Research Article

Computational uncertainty quantification for random time-discrete epidemiological models using adaptive gPC

J. Calatayud[a], J.-C. Cortés[a,*], M. Jornet[a], R.-J. Villanueva[a]

Population dynamics models consisting of non-linear difference equations allow us to get a better understanding of the processes involved in epidemiology. Usually, these mathematical models are studied under a deterministic approach. However, in order to take into account the uncertainties associated with the measurements of the model input parameters, a more realistic approach would be to consider these inputs as random variables. In this paper, we study the random time-discrete epidemiological models SIS, SIR, SIRS and SEIR using a powerful unified approach based upon the so-called adaptive generalized Polynomial Chaos (gPC) technique. The solution to these random difference equations is a stochastic process in discrete time, which represents the number of susceptible, infected, recovered, etc. individuals at each time step. We show, via numerical experiments, how adaptive gPC permits quantifying the uncertainty for the solution stochastic process of the aforementioned random time-discrete epidemiological model and obtaining accurate results at a cheap computational expense. We also highlight how adaptive gPC can be applied in practice, by means of an example using real data. Copyright © 2018 John Wiley & Sons, Ltd.

Keywords: Stochastic difference equations; Computational methods for stochastic equations; Random time-discrete epidemiological model; Random non-linear difference equations model; Random population dynamics model; Computational uncertainty quantification; Adaptive gPC

1. Introduction

Discrete models, usually expressed via finite difference equations, and continuous models, often expressed by means of ordinary and partial differential equations, allow us to get a better understanding of the processes involved in epidemiology [1, 2, 3, 4, 5]. These models for population dynamics have been usually studied in a deterministic sense, treating the involved input parameters (initial conditions, forcing term and/or coefficients) as constants. Recent examples in this regard, dealing with important epidemiological models by applying new deterministic approaches, like nonstandard finite difference schemes, modal infinite series expansions, analysis of bifurcations, for example, include [6, 7, 8]. However, due to the inherent uncertainty associated with epidemiological phenomena, it would be better to treat the input parameters in a random sense. For instance, the coefficient that describes the proportion of individuals that recover from a disease and become susceptible again should take into account the uncertainties involved in the measurements, due to errors in the collection of data, missed individuals, lack of information, etc.

The most well-known method to deal computationally with stochastic systems is Monte Carlo simulation [9]. Although it is an effective and easy to implement approach to quantify the uncertainty, the slowness to get accurately the digits in the computations makes this technique computationally expensive.

Only recently, some random continuous-time epidemic models have been studied using generalized Polynomial Chaos (gPC) [10, 11, 12]. In [12], the authors quantified the effect of randomizing the coefficients in a SIR and SIRS model (without vital dynamics, i.e., they do not include birth and death rates in their formulation) using gPC. Since gPC approach uses orthogonal

[a] Instituto Universitario de Matemática Multidisciplinar,
Universitat Politècnica de València,
Camino de Vera s/n, 46022, Valencia, Spain
*Correspondence to: J.-C. Cortés, Instituto Universitario de Matemática Multidisciplinar, Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia, Spain
†E-mail: jucagre@alumni.uv.es; jccortes@imm.upv.es; marjorsa@doctor.upv.es; rjvillan@imm.upv.es

polynomials from the Askey-Wiener scheme, the applications were limited to parameters following only some standard probability distributions.

Recently, in [13, 14, 15], the authors modified gPC to an adaptive gPC computational approach. In the adaptive gPC approach, the orthogonal polynomials do not belong to the Askey-Wiener scheme. Instead of this, these polynomials are constructed directly from the random parameter inputs, by a Gram-Schmidt orthonormalization procedure (if the random inputs are independent) or by using polynomial canonical bases (if the random inputs are not independent). The numerical experiments from [13, 14, 15] showed that adaptive gPC provides reliable results for small basis orders of the solution process. This allows us to obtain the main statistical moments associated with the solution process: expectation, variance, etc. It must be pointed out that different versions of adaptive gPC have been successfully proposed [16, 17] and applied to deal with complex problems [18, 19] and other important problems arising in areas like Fluid Mechanics, Reliability, Wave Propagation, etc., [20, 21, 22].

In this paper, we study relevant non-linear time-discrete epidemiological models, in which the parameters are assumed to be random variables. To the best of our knowledge, although particular random differential equation models for population dynamics have been extensively analyzed via gPC [12, 23, 24], random difference equations modeling epidemiological processes have not been studied using a polynomial chaos approach. Moreover, the novel variation of gPC that we deal with, adaptive gPC, permits covering a wide range of epidemiological models in which random input parameters may take any probability distributions with no independence assumptions. Our study will comprehend the time-discrete SIS, SIR (without and with vital dynamics), SIRS and SEIR models, although other epidemiological systems, like SEIQV or SEIVR models, could be considered (see [25, 26]). As it shall be seen later, when applying the adaptive gPC approach to these models, a deterministic system of non-linear recursive equations is obtained for the coefficients of the expansions. After iterating the deterministic difference system, approximations for the expectation and variance of the solution process will be obtained. This approach will allow us to quantify the uncertainties in the number of susceptible, infected, recovered, etc., individuals. More specifically, for a real epidemic with available measured data, the coefficients in the time-discrete epidemic model may be estimated in a deterministic manner, using a minimization procedure. This key issue will be illustrated via an appropriate example later. To analyze uncertainty associated with the modeling, one could set a non-informative distribution to each one of the parameters, with mean value being the deterministic estimate obtained, and with small variance to avoid instability problems [12]. The meaning of instability, in the setting of epidemiological models, refers to the case where the model outputs may take values that do not make sense from an epidemiological standpoint. For example, in the SIR model this situation arises when the number of individuals (susceptible, infected, recovered, etc.) becomes negative or larger than the total population size. This might occur if one of the model coefficients has large variability. This sense of instability has been used by other authors in the context of different mathematical models (see for example [27, Sec. 4] and [12, Sec. 4]).

The structure of the paper is the following. In Section 2, we will explain the time-discrete epidemiological models that we are going to deal with: SIS, SIR, SIRS and SEIR. In Section 3, the parameters in these epidemiological models will be considered as random variables, so that the solution of the non-linear system of difference equations will become a stochastic process. The solution process will be expanded by using adaptive gPC and we will develop the necessary computations until we obtain the deterministic non-linear system of recursive equations satisfied by the coefficients of the polynomial expansion. In Section 4, numerical experiments will be performed to show the wide range of applications of our theoretical findings. We will also show how adaptive gPC can be used in the context of an application with real data. Finally, in Section 5 conclusions will be drawn.

2. Time-discrete epidemic models

In the time-discrete epidemiological models SIS, SIR, SIRS and SEIR, the population is divided into compartments according to the disease stage of the individual: $S(m)$ denotes the number of susceptible individuals, $E(m)$ the number of exposed individuals, $I(m)$ the number of infected individuals and $R(m)$ the number of recovered individuals, at time step m. These compartmental models assume homogeneous mixing, i.e., all individuals are equally likely to contact any other individual [4]. The total population is assumed to be a constant, say N, either by assuming no birth or death (model without vital dynamics) or assuming that the birth rate equals the death rate (model with vital dynamics, but no demographic effects).

The SIS model corresponds to diseases that do not confer immunity just after infection. The SIR model refers to epidemics where there is lasting immunity after infection. If we assume vital dynamics, we add a birth rate equal to the death rate. In the SIRS model, after infection, the individuals become recovered and immune, but over time some individuals may become susceptible again. Finally, the SEIR model assumes an exposition (incubation) period before infection and lasting immunity after infection.

The mathematical formulation for these models is the following:

SIS model	SIR model	SIR model (with vital dynamics)
$S(m+1) = S(m) - bS(m)I(m) + aI(m),$	$S(m+1) = S(m) - bS(m)I(m),$	$S(m+1) = S(m) + \mu N - \mu S(m) - bS(m)I(m),$
$I(m+1) = I(m) + bS(m)I(m) - aI(m),$	$I(m+1) = I(m) + bS(m)I(m) - aI(m),$	$I(m+1) = I(m) + bS(m)I(m) - aI(m) - \mu I(m),$
	$R(m+1) = R(m) + aI(m),$	$R(m+1) = R(m) + aI(m) - \mu R(m),$

2
Prepared using mmaauth.cls

Math. Meth. Appl. Sci. **2018**, 00 1–9

SIRS model

$$S(m+1) = S(m) + \mu N - \mu S(m) - bS(m)I(m) + cR(m),$$
$$I(m+1) = I(m) + bS(m)I(m) - aI(m) - \mu I(m),$$
$$R(m+1) = R(m) + aI(m) - \mu R(m) - cR(m),$$

SEIR model

$$S(m+1) = S(m) + \mu N - \mu S(m) - bS(m)I(m),$$
$$E(m+1) = E(m) + bS(m)I(m) - \mu E(m) - cE(m),$$
$$I(m+1) = I(m) + cE(m) - aI(m) - \mu I(m),$$
$$R(m+1) = R(m) + aI(m) - \mu R(m),$$

where a is the proportion of infective that leave this compartment, b is the proportion of contacts between susceptible and infective that give rise to a new infected or exposed individual, μ is the birth rate (equal to the death rate), and c is the proportion of recovered individuals that lose immunity (SIRS model) or the proportion of exposed individuals that become infective (SEIR model).

Since the population is assumed to be constant, the following relationship $S(m) + E(m) + I(m) + R(m) = N$ fulfills, so we may remove the last equation of each one of the previous non-linear systems of difference equations.

In what follows, we will consider the input parameters a, b, c and μ as random variables, due to possible errors in their measurements and the inherent complexity and uncertainty often involved in the biological factors that determine these model parameters. Although in reality these parameters are independent [12], we think that from a mathematical standpoint it is interesting to study the population dynamics under a dependence assumption, since it will provide more generality to our subsequent analysis. On the other hand, the initial conditions S_0, E_0, I_0 and R_0 will be considered as constants.

3. Randomization and adaptive gPC

We proceed to analyze the previous randomized epidemiological models. Since the SEIR model is the most complex, our theoretical development will focus on it. For the other four epidemic models, the final results are analogous. In any case, the numerical experiments carried out to show the capability of adaptive gPC to quantify uncertainty propagation in discrete epidemiological models will be tested in all the aforementioned models.

Using the fact that the population is constant, $S(m) + E(m) + I(m) + R(m) = N$, we delete the equation for $R(m)$ and work with the first three equations of the SEIR system.

We assume that the input parameters a, b, c and μ are random variables defined in a common complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$, whose measurements depend on the outcome $\omega \in \Omega$ of an experiment: $a = a(\omega)$, $b = b(\omega)$, $c = c(\omega)$ and $\mu = \mu(\omega)$. Thus, S, E, I and R become time discrete stochastic processes: $S(m) = S(m, \omega)$, $E(m) = E(m, \omega)$, $I(m) = I(m, \omega)$ and $R(m) = R(m, \omega)$. We will assume the proportions a, b, c and μ to follow a pre-established absolutely continuous probability distribution with finite moments (not necessarily the same distributions, and with no independence assumption) and support in $[0, 1]$. In practice, in order to avoid instability problems, the variances of a, b, c and μ must be chosen small enough. As it has been pointed out at the end of Section 1, in the context of random discrete epidemiological modeling, instability applies to the physical model itself: it refers to discrete dynamical systems in which one of the variables evolves without bounds. Specifically, in epidemic models, instability arises when the number of susceptible, infected, recovered, etc., individuals becomes a negative number or is larger than the total population size. This situation might occur if one of the model coefficients has large variability. It must be remarked that, although the theoretical support of the model coefficients is contained in $[0, 1]$, a distribution that may leave $[0, 1]$ but with a negligible probability (for instance, a normal or gamma distributions with mean in $[0, 1]$ and a very small variance) could be possible in practice. The goal is to quantify the uncertainty for $S(m)$, $E(m)$, $I(m)$ and $R(m)$ using adaptive gPC [13, 14, 15].

We will work in the Hilbert space of random variables whose variance exists, $(\mathrm{L}^2(\Omega), \langle, \rangle)$, where the inner product $\langle, \rangle$ is defined as $\langle X, Y \rangle = \mathbb{E}[XY]$, being $\mathbb{E}[\cdot]$ the expectation operator. Denote $\zeta_1 = a$, $\zeta_2 = b$, $\zeta_3 = c$ and $\zeta_4 = \mu$, and $\zeta = (\zeta_1, \zeta_2, \zeta_3, \zeta_4)^\top$ ($\top$ is the transpose operator), so that the notation matches with [13, 14, 15]. Let $\mathcal{C}_i^p = \{1, \zeta_i, \zeta_i^2, \ldots, \zeta_i^p\}$ be the canonical basis of the polynomials in ζ_i up to degree p, $1 \leq i \leq 4$. We distinguish two cases depending on whether $\zeta_1, \ldots, \zeta_4$ are independent or not:

(i) Independent: Consider the inner product with respect to the density of ζ_i, f_{ζ_i}: $\langle g(\zeta_i), h(\zeta_i) \rangle = \mathbb{E}[g(\zeta_i) h(\zeta_i)] = \int_{\mathbb{R}} g(\zeta_i) h(\zeta_i) f_{\zeta_i}(\zeta_i) \, d\zeta_i$. Using a Gram-Schmidt procedure, we orthonormalize $\mathcal{C}_i^p$: $\Xi_i^p = \{\phi_0^i(\zeta_i), \ldots, \phi_p^i(\zeta_i)\}$. By applying a simple tensor product [13], we obtain an orthonormal basis $\Xi = \{\phi_1(\zeta), \ldots, \phi_P(\zeta)\}$ with respect to the inner product $\langle g(\zeta), h(\zeta) \rangle = \mathbb{E}[g(\zeta) h(\zeta)] = \int_{\mathbb{R}^4} g(\zeta) h(\zeta) f_\zeta(\zeta) \, d\zeta$, where $\phi_1 = 1$ and $P = (p+4)!/(p!4!)$.

(ii) Dependent: Using a simple tensor product directly on the canonical bases $\mathcal{C}_i^p$ [15], we obtain the (non-orthonormal) basis $\Xi = \{\phi_1(\zeta), \ldots, \phi_P(\zeta)\}$, where $\phi_1 = 1$ and $P = (p+4)!/(p!4!)$.

Consider the basis $\Xi = \{\phi_1(\zeta), \ldots, \phi_P(\zeta)\}$ constructed in any of the two previous cases (i) and (ii). We want to find deterministic coefficients $\hat{S}_i(m)$, $\hat{E}_i(m)$ and $\hat{I}_i(m)$, $1 \leq i \leq P$, so that

$$S(m) \approx \sum_{i=1}^{P} \hat{S}_i(m) \phi_i(\zeta), \quad E(m) \approx \sum_{i=1}^{P} \hat{E}_i(m) \phi_i(\zeta), \quad I(m) \approx \sum_{i=1}^{P} \hat{I}_i(m) \phi_i(\zeta).$$

Math. Meth. Appl. Sci. 2018, 00 1–9
Prepared using mmaauth.cls

Copyright © 2018 John Wiley & Sons, Ltd.   3

We impose these approximations to be a solution of the SEIR model:

$$\sum_{i=1}^{P} \hat{S}_i(m+1)\phi_i(\zeta) = \sum_{i=1}^{P} \hat{S}_i(m)\phi_i(\zeta) + \mu N - \mu \sum_{i=1}^{P} \hat{S}_i(m)\phi_i(\zeta) - b\sum_{i=1}^{P}\sum_{j=1}^{P} \hat{S}_i(m)\hat{I}_j(m)\phi_i(\zeta)\phi_j(\zeta),$$

$$\sum_{i=1}^{P} \hat{E}_i(m+1)\phi_i(\zeta) = \sum_{i=1}^{P} \hat{E}_i(m)\phi_i(\zeta) + b\sum_{i=1}^{P}\sum_{j=1}^{P} \hat{S}_i(m)\hat{I}_j(m)\phi_i(\zeta)\phi_j(\zeta) - \mu \sum_{i=1}^{P} \hat{E}_i(m)\phi_i(\zeta) - c\sum_{i=1}^{P} \hat{E}_i(m)\phi_i(\zeta),$$

$$\sum_{i=1}^{P} \hat{I}_i(m+1)\phi_i(\zeta) = \sum_{i=1}^{P} \hat{I}_i(m)\phi_i(\zeta) + c\sum_{i=1}^{P} \hat{E}_i(m)\phi_i(\zeta) - a\sum_{i=1}^{P} \hat{I}_i(m)\phi_i(\zeta) - \mu \sum_{i=1}^{P} \hat{I}_i(m)\phi_i(\zeta).$$

The initial conditions become

$$\sum_{i=1}^{P} \hat{S}_i(0)\phi_i(\zeta) = S_0, \quad \sum_{i=1}^{P} \hat{E}_i(0)\phi_i(\zeta) = E_0, \quad \sum_{i=1}^{P} \hat{I}_i(0)\phi_i(\zeta) = I_0.$$

Notice that, although the initial conditions are assumed to be deterministic, for the sake of generality a polynomial chaos expansion is used for each one of them. Clearly, in this case, as $\phi_1(\zeta) = 1$, then $\hat{S}_1(0) = S_0$, $\hat{E}_1(0) = E_0$ and $\hat{I}_1(0) = I_0$. We then multiply by $\phi_k(\zeta)$ and apply the expectation operator, $k = 1, \ldots, P$, which yields:

$$\sum_{i=1}^{P} \hat{S}_i(m+1)\mathbb{E}[\phi_i(\zeta)\phi_k(\zeta)] = \sum_{i=1}^{P} \hat{S}_i(m)\mathbb{E}[\phi_i(\zeta)\phi_k(\zeta)] + N\mathbb{E}[\mu\phi_k(\zeta)] - \sum_{i=1}^{P} \hat{S}_i(m)\mathbb{E}[\mu\phi_i(\zeta)\phi_k(\zeta)]$$
$$- \sum_{i=1}^{P}\sum_{j=1}^{P} \hat{S}_i(m)\hat{I}_j(m)\mathbb{E}[b\phi_i(\zeta)\phi_j(\zeta)\phi_k(\zeta)],$$

$$\sum_{i=1}^{P} \hat{E}_i(m+1)\mathbb{E}[\phi_i(\zeta)\phi_k(\zeta)] = \sum_{i=1}^{P} \hat{E}_i(m)\mathbb{E}[\phi_i(\zeta)\phi_k(\zeta)] + \sum_{i=1}^{P}\sum_{j=1}^{P} \hat{S}_i(m)\hat{I}_j(m)\mathbb{E}[b\phi_i(\zeta)\phi_j(\zeta)\phi_k(\zeta)] - \sum_{i=1}^{P} \hat{E}_i(m)\mathbb{E}[\mu\phi_i(\zeta)\phi_k(\zeta)]$$
$$- \sum_{i=1}^{P} \hat{E}_i(m)\mathbb{E}[c\phi_i(\zeta)\phi_k(\zeta)],$$

$$\sum_{i=1}^{P} \hat{I}_i(m+1)\mathbb{E}[\phi_i(\zeta)\phi_k(\zeta)] = \sum_{i=1}^{P} \hat{I}_i(m)\mathbb{E}[\phi_i(\zeta)\phi_k(\zeta)] + \sum_{i=1}^{P} \hat{E}_i(m)\mathbb{E}[c\phi_i(\zeta)\phi_k(\zeta)] - \sum_{i=1}^{P} \hat{I}_i(m)\mathbb{E}[a\phi_i(\zeta)\phi_k(\zeta)]$$
$$- \sum_{i=1}^{P} \hat{I}_i(m)\mathbb{E}[\mu\phi_i(\zeta)\phi_k(\zeta)].$$

$$\sum_{i=1}^{P} \hat{S}_i(0)\mathbb{E}[\phi_i(\zeta)\phi_k(\zeta)] = S_0\mathbb{E}[\phi_k(\zeta)], \quad \sum_{i=1}^{P} \hat{E}_i(0)\mathbb{E}[\phi_i(\zeta)\phi_k(\zeta)] = E_0\mathbb{E}[\phi_k(\zeta)], \quad \sum_{i=1}^{P} \hat{I}_i(0)\mathbb{E}[\phi_i(\zeta)\phi_k(\zeta)] = I_0\mathbb{E}[\phi_k(\zeta)].$$

To rewrite these expressions in matrix form, consider the vectors h, g and t whose components, respectively, are

$$h_k = N\mathbb{E}[\mu\phi_k(\zeta)], \quad g_k(m) = \sum_{i=1}^{P}\sum_{j=1}^{P} \hat{S}_i(m)\hat{I}_j(m)\mathbb{E}[b\phi_i(\zeta)\phi_j(\zeta)\phi_k(\zeta)], \quad t_k = \mathbb{E}[\phi_k(\zeta)], \quad k = 1, \ldots, P,$$

and the matrices T, V, W and U whose entries, respectively, are

$$T_{ik} = \mathbb{E}[\phi_i(\zeta)\phi_k(\zeta)], \quad V_{ik} = \mathbb{E}[\mu\phi_i(\zeta)\phi_k(\zeta)], \quad W_{ik} = \mathbb{E}[c\phi_i(\zeta)\phi_k(\zeta)], \quad U_{ik} = \mathbb{E}[a\phi_i(\zeta)\phi_k(\zeta)], \quad i, k = 1, \ldots, P.$$

In this way, we obtain a deterministic system of non-linear difference equations:

$$T\hat{S}(m+1) = (T-V)\hat{S}(m) + h - g(m),$$
$$T\hat{E}(m+1) = (T-V-W)\hat{E}(m) + g(m),$$
$$T\hat{I}(m+1) = (T-U-V)\hat{I}(m) + W\hat{E}(m),$$

$$T\hat{S}(0) = S_0 t, \quad T\hat{E}(0) = E_0 t, \quad T\hat{I}(0) = I_0 t,$$

where $\hat{S}(m) = (\hat{S}_1(m), \ldots, \hat{S}_P(m))^\top$, $\hat{E}(m) = (\hat{E}_1(m), \ldots, \hat{E}_P(m))^\top$ and $\hat{I}(m) = (\hat{I}_1(m), \ldots, \hat{I}_P(m))^\top$. By iteration, we can obtain $\hat{S}(m)$, $\hat{E}(m)$ and $\hat{I}(m)$ for time steps m as large as we desire. Numerically, we have observed that, for large order of canonical bases p, the matrix T is ill-conditioned, so the computations in the recursion may become inaccurate.

4   Copyright © 2018 John Wiley & Sons, Ltd.

Math. Meth. Appl. Sci. **2018**, 00 1–9

The expectation and variance of $S(m)$, $E(m)$ and $I(m)$ are estimated as follows:

$$\mathbb{E}[S(m)] \approx \sum_{i=1}^{P} \hat{S}_i(m)\mathbb{E}[\phi_i(\zeta)], \quad \mathbb{V}[S(m)] \approx \sum_{i,j=1}^{P} \hat{S}_i(m)\hat{S}_j(m)\mathbb{C}\text{ov}[\phi_i(\zeta),\phi_j(\zeta)],$$

$$\mathbb{E}[E(m)] \approx \sum_{i=1}^{P} \hat{E}_i(m)\mathbb{E}[\phi_i(\zeta)], \quad \mathbb{V}[E(m)] \approx \sum_{i,j=1}^{P} \hat{E}_i(m)\hat{E}_j(m)\mathbb{C}\text{ov}[\phi_i(\zeta),\phi_j(\zeta)],$$

$$\mathbb{E}[I(m)] \approx \sum_{i=1}^{P} \hat{I}_i(m)\mathbb{E}[\phi_i(\zeta)], \quad \mathbb{V}[I(m)] \approx \sum_{i,j=1}^{P} \hat{I}_i(m)\hat{I}_j(m)\mathbb{C}\text{ov}[\phi_i(\zeta),\phi_j(\zeta)].$$

From the relation $R(m) = N - S(m) - E(m) - I(m)$, we can approximate the expectation of $R(m)$,

$$\mathbb{E}[R(m)] = N - \mathbb{E}[S(m)] - \mathbb{E}[E(m)] - \mathbb{E}[I(m)],$$

and the variance of $R(m)$,

$$\mathbb{V}[R(m)] = \mathbb{V}[S(m)] + \mathbb{V}[E(m)] + \mathbb{V}[I(m)] + 2\mathbb{C}\text{ov}[S(m),E(m)] + 2\mathbb{C}\text{ov}[S(m),I(m)] + 2\mathbb{C}\text{ov}[E(m),I(m)],$$

where

$$\mathbb{C}\text{ov}[S(m),E(m)] \approx \sum_{i,j=1}^{P} \hat{S}_i(m)\hat{E}_j(m)\mathbb{C}\text{ov}[\phi_i(\zeta),\phi_j(\zeta)],$$

$$\mathbb{C}\text{ov}[S(m),I(m)] \approx \sum_{i,j=1}^{P} \hat{S}_i(m)\hat{I}_j(m)\mathbb{C}\text{ov}[\phi_i(\zeta),\phi_j(\zeta)],$$

$$\mathbb{C}\text{ov}[E(m),I(m)] \approx \sum_{i,j=1}^{P} \hat{E}_i(m)\hat{I}_j(m)\mathbb{C}\text{ov}[\phi_i(\zeta),\phi_j(\zeta)].$$

When the random inputs $\zeta_1, \ldots, \zeta_4$ are independent (case (i)), the previous formulas may be simplified. Indeed, notice that, by orthonormality of Ξ, it is derived that $T = \text{Id}_P$ and $t = e_1$, where Id_P is the $P \times P$ identity matrix and $e_1 = (1, 0, \ldots, 0)^\top$. Thus, in this case no problem of ill-conditioning of T appears. Moreover, the approximations for the expectation and variance of $S(m)$, $E(m)$ and $I(m)$ can be simplified in the following form:

$$\mathbb{E}[S(m)] \approx \hat{S}_1(m), \quad \mathbb{V}[S(m)] \approx \sum_{i=2}^{P} \hat{S}_i(m)^2,$$

$$\mathbb{E}[E(m)] \approx \hat{E}_1(m), \quad \mathbb{V}[E(m)] \approx \sum_{i=2}^{P} \hat{E}_i(m)^2,$$

$$\mathbb{E}[I(m)] \approx \hat{I}_1(m), \quad \mathbb{V}[I(m)] \approx \sum_{i=2}^{P} \hat{I}_i(m)^2.$$

Naturally, $\mathbb{E}[R(m)] \approx N - \hat{S}_1(m) - \hat{E}_1(m) - \hat{I}_1(m)$, while for computing the approximation of $\mathbb{V}[R(m)]$, the necessary covariances can be estimated via $\mathbb{C}\text{ov}[S(m),E(m)] \approx \sum_{i,j=2}^{P} \hat{S}_i(m)\hat{E}_j(m)$, $\mathbb{C}\text{ov}[S(m),I(m)] \approx \sum_{i,j=2}^{P} \hat{S}_i(m)\hat{I}_j(m)$ and $\mathbb{C}\text{ov}[E(m),I(m)] \approx \sum_{i,j=2}^{P} \hat{E}_i(m)\hat{I}_j(m)$.

4. Numerical experiments

In the following numerical examples, we will specify probability distributions for the random input parameters in each one of the five epidemiological models previously considered. For a real epidemic with available measured data, the coefficients in the time-discrete epidemiological model may be estimated in a deterministic manner, using a minimization procedure. To analyze uncertainty associated with the modeling, one could set a non-informative distribution to each one of the parameters, with mean value being the deterministic estimate obtained. In this article we do not deal with inverse parameter estimation, see for example references [28, 29] related with Bayesian inference and gPC.

As we have previously indicated, the distributions of the random input parameters must be absolutely continuous with finite moments. In order to avoid instability problems (i.e., large variability of a coefficient that makes the size of a compartment of the epidemiological model leave the interval $[0, N]$), the density functions have to be concentrated around the mean value. As we are working with proportions, the support of the random inputs must be bounded. However, distributions with an unbounded support

Math. Meth. Appl. Sci. 2018, 00 1–9
Prepared using mmaauth.cls

Copyright © 2018 John Wiley & Sons, Ltd.   5

Year	1995	1997	2001	2003	2006	2011	2014
j	0	2	6	8	11	16	19
S_j	0.5298	0.5514	0.5783	0.6244	0.6467	0.6863	0.6957

Table 1. Percentage of non-smokers Spanish men aged over 16 years old during the period 1995–2014. Source [30]. Case 1.

but with small variance are also possible in practice. On the other hand, as we have worked out in the theoretical development, the independence assumption is not required.

In the following numerical simulations, the order p of the canonical bases $\mathcal{C}_i^p$, $1 \leq i \leq 4$, will be $p = 2$ and $p = 3$. We will observe in the plots that very similar results are obtained for both orders, then showing convergence. However, if more accuracy is desired, a larger p has to be used, but this will increase the computational expense and the ill-conditioning of the matrix T.

In the following we present several examples corresponding to the five epidemiological models previously stated, in order to illustrate the capability of adaptive gPC to quantify the propagation of uncertainties associated with the model outputs when the model parameters are random variables. The numerical results are shown by means of appropriate graphical representations, where approximations of the mean value of $S(m)$, $E(m)$, $I(m)$ and $R(m)$, together with a confidence interval with the rule [mean $\pm$ standard deviation], are plotted. It is important to point out that univariate distributions (Triangular, Uniform, Gamma, Normal, etc.) and multivariate distributions (Dirichlet, bivariate Normal, etc.) have been chosen for the random model parameters. In this manner, both independent and dependent random variables are treated in the examples via a wide variety of probabilistic distributions. Although the examples are presented via numerical experiments involving simulations, we will comment their epidemiological meaning. In this regard, the first example has been devised so that it illustrates how to choose, in practice, the parameters of the probability distributions for the involved random input parameters in a SIS model using real data. In the rest of examples (Cases 2–5), the parameters can be chosen following an analogous strategy.

- Case 1. SIS model with dependent random inputs a and b. We consider a SIS model to treat the social habit of smoking. The population is divided into two groups: non-smokers (susceptible) and smokers (infective). Table 1 presents real data, denoted by S_j, $j \in \{0, 2, 6, 8, 11, 16, 19\}$, on the percentage of non-smokers Spanish men during the period 1995-2014, [30]. Then the total population is considered as $N = 1$. Using a least-squares fitting procedure, we have obtained the deterministic estimates $\hat{a} = 0.0378749$ (people giving up tobacco) and $\hat{b} = 0.0225118$ (persons who start smoking). To assess uncertainty associated with the data measurements, we specify probability distributions for the random input parameters: we suppose that (a, b) follows a bivariate normal distribution with mean vector and covariance matrix given by

$$\begin{pmatrix} 0.0378749 \\ 0.0225118 \end{pmatrix}, \quad \begin{pmatrix} 0.000025 & -0.00001 \\ -0.00001 & 0.00001 \end{pmatrix},$$

  respectively. It is clear that the covariances are negative because of the opposite behavior of the random variables a and b. In Figure 1, the results of the adaptive gPC expansions are shown. We have plotted model outputs for $p = 2$ and $p = 3$ in black and blue colors, respectively. In this plot we can see that they completely match, then showing rapid convergence of adaptive gPC. We can observe that the confidence intervals capture the uncertainty associated with the outputs.

- Case 2. SIR model with independent random inputs a and b. Assume an epidemic in a population of size $N = 65$, where the recovery and infection parameters have been estimated via a least-squares fitting: $\hat{a} = 0.2$ and $\hat{b} = 0.01$. To study the time evolution of the epidemic, we set test probability distributions to a and b, being the mean values $\hat{a}$ and $\hat{b}$, respectively: $a \sim \text{Gamma}(4, 0.05)$ (shape-scale notation) and $b \sim \text{Triangular}(0.008, 0.012)$. Let us assume that the initial conditions are $S_0 = 5$, $I_0 = 60$ and $R_0 = 0$. The estimates computed via adaptive gPC are presented in Figure 2. Both plots are nearly identical, then showing convergence. Epidemiologically, we start in a scenario where most of the population is infected and a few individuals are susceptible. As the time goes on, the infected individuals become recovered and finally, the whole population is recovered. As a consequence, the number of susceptible decreases very slowly. Note that the uncertainty in the evolution of the infected and recovered subpopulations may lead to a delay of more that 15 days in the disappearance of the infective, because the higher percentile of recovered is reached between days 10–15, meanwhile the same level is reached by the infective at the lower percentile between the days 25–30.

- Case 3. SIR model with vital dynamics, being the random inputs a, b and μ non-independent. We suppose an epidemic in a population of size $N = 65$. Let us assume that the recovery, infection and birth parameters have been estimated via a deterministic procedure, obtaining $\hat{a} = 0.2$, $\hat{b} = 0.01$ and $\hat{\mu} = 0.1$, respectively. For uncertainty quantification, we set the following test distributions: $(a, b) \sim \text{Dirichlet}(80, 4, 316)$ and $\mu \sim \text{Uniform}(0.04, 0.16)$, whose respective expectations match the numerical values $\hat{a} = 0.2$, $\hat{b} = 0.01$ and $\hat{\mu} = 0.1$. The random vector (a, b) and the random variable μ are assumed to be independent. We take $S_0 = 5$, $I_0 = 60$ and $R_0 = 0$ as the initial conditions. The numerical results are shown in Figure 3. Both plots are nearly identical, then showing convergence. Starting in the same scenario as Figure 2, now we observe that an endemic situation is reached. This difference happens because the inclusion of the vital dynamics makes that old recovered people leave the system and new susceptible people enter into the system.

- Case 4. SIRS model with independent random inputs a, b, c and μ. In a population of size $N = 65$, we suppose the following deterministic estimates for the recovery, infection, decline and birth parameters: $\hat{a} = 0.2$, $\hat{b} = 0.01$, $\hat{c} = 0.1$ and
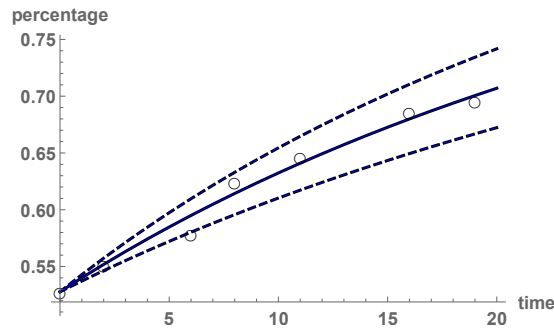
Figure 1. Mean value estimates (solid line) and confidence intervals (in dashed lines) for Case 1. Real data from Table 1 is denoted with black circles.
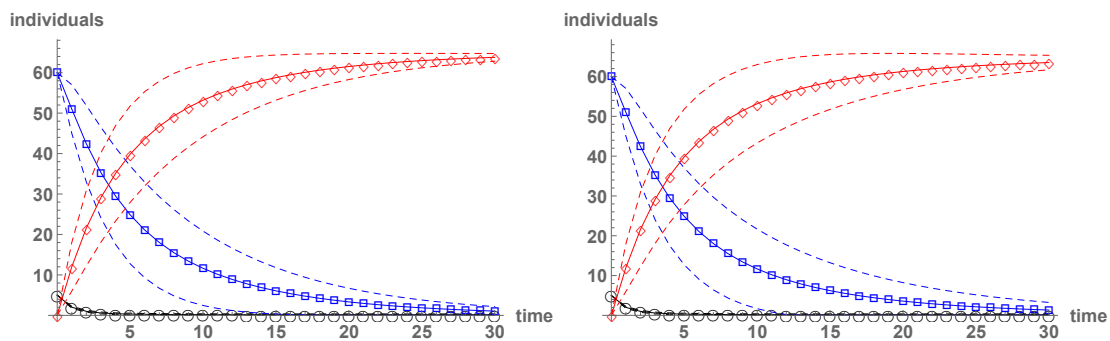

Figure 2. Mean value estimates and confidence intervals (in dashed lines) for Case 2. Left: $p = 2$. Right: $p = 3$. S is denoted by $\bigcirc$ in black color, I is denoted by $\square$ in blue color, and R is denoted by $\diamond$ in red color.

$\hat{\mu} = 0.1$. The test probability distributions given to the random inputs are: $a \sim \text{Gamma}(4, 0.05)$ (shape-scale notation), $b \sim \text{Triangular}(0.008, 0.012)$, $c \sim \text{Normal}(0.1, 0.01)$ (mean-deviation notation) and $\mu \sim \text{Uniform}(0.04, 0.16)$. As in the previous examples, the values of the parameters of each one of these probability density functions have been chosen so that the theoretical expectation of a, b, c and μ coincides with the numerical values $\hat{a} = 0.2$, $\hat{b} = 0.01$, $\hat{c} = 0.1$ and $\hat{\mu} = 0.1$, respectively. The initial iterates are assumed to be $S_0 = 5$, $I_0 = 60$ and $R_0 = 0$. The computational results are presented in Figure 4. Both plots are nearly identical, then showing convergence. Analogously to Figure 3, we start with a scenario where most of the people is infected and, after a while, the subpopulations stabilize in an endemic situation with some variations due to the uncertainty (confidence intervals). In this case, apart from the vital dynamics, recovered people may become susceptible before leaving the system.

- Case 5. We consider a SEIR model with independent inputs, having as distributions $a \sim \text{Gamma}(4, 0.05)$ (shape-scale notation), $b \sim \text{Triangular}(0.008, 0.012)$, $c \sim \text{Normal}(0.1, 0.01)$ (mean-deviation notation) and $\mu \sim \text{Uniform}(0.04, 0.16)$. The population has size $N = 65$ and the initial conditions are $S_0 = 5$, $E_0 = 0$, $I_0 = 60$ and $R_0 = 0$. The numerical results are shown in Figure 5. Both plots are nearly identical, then showing convergence. We have plotted the subpopulations and their confidence intervals, included the exposed. Here, the level of susceptibles is higher than in the previous cases, because during the exposed time, the individuals are not infective, reducing the number of people capable to infect.

5. Conclusions

Epidemic models consisting of non-linear difference equations allow us to get a better understanding of the spread diseases. The time-discrete models SIS, SIR, SIRS and SEIR have been commonly utilized to describe epidemics successfully. We have analyzed the effect of randomness in the input parameters, by using absolutely continuous distributions with no independence assumption. The solution of the random system of non-linear difference equations becomes a stochastic process in discrete time. Adaptive gPC allows us to approximate the solution in terms of multivariate polynomials, with coefficients that satisfy a deterministic system of non-linear recursive equations which is numerically solvable. Numerical experiments permit quantifying the uncertainty for the number of susceptible, exposed, infected and recovered individuals at each time-discrete step. The convergence is achieved quickly with a cheap computational expense. The analysis performed in this paper can be extended to other epidemiological models, like SEIQV or SEIVR models, where quantifying uncertainty can be a major goal.
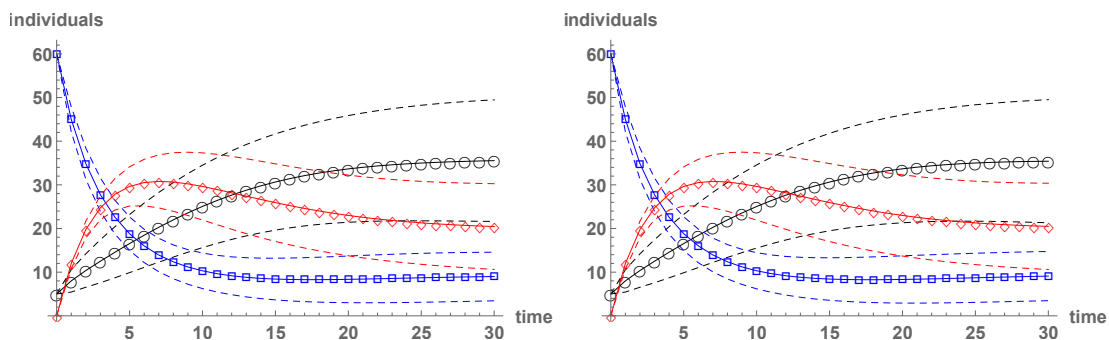
Math. Meth. Appl. Sci. 2018, 00 1–9
Prepared using mmaauth.cls

Copyright © 2018 John Wiley & Sons, Ltd.

7

Figure 3. Mean value estimates and confidence intervals (in dashed lines) for Case 3. Left: $p = 2$. Right: $p = 3$. S is denoted by $\bigcirc$ in black color, I is denoted by $\square$ in blue color, and R is denoted by $\diamond$ in red color.
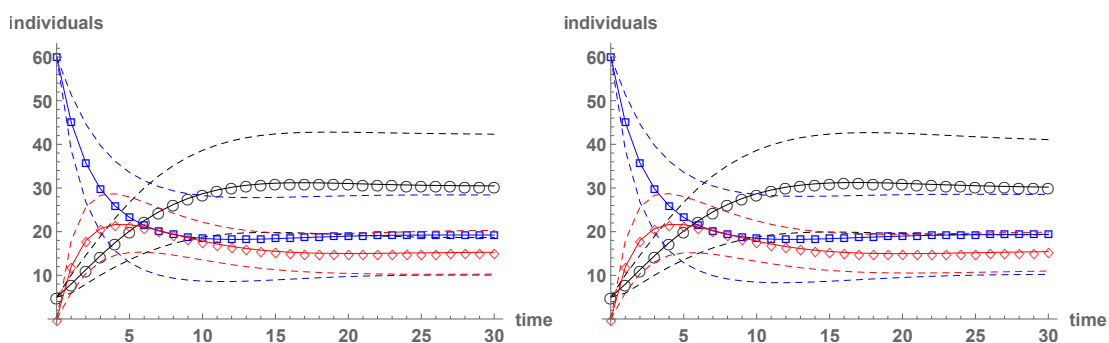

Figure 4. Mean value estimates and confidence intervals (in dashed lines) for Case 4. Left: $p = 2$. Right: $p = 3$. S is denoted by $\bigcirc$ in black color, I is denoted by $\square$ in blue color, and R is denoted by $\diamond$ in red color.
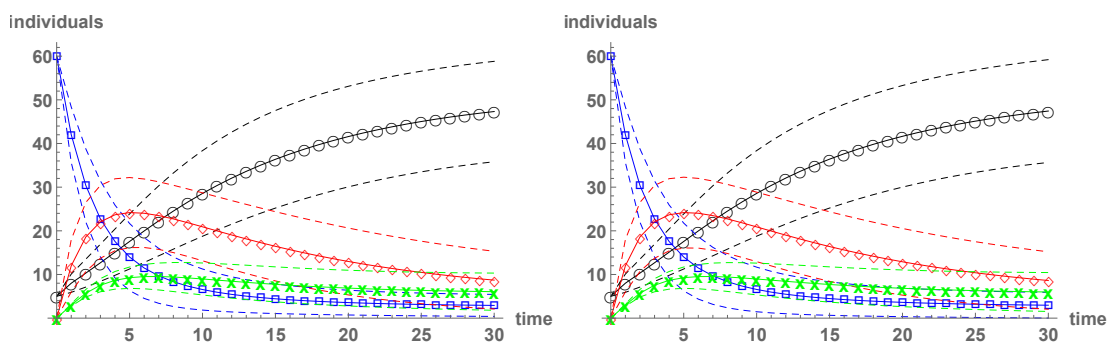

Figure 5. Mean value estimates and confidence intervals (in dashed lines) for Case 5. Left: $p = 2$. Right: $p = 3$. S is denoted by $\bigcirc$ in black color, E is denoted by $\times$ in green color, I is denoted by $\square$ in blue color, and R is denoted by $\diamond$ in red color.

Acknowledgements

Conflict of Interest Statement

The authors declare that there is no conflict of interests regarding the publication of this article.

8   Copyright © 2018 John Wiley & Sons, Ltd.

Math. Meth. Appl. Sci. **2018**, 00 1–9

Prepared using mmaauth.cls

References

1. Allen LJS. Some discrete-time SI, SIR, and SIS epidemic models. *Mathematical Biosciences*. 1994; 124 (1): 83–105.
2. Hethcote HW. The mathematics of infectious diseases. *SIAM review*. 2000; 42 (4): 599–653.
3. Levin SA, Hallam TG, Gross LJ. *Applied Mathematical Ecology*. Springer Science & Business Media; 2012.
4. Brauer F. *Compartmental models in epidemiology*. Mathematical epidemiology, pp. 19–79, Springer; 2008.
5. Huppert A, Katriel G. Mathematical modelling and prediction in infectious disease epidemiology. *Clinical Microbiology and Infection*. 2013; 19 (11): 999–1005.
6. Acedo L, Arenas AJ, González-Parra G. An exact global solution for the classical SIRS epidemic model. *Nonlinear Analysis: Real World Applications* 2016; 11 (3): 1819–1825.
7. Arenas AJ, González-Parra G, Chen-Charpentier BM. Construction of nonstandard finite difference schemes for the SI and SIR epidemic models of fractional order. *Mathematics and Computers in Simulation* 2016; 121: 48–63.
8. Khan MA, Khan Y, Islam S. Complex dynamics of an SEIR epidemic model with saturated incidence rate and treatment. *Physica A: Statistical Mechanics and its Applications*. 2018; 493: 210–227.
9. Fishman G. *Monte Carlo: Concepts, Algorithms, and Applications*. Springer Science & Business Media; 2013.
10. Xiu D. *Numerical Methods for Stochastic Computations. A Spectral Method Approach*. Princeton University Press; 2010.
11. Xiu D, Karniadakis GE. The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM Journal of Scientific Computation* 2002; 24 (2): 619–644.
12. Chen-Charpentier BM, Stanescu D. Epidemic models with random coefficients. *Mathematical and Computer Modelling*. 2010; 52 (7-8): 1004–1010.
13. Chen-Charpentier BM, Cortés JC, Licea JA, Romero JV, Roselló MD, Santonja FJ, Villanueva RJ. Constructing adaptive generalized polynomial chaos method to measure the uncertainty in continuous models: A computational approach. *Mathematics and Computers in Simulation*. 2015; 109: 113–129.
14. Cortés JC, Romero JV, Roselló MD, Villanueva RJ. Improving adaptive generalized polynomial chaos method to solve nonlinear random differential equations by the random variable transformation technique. *Communications in Nonlinear Science and Numerical Simulations*. 2017; 50: 1–15.
15. Cortés JC, Romero JV, Roselló MD, Santonja FJ, Villanueva RJ. Solving continuous models with dependent uncertainty: A computational approach. *Abstract and Applied Analysis*. 2013; 2013.
16. Wan X, Karniadakis GE. An adaptive multi-element generalized polynomial chaos method for stochastic differential equations. *Journal of Computational Physics*. 2005; 209 (2): 617–642.
17. Blatman G, Sudret B. An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis. *Probabilistic Engineering Mechanics*. 2010; 25 (2): 183–197.
18. Petterson MP, Iaccarino G, Nordström J. *Polynomial chaos methods for hyperbolic partial differential equations: Numerical Techniques for Fluid Dynamics Problems in the Presence of Uncertainties*. Mathematical Engineering, Springer International Publishing; 2015.
19. Pettersson P, Nordström J, Doostan A. A well-posed and stable stochastic Galerkin formulation of the incompressible NavierStokes equations with random data. *Journal of Computational Physics*. 2016; 306: 92–116.
20. Tang K, Congedo PM, Abgrall R. Adaptive surrogate modeling by ANOVA and sparse polynomial dimensional decomposition for global sensitivity analysis in fluid simulation. *Journal of Computational Physics*. 2016; 314: 557–589.
21. Paffrath M, Wever U. Adapted polynomial chaos expansion for failure detection. *Journal of Computational Physics*; 2007. 226 (1): 263–281.
22. Lucor D, Witteveen J, Constantine P, Schiavazzi D, Iaccarino G. Comparison of adaptive uncertainty quantification approaches for shock wave-dominated flows. In *Proceedings of the Summer Program*, p. 219; 2012.
23. Stanescu D, Chen-Charpentier BM. Random coefficient differential equation models for bacterial growth. *Mathematical and Computer Modelling*. 2009; 50 (5-6): 885–895.
24. Santonja F, Chen-Charpentier BM. Uncertainty quantification in simulations of epidemics using polynomial chaos. *Computational and Mathematical Methods in Medicine*. 2012; 2012.
25. Khan MA, Khan Y, Khan TW, Islam S. Dynamical system of a SEIQV epidemic model with nonlinear generalized incidence rate arising in biology. *International Journal of Biomathematics*. 2017; 10 (7): 1–19.
26. Khan MA, Khan Y, Badshah Q, Islam S. Global stability of SEIVR epidemic model with generalized incidence and preventive vaccination. *International Journal of Biomathematics*. 2015; 8 (6): 1–19.
27. Chen-Charpentier BM, Stanescu D. Biofilm growth on medical implants with randomness. *Mathematical and Computer Modelling*. 2011; 54 (7–8): 1682–1686.
28. Marzouk YM, Najm HN, Rahn LA. Stochastic spectral methods for efficient Bayesian solution of inverse problems. *Journal of Computational Physics*. 2007; 224 (2): 560–586.
29. Marzouk Y, Xiu D. A stochastic collocation approach to Bayesian inference in inverse problems. *Communications in Computational Physics*. 2009; 6 (4): 826–847.
30. National Spanish Health Survey (Encuesta Nacional de Salud de España, ENSE). `http://pestadistico.inteligenciadegestion.msssi.es/publicoSNS/comun/ArbolNodos.aspx` [accessed 28.05.18].

Math. Meth. Appl. Sci. 2018, 00 1–9
Prepared using mmaauth.cls

Copyright © 2018 John Wiley & Sons, Ltd.   9