Nonstationarity in Multifactor Models of Discrete Jump Processes, Memory, and Application to Cloud Modeling

ILLIA HORENKO

Universita della Svizzera Italiana, Lugano, Switzerland

(Manuscript received 7 October 2010, in final form 17 December 2010)

ABSTRACT

Because of the mathematical and numerical limitations, standard statistical methods known from the literature are not applicable to inferring jump processes under exogenous influence. Such processes can be considered, for example, in the atmosphere (transitions between different cloud types) and in the ocean (phase transitions between water and ice). Reasons for these intrinsic limitations of standard methods are investigated and a method for the inference of discrete microscopic jump models based on macroscopic ensemble observations is presented. It significantly extends the recently developed methods of nonstationary Markov model parameterization (which are constrained to a single exogenous factor and to direct individual observations of the jump process realizations). The main advantage of the new method is the possibility of inference from indirect ensemble observations with multiple exogenous factors. Moreover, this method allows for a new possibility to test whether the available time series is best described via stationary or nonstationary and Markovian (with memory) or independent (without memory) processes. It also allows estimation of the relative significance of the exogenous factors and their impact on the jump probabilities. The new framework provides a unified toolkit for data analysis of jump processes with the same level of detail now possible for standard continuous state space tools. The resulting numerical algorithm is applied to analysis of the total relative cloud cover data in the midlatitudes and in the tropics under the influence of some meteorologically relevant local and global exogenous factors.

1. Introduction

Discrete probabilistic models are becoming increasingly important mathematical instruments helping to enhance our understanding of many important phenomena in atmospheric science and climate research. Homogeneous (stationary) Markov processes with discrete state spaces and time-independent probability transition matrices, switching between a finite number of regimes, represent one of the most widely exploited families of processes in this class of probabilistic models. Some of the recent applications of stationary discrete Markov processes (or Markov chains) in the area of turbulence mechanics and atmospheric science include the Markovian subgrid-scale models (Crommelin and Vanden-Eijnden 2006, 2009), models of transitions between different atmospheric circulation regimes (Majda et al. 2006; Franzke et al. 2009; Horenko et al. 2008a,b), and stochastic multicloud models (Khouider et al. 2003; Majda et al. 2006; Khouider et al. 2010). However, in most of the cases of geophysical data analysis, discrete transitions between the quantities of interest may be influenced by exogenous (or external) macroscopic factors, such as temperature, circulation regime, etc. Strictly speaking, because of that they cannot be treated by standard tools of jump process analysis (i.e., homogeneous Markov chains).

Most of the available time series data from such processes share the four following properties: (i) the data have only a short record since observations are usually available only on some relatively short time intervals; (ii) the underlying dynamics usually cannot be considered to be independent of some external factors and cannot be assumed stationary (i.e., in the Markovian case, the transition matrix cannot a priori be assumed to be time independent); (iii) it is not a priori clear whether the available data are more adequately represented via the probabilistic model without any memory on the previous realizations (as for independent identically
distributed random variables; e.g., Gaussian processes) or the memory is an important issue (as in the case of Markov processes); and (iv) direct observations of individual processes (such as time-resolved dynamics of the single cloud in the atmosphere) are not available and only the observables for the ensembles of such systems (such as relative cloud cover over a certain geographic area) can be observed.

Because of the implicit structural mathematical limitations associated with the preservation of the stochasticity of the Markov transition matrix (i.e., that the row sums are all equal to one and that all elements are non-negative for all possible allowed combinations of external factors), standard methods of statistical inference and regression time series analysis, such as continuous processes including autoregression (AR), autoregression with exogenous inputs (ARX), vector autoregression (VAR), vector autoregression with moving average (VARMA), multivariate autoregression with exogenous inputs (VARX), autoregressive conditional heteroscedasticity (ARCH), generalized autoregressive conditional heteroscedasticity (GARCH), and their modifications known from the literature (Box and Jenkins 1976; Brockwell and Davis 2002) (and available in the major statistical software packages such as MATLAB, S, and R) are not applicable for inference of jump processes.

This paper describes a unified numerical framework for data-based inference of jump-process Markovian (with one-step memory) and independent (with no memory) models under the influence of multiple exogenous factors. In the context of cloud modeling, such models will describe probabilistic transitions between a fixed number of cloud types (e.g., “cloud” and “no cloud” states) at the microscopic scale (by contrast with the macroscopic scale, where the averaged observation quantities such as the total cloud cover over a certain area are recovered). The microscopic dynamics is considered as a jump process under the influence of certain local and global external factors. In contrast to previous work (Khoudier et al. 2003; Majda et al. 2006; Khoudier et al. 2010), where such a Markov chain model was elegantly but “manually” designed based on the physical properties of the atmospheric system, in the presented framework the Markov assumption is not implied a priori but is tested a posteriori, based on the available observation data. The optimal data-based cloud model (independent or Markovian, stationary or nonstationary, with or without certain external factors) is then estimated applying the locally stationary multifactor jump-process estimation technique elaborated in the present work, which is embedded in the nonstationary data-analysis framework developed by the author in the previous papers (Horenko 2010a,b). Information theory is deployed to allow for the statistical discrimination between the models; that is, the optimal model for given observational data is inferred via a continuous minimization of the regularized variational parameter functional combined with the discrete minimization of the number of involved model parameters.

Moreover, the technique for multifactor jump-process parameterization developed in the paper may be applied repeatedly to results of the primal data analysis (e.g., to the inferred persistent transition process that governs the parameter change of transition probability matrices). Since this process is again derived in the form of the discrete probability distribution time series, the introduced framework allows us to infer its regressive discrete jump model and to construct self-contained coupled discrete nonstationary models that can be applied online. This idea also opens the way of applying the presented method to postprocessing and prediction of the cluster affinities obtained by other clustering and time series analysis methods such as finite element method (FEM)-VARX (Horenko 2010b), K-means, fuzzy C-means (Bezdek 1981; Christiansen 2007), and other techniques used for the analysis of time series data in geophysics.

This paper begins with a description of the multifactor discrete stochastic model for indirectly observed jump processes, followed by the combination of the introduced concepts with the approach based on nonstationary clustering and information theory context introduced in the previous works (Horenko 2010a,b). The final section deals with an application of the presented techniques to the analysis of National Centers for Environmental Prediction (NCEP) total cloud cover resimulation data subject to different local and global quantities such as temperature, pressure, and humidity; their advection and temporal tendencies and global circulation indices such as the Arctic Oscillation (AO), Antarctic Oscillation (AAO), and Madden–Julian oscillation (MJO). The time series analysis based on the presented methodology is conducted for two different geographical locations: one in the midlatitudes and one in the tropics. The obtained self-contained coupled models are interpreted and compared with respect to their online prediction quality.

2. Methods and approach

a. Derivation of the stochastic master equation

In the following we will consider dynamical systems that are discrete in time and space, that is, \( \{X_t\}_{t=1,\ldots,T} \), taking values from some fixed set of \( m \) distinct quantities \( s_1, \ldots, s_m \) and are dependent on a vector of external factors \( \mathbf{u}_t = (u_{t-1}, \mathbf{u}_{t-1}^{\text{ext}}) \), where \( u_{t-1} \) is a \( k \)-dimensional...
vector of the known (or explicit) external factors and $\mathbf{u}_{t-1}^{\text{res}}$ is an $l$-dimensional vector of unresolved (or implicit) external factors acting on the process $X_t$ at time $t$. The jump process underlying the observations is called Markovian if the probability $P$ of any current realization state $X_t$ of the process at time $t$ depends only upon the previous state $X_{t-1}$ and a value $\mathbf{u}_{t-1}$ of all external factors at time $t-1$ and does not depend on any other previous state. Mathematically this property can be expressed as

$$P\left[X_t = s_j | X_{t-1} = s_i, \mathbf{u}_{t-1} \right] = P_{ij}(\mathbf{u}_{t-1}),$$

which denotes the probability of transition from state $s_i$ to state $s_j$ in one step for a fixed value $\mathbf{u}_{t-1}$ of all explicit and implicit external factors. These probabilities can be put together into an $m \times m$ stochastic transition matrix $P(\mathbf{u}_{t-1})$, that is, $\sum_{j=1}^{m} P_{ij}(\mathbf{u}_{t-1}) = 1$ for any $i$ and $\mathbf{u}_{t-1}$ as functions of time belong to some class of functions, namely

$$\{ \mathbf{u}_{t-1} \}_{t=1,...,T} \in \mathcal{U}$$

(where $\mathcal{U}$ is in the following assumed to be a bounded set of allowed factor functions, representing all possible external influences for the considered process $\{X_t\}_{t=1,...,T}$). The left panels of Fig. 1 exemplify some of the most commonly allowed factor sets $\mathcal{U}$ such as independent (top left, rectangular domain $\mathcal{U}$) or dependent factor sets (bottom left).

In the following, we will consider a case when the time series of the single individual process $\{X_t\}_{t=1,...,T}$ is not available and only an ensemble of such systems can be observed. In such a case the observable quantities are, for example, the total numbers $n_i(t)$ of the ensemble members being in the same state $s_i$ at time $t$. If the total number of the observed ensemble members is $n = \sum_{i=1}^{m} n_i(t)$, then the corresponding discrete empirical state probabilities $\pi_i^{(n)}(t)$ are computed as

$$\pi_i^{(n)}(t) = \frac{n_i(t)}{n},$$

and, obviously,

$$\sum_{i} \pi_i^{(n)}(t) = 1$$

As the number of the ensemble members grows, this quantity approximates the discrete state probabilities $\pi_i(t)$.
\[ \pi_i(t) = \mathbb{P}[X_i = s_i, \mathbf{u}_{i-1}] = \lim_{n \to \infty} \pi_i^{(n)}(t), \quad (4) \]

and for strictly Markovian systems the dynamics of the state probability distribution vector \( \pi(t) = [\pi_1(t), \ldots, \pi_m(t)] \) is described by the discrete analog of the Fokker–Planck equation, namely by the master equation (Gardiner 2004)

\[ \pi(t + 1) = \pi(t) \bar{P}(\mathbf{u}_t). \quad (5) \]

Let us assume that the Markovian transition matrix is continuously differentiable and has bounded second derivatives with respect to (w.r.t.) the unresolved external factors \( \mathbf{u}_t^{\text{unres}} \). Then, in the neighborhood of the statistical equilibrium of the unresolved degrees of freedom \( \mu_i^{\text{unres}}(t) = \mathbb{E}[\mathbf{u}_t^{\text{unres}} | \mathbf{u}_t] \), a Taylor expansion of the function \( \bar{P}(\mathbf{u}_t) \) can be written as

\[
\bar{P}(\mathbf{u}_t) = \bar{P}[\mathbf{u}_t, \mu_i^{\text{unres}}(t)] \\
+ \sum_{i=1}^l \frac{\partial \bar{P}}{\partial \{\mathbf{u}_t^{\text{unres}}\}_i} \left[ (\mathbf{u}_t^{\text{unres}})_i - \mu_i^{\text{unres}}(t) \right] \\
+ O\left[ ||(\mathbf{u}_t^{\text{unres}})_i - \mu_i^{\text{unres}}(t)||^2 \right]. \quad (6)
\]

If the factors \( (\mathbf{u}_t^{\text{unres}})_i - \mu_i^{\text{unres}} \) are statistically independent in \( i \) and \( t \), have zero expectation, and their number \( l \) is large enough, then the central limit theorem for independent processes (Feller 1974) will guarantee that

\[ \bar{P}(\mathbf{u}_t) = P(t, \mathbf{u}_t) + \epsilon_t, \quad (7) \]

where \( P(t, \mathbf{u}_t) = \bar{P}[\mathbf{u}_t, \mu_i^{\text{unres}}(t)] \) is now a time-dependent (or nonstationary) stochastic matrix and \( \epsilon_t \) is a matrix-valued constrained \( m \times m \)-dimensional stationary independent and identically distributed (i.i.d.) Gaussian random process with \( \mathbb{E}[\epsilon_t] = 0, \forall t \). Assuming that the number of ensemble members \( n \) is large enough, we insert the observed discrete probability distribution \( \{\pi^{(n)}(0), \ldots, \pi^{(n)}(0)\} \) into (5), and deploying (7) we get the following stochastic version of the above master equation:

\[ \pi^{(n)}(t + 1) = \pi^{(n)}(t) P(t, \mathbf{u}_t) + \epsilon_{t+1}, \quad (8) \]

where \( \epsilon_{t+1} = \pi^{(n)}(t) \epsilon_{t+1} \) is now a vector-valued random process. It is easy to verify that the new stochastic process \( \epsilon_t \) is again i.i.d. and has a zero expectation and a bounded variance. Therefore, the Gauss–Markov theorem (Markov 1912) can be deployed to estimate the transition matrix \( P^{opt}(t, \mathbf{u}_t) \) based on the observation time series \( \{\pi^{(n)}(0), \ldots, \pi^{(n)}(0)\} \); that is,

\[ P^{opt}(t, \mathbf{u}_t) = \arg \min_{P(t, \mathbf{u}_t)} \sum_t g[\pi^{(n)}(t + 1), \pi^{(n)}(t), P(t, \mathbf{u}_t)], \quad (9) \]

where \( g[\pi^{(n)}(t + 1), \pi^{(n)}(t), P(t, \mathbf{u}_t)] = ||\pi^{(n)}(t + 1) - \pi^{(n)}(t) P(t, \mathbf{u}_t)||^2 \). However, the inverse problem (9) is ill posed since the amount of unknowns is much higher than the amount of the known information [i.e., the available time series \( \{\pi^{(n)}(0), \ldots, \pi^{(n)}(0)\} \)]. Moreover, additional constraints should be imposed on (9) because the stochasticity of the transition matrix should be preserved for all times and all allowed combinations of control variables \( \mathbf{u}_t \). As will be demonstrated later in the text, fulfillment of these constraints represents the main numerical bottleneck since their amount grows linearly in time and exponentially with the dimensionality of the control vector \( \mathbf{u}_t \). The standard approach to this problem will be to assume the time independence (stationarity) and control independence—in other words, the homogeneity—of the transition matrix. But this is obviously an oversimplification in the cases when the influence of the control variables is significant. Before we approach the numerical solution of the problem (9), we will consider an important scenario of the above model for the cases in which the underlying discrete jump dynamics has no memory.

b. Special case of the Markov model: Processes with independent realizations

In many practical applications the considered processes are memoryless; that is,

\[ P[X_t = s_j | X_{t-1}, \ldots, X_1] = s_j | \mathbf{u}_t, \mathbf{u}_{t-1}, \ldots, \mathbf{u}_1] \]

\[ = P[X_t = s_j | \mathbf{u}_{t-1}] = \mu_j(t - 1, \mathbf{u}_{t-1}). \]

This processes are called processes with independent realizations (in a nonstationary case) or i.i.d. processes (in a stationary case) and \( \mu(t, \mathbf{u}_t) = [\mu_1(t, \mathbf{u}_t), \ldots, \mu_m(t, \mathbf{u}_t)] \) is the corresponding discrete probability distribution function. The most widely used representatives from this family are the Gaussian white noise process and the Poisson process. If

\[ P_{ij}(t, \mathbf{u}_t) = P_{ij}(t, \mathbf{u}_t) = \ldots = P_{mij}(t, \mathbf{u}_t) = \mu_j(t, \mathbf{u}_t) \]

for all \( j \in \mathbf{J}, t, \mathbf{u}_t \in \mathcal{U} \)

(10)

(where \( \mathbf{J} \) is the set of states with independent dynamics; \( \mathbf{J} = \{1, \ldots, m\} \) if all states are independent), then by applying (3) and (8) is easy to verify that the stochastic

---

1 Application of the central limit theorem at this point also relies on the mild assumption that the Lindberg condition (Feller 1974) is fulfilled. This assumption is a weakest form that expresses the requirement that the probability for \( ||(\mathbf{u}_t^{\text{unres}})_i - \mu_i^{\text{unres}}|| \) to be very large is very small.
master equation for the processes with independent realizations has a form

\[ \pi^{(n)}(t + 1) = \mu(t, u_i) + \epsilon_{t+1}. \]  
(11)

Above, (11) is a special case of Markovian stochastic master equation (8) if the constraints (10) are fulfilled. This property will be used later in this paper to formulate and to apply a numerical criterion for discrimination between Markovian and independent models based on observation time series. It also shows that the parameter identification for both Markovian and independent families can be done in a unified way, deploying the universal form (8). The distinction can be done by switching on and off the constraints (11). This issue will be elaborated further below; the structural difference between the i.i.d. and the Markov models with external factors is graphically represented in Fig. 2.

Note also that the formalism (10) and (11) opens the way to description of the mixed Markovian-independent models that are independent in certain states \( J \) and are Markovian in all other states.

c. Inverse problem and information criterion:  
Extension to multifactor Markov models

To estimate the a priori unknown transition probabilities \( P_j(t, u_i) \) from the observed time series of empirical probability distributions \( \pi^{(n)}(t) \) and external factors \( u_i \), an inverse problem can be formulated as a minimization of the least squares functional

\[ L[P(\cdot, \cdot)] = \sum_{t=1}^{T-1} g[\pi^{(n)}(t + 1), \pi^{(n)}(t), P(t, u_i)] \rightarrow \min, \quad P(\cdot, \cdot) \]  
(12)

where \( g[\pi^{(n)}(t + 1), \pi^{(n)}(t), P(t, u_i)] \) is an appropriate functional describing a distance between the observed data \( \pi^{(n)}(t + 1) \) and their one-step predictions based on the model (8) and the data at time \( t \). To guarantee that the resulting transition matrix \( P \) is stochastic, problem (12) should be subject to the constraints

\[ \sum_{j=1}^{m} P_j(t, u_i) = 1 \quad \text{for all} \quad t, u_i, i, \]  
(13)

\[ P_j(t, u_i) \geq 0, \quad \text{for all} \quad t, i, j \quad \text{and} \quad \{u_i\}_{t=1,...,T} \in \mathcal{U}. \]  
(14)

Note that the above parameter constraints are a characteristic feature of the Markovian jump processes with discrete state spaces and are in general not present in the context of the space-continuous system [e.g., for VARX models, cf. section 2a of Horenko (2010b)]. It is straightforward to verify that the global optimizer \( P_{\text{opt}}(t, \cdot) \) of the problem (12)–(14) at time \( t \) \{with \( L[P_{\text{opt}}(t, \cdot)] = 0 \)\} in such a case will be nonunique and independent of the external factors, resulting in a meaningless (from the practical point of view) estimation. The second, practical and numerical difficulty, is that the number of constraints (13) and (14) in the above minimization problem (12) is growing linearly in \( t \) and exponentially in the dimensionality of \( u_i \). This fact can best be illustrated...
considering the quadratic or rectangular domain $\mathcal{U}$: if per every dimension of $\mathbf{u}$, the number of allowed values at time $t$ is $N(t)$, then the total number of inequality constraints will be $O[\sum N(t)^k]$. It means that the numerical cost of the optimization procedure—for example, the interior point method (Nocedal and Wright 2000), which is one of the most efficient and frequently used constrained minimization methods)—will scale as $O\{\exp[\sum N(t)^k]\}. In the previous work (Horenko 2009) it was demonstrated how this second problem can be solved by replacing (13) and (14) via a set of equivalent time-independent constraints in the case of a single factor model (i.e., for $k = 1$). However, in the case of multiple external factors, the procedure described in Horenko (2009b) does not work. One of the key enhancements of the current work is that it demonstrates how the above parameter constraints can be substituted by equivalent time-independent constraints in the case of multiple external factors (i.e., for $k > 1$).

In the following, three steps of the regularization procedure will be described. It will be demonstrated how the set of constraints (13) and (14) can be approximated via a convex set of time-independent equations.

1) FIRST STEP: INTERPOLATION

Along the lines first introduced in (Horenko 2009, 2010b), we will assume that for any $t \in [1, T]$ the scalar-valued model distance functional $g$ from (12) can be represented as a convex linear combination of

$$1 \leq K \leq K_{\text{max}}$$

stationary model distance functionals—that is, model functionals dependent on time-dependent Markov models $P(t, u_i)$ of the following multifactor form:

$$P(t, u_i) = P(0) + \sum_{l=1}^{K} P(l)u_i^l, \quad i = 1, \ldots, K,$$

$$g[\ldots, P(t, u_i), \ldots] = \sum_{l=1}^{K} \gamma(t)g[\ldots, P(t, u_i), \ldots],$$

where $P(0), P(1), \ldots, P(K) \in \mathbb{R}^{n \times m}$ are matrices and $\gamma(t) \in \mathbb{R}$ are the scalar time-dependent model affiliations fulfilling the convexity condition

$$\sum_{l=1}^{K} \gamma(t) = 1, \quad \forall t \in [0, T],$$

$$\gamma(t) \geq 0, \quad \forall t \in [0, T], \quad i = 1, \ldots, K.$$  

Note that the functional $g$ on the left-hand side (lhs) of (17) is the same functional as on the right-hand side (rhs) of this expression—the difference is in argument $\mathbf{P}$ for which it is evaluated: on the lhs it is a time-dependent transition matrix $P(t, u_i)$, whereas on the rhs these are the explicitly time-independent transition matrices $P(i)(u_i)$ (dependent only on the vector of external factors $\mathbf{u}$).

In another words, we assume here that at any $t$ the global time-dependent (or nonstationary) model distance functional can be approximated by a combination of $K$ local Markov model distance functionals chosen according to some time-dependent probabilities (or model affiliations) $\Gamma(t) = [\gamma_1(t), \ldots, \gamma_K(t)]$. This idea for inverse numerical problems is widely used in the context of data clustering (Höppner et al. 1999); the presented general form was introduced in Horenko (2010a) and stems from the classical spline interpolation approach for direct numerical problems (see, e.g., Deuflhard 2004).

The main bottleneck for the direct numerical solution of the inverse problem [(12)–(14)] results from the fact that the number of constraints [(13) and (14)] grows exponentially with the dimension of the control vector $\mathbf{u}$.

This fact is illustrated in Fig. 1 for the case of two external factors: black dots in the two left panels of Fig. 1 represent the allowed values for the pairs of control variables, and the black dots on the two right panels represent the values of matrix elements $\{P(0)(0), P(0)(1), P(1)(0), P(1)(1), \ldots, P(K)(0), P(K)(1), \ldots\}$.

As already mentioned above, the same procedure applied to the set of inequality constraints (14) will be not so straightforward if the number $k$ of external factors is larger than 1.

It turns out (and this is basically the main mathematical result of the current paper) that when applying the induction in the dimensionality of the control variable $\mathbf{u}$, in the case of the rectangular domain $\mathcal{U}$, the set of $O[\sum N(t)^k]$ inequality constraints (14) [where $N(t)$ is the number of allowed values for a single dimension of $\mathbf{u}$ at time $t$; black dots on the top right panel of Fig. 1] is equivalent to a set of $2^K$ linear inequality constraints defined by all possible combinations of supremum and
infinuma of the $k$ external factor components $u^i_l$, $l=1,\ldots,k$ (visualized as the gray planes in the same panel of Fig. 1):

\[
\{P^{(i)}_{(0)}\}_{ke} + \sum_{l=1}^{k} \{P^{(i)}_{(l)}\}_{jn} \left[ \inf_{r \in (1, T]} u^i_l \right] = 0, \quad \text{for all } i,j,n.
\] (22)

Therefore, the numerical optimization of the original inverse problem subject to (13) and (14) will cost $O(\exp[\sum_i N_i])$ operations, whereas its equivalent formulation subject to (20)–(22) will cost only $O(\exp(2^k))$ operations.

As demonstrated in the lower left panel of Fig. 1, if the single dimensions of the vector $u_i$ are not independent of each other, the set $U$ of allowed external factors can build a domain with nonlinear boundaries, resulting in a complex geometrical structure of the set of allowed parameter values $(P_{(0)}, \ldots, P_{(k)})$ that fulfill (13) and (14) (bottom right panel of Fig. 1). In such a case, to avoid the bad numerical scaling of the resulting optimization problem with the dimension of $u$, identification of the set of allowed parameters $\Theta$ and its optimal (if possible, time independent) approximation should be done prior to the solution of the parameter identification. In the case of the nonlinear domain $U$, a set of time-independent inequality constraints can be approximated via a linear convex hull of time-independent constraints, applying standard tools of computational geometry (e.g., the Quickhull algorithm; Barber et al. 1996) (see the top right panel of Fig. 1).

Inserting the ansatz (17) in (12) results in the minimization of the averaged clustering functional

\[
\mathbf{L}[\Gamma(\cdot), \Theta] = \sum_{i=1}^{K} \sum_{t=1}^{T-1} \gamma_i(t) g[\cdot, P_j^i(u_i)] \rightarrow \min_{\Gamma(\cdot)/\Theta},
\] (23)

subject to (18) and (19) and to (20)–(22). Direct numerical treatment of the problem (23) is still hampered by the ill posedness.

In the next step, to handle the nonstationary $y$-dependent part of the optimization problem, concepts of bounded variation (BV) regularization and FEM clustering introduced in the previous work will be shortly described in the new context of discrete jump-process parameterization. More details on this issue can be found in Horenko (2010b).

2) SECOND STEP: PERSISTENCE

Many processes in the real world have multiscale character (e.g., for inverse problems very often it is reasonable to assume that the change of the model parameters is much slower than the change of the observed data). In other words, the dynamics of model parameters [i.e., the Markov matrix $P(t, u_i)$ in our case] can be assumed to be slow or persistent w.r.t. the first argument $t$. This means that in the context of (23), a persistency condition for the cluster affiliation function $\Gamma_i(t)$ (switching between the local model parameters) in the time-discrete bounded variation sense can be imposed

\[
|\gamma_i|_{BV(1,T)} = \sum_{t=1}^{T-1} |\gamma_i(t+1) - \gamma_i(t)| = \|D\gamma_i\|_1 \leq C,
\] (24)

where the persistency parameter $C$

\[
1 \leq C \leq T
\] (25)

measures the maximal number of transition between the local model $i$ and all other models in the time interval $(1, T)$. $\gamma_i = [\gamma_i(1), \ldots, \gamma_i(T)] \in \mathbb{R}^T$, $\dagger$ is the transposition operation, and $\| \cdot \|_1$ is the 1-norm (Horenko 2010b). Condition (24) considered together with other constraints (18), (19), and (20)–(22) allows one to find all those solutions of the problem (23) that are more persistent than $C$ in the BV sense.

3) THIRD STEP: MODEL DISCRIMINATION

In addition to the regularity problems described above, several other problems arise when dealing with realistic systems. First of all, there remains a problem of identification of the optimal parameters $K$ and $C$ used in (23). Since Markov models (16) with $k$ external scalar factors are used in the parameterization, another challenge will be the data-based identification and comparison of their relative significance and identification of redundant external factors $u_i$ out of a fixed countable set of possible factors $U$. Also, as was already mentioned above, it is not a priori clear which states can be assumed to be Markovian and which independent [i.e., what is $J$ in expression (10)]. Another problem will arise if the number $m$ of the system’s states is growing: the number of the unknown Markov parameters will grow as $2m(m - 1)$, resulting in a growing uncertainty of the estimated Markov model parameters $P_i^j_{(0)}$, $P_i^j_{(1)}$ if the number $T$ of observations is kept constant. In other words, two optimal sets of matrix indices $I_0$ and $I_1$ (denoting the matrix elements a priori assumed to be zero)
\[ I_0 = \{ (i_0, j_0) \mid \{ P_{ij}^{(0)} \}_{i,j=0} = 0, \forall i \}, \]

\[ I_1 = \{ l \{ P_{ij}^{(1)} \}_{i,j=1} = 0, \forall i_1, j_1 \}, \] (26)

should also be identified during the data analysis.

In the following, the problem of identification (or, in the language of statistics, discrimination) of the most appropriate model will be discussed. The main aim thereby will be to identify the models that best explain the data with a minimum number of involved parameters in order to avoid the overfitting effects.

Consider the nonnegative scalar values of the model distance functionals \( r_i(t) = g[\cdot, \cdot, P_{ij}^{(0)}(u)] \) calculated in the minima of the (23) for fixed values of \( K, C, J, I_0, \) and \( I_1 \). Given (7), they represent the variances of the cumulative sums of the unresolved independent external factors on the jump process dynamics in the model state \( i \), for example,

\[ r_i(t) = \epsilon_{i+1}^t \epsilon_{i+1}^t = \text{Var}[\epsilon_{i+1}^t]. \] (27)

In the derivation of jump-process model (7), \( \epsilon_{i+1} = \sum_{r=1}^K y_i(t) \gamma_i(t) \) was assumed to be independent in \( t \) and in dimension components. Hence, values of \( \epsilon_{i+1} \) are independent and therefore optimization residuals (27) are scalar i.i.d. processes. The optimal parametric probability distribution family \( \psi_i \) (i.e., Gaussian, exponential, etc.) for each of the scalar processes \( r_i(t) \) can be inferred by applying standard hypothesis tests from the statistics. Then the log-likelihood \( \mathcal{L}(K, C) \) of the residual process is written as a log-likelihood of the following mixture model:

\[ \mathcal{L}(K, C) = \sum_{i=0}^T \log \left\{ \sum_{i=1}^K \gamma_i(t) \psi_i(r_i(t)) \right\}. \] (28)

Defining the number of parameters for each single scalar probability distribution as \( N \) (\( N = 2 \) for Gaussian and \( N = 1 \) for exponential distribution), we can deploy the standard Akaike information criterion (AIC) functional known from the literature (Akaike 1974) in the new structural form

\[ \text{AIC}[\Gamma(\cdot), \Theta, K, C, u, J, I_0, I_1] = -2\mathcal{L}(K, C) + 2[KN + \mathcal{N}(K, C, I_0, I_1, J)], \] (29)

where \( \mathcal{N}(K, C, I_0, I_1, J) \) is a functional measuring the total number of involved jump-process model parameters. Since \( K \) BV functions \( \gamma_i(t) \) fulfilling the persistency constraint (24) can be coded with the help of up to \( (K - 1)(2C + 1) \) parameters [where coding involves the initial values \( \gamma_i(1), i = 1, \ldots, K - 1 \) together with up to \( 2C \) times and values of the \( \gamma_i(t) \) jumps], function \( \mathcal{N} \) can be written in the form

\[ \mathcal{N}(K, C, I_0, I_1, J) = K[2m(m - 1) - J(m - 1) - I_0 \]

\[ - m(m - 1)I_1] + (K - 1)(2C + 1), \] (30)

where \( || \cdot || \) measures the number of elements in the respective sets.

Discrete minimization of the information functional (29) w.r.t. \( K, C, J, \) and \( C \) subject to (15) and (25) can be approached numerically, simultaneously with the continuous minimization of the averaged clustering functional (23), (18), (19), (20)–(22), (18), (19) [continuous minimization algorithm is described in detail described in Horenko (2010b)]. The result will provide the optimal persistent nonstationary and nonhomogeneous inverse model for the analyzed data with a minimal number of involved parameters.

d. Self-contained data-based models and online predictions of jump processes

Discrete-continuous minimization of (29) for a given time series of discrete probability distributions \( \{ \pi^{(0)}(0), \ldots, \pi^{(0)}(T) \} \) (e.g., total cloud cover), subject to a given set of explicitly known exogenous factors \( u_0, \ldots, u_T \), will result in identification of the i.i.d set of states \( J \), statistically insignificant subsets of Markovian transition probabilities \( I_0 \) and subset of irrelevant external factors \( I_1 \), together with the optimal number of local multifactor models \( K \) and optimal persistency \( C \). The continuous part of this minimization will also provide the \( K \) sets of local multifactor model parameters \( \{ P_{ij}^{(0)}, \ldots, P_{ij}^{(n)} \} \) \((i = 1, \ldots, K)\) and \( K \) time-dependent model affiliations \( \{ \gamma_i(t), \ldots, \gamma_k(t) \} \), which according to (5)–(7) describe the net influence of the unresolved external factors on the effective jump-process dynamics subject to the resolved factors \( u \) only. Given this information, the data-based predictive model for the underlying jump process \( X_{t+1} \) will take the form

\[ E[X_{t+1} | X_t, u_t] = E[\pi^{(0)}(t + 1)] \]

\[ = \sum_{i=1}^K \gamma_i(t) \pi^{(0)}(t) \left[ P_{ij}^{(0)} + \sum_{l \in I_1} P_{ij}^{(l)} u_t \right]. \] (31)

To be able to use the above model for online predictions in a self-contained way beyond the time span where the model was originally identified (or, in the statistics language, trained), one needs to solve two problems associated with the model affiliation function \( \Gamma(t) = \{ \gamma_i(t), \ldots, \gamma_k(t) \} \): (i) it is necessary to be able to assign the new values of the vector \( \Gamma(T) \) based on the new
value \( \pi^{(n)}(T + 1) \) and on the previously estimated parameters \( \{ P^{(i)}_0, \ldots, P^{(i)}_k \} \) \( i = 1, \ldots, K \); and (ii) to go beyond the one-step predictions, one needs to construct a dynamical model for model affiliations \( \Gamma(t) \). The brute-force possibility to solve the problem (i) and to get the value \( \Gamma(T) \) if the original time series is updated with a new value \( \pi^{(n)}(T + 1) \) would be to repeat the whole minimization procedure (29) again for the updated time series. Another computationally less expensive way can be taken if we consider the minimization of the updated version of (23) w.r.t. \( \Gamma(T) \) only, that is,

\[
\sum_{i=1}^{K} \sum_{t=1}^{T-1} \gamma_j(t) g[\pi^{(n)}(t + 1), \pi^{(n)}(t), P^{(i)}(u_t)] + \sum_{i=1}^{K} \gamma_j(T) g[\pi^{(n)}(T + 1), \pi^{(n)}(T), P^{(i)}(u_T)] \rightarrow \min_{\Gamma(T)}.
\]

The solution of this minimization problem subject to conditions (18) and (19) for \( \Gamma(T) \) is then given by

\[
\gamma_j(T) = \begin{cases} 1, & i = \arg\min_j g[\pi^{(n)}(T + 1), \pi^{(n)}(T), P^{(i)}(u_T)], \\ 0, & \text{else} \end{cases}
\]

Conditions (18) and (19) guarantee that at any time \( t \) the affiliation vector \( \Gamma(t) \) resulting from the minimization of (29) can be considered as a discrete probability distribution of a certain jump process. Hence, if this discrete process could be assumed to be i.i.d. or Markovian, the dynamics of \( \Gamma(t) \) could then be approximated as a multifactor model

\[
\mathbb{E}[\Gamma(t + 1)] = \Gamma(t) \left( \sum_{i=1}^{K} P^{(i)}_0 + \sum_{i=1}^{K} P^{(i)}(u_T) \right),
\]

where the parameters \( \{ P^{(i)}_0, \ldots, P^{(i)}_k \} \) \( i = 1, \ldots, K \) are inferred by minimizing (23), with \( K = 1 \) and for \( \{ \Gamma(0), \ldots, \Gamma(T) \} \) at the place of the analyzed time series \( \pi^{(n)}(t) \).

Equations (31) and (34), together with the affiliation assignment equation (33), constitute a self-contained dynamical model for nonstationary jump-process prediction. Its inference and applications to the cloud data analysis will be demonstrated in the next section.


In the following we consider as \( \pi^{(n)}(t) \) the daily mean total cloud cover time series (retrieved from http://www.esrl.noaa.gov) for the two geographic locations: one in the midlatitudes (47.5°N, 7.5°E, on the French–Swiss border northern of the Alps) and one in the tropics (7.5°N, 7.5°E, in the rain forest area of Nigeria). Both time series are taken in the time interval between 1 January 1990 and 1 January 2000, with 3653 data points. To illustrate an application of the method described above, the total cloud cover \( \pi_1(t) \) (with the values changing between 0 and 1) is interpreted as the first component of the discrete macroscopic probability density of the ensemble of “microscopic” birth–death cloud processes (i.e., jump processes \( X_i \) switching between two states: “cloud state” 1 and “no cloud state” 2). The second component of the discrete probability density is then simply defined as \( \pi_2 = 1 - \pi_1 \).

As was demonstrated in Khouider et al. (2010), convective available potential energy (CAPE; correlates with temperature) and air dryness can be successfully used as external factors to design the physically motivated discrete Markovian model of cloud dynamics. In the following analysis we will consider a set of the following NCEP time series as explicit external factors: surface temperature \( u_1^1 \), horizontal surface temperature advection \( u_2^1 \), horizontal surface humidity advection \( u_3^1 \), surface pressure \( u_4^1 \), surface pressure tendency \( u_5^1 \), 500-mb vertical velocity \( u_6^1 \), Arctic oscillation index \( u_7^1 \), Antarctic oscillation index \( u_8^1 \), and Madden–Julian oscillation \( u_9^1 \). Whereas the group \( u_1^1 - u_9^1 \) represents the local synoptic factors retrieved for the same geographical locations, factors \( u_7^1 - u_9^1 \) represent the influence of major global teleconnection patterns. To be able to interpret the relative influences of the factors on the transition probabilities of the jump process, all of the external factors are made dimensionless and are scaled, subtracting the respective time averages and dividing with the maximal absolute values of the respective external factor series.\(^2\)

It is a priori assumed that (i) the long term-memory effects are negligible for the local values of the total cloud cover and (ii) the cumulative effect \( \tau_i \) of the unresolved external factors \( u_{i}^{\text{res}} \) is an i.i.d. process with zero expectation. Assumption (i) is supported by the

\(^2\)Since some of the considered factors are available only in the daily averaged form on the National Oceanic and Atmospheric Administration (NOAA) data server, all of the data including the total cloud cover time series were retrieved and analyzed in the daily averaged form. Since the latitude discretization grid of the NCEP total cloud cover is slightly shifted w.r.t. the latitude grids of all other analyzed quantities and since the NCEP daily averaged total cloud data fields exhibit enough of the spatial smoothness, daily 2D spatial interpolation on the sphere is deployed in every time step to retrieve the point values of total cloud cover at the positions of interest.
observed quick equilibration of the process and rapid decay of the autocorrelation function on the considered local scale; assumption (ii) is positively verified a posteriori by the standard statistical tests of independence applied to the model residuals (27). Hence, the stochastic master equation (7) can be considered as an appropriate model of the cloud process dynamics.

An information functional (29) is minimized for both datasets deploying the discrete-continuous minimization procedure described in Appendix A for $\text{tol} = 10^{-8}$ and $N_{\text{SP}} = 50$ with no optimization w.r.t. $I_1$ because of the run-time limitations.\(^3\)

2) RESULTS FOR MIDLATITUDE LOCATION

The top panel of Fig. 3 shows the resulting AIC profiles of (29) for different values of the persistency threshold $C$ and the number of local models $K$, as calculated for the location in midlatitudes. In contrast to the stationary case ($K = 1$), Markovian models [where $J$ from (10) is empty] are not superior to the independent distribution models (where $J = \{1, 2\}$) in the nonstationary case (for $K = 2$). The optimal model is therefore the independent nonstationary data model with $K = 2$, $C = 1250$. The resulting self-consistent model [(31) and (34)] has the form

$$E[X_{t+1}] = \Gamma_1(t)\mu_1(u_t) + \Gamma_2(t)\mu_2(u_t)$$

$$E[\Gamma(t+1)] = \Gamma(t)P^\Gamma(u_t), \quad (35)$$

and the data-based parameterization of $P^{(1)}(u_t)$, $P^{(2)}(u_t)$, and $P^\Gamma(u_t)$ is given in Appendix B [(B1)].

As can be seen from (B1), overall dynamics of the midlatitudes cloud cover can be represented as the Markovian process $\Gamma(t)$ switching between the clear-sky i.i.d. model 1 and a cloudy-sky i.i.d. model 2. The switching process itself is predominantly governed by such local factors as $u_4$ (most significant influence), $u_2$, $u_1$, and $u_6$ (least significant). Among the global teleconnection factors, AO and MJO have the biggest impact on $\Gamma(t)$. For the clear-sky model surface pressure and temperature (i.e., seasonal components) are most significant, and all of the investigated teleconnection factors are insignificant. In the case of a cloudy-sky model, besides the seasonal pressure influence, the local temperature advection plays the most significant role. Among the global factors only the AO plays some role in this model case.

To verify the models derived by the minimization (29), we can check if the (29)-optimal models are also better in their predictive skill compared with other methods. Since, as discussed above, standard techniques of regression analysis [such as continuous VARX($p$) models] are conceptually inapplicable to analysis and prediction of jump processes, we will compute the online prediction errors obtained with the stationary and nonstationary, i.i.d. (top panel of Fig. 2) and Markovian (bottom panel of Fig. 2) methods described in this paper. As a benchmark to these approaches we will take the “zero” model prediction frequently used in the meteorological literature (i.e., predicted cloud cover for tomorrow is the same as today); crosses and prediction based on the nonstationary (squares) and stationary multifactor models (circles).
optimal nonstationary i.i.d. model obtained by minimizing (29) for the first half of the data is the one that also has the smallest mean prediction error. One-day-ahead online predictions generated with it have a mean error that is almost 2 times smaller than the one obtained with the zero model.

3) RESULTS FOR THE TROPICAL LOCATION

The top panel of Fig. 4 shows the profiles of the AIC-functional minima (29) obtained for different values of the regularization constraint $C$ and different numbers $K$ of local i.i.d. and Markovian multifactor models (see Fig. 2 for the graphical representation of the causal structures of the respective models). In contrast to the midlatitude case above, compared to other descriptions, the nonstationary Markov model with $K = 2$ and $C = 500$ reproduces the underlying data in the best way with a minimal number of parameters involved. This fact is in line with the general notion of the tropical cloud meteorology and, based on the available data, statistically demonstrates that Markovian models can be deployed for the parameterization of the cloud processes on the daily scales in the tropics (Khoudier et al. 2003; Majda et al. 2006; Khoudier et al. 2010).

Markov model equations (31) and (34) for the analyzed location in the tropics have the form

$$E[\pi^{(n)}(t + 1)] = \Gamma_1(t)\pi^{(n)}(t)P^{(1)}(u_i) + \Gamma_2(t)\pi^{(n)}(t)P^{(2)}(u_i),$$

$$E[\Gamma(t + 1)] = \Gamma(t)P^\Gamma(u_i),$$  \hspace{1cm} (36)

and the parameterized form of the functions $P^{(1)}(u_i)$, $P^{(2)}(u_i)$, and $P^\Gamma(u_i)$ is presented in Appendix B (B2). Comparison of the matrix elements of (B2) reveals that, similar to the midlatitude cloud cover dynamics, the overall dynamics of the tropical cloud cover can be represented as the Markovian process $\Gamma(t)$ switching between the clear-sky Markovian model 1 and a cloudy-sky Markovian model 2. However, the switching process itself is predominantly governed by a subset of three external factors that is different to the set identified for the midlatitudes: seasonal $u_1^t$ change (most significant influence), $u_3^t$, and $u_7^t$ (least significant). Also, the optimal local models 1 and 2 are significantly different. For the clear-sky model 1, seasonal temperature change, humidity advection, 500-mb vertical velocity, and the MJO are most relevant. In addition to these factors, temperature advection, pressure tendency, and the AO and AAO have significance for the cloudy-sky model 2. It is interesting to observe that the MJO factor has the highest significance among all other tested teleconnection influences in the case of the tropical cloud cover dynamics; this finding also agrees well with the notion in the literature (Khoudier et al. 2010). As demonstrated in the bottom panel of Fig. 4, the optimal nonstationary Markovian model obtained by minimizing (29) for the first half of the data is the one that also has the smallest mean prediction error. As in the previous midlatitude example, one-day-ahead online predictions are almost twice as good as the zero predictions.

3. Concluding discussion

Computational framework for identification of nonstationary discrete independent and Markovian models with impact from implicit and explicit external factors is presented. The presented methodology is designed to close the methodological gap in the current literature and available repository of statistical analysis tools: whereas there is a huge variety of existing methods for regressive analysis of continuous vector-valued systems (i.e., AR, VARX, VARMAX, GARCH, and many others), there is almost nothing (besides homogeneous stationary Markov chain techniques) available for the time series analysis of discrete jump-process observables.

The presented framework allows us to estimate the transition probabilities between the states, to identify the Markovian states in the system and to distinguish them from the independent ones, to identify the probabilities that are nonstationary and dependent on external factors, to compare
the external factors w.r.t. their informational impact on the resulting models and, as was demonstrated in the application to cloud cover data analysis, to construct predictive persistent probabilistic models that are optimal with respect to the long-range online predictions. A discrete-continuous algorithm for minimization of the respective information functional is presented, and its basic MATLAB version can be provided by the author on e-mail request.

Besides the analyzed ensemble time series and a specified discrete set of possible external factors, the presented framework relies only on two free input parameters: the optimization tolerance $\text{tol}$ and the number of random starting points $N_{SP}$. All other quantities (such as the number $K$ of local models, optimal persistency $C$, set of independent realizations states $J$ and the sets of indexes $I_0$, $I_1$ of a priori zero matrix elements) are determined automatically during the optimization.

As demonstrated in the application examples, the resulting models with external factors can be used both to get interesting insights into the dynamics and to construct self-consistent nonstationary predictive models for discrete jump processes.

One of the main conceptual difficulties of the presented framework is associated with the data exhibiting long-term memory effects [i.e., data where the Markovian assumption (5) is not justified]: embedding of such data in order to make them Markovian in the extended space will result in exponential increase of the total number of transition probabilities that have to be estimated for a time series of the same fixed length. This will result in a decreasing robustness of the estimates for time series that are relatively short (which is the case in most of the atmospheric observations). Longer observation series and additional regularity constraints on the memory decay could help in such cases. The biggest challenge, however, will be to test the performance of such probabilistic data-driven cloud and multicloud models (including several possible cloud types such as cumulus congestus clouds, deep convection clouds, etc.) in the form of subgrid-scale parameterization engines for GCMs, taking explicitly all the relevant physical properties of the underlying processes into account, not only in the “no physics, but math and learning from the data” manner as implemented in the current framework. This issue is a matter of future research and collaboration.

Acknowledgments. The author thanks the staff and the organizers of the IPAM program “Model and Data Hierarchies for Simulating and Understanding Climate,” 8 March–11 June 2010, for organizing and running a great program. A lot of input concerning the understanding of atmospheric and cloud physics is due to other participants of the program (special thanks to Boualem Khouider, Andy Majda, and Peter Spichtinger). The work was supported by the Swiss HP2C initiative “Swiss Platform for High-Performance and High-Productivity Computing,” the German DFG SPP 1276 (MetStroem) “Meteorology and Turbulence Mechanics,” and research project AnaGraph [founded by the Swiss National Fund (SNF)].

APPENDIX A

Discrete-Continuous Numerical Optimization Algorithm

Algorithm

Set the minimization tolerance $\text{tol}$ and number of random start points $N_{SP}$

Go through the finite discrete set of all possible combinations of $K$, $C$, $u$, $J$, $I_0$, and $I_1$

For any fixed $K$, $C$, $u$, $J$, $I_0$, and $I_1$ set $i = 1$

while $i \leq N_{SP}$

choose a random $\Gamma^{i-1}_0(t)$ satisfying (18) and (19)

set $s = 1$

while $\Delta L^{s}_{(i)} \approx \text{tol}$ repeat

step 1: minimize (29) w.r.t. $\Theta$ for a fixed $\Gamma^{s}_{(i)}(t)$ and identify $\Theta^{s}_{(i)}$

step 2: minimize (29) w.r.t. $\Gamma(t)$ (via linear programming; cf. Horenko 2010b) and identify $\Gamma^{s+1}_{(i)}(t)$

step 3: calculate $L^{s+1}_{(i)}(t) = L[\Gamma^{s+1}_{(i)}(t), \Theta^{s+1}_{(i)}(t)]$

and $\Delta L^{s}_{(i)} = ||L^{s+1}_{(i)}(t) - L^{s}_{(i)}||$

set $s = s + 1$

set $i = i + 1$

set $[\Gamma(t), \Theta] = \arg \min \Delta L^{s}$

Find $[\Gamma(t), \Theta, K, C, u, J, I_0, I_1] = \arg \min \text{AIC}(\Gamma(t), \Theta, K, C, u, J, I_0, I_1)$

As can be seen from the pseudocode above, there are two free parameters in the optimization procedure: (i) the optimization tolerance (controlling the convergence toward the local minima) and (ii) the number of random starting points (increasing this number decreases the probability of getting trapped in the local minimum, simultaneously increasing the overall computational cost of the scheme). They control the numerical performance of the framework and their choice depends on the available computer resources.

APPENDIX B

Optimal Parameters for the Cloud Cover Models of Locations in the Midlatitudes and Tropics

The parameterizations for the functions in midlatitude and tropical locations [35 and (36), respectively] follow. See the main text for further discussion.
a. Optimal parameterization of (35), midlatitude location in Europe

\[
\begin{align*}
\mu_1(u) &= [0.26 \ 0.74] + 10^{-1}\begin{bmatrix} 0.96 & -0.96 \ 0.16 & -0.16 \ 0.67 & -0.67 \end{bmatrix}u^1 + 10^{-1}\begin{bmatrix} 0.16 & -0.16 \ 0.7 & -0.7 \ 10^{-1} & 0.3 \end{bmatrix}u^2 + 10^{-1}\begin{bmatrix} 0.12 & -0.12 \ 0.1 & -0.1 \end{bmatrix}u^3 \\
\mu_2(u) &= [0.6 \ 0.4] + 10^{-2}\begin{bmatrix} 0.25 & -0.25 \ 0.98 & -0.98 \end{bmatrix}u^1 + 10^{-1}\begin{bmatrix} 0.16 & -0.16 \ 0.37 & -0.37 \ 0.1 & -0.1 \end{bmatrix}u^2 + 10^{-2}\begin{bmatrix} -0.2 & 0.2 \ -0.1 & 0.1 \end{bmatrix}u^3 \\
\end{align*}
\]

\[
\begin{align*}
P^\Gamma(u) &= \begin{pmatrix} 0.64 & 0.36 \\ 0.36 & 0.64 \end{pmatrix} + 10^{-1}\begin{pmatrix} -0.6 & 0.6 \\ -0.2 & 0.2 \end{pmatrix}u^1 + 10^{-1}\begin{pmatrix} -0.2 & 0.2 \\ -1.2 & 1.2 \end{pmatrix}u^2 + 10^{-1}\begin{pmatrix} -0.2 & 0.2 \\ -0.1 & 0.1 \end{pmatrix}u^3 \\
&+ 10^{-1}\begin{pmatrix} 1.5 & -1.5 \\ -0.9 & 0.9 \end{pmatrix}u^4 + 10^{-2}\begin{pmatrix} 0 & 0 \\ 0.2 & -0.2 \end{pmatrix}u^5 + 10^{-1}\begin{pmatrix} 0.3 & -0.3 \\ 0 & 0 \end{pmatrix}u^6 + 10^{-2}\begin{pmatrix} -6 & 6 \\ -4 & 4 \end{pmatrix}u^7 \\
&+ 10^{-2}\begin{pmatrix} -1 & 1 \\ 8 & -8 \end{pmatrix}u^8 + 10^{-2}\begin{pmatrix} -2 & 2 \\ 3 & -3 \end{pmatrix}u^9.
\end{align*}
\]

b. Optimal parameterization of (36), tropical location in Nigeria

\[
\begin{align*}
P^{(1)}(u) &= \begin{pmatrix} 0.74 & 0.26 \\ 0.16 & 0.84 \end{pmatrix} + 10^{-1}\begin{pmatrix} -1.0 & 1.0 \\ 0.0 & 0.0 \end{pmatrix}u^1 + 10^{-1}\begin{pmatrix} -0.2 & 0.2 \\ -0.2 & 0.2 \end{pmatrix}u^2 + 10^{-1}\begin{pmatrix} 0.0 & 0.0 \\ 1.1 & -1.1 \end{pmatrix}u^3 \\
&+ 10^{-16}\begin{pmatrix} -0.2 & 0.2 \\ -0.1 & 0.1 \end{pmatrix}u^4 + 10^{-1}\begin{pmatrix} -0.1 & 0.1 \\ 0.0 & 0.0 \end{pmatrix}u^5 + 10^{-1}\begin{pmatrix} -1.4 & 1.4 \\ 0.0 & 0.0 \end{pmatrix}u^6 \\
&+ 10^{-3}\begin{pmatrix} -1.5 & 1.5 \\ 0.0 & 0.0 \end{pmatrix}u^7 + 10^{-1}\begin{pmatrix} 0.0 & 0.0 \\ -0.4 & 0.4 \end{pmatrix}u^8 + 10^{-15}\begin{pmatrix} -0.1 & 0.1 \\ 0.0 & 0.0 \end{pmatrix}u^9,
\end{align*}
\]

\[
\begin{align*}
P^{(2)}(u) &= \begin{pmatrix} 0.82 & 0.18 \\ 0.37 & 0.63 \end{pmatrix} + 10^{-1}\begin{pmatrix} -1.4 & 1.4 \\ 0.3 & -0.3 \end{pmatrix}u^1 + 10^{-1}\begin{pmatrix} 0.2 & -0.2 \\ -0.4 & 0.4 \end{pmatrix}u^2 + 10^{-1}\begin{pmatrix} 0.1 & -0.1 \\ 0.6 & -0.6 \end{pmatrix}u^3 \\
&+ 10^{-2}\begin{pmatrix} 1.0 & -1.0 \\ 0.5 & -0.5 \end{pmatrix}u^4 + 10^{-1}\begin{pmatrix} -0.1 & 0.1 \\ -0.5 & 0.5 \end{pmatrix}u^5 + 10^{-1}\begin{pmatrix} -0.2 & 0.2 \\ -0.3 & 0.3 \end{pmatrix}u^6 \\
&+ 10^{-1}\begin{pmatrix} 0.1 & -0.1 \\ -0.5 & 0.5 \end{pmatrix}u^7 + 10^{-1}\begin{pmatrix} 0.0 & 0.0 \\ -0.9 & 0.9 \end{pmatrix}u^8 + 10^{-1}\begin{pmatrix} 0.0 & 0.0 \\ -0.4 & 0.4 \end{pmatrix}u^9,
\end{align*}
\]

\[
\begin{align*}
P^\Gamma(u) &= \begin{pmatrix} 0.78 & 0.22 \\ 0.12 & 0.88 \end{pmatrix} + 10^{-1}\begin{pmatrix} 0.7 & -0.7 \\ 1.5 & -1.5 \end{pmatrix}u^1 + 10^{-16}\begin{pmatrix} -0.6 & 0.6 \\ 0.2 & -0.2 \end{pmatrix}u^2 + 10^{-1}\begin{pmatrix} -1.1 & 1.1 \\ -0.1 & 0.1 \end{pmatrix}u^3 \\
&+ 10^{-16}\begin{pmatrix} -0.3 & 0.3 \\ -0.1 & 0.1 \end{pmatrix}u^4 + 10^{-16}\begin{pmatrix} 0 & 0 \\ 0.4 & -0.4 \end{pmatrix}u^5 + 10^{-2}\begin{pmatrix} 0.0 & -0.0 \\ 0.6 & -0.6 \end{pmatrix}u^6 \\
&+ 10^{-1}\begin{pmatrix} 0.6 & -0.6 \\ -0.2 & 0.2 \end{pmatrix}u^7 + 10^{-16}\begin{pmatrix} 0.5 & -0.5 \\ 0.0 & 0.0 \end{pmatrix}u^8 + 10^{-1}\begin{pmatrix} 0.2 & -0.2 \\ 0.0 & 0.0 \end{pmatrix}u^9.
\end{align*}
\]

REFERENCES