On the physical nudging equations

Giovanni Conti[1] · Ali Aydoğdu[1] · Silvio Gualdi[1] · Antonio Navarra[1,2] · Joe Tribbia[3]

Abstract

In this work we show how it is possible to derive a new set of nudging equations, a tool still used in many data assimilation problems, starting from statistical physics considerations and availing ourselves of stochastic parameterizations that take into account unresolved interactions. The fluctuations used are thought of as Gaussian white noise with zero mean. The derivation is based on the conditioned Langevin dynamics technique. Exploiting the relation between the Fokker–Planck and the Langevin equations, the nudging equations are derived for a maximally observed system that converges towards the observations in finite time. The new nudging term found is the analog of the so called quantum potential of the Bohmian mechanics. In order to make the new nudging equations feasible for practical computations, two approximations are developed and used as bases from which extending this tool to non-perfectly observed systems. By means of a physical framework, in the zero noise limit, all the physical nudging parameters are fixed by the model under study and there is no need to tune other free ad-hoc variables. The limit of zero noise shows that also for the classical nudging equations it is necessary to use dynamical information to correct the typical relaxation term. A comparison of these approximations with a 3DVar scheme, that use a conjugate gradient minimization, is then shown in a series of four twin experiments that exploit low order chaotic models.

1 Introduction

Nudging, or Newtonian relaxation, (Hoke and Anthes 1976; Kistler 1974) is an empirical technique that consists in adding to the prognostic equations a term that nudges the solution towards the observations (Kalnay 2002). It is a widely used method for data assimilation to avoid more complex schemes since it is easy to implement and computationally more efficient than variational and ensemble methods (Asch et al. 2016).

The model equations, discretized in space, can be written in the form of a multi-dimensional dynamical system

$$\begin{cases} \dfrac{d\mathbf{x}(t)}{dt} = -\dfrac{1}{2}\mathbf{f}(\mathbf{x}(t)), & t \in [0,\, t_f], \\ \mathbf{x}(0) = \mathbf{x}_0, \end{cases} \tag{1}$$

with the state vector and the model dynamics, $\mathbf{x}$ and $\mathbf{f} \in \mathbb{R}^n$, respectively. Here we consider an arbitrary constant factor $-1/2$ on the right hand side of (1) in order to simplify the results that will follow in the next section. When observations are available at a discrete time, say e.g. at $t = t_f$, the nudging equation can be written as

$$\begin{cases} \dfrac{d\mathbf{x}(t)}{dt} = -\dfrac{1}{2}\mathbf{f}(\mathbf{x}(t)) + \mathbf{K}(\mathbf{y}(t_f) - \mathscr{H}[\mathbf{x}(t)])\delta(t_f - t), & t \in [0,\, t_f] \\ \mathbf{x}(0) = \mathbf{x}_0 \end{cases} \tag{2}$$

where $\mathbf{y}(t_f) \in \mathbb{R}^d$ are the observations, $\mathscr{H}$ is the operator that transforms the state vector in the observational space, $\delta(t_f - t)$ is the Dirac delta function, and $\mathbf{K} \in \mathbb{M}^{n \times d}$ is the nudging, or gain, matrix. In practical application, where the observations are discrete in time and space, the delta function in (2) is neglected and the the same observations are used for longer time intervals and they can also be interpolated in space.

It is important to remark that the simplicity of the nudging method brings advantages and drawbacks. As a counterpart of the ease of implementation, the nudging method gives suboptimal results, in the sense that it does not satisfy any minimum error criterion. The Kalman gain matrix, for example, is chosen in order to minimize the a posteriori error

✉ Giovanni Conti
  giovanni.conti@cmcc.it

1. Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC), Bologna, Italy

2. Department of Biological, Geological, and Environmental Sciences (BIGEA), University of Bologna, Bologna, Italy

3. National Center for Atmospheric Research (NCAR), Boulder, Colorado, USA

covariance, assuming that the analysis can be written as a linear weighted sum of a background field and observations. This pure statistical viewpoint can be perfectly reproduced also in the 3DVar framework (Asch et al. 2016). Nevertheless, in many applications where the temporal and/or spatial homogeneity of the nudged dataset is a primary concern, e.g. initialization of long-term prediction systems, multi-decadal reanalyses, nudging may provide more consistent trajectories with respect to sequential schemes. Compared to variational methods, nudging does not require any adjoint model computation, notably simplifying the computational scheme, compared to the Kalman filter it does not require any complex computation of covariance matrices.

The implementation of the nudging consists in, essentially, coding the observation operator, as for variational and Kalman schemes, slightly modifying the model equations to incorporate the relaxation term, and tuning the **K** matrix, which can be chosen diagonal or even scalar. The gain matrix, that can also depend on the state variable, contains information about the relaxation time scales for the components of the state vector. There are several ways to choose the gain matrix, its tuning could be automated and optimized, however the ease of implementation would be lost. Usually its coefficients are determined via numerical experimentation so that the nudging term is kept small with respect to the model equations, but large enough to force the model towards the observations (Hoke and Anthes 1976; Stauffer and Seaman 1990). If **K** is very small, then **x** would solve only the model equations (1) without nudging the observations. If **K** is too large then $\mathscr{H}[\mathbf{x}(t)]$ remains close to $\mathbf{y}(t_f)$, but the dynamics of the model would be neglected. The gain matrix can be chosen to be diagonal and the relaxation time scales can be statistically estimated by means of the correlation time scales of the variables studied. The difference between (2) and the Kalman filter lies in the choice of **K**. If this matrix is chosen to be equal to the Kalman gain matrix the solution for (2) is optimal for linear and Gaussian model, but its implementation would be more time consuming. Then, the gain matrix can be seen as a simplified version of the Kalman filter gain matrix.

The parameters of **K**, can also be explored by means of optimal parameters estimation (Zou et al. 1992; Stauffer and Bao 1993). An iterative variant of the nudging equations is the Back and Forth Nudging (BFN; Auroux and Blum 2005, 2008). In the BFN algorithm, the model is integrated repeatedly forwards and backwards by means of the relaxation terms, which change their sign in the direct and inverse integration to stabilize the algorithm. The BFN has some common features with the 4DVar algorithm (Le Dimet and Talagrand 1986) which also consists of a series of forwards and backwards integrations using the adjoint of the model. The comparison between the Kalman filter (Kalman 1960) and the optimal nudging has been studied by Vidard et al.

(2003). It is important to mention that the nudging method has been known since the 1960s in the context of automatic control theory, in the linear case, where it is called Luenberger observer (Luenberger 1964) or asymptotic observer. In fact, under some hypotheses, it is possible to show that the error between the true model state and the nudging equation solution tends to zero when t goes to infinity. Auroux and Blum (2005) proved, in a very elementary case, the convergence in finite time of the iterative BFN algorithm.

Due to the wide use of the nudging technique in the geophysical sciences it is legitimate to ask:

(i) *Is there a physical framework that allows for the theoretical derivation of the nudging equations ?*

(ii) *What happens if for example unresolved processes, uncertain model parameters, boundary conditions, are explicitly taken into account in (1) by means of stochastic parameterizations? Can these fluctuations be used in order to determine the classical nudging equations in the zero-noise limit?*

(iii) *Why, in the nudging system, can we assume that the adjustment of the state vector is correct just because we force the system towards the observations? In order to simulate a realistic adjustment through the observations, when we add a new forcing term into the equations, like a relaxation, we could also need to modify the model operator **f** adding other corrections.*

(iv) *Is there a formulation that allows for the convergence in finite time to the true evolution?*

The purpose of this work is to answer these questions.

Question (ii) automatically leads to the choice of a statistical physics framework as a candidate answer for (i). For this reason the equations derived could be thought as the physical nudging system of equations. Indeed, in this context we derive a modified version of the nudging equation, also answering to (iii), and then we explore the limit for negligible noise to recover the deterministic model limit, that can be thought as the limit for which there are not unresolved scales interactions. Stochastic parameterization is commonly used in geophysical dynamics for example in turbulent closure problems (Farrell and Ioannou 2007), for large eddy simulation of geophysical flow (Duan and Nadiga 2007), or to explain zonostrophic instability (Srinivasan and Young 2012). In the limit of zero-noise then we can also gain insight on the classical nudging equations. The procedure we present is also called conditioned Langevin dynamics (Delarue et al. 2017) and it is used as a method to generate conditioned Brownian paths. It makes use of the relation between the Langevin and Fokker–Planck equations, the latter playing the role of a bridge equation that connects the probability space with the model space, and it exploits the

path integral formalism for the probability density function to express the final bridge equation in a simpler manner. For our purposes, we extend the formalism to multidimensional spaces subjected to general forcing terms not necessarily derived from a potential function. For the use of path integral and Fokker–Planck equation in the context of climate dynamics or data assimilation we refer to for example Navarra et al. (2013a, b) and Abarbanel et al. (2017). The general system of equations found, by construction, converges in finite time through the observations, partially answering to (iv). In fact the observations bring errors due to the measurement process, so the final state of the evolution of the system we find may differ from the truth up to the observational errors.

However, solving the derived nudging equation would imply the knowledge of a probability density function that could be computed by means of a Kolmogorov backward equation, or directly by the evaluation of a path integral. The computation of such a function in an operational system, with many degrees of freedom, is basically impossible. In order to make the computation feasible we perform a small time approximation of the probability density function that allows for an easier analytical evaluation of the path integral emerging from the computation. The effect of neglecting the dynamical part in the path integral evaluation is also considered.

An important consequence of this derivation is that in the final system of equations all the physical nudging parameters, when noise tends to zero, are fixed by the model and there are not free physical parameters that must be tuned such as the choice of the correct relaxation time scales. Clearly, all the other technical parameters regarding for example the discretization of the equation studied, the evolution algorithm adopted, the number of the observations used, still remain.

It is important to point out that the conditioned Langevin dynamics have been used over the past decade in order to derive a linear theory for filtering with model error from unresolved scales. In particular Harlim (2017) provides an in-depth discussion of the subject. An elegant way to formulate the problem makes use of the Mori–Zwanzig formalism (Zwanzig 1961; Mori 1965; Zwanzig 1973). In this framework a system of differential equations, involving variables representing different scales processes, can be rewritten in a closed form known as Generalized Langevin Equation (GLE) involving only the coarse grained variables. The resulting equation requires a terms resulting from the projection of the dynamics onto a function of the coarse grained variables only. The projection is not-unique in general. It also appears a memory terms in the form of an integral that represents the feedback from unresolved scales. This memory terms can be seen as a sort of nudging component also if the information of the observations in this case is not explicitly required. Finally, the last term in the GLE describes the orthogonal dynamics, introduced as noise that depends on the randomness of the initial conditions of the unresolved scales. In our derivation we assume that the unresolved scales interactions and feedback can be directly parameterized with small noise. In this way we can avoid the non unique projection procedure, and a possible lack of information, about neglecting the explicit feedback information of unresolved scales dynamics, can be mitigated by means of the observations information propagated backwards in time thanks to a new forcing term that constrains the dynamics. Then, since the dynamics is explicitly constrained by the observations, the new nudging term does not take only into account unresolved scale interactions, but also other possible missing interaction in the coarse grained space due to an excessive model simplification. In the zero-noise limit, that should be interpreted as the limit in which all the scales interactions are resolved, the memory term for the unresolved scales interactions does not represent a problem anyway.

It is worth remarking that the set of equations we explore are derived for a maximally observed system, i.e. a perfect system, in which we can observe all the components. We propose a possible way to extend the new sets of approximated physical nudging equations to the study of non-perfect systems, for which the components are only partially observed. Numerical comparisons between these approximations and a 3DVar data assimilation scheme are performed using low order models, the Lorenz system (Lorenz 1963, hereafter L63) and the Molteni system (Molteni et al. 1993, hereafter M93). The L63 system, being a simplification of the convection model, shows important characteristics that also impact the atmosphere. The M93 system can be seen as a simplification of the coupled ocean-atmosphere model. Although simplistic, this model has allowed to address in physical terms important issues related to predictability (Palmer 1993).

In Sect. 2, we present the statistical framework for the derivation of the physical nudging system of equations and we consider the zero-noise limit. We explore the analytical approximations of the path integral involved in the definition of the equations and propose how to extend the use of these equations to non-perfect systems. Furthermore we consider the reduction to the zero noise limit, the deterministic scheme. Section 3 is devoted to the application of the approximated physical nudging equations to simple dynamical systems relevant for the geoscience, the L63 nd M93 models. Finally the discussions and conclusions are given in Sect. 4.

2 Fokker–Planck bridge and nudging equations

In order to address the issue (ii), we start by considering the following stochastic system of first order in time differential equations, Langevin equations, here given in terms of components

$$\dot{x}^i(t) = -\frac{1}{2} f^i(\mathbf{x}(t)) + e^i_a(\mathbf{x}(t)) v^a(t), \qquad t \in [0, t_f], \tag{3}$$

where $\dot{x} = \dfrac{\mathrm{d}x}{\mathrm{d}t}$, $\mathbf{x}(t)$ is a trajectory in $\mathbb{R}^n$, $f^i(\mathbf{x})$ a differentiable function of $\mathbf{x}$, $v \in \mathbb{R}^m$ is a set of rapidly fluctuating stochastic terms to which we will refer as noise, and $\mathbf{e}_a$ are column vectors that belong to a matrix $\mathbf{E} \in \mathbb{R}^{n \times m}$. The Einstein summation over repeated lower and upper indices is implied, with the first letters of the alphabet, $a, b, \dots$ indicating indices corresponding to Euclidean metric and letters $i, j, \dots$ to represent tensor indices. In particular, we specialize on Gaussian white noise, that is

$$\langle v^a(t) \rangle_v = 0 \quad \text{and} \quad \langle v^a(t) v^b(t') \rangle_v = \Omega \, \delta^{ab} \, \delta(t' - t), \tag{4}$$

with δ^{ab} the Kronecker delta, Ω the noise correlation strength, and $\langle \rangle_v$ denotes the average over the noise. We also assume for the moment that every component of the system is observed, and since it is easier to work in the model space using the following tools, we set $\mathscr{H}^{-1}[\mathbf{y}(t_f)] = \mathbf{x}_f$, where $\mathbf{x}_f$ then represents the final state vector at time t_f and is entirely observed. The initial state vector is $\mathbf{x}(0) = \mathbf{x}_0$. Hereafter, to simplify the computation, we will consider the diffusion matrix $\mathbf{E}$ as constant, that is independent from the state vector $\mathbf{x}(t)$, in the time interval investigated. In this way the Ito and Stratonovich interpretations of the stochastic equation (3) coincide (Jazwinski 1970; Gardiner 1985).

A Langevin equation implies a partial differential equation (e.g. Risken and Haken 1989; Zinn-Justin 2002; Navarra et al. 2013a, b), for the time-dependent distribution $P(\mathbf{x}, t)$. This function represents the probability function for a particle to be at point $\mathbf{x}$ at time t starting from $\mathbf{x}_0$ at time $t = 0$. This equation for the probability distribution is known by several names, Kolmogorov Forward Equation, Smoluchowsky Equation, and Fokker–Planck equation (Gardiner 1985, for more details). Without assessing the merit of this terminology, we will refer to this equation in the rest of the manuscript with the term Fokker–Planck equation.

Introducing the positive symmetric matrix G^{ij}, that can be seen as a metric tensor, and its inverse determinant so that

$$G^{ij} = e^i_a e^j_a,$$
$$g = 1 / \det(G^{ij}),$$

and set the inverse metric tensor G_{ij} such that

$$G^{ij} G_{jk} = \delta^i_k,$$

the evolution equation for the probability is then

$$\frac{\partial P(\mathbf{x}, t)}{\partial t} = \frac{1}{2} \frac{\partial}{\partial x^i} \left[\Omega \, G^{ij} \frac{\partial P(\mathbf{x}, t)}{\partial x^j} + f^i(\mathbf{x}) P(\mathbf{x}, t) \right]. \tag{5}$$

This equation must be completed with the following initial condition

$$P(\mathbf{x}, 0) = \delta(\mathbf{x} - \mathbf{x}_0), \tag{6}$$

since we assume that the initial state vector is in $\mathbf{x}_0$ at $t = 0$. The probability function P must be intended as a conditioned probability density, that is $P(\mathbf{x}, t) = P(\mathbf{x}, t | \mathbf{x}_0, 0)$. Then, the probability $\mathscr{P}(\mathbf{x}, t)$ to find the system in $\mathbf{x}$ at time $t \in [0, t_f]$, over all possible paths starting in $\mathbf{x}_0$ at time $t = 0$ and conditioned to end at a given, observed, state vector $\mathbf{x}_f$ at time t_f, is nothing more than the product between two probability functions opportunely normalized. Setting

$$P(\mathbf{x}, t) = P(\mathbf{x}, t | \mathbf{x}_0, 0)$$
$$Q(\mathbf{x}, t) = P(\mathbf{x}_f, t_f | \mathbf{x}, t),$$

leads to

$$\mathscr{P}(\mathbf{x}, t) = \frac{Q(\mathbf{x}, t) P(\mathbf{x}, t)}{P(\mathbf{x}_f, t_f | \mathbf{x}_0, 0)}, \tag{7}$$

in other words, $\mathscr{P}$ is the product between the probability to go to $\mathbf{x}$ at time t starting in $\mathbf{x}_0$ at time $t = 0$ by the probability for the system to end in $\mathbf{x}_f$ at time t_f, conditioned to start at $\mathbf{x}$ at time t. We need to find an evolution equation for $\mathscr{P}$, and to do this, at first we introduce the evolution for Q, also called adjoint equation, or Kolmogorov backward equation (Jazwinski 1970; Risken and Haken 1989)

$$\frac{\partial Q(\mathbf{x}, t)}{\partial t} = \frac{-\Omega \, G^{ij}}{2} \frac{\partial^2 Q(\mathbf{x}, t)}{\partial x^i \partial x^j} + f^i(\mathbf{x}) \frac{\partial Q(\mathbf{x}, t)}{\partial x^i}, \tag{8}$$

that is the formal adjoint of a forward equation. Equation (8) must be completed with

$$Q(\mathbf{x}, t_f) = \delta(\mathbf{x} - \mathbf{x}_f). \tag{9}$$

From Eqs. (5) and (8) we can derive an evolution equation satisfied by the conditional probability $\mathscr{P}(\mathbf{x}, t)$,

$$\frac{\partial \mathscr{P}(\mathbf{x}, t)}{\partial t} = \frac{1}{2} \frac{\partial}{\partial x^i} \left[\Omega \, G^{ij} \frac{\partial \mathscr{P}(\mathbf{x}, t)}{\partial x^j} + \left(f^i(\mathbf{x}) - 2\Omega \, G^{ij} \frac{\partial \log Q(\mathbf{x}, t)}{\partial x^j} \right) \mathscr{P}(\mathbf{x}, t) \right]. \tag{10}$$

The latter equation plays the role of a bridge between the space of probability functions and the configuration space

spanned by paths that satisfy the following set of Langevin equations

$$\dot{x}^i(t) = -\frac{1}{2}f^i(\mathbf{x}(t)) + \Omega\, G^{ij}\frac{\partial \log Q(\mathbf{x}(t), t)}{\partial x^j} + e^i_a v^a(t). \qquad (11)$$

The system above is nothing more than the system representing the nudging equation when unresolved scales interactions are taken into account by means of additive stochastic parameterizations. Note that the original system is modified by the presence of an additional force. This new forcing is the analog of the so called quantum potential of the Bohmian mechanics, see for example (Benseny et al. 2016; Stauffer and Seaman 1990). It plays the role of transporting the information of the observations backwards in time constraining the solution of the dynamical system to converge to the observations in finite time. Although the observations are retrieved at discrete time, they influence the trajectory of the model also at previous time. This feature makes the physical nudging more similar to smoothers algorithms than filters.

Since we are treating a stochastic system, to properly simulate the true trajectory we have to employ an ensemble of simulations. We are interested not only in the final state, the observed state that is known here, but also in the rest of the trajectory between the observations. We can start with an ensemble of initial conditions, obtained by means of a Gaussian perturbation of the observed system, and consider the averaged trajectory evolution as a closer approximation of the true trajectory. The Gaussian perturbation, after every observation convergence, can be seen as a sort of the inflation used in the algorithms for the Ensemble Kalman Filters (Anderson and Anderson 1999) to prevent the collapsing of the ensemble spread.

In order to derive a more familiar form for the nudging equations system, and possibly to reduce Eq. (11) to the usual form in the limit of negligible noise, we need to explicitly introduce a solution for the conditional probability Q. This procedure is also necessary in order to approach realistic models with several degrees of freedoms. The computations of Q could be difficult, if not impossible, making (11) a purely formal system of equations. The analytical investigation of Q can be done by means of the path integral formalism, that is with an integral over all the possible configurations of a system, but in our case, with fixed initial and final conditions.

The conditional probability Q can be formally written as (Zinn-Justin 2002)

$$Q(\mathbf{x}, t) = P(\mathbf{x}_f, t_f|\mathbf{x}, t) = \int_{\mathbf{x}(t)=\mathbf{x}}^{\mathbf{x}(t_f)=\mathbf{x}_f} \left[d\mathbf{x}(\tau)g^{1/2}\right] \\ \exp(-\mathscr{S}[\mathbf{x}]/\Omega), \qquad (12)$$

with

$$\mathscr{S}[\mathbf{x}] = \frac{1}{2}\int_t^{t_f} d\tau \left[\left(\dot{x}^i + \frac{1}{2}f^i(\mathbf{x})\right)G_{ij}\left(\dot{x}^j + \frac{1}{2}f^j(\mathbf{x})\right) \\ - \frac{1}{2}\Omega\,\frac{\partial f^i(\mathbf{x})}{\partial x_i} \right], \qquad (13)$$

and $\left[d\mathbf{x}(\tau)g^{1/2}\right]$ an opportunely normalized functional measure of the path integral so that we can interpret (12) as the ratio between all the possible s-steps paths with fixed initial and final points, $\mathbf{x}$ and $\mathbf{x}_f$, and the s-steps paths that originate in $\mathbf{x}$. Here s indicates the discretization with which we have to interpret the functional measure, $\left[d\mathbf{x}(\tau)g^{1/2}\right] = g^{1/2}\,d\mathbf{x}_1 \ldots \mathbf{x}_{s-1}\mathscr{N}$, with $\mathscr{N}$ the right normalization as described above. The solution of the path integral (12) is only formal, it is difficult to find a solution of the multidimensional integral due to the presence of the usually non-linear force f in Eq. (13). Introducing the free Lagrangian function

$$\mathscr{L}_{free}(\mathbf{x}) = \dot{x}^i G_{ij}\dot{x}^j, \qquad (14)$$

and the so called interacting Lagrangian

$$\mathscr{L}_{int}(\mathbf{x}) = \dot{x}^i G_{ij}f^j(\mathbf{x}) + \frac{1}{4}f^i(\mathbf{x})G_{ij}f^j(\mathbf{x}) - \frac{1}{2}\Omega\,\frac{\partial f^i(\mathbf{x})}{\partial x^i}, \qquad (15)$$

we can define P_0, the probability distributions for the free particle, as

$$Q_0(\mathbf{x}, t) = P_0(\mathbf{x}_f, t_f|\mathbf{x}, t) \\ = \int_{\mathbf{x}(t)=\mathbf{x}}^{\mathbf{x}(t_f)=\mathbf{x}_f} \left[d\mathbf{x}(\tau)g^{1/2}\right] \\ \exp\left(-\frac{1}{2\Omega}\int_t^{t_f} d\tau\,\mathscr{L}_{free}\right), \\ = \frac{1}{g^{1/2}\left[2\pi\Omega(t_f - t)\right]^{n/2}} \\ \exp\left[-\frac{1}{2\Omega(t_f - t)}(\mathbf{x}_f - \mathbf{x})^\top \mathbf{G}^{-1}(\mathbf{x}_f - \mathbf{x})\right]. \qquad (16)$$

This probability does not depend on $\mathscr{L}_{int}$ and then on the dynamical forcing $\mathbf{f}$. Note that if we use Q_0 in (11) instead of the whole Q we obtain

$$\dot{x}^i(t) = -\frac{1}{2}f^i(\mathbf{x}(t)) + \frac{x^i_f - x^i}{t_f - t} + e^i_a v^a(t), \qquad (17)$$

that has the same form of (2). To obtain (17) all the information deriving from the interaction Lagrangian $\mathscr{L}_{int}$ has been totally neglected. Other terms correcting the Gaussian shape of the Q_0 distribution must be taken into account.

It is interesting to note that in (12) all the information coming from the observations is hidden in the boundary term of the path integral. In this physical derivation a term in the action that weights the innovations does

not appear, similarly to the one used in the variational schemes and derived by pure statistical, although reasonable, assumptions.

Note that the relaxation term in (17) diverges in time as $1/\sqrt{t}$ as the derivative of a Wiener process. Since we interpret the stochastic equation above by means of the framework by Ito or Stratonovich for the stochastic differential equation, this divergence does not generate any problem.

2.1 The small time approximation

If $\mathbf{f}$ is linear, or if the evolution window is small enough to allow a linear approximation of the dynamics, the path integral (12) is still Gaussian and an attempt to solve it can be done. However, the evaluation of the determinants rising from the computation is not always feasible for practical/operational purposes. If the linear approximation of the dynamics is not valid and the non-linear terms can not be absorbed in the noise forcing, the computation of the path integral is also more difficult.

In order to exploit the information contained in (15) avoiding difficult, if not impossible, computation of determinant of large matrices, or Monte Carlo integrations in space with billions of dimensions, we need a compromise, an approximation that makes the computation suitable for a possible operational use.

An attempt to find an approximated analytical solution for $Q(\mathbf{x}, t)$ can still be done considering a small $(t_f - t)$ evolution. In fact, in this case the solution of (8) will still be sharply peaked and we could drop the time dependence from the drift term. Then, for a small time evolution we are reduced to solving

$$\frac{\partial Q(\mathbf{x}, t)}{\partial t} = \frac{-G^{ij}\Omega}{2} \frac{\partial^2 Q(\mathbf{x}, t)}{\partial x^i \partial x^j} + \frac{f^i(\mathbf{x}_f)}{2} \frac{\partial Q(\mathbf{x}, t)}{\partial x^i}, \qquad (18)$$

subject to the initial (final) condition $Q(\mathbf{x}, t) = \delta(\mathbf{x} - \mathbf{x}_f)$. This backwards equation for $Q(\mathbf{x}, t) = P(\mathbf{x}_f, t_f | \mathbf{x}, t)$ is related to a correspondent forward equation, for $P(\mathbf{x}_f, t_f) = P(\mathbf{x}_f, t_f | \mathbf{x}, t)$ that can be written as

$$\frac{\partial P(\mathbf{x}_f, t_f)}{\partial t_f} = \frac{G^{ij}\Omega}{2} \frac{\partial^2 P(\mathbf{x}_f, t_f)}{\partial x_f^i \partial x_f^j} + \frac{f^i(\mathbf{x}_f)}{2} \frac{\partial P(\mathbf{x}_f, t_f)}{\partial x_f^i} \qquad (19)$$

where for the same reason explained above we have neglected the derivatives of the drift term, being small compared to the derivatives of the sharply peaked distribution. The solutions for the two equations (18) and (19) are related (Risken and Haken 1989), and we can get one solution from the other one just interchanging the set of variables considered $\mathbf{x}$, or $\mathbf{x}_f$. In particular we have

$$Q(\mathbf{x}, t) = \frac{1}{g^{1/2} \left[2\pi\Omega(t_f - t)\right]^{n/2}}$$
$$\exp\left[-\frac{1}{2\Omega(t_f - t)} \left(\mathbf{x} - \mathbf{x}_f + \frac{1}{2}\mathbf{f}(\mathbf{x}_f)(t_f - t)\right)^{\top} \mathbf{G}^{-1} \qquad (20)\right.$$
$$\left.\left(\mathbf{x} - \mathbf{x}_f + \frac{1}{2}\mathbf{f}(\mathbf{x}_f)(t_f - t)\right)\right].$$

This solution is a Gaussian with a covariance matrix given by $\mathbf{G}\Omega(t_f - t)$ and a mean given by $\mathbf{x}_f - \mathbf{f}(\mathbf{x}_f)(t_f - t)/2$. We have the picture of the system moving with a systematic drift, whose velocity is $-\mathbf{f}(\mathbf{x}_f)/2$, on which a Gaussian fluctuation is superimposed (Gardiner 1985). By using this result in (11) we obtain another correction to the nudging equations system, that is

$$\dot{x}^i(t) = -\frac{f^i(\mathbf{x}(t))}{2} + \frac{x_f^i - x^i}{t_f - t} - \frac{f^i(\mathbf{x}_f)}{2} + e_a^i v^a(t). \qquad (21)$$

The new constant term introduces a systematic drift, dependent on the dynamics, that helps the system adjustment towards the observations. The new constant terms can be seen as the answer to (iii), so along with the relaxation terms other corrections must be taken into account. Clearly (21) is just an approximation of (11), and we can not expect that in finite time the solution of the model converges exactly towards the observations. Note that, although the physical nudging has to be considered an ensemble method, no huge covariance matrices must be computed. What is required is the evolution of N members with the additional terms in (21).

In reality we have access only to portions of the system, that is we can measure only a limited amount of components of the whole state vector. The presented formulation is strictly derived only for a perfect system. The constant term $f^i(\mathbf{x}_f)/2$, that adjusts the relaxation of the i component of the state vector in (21), depends on the whole state vector at time t_f, that must be interpreted as the time for which we have the observations of the system. In order to extend the formulation we have presented to non-perfect systems, we need to introduce another assumption.

It is reasonable to think that if only p out of n components, x^i with $i = 1, \ldots, p$, are observed, but the variability of the system is slow enough, we can attempt, at a first guess, to approximate the non observed $n - p$ components with their mean values computed on the last assimilation window, that is $\bar{x}^j$ with $j = p + 1, \ldots, n$. Defining a new state vector computed at the observational time as $\tilde{\mathbf{x}}_f = (x_f^1, \ldots, x_f^p, \bar{x}^{p+1}, \ldots, \bar{x}^n)$ and introducing the further assumption that

$$\mathbf{x}_f \approx \tilde{\mathbf{x}}_f \quad \text{and} \quad \mathbf{f}(\mathbf{x}_f) \approx \mathbf{f}(\tilde{\mathbf{x}}_f) \qquad (22)$$

we can still directly drive the p observed components, and indirectly the remaining $n - p$ components, with the approximated nudging equation, adding the guessed nudging terms only on the observed components. The slow variability can be obtained reducing the time interval between assimilations, if the frequencies of the observations permit it. This procedure, that exploits the average information for each component computed during the last forecast evolution, can be thought as a background regularization, similarly to the one used, for example, in the usual variational methods.

2.2 A note on the zero-noise limit and the observational error treatment

In the zero-noise limit we consider $\Omega \to 0$ and we can neglect the noise appearing in Eq. (21). Nevertheless, part of the new forcing term still remains, and a clear reduction to the usual empirical system is not possible. It is interesting to remark that the remaining relaxation term in (17) and (21) diverges in time as $1/\sqrt{t}$, and the deterministic version of the equation is then only formal. In order to give some meaning to this approximation we need to consider again the discretized version of the equation as in the pure stochastic context. Note also that since the fluctuations appearing in (17) and (21) are Gaussian with zero mean (4), the average of the ensemble evolution for these systems of perturbed equations gives the same results of the zero-noise limit considered above. To study a deterministic model, or if you prefer a model where the dynamics of the unresolved scales can be neglected, there is no need for an ensemble of simulations and the computational cost of the schemes can be largely reduced.

In reality the observations possesses an uncertainty. An extension to the observational error treatment is however, at least formally, straightforward. In fact, if we assume that the observations are distributed according to the distribution $O(\mathbf{x}, \mathbf{x}_f, \boldsymbol{\epsilon})$ at time t_f, where $\boldsymbol{\epsilon}$ are parameters of the distribution, the final condition for the Kolmogorov backward equation is no longer a delta function centered on $\mathbf{x}_f$, but

$$Q(\mathbf{x}, t_f) = O(\mathbf{x}, \mathbf{x}_f, \epsilon). \tag{23}$$

If we indicate the fundamental solution (20) with $Q_d(\mathbf{x}_f, t_f | \mathbf{x}, t)$, the new solution is then

$$Q(\mathbf{x}, t) = \int O(\mathbf{y}, \mathbf{x}_f, \epsilon) Q_d(\mathbf{y}, t_f | \mathbf{x}, t) \, d\mathbf{y}. \tag{24}$$

Using this expression, we could recover a nudging equation for a small time that also exploits the observational error. Although the procedure is simple, unless we manage particular situations, in which for example the dynamic is linear $\mathbf{f}(\mathbf{x}) = \mathbf{F}\mathbf{x}$ with $\mathbf{F}$ symmetric and O Gaussian, (24) can not be easily solved.

3 Twin experiments with low dimensional models

3.1 Experiments with the L63 model

In the following we explore the quality of the physical nudging approximations derived in the previous sections by means of a well known low dimensional L63 model

$$\dot{x}(t) = \sigma(y(t) - x(t))$$
$$\dot{y}(t) = x(t)(\rho - z(t)) - y(t) \tag{25}$$
$$\dot{z}(t) = x(t)y(t) - \beta z(t),$$

which exhibits chaotic behavior when, for example, $\sigma = 10$, $\rho = 28$ and $\beta = 8/3$. This system represents a simplification for the convection in atmosphere, and due to its chaotic behavior there is lack of predictability.

For our purpose, to test the GN/PN approximations, the system is augmented with a stochastic forcing. For simplicity let us consider $\mathbf{E} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix, and then also $\mathbf{G} = \mathbf{I}$. The stochastic system can then be written simply adding a fast fluctuating term, the noise $\nu(t)$ satisfying Eq. (4), to the three components of the system (25).

This relatively simple model enables us to compare the approximated physical nudging (hereinafter, PN) (21) with the physical nudging approximation that does not consider the dynamical correction but only the gaussian terms of the free Lagrangian (hereinafter, GN) (17), with the deterministic physical nudging (hereinafter, PND), that is Eq. (21) in the limit of zero noise, and finally with a robust and well known data assimilation technique, 3DVar. The aim of this section is not to make a list of the best nudging methods, that should be done using more realistic models, but to understand the behavior of the physical nudging approximations under different situations, such as, for example, the partial knowledge of the components of the system at fixed times. We propose three different experiments.

In the first experiment, E_1, we consider a maximally observed system, that is at the analysis times we observe all the components of the system. In the second experiment E_2, we take into account only partial information about the observed system components at analysis time, that is we observe only two components of the system, in particular y and z. The first component, x, is not constrained by the observations, but it is constrained directly by y that appears in the first equation (25), and indirectly by z that does not appear in the first equation if not only through y. Finally we consider the worst case, the minimally observed system, E_3, in contrast with E_1. In this last experiment we study the behavior of the system observing only the z component. This means that the y component is

constrained directly by the presence of z into the second equation, although there are not direct observation for this second component, and the x component is constrained only indirectly by the observations. Experiments E_2 and E_3 allow to test the regularization introduced in order to manage non-maximally observed systems. The experiments differ in the number of the observed components, but the rest of the configuration is kept fixed, including the observations. In particular, we use a simple Euler-Maruyama scheme for the evolution of the stochastic ensemble of the PN and GN procedure, and a forward Euler scheme to evolve the true solution and the PND and 3DVar trajectories. For the stochastic evolutions, we fix $\Omega = 4 \cdot 10^{-1}$ and we make use of $N_{ens} = 50$ ensemble members. The time step is d$t = 2.5 \cdot 10^{-3}$ and the analysis window length is $\Delta t_A = 6 \cdot 10^{-2}$. We explore $N_A = 100$ assimilation cycles and a forecast of the same duration after the last assimilation. Starting from an initial condition inside the strange attractor of the system, we evolve the dynamical system (25) in order to find the true solution around which we measure the quality of the different assimilation procedures. The observations are created perturbing, every Δt_A time interval, the true solution by means of Gaussian noise with standard deviation $\epsilon = 2$, the same order of magnitude used also in (Lei and Bickel 2011; Tandeo et al. 2015; Goodliff et al. 2015). After each assimilation cycle the PN and GN ensemble are recreated perturbing the mean of the ensemble analysis using a standard deviation $\sigma_{inf} = 2 \cdot 10^{-1}$. This parameter acts as a sort of inflation that restores the ensemble collapsed near to the last assimilated observations. This process helps to explore the possible true trajectory of the system better. The covariance matrix **B**, proper of the 3DVar scheme, is computed directly from the covariances between the system components after a long evolution run inside the strange attractor, and is kept fixed for all the experiments. The observations error covariance matrix is $\mathbf{R} = \epsilon^2 \cdot \mathbf{I}$. We use a conjugate gradient method for the minimization of the 3DVar cost function. The initial conditions for the single trajectory of 3DVar and PND are computed as the mean of the initial conditions for the ensemble of PN/GN.

Since we are performing Observing System Simulation Esperiments (OSSEs) we know the true trajectory of the system. For each experiment we compute, for every time step, the squared root of the quadratic distance between the

analysis trajectory of 3DVar/PND, or the mean of the analysis trajectories for PN/GN, and the true trajectory. The error is computed for each component separately and also as an aggregate quantity, that is as the squared root of the mean of the variances between the components and the truth. This approach allows us to compare the quality of the analysis between experiments with a variable number of observed components. A summary of the experiments is depicted in table 1 where we also show the ratio between the root mean square error (rmse) computed for the PND scheme and the rmse of the 3DVar method. The bold numbers in the table underline where PND performs better that 3DVar. The ratio between the rmse of the PN method and 3DVar is not shown because it gives the same results up to two digits. The rmse is computed for the single components and as an aggregate quantity as described above. For the experiment E_1 we also explore the effect of the ensemble size, the impact of the noise correlation strength Ω and the residuals at the assimilation time, that is the square of the quadratic distance between analysis and observations.

3.2 E_1

In the first experiment, all the explored assimilation methods behave similarly in terms of the ability to give a reliable long forecast close to the true trajectory (Fig. 1a). However, if we give a closer look at the evolution during the assimilation period, for example zooming as in Fig. 1b in the time window $t = [3.6, 4]$, the differences can be noticed more clearly. The GN approximation, that does not use the dynamical term that brings backwards the information related to the observations, generates big jumps in the evolution at the assimilation time. This seesaw, although reduced, is present also for 3DVar. It is enough to add the constant dynamical term depending on the observed state to correct this bad behavior. In fact, the mean trajectory given by the PN approximation gives much smoother results compared to the ones obtained with the 3DVar. The deterministic version, PND, gives a trajectory that is almost overlapped to the mean trajectory of the PN methods that remain hidden in the first panel of Fig. 1. Although the full Eq. (11) is approximated using only the linear term in the small time expansion, the results of the assimilation are interesting. Figure 1c shows the behaviour in time of the error between the true trajectory and the

Table 1 Summary of the experiments' configuration and results

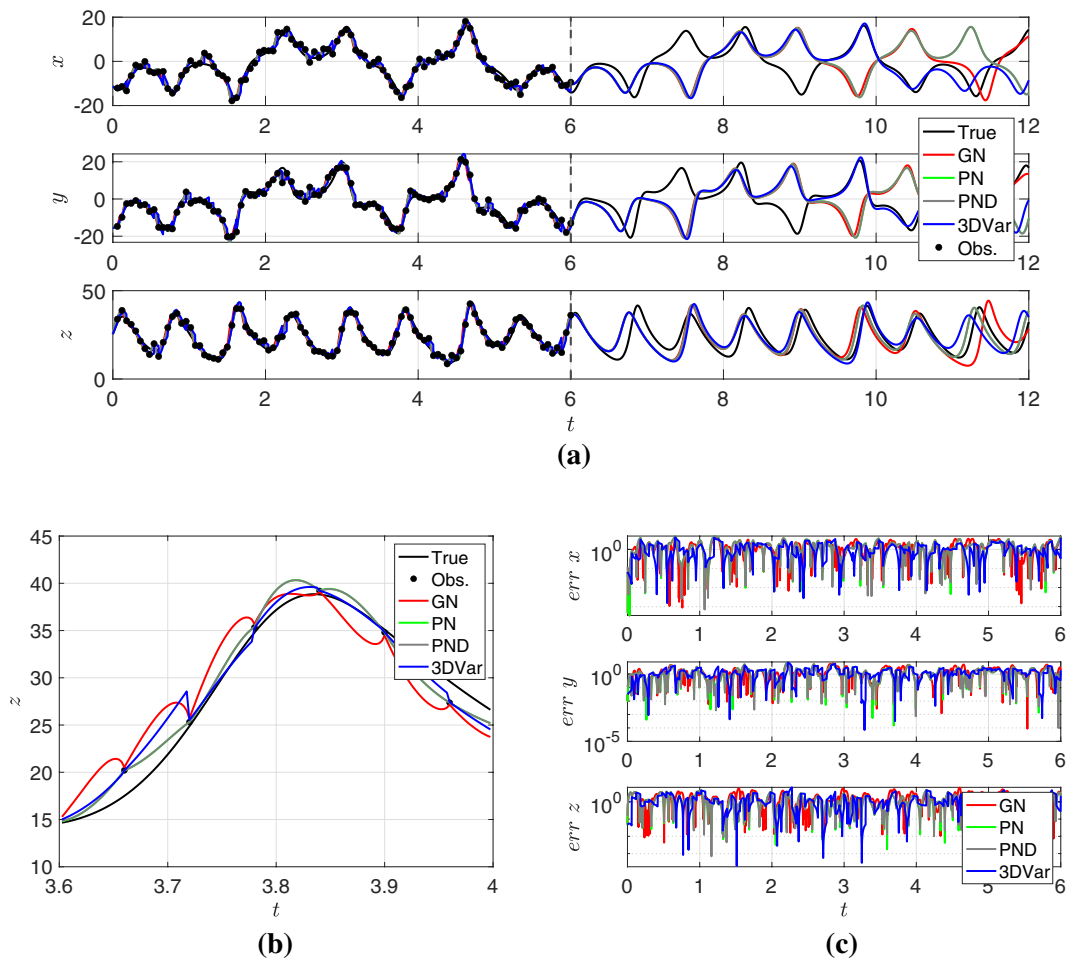| Exp.	Observations			rmse PND / rmse 3DVar			
	x	y	z	x	y	z	*tot.*
E_1	✓	✓	✓	1.22	**0.94**	**0.93**	**1.00**
E_2	–	✓	✓	**0.62**	**0.87**	**0.91**	**0.82**
E_3	–	–	✓	–	–	–	–

(a)

(b)

(c)

Fig. 1 Summary of E_1. In panel **a** it is shown the evolution of the trajectories for the assimilation period and the long forecast run without assimilation. The two time windows are separated by the vertical dashed line. The true solution is in black, the observations that are present only in the first part of the evolution are depicted as filled circle, and the red, green, gray, blue lines represent respectively the evolution of the mean trajectory for GN, PN, the deterministic physical nudging trajectory and the evolution for the system that uses 3DVar.

The green and the gray lines are almost overlapped. In panel **b** there is a zoom of the evolution of the z component during $t = [3.6, \ 4]$ to highlight the jumps of the GN method and 3DVar, and the smoother trajectories from PN, PND. Finally, in panel **c** is shown the squared root of the quadratic distance between the true trajectory and the trajectories obtained with the different assimilation schemes for every single component at every time step

different analysis trajectories at every time step for the single system components. In particular for this experiment, as also highlighted in Table 1, PN/PND and 3DVar have comparable performance in terms of rmse. Values smaller than one for the rmse ratio are found for y and z, indicating that for these two components the PN/PND schemes give better results than 3DVar. Nevertheless it is important to remind that this is a maximally observed system. Note that the error axis in Fig. 1c is in logarithmic scales.

The improvements of the PN/PND approximation with respect to GN are still more evident looking at the residual, computed at the observations times, Fig. 2a. The residuals generated by the GN approximation are comparable with the one generated by means of 3DVar and they are always bigger than the residual derived with the PN/PND schemes.

For completeness, in order to emphasize the effects of the ensemble size and of the noise correlation strength on the PN method, we show in Fig. 2b the behaviour of the total rmse varying the ensemble size for different $\mathbf{\Omega}$ values. The rmse quickly reach a plateau increasing the number of the instances in the ensemble, nevertheless the convergence speed and the size of the error depend on the magnitude of $\mathbf{\Omega}$. The worst result, in terms of rmse, is obtained using only one member. It is important to stress that the PN method with one member is not equivalent to PND. Indeed, although both methods give a single trajectory, the one obtained with the PN scheme is affected by noise, that is a non-perfect model taken into account.
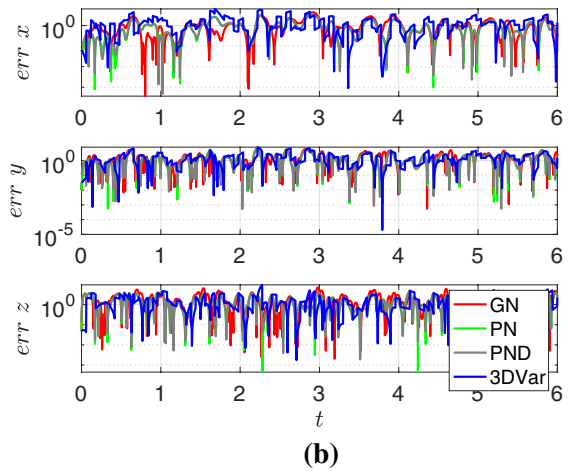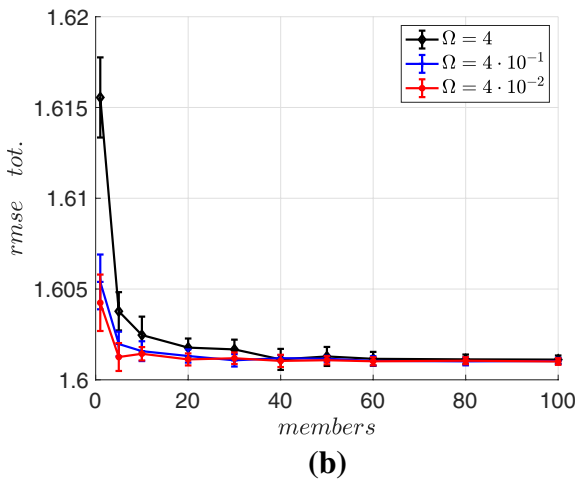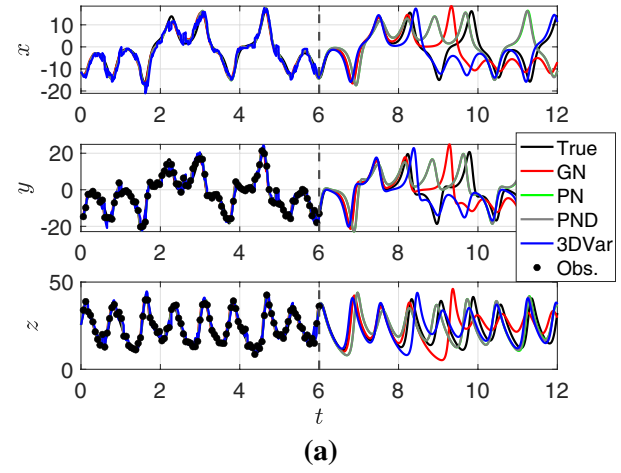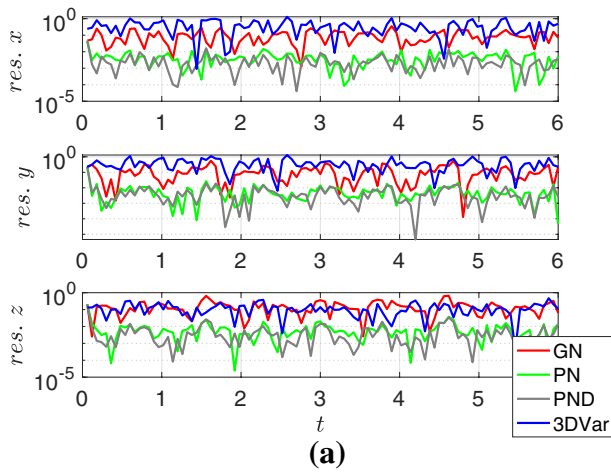
(a)

(b)

Fig. 2 On the top panel there is a comparison of the squared root of the quadratic distance, for every single component, computed between the observations and the analysis, the residual, obtained with the different assimilation schemes. On the bottom panel is shown the behaviour of the total rmse varying the number of ensemble members and different noise correlation strength, $\Omega = 4, 4 \cdot 10^{-1}, 4 \cdot 10^{-2}$

3.3 E_2

In this experiment only two components, y and z, out of three are observed. This means that the x component evolves being constrained only indirectly by the observations through the y component, as highlighted by the first equation of the system (25). As shown in Fig. 3 all the three methods are able to describe the evolution of the true solution during the assimilation period, the first evolution window of 6 time units. The x component is also well driven, but the trajectory of 3DVar exhibits strong corrections and the path it is not smooth at all. It is interesting to note that the PN/PND analysis trajectories are closer to the observations at the end of the assimilation window with respect to 3DVar. However, since the observations during the last assimilation cycle are particularly perturbed,

(a)

(b)

Fig. 3 As Fig. 1a, c but for the experiment E_2

the 3DVar analysis, that remains closer to the truth, gives a better forecast. The errors, Fig. 3b, at every time step show that the performance of the different assimilation schemes are still comparable. This is also highlighted by means of the rmse ratio appearing in Table 1. For this experiment, generally, for all the components, the PN/PND schemes work better than 3DVar. This suggests that the regularization procedure introduced, based on the slow variability assumption, can be used in order to manage non-maximally observed systems, extending the usability of the developed tool that was strictly derived for maximally observed system.

3.4 E_3

This experiment represents the extreme case, opposed to E_1, in which we have a minimally observed system. In this case only the z component is observed. This is the worst case we can imagine since the y component is driven indirectly by the observations through z, while x is not affected by z.

While the z component tends to pass towards the observations, the other two components are poorly constrained, and they tend to drift away from the true solution. This implies big jumps for the z component, that wants to restore the right dynamics, creating the instability that makes the full trajectory of the system diverge from the truth. For some set of observations the PN/PND schemes exhibit particular instabilities that sometimes drive the system outside the strange attractor region, not shown here. When the assimilation is switched off the trajectory still collapses on the strange attractors, but the possibility to do a forecast is clearly null. Also 3DVar exhibits a clear seesaw in the z trajectory during the assimilation period, Fig. 4b. This makes the other two components drift away from the true trajectory. For this particular experiment none of the assimilation schemes used in this work is able to perform a correct analysis, and therefore neither the regularization procedure can be tested.
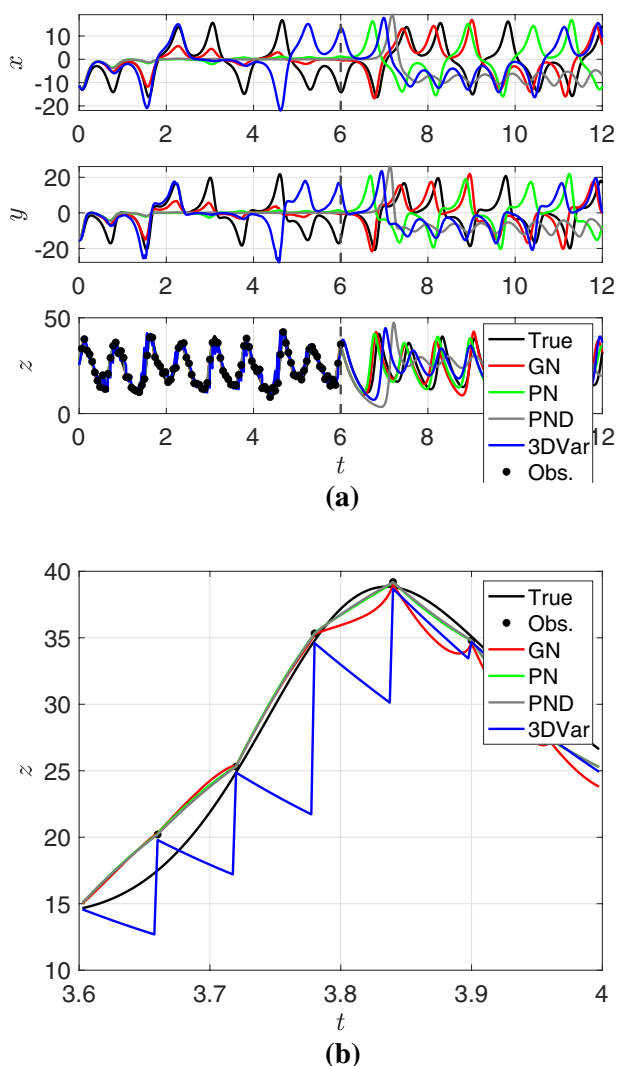

(a)

(b)

Fig. 4 As Fig. 1a, b but for the experiment E_3

3.5 An example with a simplified coupled model for "atmosphere" and "ocean", M93

Using the L63 model we explored the behavior of PN/PND applied to a low order chaotic dynamical system that mimics the main features of the atmosphere. However, the climate system is characterized by several different interacting components which dynamics span different time and spatial scales. The aim of the following example is the investigation of the PN/PND behavior applied to the L63 model forced by a system evolving with a different temporal scale. Once again these methods are compared with the 3DVar scheme. For completeness, in the plots the GN is also shown, although its performance is worse than the one of the PN/PND schemes.

The low dimensional system under examination has been introduced in Molteni et al. (1993). The L63 system is coupled with a two dimensional linear set of equations in order to simulate the equatorial sea surface temperature (SST) anomalies influence on the global system. The resulting system is the following

$$
\begin{aligned}
\dot{x}(t) &= \sigma(y(t) - x(t)) + v(t) \\
\dot{y}(t) &= x(t)(\rho - z(t)) - y(t) + w(t) \\
\dot{z}(t) &= x(t)y(t) - \beta z(t) \\
\dot{w}(t) &= \gamma v(t) - k(w(t) - w*) - y(t) \\
\dot{v}(t) &= \gamma(w(t) - w*) - kv(t) - x(t),
\end{aligned}
\tag{26}
$$

where $\sigma = 10$, $\rho = 30$, $\beta = 8/3$, $k = 0.1$, $\Omega = 2\pi/20$ and w^* will assume three different values for the tests we perform as described in the following. While the L63 model does not correspond directly to large scale motion there are important qualitative similarities with the large scale dynamics of the atmosphere. The linear system introduced by Molteni et al. (1993) can be seen as a conceptual model for the coupling of the tropical SST to the chaotic extratropics. Although it may not be straightforward to establish a strict connection between variables of the simplified model with the components that represent the general circulation of the atmosphere and ocean, this simple system of equations allows one to address, in physical terms, important questions related to the predictability (Palmer 1993), and to investigate the asynchronous coupling of the ocean-atmosphere system (Dubois 1999).

We explore what happens to PN/PND methods when applied only to a L63 system forced by a slower dynamical system in three different configurations. Since w^* can be thought as a representation of the Pacific SST anomaly, conceptually, we mimic the effect of a positive SST anomaly in the Pacific, El Niño condition with $w^* = 2$, a neutral state with $w^* = 0$ and finally a negative anomaly, La Niña condition with $w^* = -2$. We call these three cases

respectively as E_+, E_0 and E_-. Only the fast components, that are x, y, and z, now affected by the forcing due to w and v, are observed. This is another way to test the regularization procedure based on the slow variability assumption. Their comparison and results are discussed in the experiment E_4 in the next section.

The probability density function of the L63 model in the $x - y$ plane is represented by a symmetric double peaked function. This shape is still maintained coupling the L63 system with the two linear equations when $w^* = 0$, that is when a neutral configuration is considered. In case of positive or negative SST anomalies, that is the case of El Niño ($w^* > 0$) or La Niña ($w^* < 0$) events, the symmetry is broken and the probability density function becomes asymmetric. The oceanic forcing pulls away the trajectories of the L63 system towards preferred regions of the attractor. These changes, modifying the statistics proper of the system under study, could have some impact on the capabilities of the data assimilation algorithms to make the right corrections.

As before for our purpose the system is augmented with a stochastic forcing. For simplicity all the components are subjected to the same Gaussian white noise with $\mathbf{E} = \mathbf{I}$. The stochastic system can then be written simply by adding a fast fluctuating term, the noise $v(t)$ satisfying Eq. (4), to the five components of the system (26). The other parameters used for the integration of the trajectories are as in the previous experiments.

3.5.1 E_4

In this experiment we compare the performance of PN/PND against the 3DVar in three different configuration of the M93 model, namely E_+, E_0 and E_-, characterized by different values of w^*. The three configurations differ with the shape of the strange attractor generated and then in the probability density function characterizing the systems. Despite the fact we observe only three components (x, y, z) out of the five that describe the complete M93 model, we obtain for all the three cases similar results to the ones obtained for E_1, though, with interesting differences. Here we are considering only a partially observed system, however, all the fast components, the L63 components, are observed as in the E_1 experiment

. The similarity with the E_1 experiment can be appreciated looking at the rmse ratio reported in Table 2. The bold numbers in the table underline where PND performs better that 3DVar. As before the x component is generally better assimilated with 3DVar. Panel (a) of Figs. 5, 6, 7 shows the behaviour in time of the analysis residuals in semilog scale for different assimilation methods. At the assimilation time the PN/PND schemes bring the analysis trajectories closer to the observations than 3DVar. Thanks to the forcing terms in the PN/PND schemes, the information of the observations propagated backwards in time generally allows smoothers and more accurate evolution of the system, see panel (b) of Figs. 5, 6, 7 which is a zoom of the z component of the system in the time interval [3.6, 4] as before. The "oceanic forcing", generally improves the assimilation of the PN/PND schemes with respect to 3DVar, in particular in the case of La Niña experiment, E_-, and increase the seesaw pattern in the 3DVar analysis with respect to E_1.

For the PN/PND method the analysis residuals are clearly similar to the innovations, but this is not true for 3DVar for which the trajectories before the correction at the assimilation time tend to diverge from the truth. Then, the innovation of 3DVar is a few orders of magnitude bigger than the one for PN/PND.

Also this experiment suggest that, by means of the regularization procedure, the derived tools can be used for partially observed system. Indeed, although the w and v components are introduced in the new nudging terms as averaged quantities, computed during the last forecast evolution, the results are encouraging. Panel (c) of Figs. 5, 6, 7 shows the square of the quadratic distance between the analysis and the true trajectory, at every time step, for each component. Although the residuals are smaller for the PN/PND schemes, the global behaviour, as already highlighted by Table 2, is comparable between the different assimilation methods. It should be noted that although as before the assimilation of the z component alone is not possible, in the new forced system PN/PND no longer produce trajectories that leave the attractor as before (not shown here).

4 Discussions and conclusions

In this work we derive a new set of nudging equations starting from a statistical physics point of view.

We explicitly consider the unresolved scales interactions as stochastic parameterization in the form of Gaussian white noise, and we derived a nudging equations system highlighting the strong relation between these equations and physics. The derivation is strictly valid only for maximally observed systems in which we can observe all the components of the state vector. Nevertheless this was a starting point from which we looked for approximated nudging equations

Table 2 Results of the rmse ratio for E_4

E_4	rmse PND/rmse 3DVar			
	x	y	z	tot.
E_+	1.06	**0.92**	**0.94**	**0.96**
E_0	1.20	**0.96**	**0.93**	1.01
E_-	1.20	**0.89**	**0.83**	**0.96**

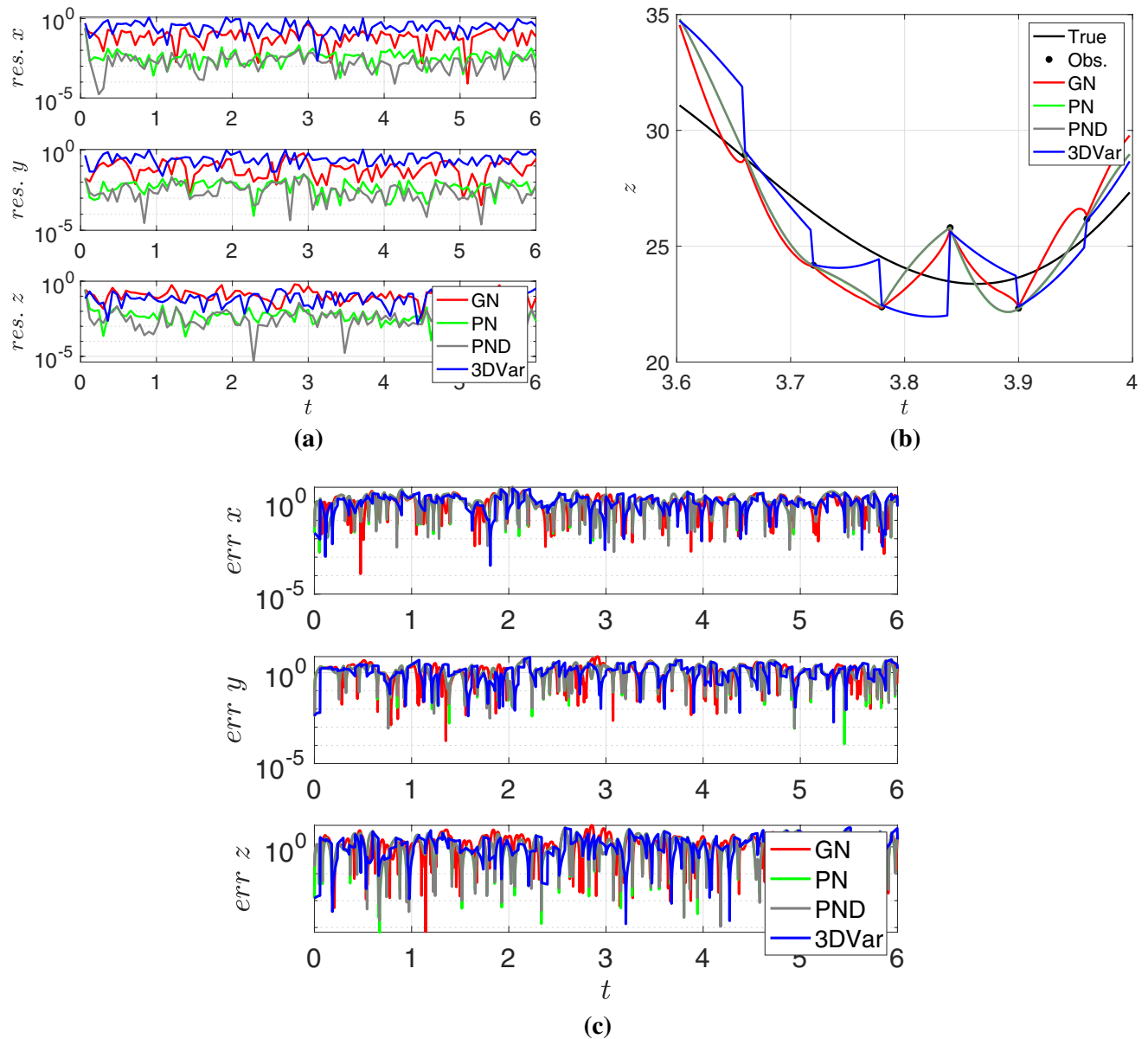All the components of the forced L63 model are observed

Fig. 5 Results from the test E_+. **a** Shows the residual for the single components x, y and z. **b** Shows a zoom of the evolution during $t = [3.6, 4]$ for the z component. **c** Shows the squared root of the quadratic distance between the truth and the trajectories computed with the other schemes for all the time steps

system valid also for non-maximally observed systems. We also explored the approximated schemes in the limit of zero-noise in order to check the differences with the classical nudging schemes.

The derivation was based on the conditioned Langevin dynamics technique (Delarue et al. 2017). Exploiting the relation between the Fokker–Planck and the Langevin equations, we defined the evolution equation for a probability distribution of a multidimensional system, conditioned to start from a specific point of the state space at $t = 0$, and to arrive at a specific location, the observations, at $t = t_f$. From

this evolution equation in the probability space we could then recognize the related Langevin equation in the coordinate space, the physical nudging set of equations searched.

The new nudging term found is the analog of the so called quantum potential of the Bohmian mechanics (Benseny et al. 2016; Stauffer and Seaman 1990). It involves a probability related to a backwards Kolmogorov equation that plays the role of transporting the observations' information backwards in time, and constraining the solution of the dynamical system to converge towards the observations in finite time. Although the observations are retrieved at a discrete time,
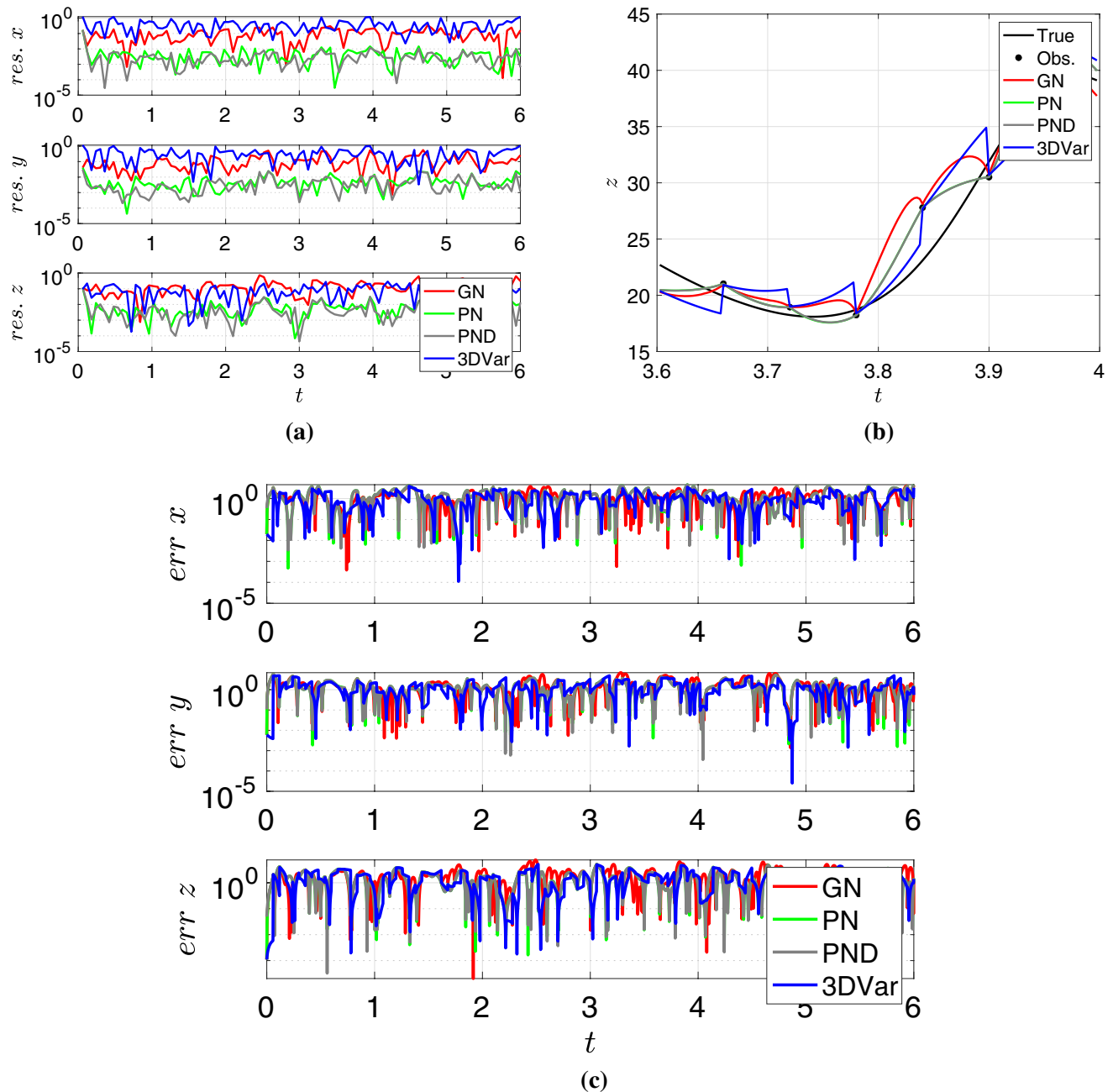
Fig. 6 As in Fig. 5 but for E_0

they influence the trajectory of the model also at a previous time, making the physical nudging, ideally, more similar to smoothers rather than filters. The physical nudging itself should be regarded as an ensemble method.

The computation of the physical nudging term is costly for high dimensional numerical models and in order to simplify its use, we explored a few approximations of the original physical nudging system. The first approximation attempt, GN, to reduce the new system to the usual nudging equation, exploited a strong truncation of the path

integral that is used to define, at least formally, the new nudging term. This truncation neglected all the effects of the model dynamics in the backwards evolution and translated in a simple relaxation term that must be added to the original model equations. The second approximation, PN, exploited an expansion for small time of the backwards probability, and it allowed keeping some dynamical information other than the typical relaxation term introducing differences with the usual nudging schemes. A deterministic version of the PN scheme has also been derived taking
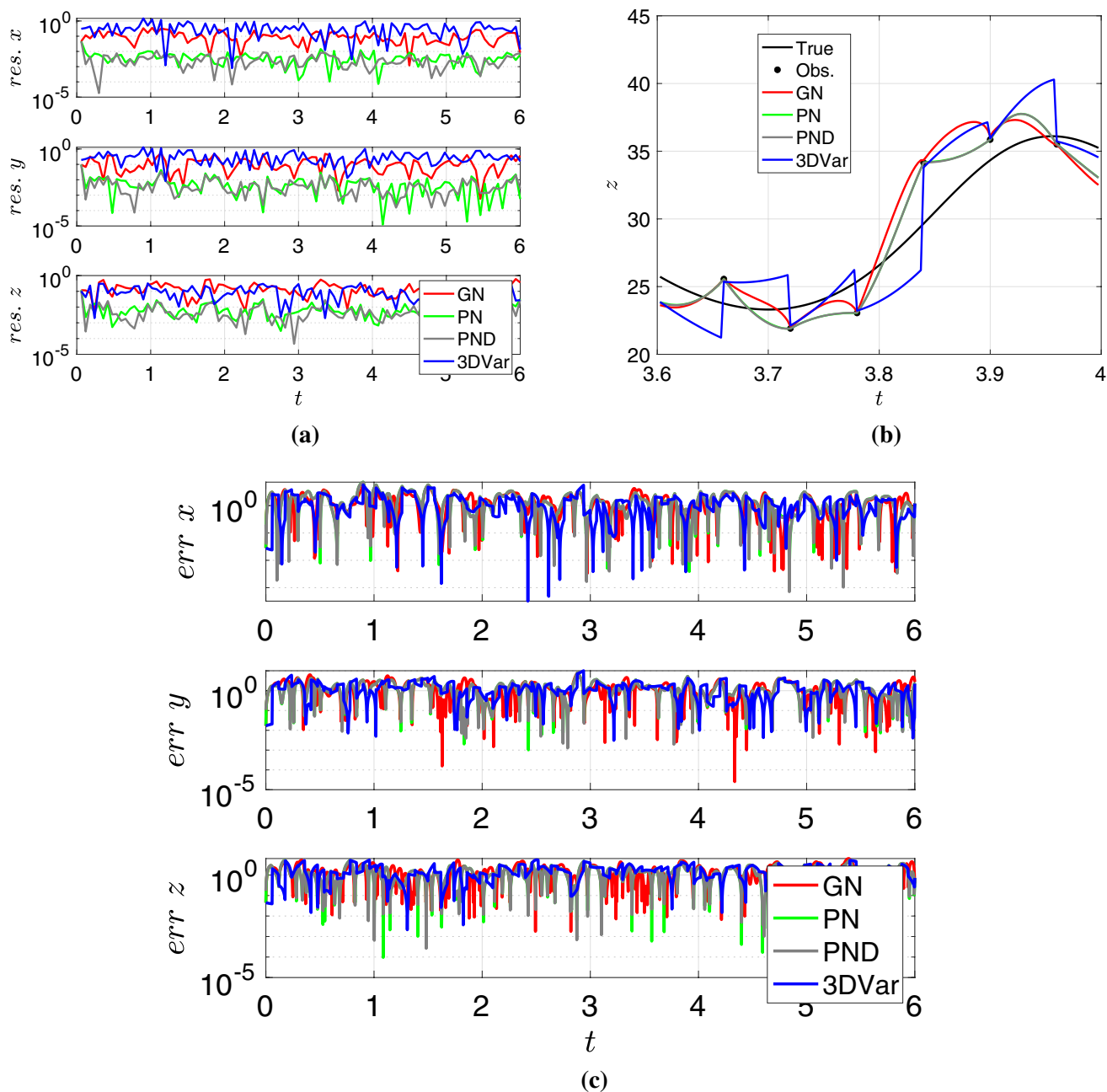
Fig. 7 As in Fig. 5 but for E_-

the zero noise limit, namely the PND approximation. A consequence of this derivation is that in the final system of equations all the physical parameters are fixed by the model and there is no need to tune any ad hoc free physical parameters such as the relaxation time scales. The deterministic approximation, PND, that has good performance with respect 3DVar/PN in the experiments taken into account, also solves a major issue of the PN scheme, that is the computational cost. Indeed, PND does not require the evolution of any ensemble, and the implementation

of the new nudging terms is straightforward. The PND approximation is then a good candidate scheme to be taken into account for problems involving nudging equations.

We also addressed another weakness of the derived scheme, that is the perfectly observed scenario assumption. In real world problems it is unthinkable to have a maximally observed system at certain time interval. However, the approximations of (11), formed the basis of an extension of the tool to non-maximally observed systems. This required a further assumption of slow variability of the studied system.

Indeed the physical nudging corrections, also in the two approximations described above, involve the information of the whole state vector at $t = t_f$. To overcome this issue, the physical nudging corrections, in this case applied only to the observed components, can be computed exploiting both the observations and the mean values computed in the forecast evolution window for the non-observed state vector components. This procedure can be seen as a sort of regularization as the one used, for example, in the variational methods. Since the number of observations are usually much less that the number of the state vector components studied, the data assimilation problems are usually ill posed. For this reason empirical regularization techniques, as the introduction of the background state and the background covariance matrix in the variational cost functions are needed. In this sense our correction method, that exploits the average information for each component in the last forecast window, can be thought of as a background regulator.

It is important to remark that no uncertainties for the observations were taken into account. A straightforward extension, at least formally, to the observational error treatment was proposed. Assuming that the observations posses an error, the final condition of the Kolmogorov backward equation is no longer a delta function, and the backward probability distribution could be expressed as an integral of the product between the distribution that describe the observations and the approximated backward probability derived by means of the small time approximation. Although the procedure is simple, unless we manage particular situations, in which for example the dynamic is governed by a linear symmetric operator and the uncertainty of the observations is Gaussian, an analytical solution can not be easily found.

Then, the approximations found were compared with a 3DVar by performing twin experiments using low-dimensional models that exhibit chaotic behaviour, the Lorenz (Lorenz 1963) and the Molteni (Molteni et al. 1993) systems.

For the first three experiments, based on the L63 model, all the model parameters were kept fixed, but the number of the observed components was different. In the first experiment, E_1, we considered a fully observed system, while in the second experiment, E_2 we observed only the y and z components. The first component, x, was not constrained by the observations but it was constrained directly by y, that appears in the first equation of the system (25), and indirectly by z that does not appear in the first equation if not only through y. Finally, we explored the case, E_3, in which the system is observed minimally in contrast with E_1. In this experiment, we observed only the z component. This means that the y component was constrained directly by the presence of z into the second equation, although there were no direct observations for this second component, and the x component was constrained only indirectly by the observations.

As summarized in Table 1, the PND approximation and the 3DVar scheme have comparable performances. The PN scheme was not considered in the table because it gave similar results, up to two digits, to PND. Looking at the rmse ratio in the table it emerges that, for the experiment E_1, PND perform a little bit better than 3DVar in the assimilation of the y and z components, while it is always better than 3DVar in the experiment E_2 where the x component was not observed. Although the behaviour in time of the error between the true and the analysis trajectories shows that the different schemes are comparable, the analysis residuals considered for E_1, the maximally observed system, show that at the assimilation time the PND analysis trajectories are closer to the observations with respect to the 3DVar trajectories. Clearly, if the observations have large errors, the PN/PND schemes tend to push away the analysis trajectory from the truth, because the observations drive the evolution in a stronger way than 3DVar. In all cases the trajectories of these new nudging schemes seem to be smooth, while the variational method, implemented using a conjugate gradient minimization, exhibits strong corrections that bring out a clear seesaw structure in the trajectories. The first approximation of (11) that does not use the dynamical information, that is GN, has poorer performance than PN/PND. In the last experiment, E_3, in which only the z component was observed no one of the schemes used were able to mimic the true solution of the system. The experiment E_2, that mimics a non-maximally observed system, suggests that the regularization procedure introduced, based on the slow variability assumption, can be used in order to manage non-maximally observed systems, extending the use of the developed tool that was strictly derived for maximally observed system.

In the last experiment, E_4, we compared the performance of PN/PND against the 3DVar in three different configurations of the M93 model, namely E_+, E_0 and E_-, characterized by different values of w^*. The three configurations differ for the shape of the strange attractor generated and then in the probability density function describing the systems. The aim of this example was the investigation of the PN/PND behaviour applied to the L63 model forced by a system evolving with a different temporal scale. M93 is a conceptual model that simplifies the ocean-atmosphere interaction. In particular w^* can be thought of as a parameter that characterizes the Pacific sea surface temperature anomaly and then, E_+, E_0 and E_- represent the atmosphere forced by an ocean in an El Niño state, a neutral state, and in a La Niña state. For these three experiments we only partially considered observed systems, in particular we observed only the L63 components, x, y and z. As before, the performance of the different schemes are comparable, but PN/PND perform better than 3DVar in the assimilation of the y and z components, as shown in Table 2. The "oceanic forcing", generally improves the assimilation of the PN/PND schemes with respect to

3DVar, in particular in the case of La Niña experiment, E_-, and increases the seesaw pattern in the 3DVar analysis with respect to E_1. Also this experiment suggests that, by means of the regularization procedure, the derived tools can be used for partially observed system. Although the w and v components are introduced in the new nudging terms as averaged quantities, computed during the last forecast evolution, the results are encouraging. The analysis residuals of the PN/PND methods correspond to the innovations, but this is not true for 3DVar for which the trajectories before the correction at the assimilation time tend to diverge from the truth. For this reason, the innovations of 3DVar are a few orders of magnitude larger than the one for PN/PND. It should be noted that, although as before the assimilation of the z component alone is not possible, in the new forced system PN/PND no longer produce trajectories that leave the attractor (it is not shown here).

We note that the derivation did not take into account possible bias of the model that would also affect the nudging system. The 3DVar scheme used in this study also does not take the bias into account but only the errors in the initial condition. Further investigation using a bias-aware/bias-corrective scheme is needed with care to correctly attribute a detected bias to its source otherwise which may degrade the solution (Dee 2005). It is also needed to explore fully the capability of the physical nudging scheme against more advanced techniques such as weak-constraint 4DVar or small-time-approximated 4DVar. The latter is shown to improve the analysis as much as the former one in certain circumstances (Carrassi and Vannitsem 2010). Parameter estimation through state-augmentation can also be considered to reduce the model errors. The small-time approximation that we adopt in this work helps to limit the error growth due to the uncertainty in the models in a reasonable range, especially in the non-maximally observed system experiments, as demonstrated by the experiment E2. However, it clearly has a limit in a chaotic dynamical systems such as the L63 when the observations does not constrain the unobserved variables directly (as in E3) through model equations. These aspects will be further investigated in the future works.

Finally, the relation between physics and physical nudging pave the way to other approximations, for example considering more terms in the small time expansion for the new forcing potential. In order to have a time dependent estimate of Q another possible approximation could exploit the Koopman Operator, that is suitable also for more realistic systems. In future works we will further explore the technique using more complex high dimensional models.

References

Abarbanel HDI, Shirman S, Breen D, Kadakia N, Rey D, Armstrong E, Margoliash D (2017) A unifying view of synchronization for data assimilation in complex nonlinear networks. Chaos 27(12):126–802

Anderson JL, Anderson SL (1999) A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. Month Weather Rev 127(12):2741–2758

Asch M, Bocquet M, Nodet M (2016) Data assimilation: methods, algorithms, and applications. SIAM

Auroux D, Blum J (2005) Back and forth nudging algorithm for data assimilation problems. Compt Rendus Math 340(12):873–878

Auroux D, Blum J (2008) A nudging-based data assimilation method: the back and forth nudging (bfn) algorithm. Nonlinear Process Geophys 15(2):305–319

Benseny A, Tena D, Oriols X (2016) On the classical Schrödinger equation. Fluctuat Noise Lett 15(03):1640011

Carrassi A, Vannitsem S (2010) Accounting for model error in variational data assimilation: a deterministic formulation. Month Weather Rev 138(9):3369–3386

Dee DP (2005) Bias and data assimilation. Q J R Meteorol Soc 131(613):3323–3343

Delarue M, Koehl P, Orland H (2017) Ab initio sampling of transition paths by conditioned Langevin dynamics. J Chem Phys 147(15):152–703

Duan J, Nadiga BT (2007) Stochastic parameterization for large eddy simulation of geophysical flows. Proc Am Math Soc 135:1187–1196

Dubois MA (1999) Testing asynchronous coupling on simple "ocean-atmosphere'' dynamic systems. Clim Dyn 15:1–7

Farrell BF, Ioannou PJ (2007) Structure and spacing of jets in barotropic turbulence. J Atmos Sci 64(10):3652–3665

Gardiner C (1985) Handbook of stochastic methods for physics, chemistry, and the natural sciences. In: Proceedings in life sciences. Springer, Berlin

Goodliff M, Amezcua J, Leeuwen PJV (2015) Comparing hybrid data assimilation methods on the Lorenz 1963 model with increasing non-linearity. Tellus A 67(1):26–928

Harlim J (2017) Model error in data assimilation. Cambridge University Press, Cambridge, pp 276–317

Hoke JE, Anthes RA (1976) The initialization of numerical models by a dynamic-initialization technique. Month Weather Rev 104(12):1551–1556

Jazwinski A (1970) Stochastic processes and filtering theory No. 64. Mathematics in science and engineering. Academic Press, New York

Kalman RE (1960) A new approach to linear filtering and prediction problems. Trans ASME J Basic Eng 82:35–45

Kalnay E (2002) Atmos model. Data assimilation and predictability. Cambridge University Press, Cambridge

Kistler RE (1974) Astudy of data assimilation techniques in an autobarotropic primitive equation channel model. M.S. thesis, Department of Metereology, Penn State University

Le Dimet F-X, Talagrand O (1986) Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. Tellus A 38A(2):97–110

Lei J, Bickel P (2011) A moment matching ensemble filter for nonlinear non-gaussian data assimilation. Month Weather Rev 139(12):3964–3973

Lorenz EN (1963) Deterministic nonperiodic flow. J Atmos Sci 20(2):130–141

Luenberger DG (1964) Observing the state of a linear system. IEEE Trans Military Electron 8(2):74–80

Molteni F, Ferranti L, Palmer TN, Viterbo P (1993) A dynamical interpretation of the global response to equatorial pacific sst anomalies. J Clim 6(5):777–795

Mori H (1965) Transport, collective motion, and Brownian motion*. Prog Theoret Phys 33(3):423–455

Navarra A, Tribbia J, Conti G (2013) Atmosphere-ocean interactions at strong couplings in a simple model of el niño. J Clim 26(23):9633–9654

Navarra A, Tribbia J, Conti G (2013) The path integral formulation of climate dynamics. PLOS ONE 8(6):1–16

Palmer TN (1993) Extended-range atmospheric prediction and the Lorenz model. Bull Am Meteorol Soc 74(1):49–66

Risken H, Haken H (1989) The Fokker–Planck equation: methods of solution and applications, 2nd edn. Springer, Berlin

Srinivasan K, Young WR (2012) Zonostrophic instability. J Atmos Sci 69(5):1633–1656

Stauffer DR, Bao J-W (1993) Optimal determination of nudging coefficients using the adjoint equations. Tellus A 45(5):358–369

Stauffer DR, Seaman NL (1990) Use of four-dimensional data assimilation in a limited-area mesoscale model. Part I: Experiments with synoptic-scale data. Month Weather Rev 118(6):1250–1277

Tandeo P, Coauthors, (2015) Combining analog method and ensemble data assimilation: application to the Lorenz-63 chaotic system. Machine learning and data mining approaches to climate science. Springer, Berlin, pp 3–12

Vidard PA, Le Dimet F-X, Piacentini A (2003) Determination of optimal nudging coefficients. Tellus A 55(1):1–15

Zinn-Justin J (2002) Quantum field theory and critical phenomena. International series of monographs in physics, 4th edn. Clarendon Press, Oxford

Zou X, Navon IM, Ledimet FX (1992) An optimal nudging data assimilation scheme using parameter estimation. Q J R Meteorol Soc 118(508):1163–1186

Zwanzig R (1961) Nonlinear generalized Langevin equations. Lect Theoret Phys 3:106–141

Zwanzig R (1973) Statistical mechanics of irreversibility. J Stat Phys 9:215–220