hooimeyer, P. Joe, D. M. Barker, P.-W. Li, B. Golding, M. Xu, and J. Pinto, 2014: Use of NWP for nowcasting convective precipitation. *Bull. Amer. Meteor. Soc.*, **95**, 409–426.

Talagrand, O., and P. Courtier, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation, I. Theory. *Quart. J. Roy. Meteor. Soc.*, **113**, 1311–1328.

Temam, R., and J. Tribbia, 2014: Uniqueness of solutions for moist advection problems. *Quart. J. Roy. Meteor. Soc.*, **140**, 1315–1318.

Tennekes, H., 1978: Turbulent flow in two and three dimensions. *Bull. Amer. Meteor. Soc.*, **59**, 22–28.

Termonia, P., and R. Hamdi, 2007: Stability and accuracy of the physics–dynamics coupling in spectral models. *Quart. J. Roy. Meteor. Soc.*, **133**, 1589–1604.

Thorpe, A. J., M. J. Miller, and M. W. Moncrieff, 1982: Two-dimensional convection in non-constant shear: a model of mid-latitude squall lines, *Quart. J. Roy. Meteor. Soc.*, **108**, 739–762.

Uboldi, F., and A. Trevisan, 2015: Multiple-scale error growth in a convection-resolving model. *Nonlinear Processes in Geoph.* , **22**, 1–13.

van Lier–Walqui, M. T. Vukicevic, and D. J. Posselt, 2014: Linearization of micropysical parameterization uncertainty using multiplicative process perturbation parameters. *Mon. Wea. Rev.*, **142**, 401–413.

M*Atmos. Sci. Let.*, **14**, 193–199.

Yano, J.–I., 2014: Formulation structure of the mass–flux convection parameterization. *Dyn. Atmos. Ocean*, **67**, 1–28.

Yano, J.-I., 2016: Subgrid–scale physical parameterization in atmospheric modelling: How can we make it consistent? *J. Phys. A: Math. Theor.*, **49**, 284001.

Yano, J.-I., L. Bengtsson, J.–F. Geleyn, and R. Brozkova, 2015: Towards a unified and self–consistent parameterization framework. *Parameterization of Atmospheric Convection*, Volume II, R. S. Plant and J.-I. Yano, Eds., World Scientific, Imperial College Press, 423–435.

Yano, J.–I., C. Liu, and M. W. Moncrieff, 2012: Self-organized criticality and homeostasis in atmospheric convective organization. *J. Atmos. Sci.*, **69**, 3449–3462.

Yano, J.-I., and E. Ouchtar, 2017: Convective initiation uncertainties without trigger or stochasticity: Probabilistic description by the liouville equation and Bayes' theorem. *Quart. J. Roy. Meteor. Soc.*, **143**, 2015–2035.

Yano, J.-I., and V. T. J. Phillips, 2016: Explosive ice multiplication induced by multiplicative–noise fluctuation of mechanical break–up in ice-ice collisions. *J. Atmos Sci.*, **73**, 4685–4697.

Yano, J.-I., and R. S. Plant, 2012: Interactions between shallow and deep convection under a finite departure from convective quasi–equilibrium. *J. Atmos. Sci.*, **69**, 3463–3470.

Zilitinkevich, S. S., T. Elperin, N. Kleeorin, I. Rogachevskii, I. Esau, 2013: A hierarchy of energy– and flux-budget (EFB) turbulence closure models for stably–stratified geophysical flows. *Bound.–Layer Meteor.*, **146** 341–373.