Research Paper

Global sensitivity analysis of a single-cell HBV model for viral dynamics in the liver

Md Afsar Ali [a], *, S.A. Means [b], Harvey Ho [c], Jane Heffernan [a]

[a] York University, Canada
[b] Massey University, New Zealand
[c] The University of Auckland, New Zealand

ABSTRACT

The predictive accuracy of mathematical models representing anything ranging from the meteorological to the biological system profoundly depends on the quality of model parameters derived from experimental data. Hence, robust sensitivity analysis (SA) of these critical model parameters aids in sifting the influential from the negligible out of typically vast parameter regimes, thus illuminating key components of the system under study. We here move beyond traditional local sensitivity analysis to the adoption of global SA techniques. Partial rank correlation coefficient (PRCC) based on Latin hypercube sampling is compared with the variance-based Sobol method. We selected for this SA investigation an infection model for the hepatitis-B virus (HBV) that describes infection dynamics and clearance of HBV in the liver [Murray & Goyal, 2015]. The model tracks viral particles such as the tenacious and nearly ineradicable covalently closed circular DNA (cccDNA) embedded in infected nuclei and an HBV protein known as p36. Our application of these SA methods to the HBV model illuminates, especially over time, the quantitative relationships between cccDNA synthesis rate and p36 synthesis and export. Our results reinforce previous observations that the viral protein, p36, is by far the most influential factor for cccDNA replication. Moreover, both methods are capable of finding crucial parameters of the model. Though the Sobol method is independent of model structure (e.g., linearity and monotonicity) and well suited for SA, our results ensure that LHS-PRCC suffices for SA of a non-linear model if it is monotonic.

1. Introduction

Liver infection with hepatitis B is a life-threatening disease—causing hepatocellular carcinoma, liver cirrhosis, liver damage and failure with an estimated death total of at least 887, 000 worldwide each year (Elgouhari et al., 2008; Tarao et al., 2019; World Health Organization, 2015). About 250 million people worldwide suffer from chronic hepatitis B virus (HBV) infection, and few effective treatment options exist (Aparna et al., 2015; Means et al., 2020).

* Corresponding author.
E-mail addresses: mali06@yorku.ca (M.A. Ali), s.means@massey.ac.nz (S.A. Means), harvey.ho@auckland.ac.nz (H. Ho), jmheffer@yorku.ca (J. Heffernan).
Peer review under responsibility of KeAi Communications Co., Ltd.

A better understanding of HBV infection in-host is needed to better inform pharmaceutical intervention. The discovery of effective treatments critically hinges on the characterization of the HBV infection and replication dynamics within hepatocytes, cells in the liver (Fausto & Campbell, 2003). HBV virus replication requires the development of closed covalent circular DNA (cccDNA) from the viral DNA that is injected into the cell upon infection (Guo & Guo, 2015). During infection, cccDNA accumulates in the cell nuclei in which it persists as a stable episome and functions as a template for the transcription of viral genes. Chronic HBV infection is maintained in cells by a replicative form of HBV cccDNA (Liu & Xin, 2019; Shi & Shi, 2009). As such, HBV cccDNA has been identified as a potential and important target for therapeutic intervention, and represents a focus for antiviral drug discovery.

In a recent study of HBV infection, Murray and Goyal (Murray & Goyal, 2015) developed a model of HBV replication in a single hepatocyte. The model explicitly considers cccDNA, and it tracks intracellular and extracellular virus particles. The model has since been extended to study cell-cell transmission (Goyal & Murray, 2016) and has been extended to a study of a hepatocyte-sinusoid model (Cangelosi et al., 2017). As such, the single-cell model developed by Murray and Goyal (Murray & Goyal, 2015) has been shown to provide an important building block over which an HBV infection model of a liver could be constructed. The model thus also provides an important base to study potential drug targets.

Herewith, we conduct a sensitivity analysis on the Murray and Goyal model (Murray & Goyal, 2015). We apply two global SA techniques, LHS PRCC (Blower & Dowlatabadi, 1994; Marino et al., 2008; Zheng & Rundell, 2006), and a variance-based Sobol method (Sobol', 1990; Sobol', 2001) to identify key parameters that most drive HBV replication in a liver cell. We pay particular attention to infection processes involving cccDNA. The outcomes of our study are two-fold. Our SA results can be used to determine key parameters that affect cccDNA, which can, in turn, be used to inform drug therapy development. Additionally, the SA results can be used to inform the future construction of a spatial model of the liver from the Murray and Goyal model building block.

2. Overview of sensitivity analysis, computing methods and mathematical model

2.1. Uncertainty, sensitivity and scale

Uncertainty in model parameters exists. While estimates of some model parameters may exist, there is no true measurement of a parameter that is universal. Uncertainty in model parameters will affect confidence in predictions of mathematical models. Uncertainty analysis (UA) aims at providing a measurable degree of confidence to address this complication (Marino et al., 2008). Falling under this umbrella of uncertainty, is the sensitivity of model outcomes to model parameters (Helton & Davis, 2002; Helton & Oberkampf, 2004). Sensitivity analysis (SA) itself can assist in efficient parameter calibration and fortify models against over-parameterization (Van Griensven et al., 2005) − further relieving modelers of excessive effort. Two scales of SA emerge on both a local (LSA) and global (GSA) (Van Griensven et al., 2005; Zhan et al., 2013) range. Whereas LSA aims at individual parameter influences, GSA grapples with entire regions for families of parameters and quantification of their interactions on output variables (Zi, 2011). Given likely multitudes of dimensions to parameter spaces and constraints of input-output realism for model predictions, SA methods are an essential addition to a modeler's toolkit for refinement and simplicity. We here consider two GSA approaches mentioned: LHS-PRCC and the variance-based Sobol method, which are briefly described in turn here, and provide modestly more detail in the Appendix [A, B, C]. Substantial introductions to these methods are available where we refer the reader as noted. General working steps of sensitivity analysis are listed concisely in the diagram presented in Fig. 1.

2.2. Latin hypercube sampling-LHS

Introduced by Mckay et al. (Helton & Davis, 2003; Iman & Conover, 1982; Iman et al., 1981; McKay et al., 1979), LHS generates samples of model inputs arrayed across a 'hypercube' whose dimension corresponds to the number of model parameters; call this dimension p. Utilizing parameter ranges partitioned into intervals, LHS deploys a selected probability density function for sampling parameter values from within these intervals that are then paired (in p-dimensional tuples) with samples, in like manner, for the entire suite of parameters. Simulations of the model are then performed iterating over all the p-sized parameter tuples. Depending on the number of interval partitions, this can result in a substantial number of p-sized parameter suites for testing − but vastly fewer than required for interrogating the entire parameter space.

Results of simulations across the LHS-sampled space are compared for influence with Pearson correlations − here a 'Partial rank correlation coefficient' (PRCC). Resulting correlations may be negative to positive in the range [-1, 1] as is usual (Marino et al., 2008) − illuminating the influence of parameters as either amplifying or dampening model outputs. Generally, PRCC analysis provides a measurement of the nonlinear, but monotonic, relationship between a model output and the model parameters. Examples in disease modelling can be found in (Blower & Dowlatabadi, 1994; Marino et al., 2008; Zheng & Rundell, 2006), and many other studies.

2.3. Sobol method

Alternatively, the output variables of a nonlinear model are amenable to Analysis of Variances used with the Sobol technique (Sobol', 1990; Sobol', 2001). An orthogonal decomposition of the model into components for each input parameter
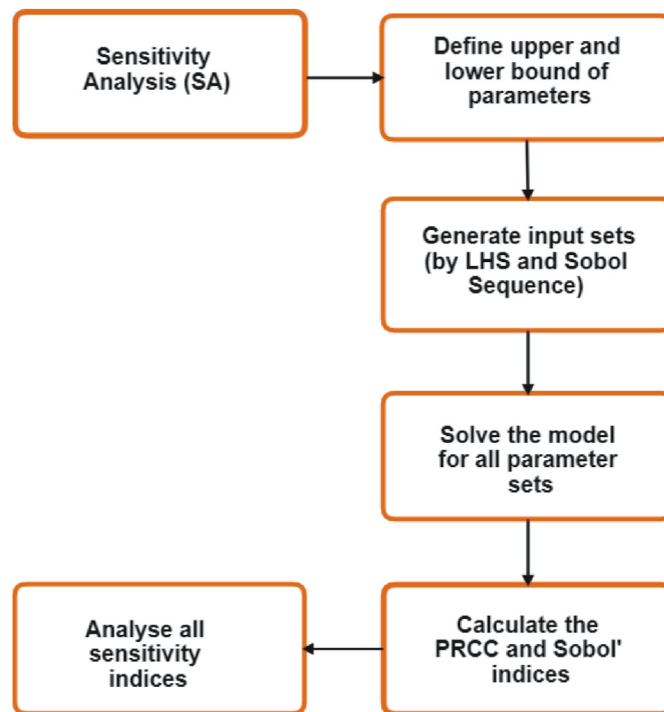
Fig. 1. Working steps of sensitivity analysis are listed concisely in the diagram, and described below: (a) Choose a sensitivity analysis method based on the number of model evaluations, the correlation structure of input-output parameters and an experimental design. (b) Identify simulation input-output parameters of interest and what input factors are needed to include in the analysis. (c) Define upper and lower bound of input parameters by taking respectively 15 − 20% more and less from the mean value of each parameter and generate input sets of parameters by LHS method and Sobol sequence. An input set has the form of N strings of input factor values on which the model is evaluated. (d) Evaluate the model on the generated input samples and produce the outputs, which contain N output values in the desired form. (e) Analyse the model outputs and draw conclusions depending on PRCC and Sobol indices calculated using the outputs. (f) Based on the observation of model outputs parameter sample size is determined. More detail on sample size can be found in Appendix D.

is assembled across a p − dimensional normalized hypercube, I^p, with I over [0,1] and p again the dimension of parameter space. With such a decomposition of the model combined with a uniform sampling of the p-dimensional hypercube, variances per model output are computed. These variances are in turn utilized for calculation of the Sobol index − a metric of influence for a given parameter. Two such indices are computed revealing the individual and the total influence of a parameter, denoted S_i and ST_i, whose values fall between [0, 1]. When, for instance, $S_i = ST_i = 0$, the parameter influence over the model is nil, but, by contrast, supreme when $S_i = ST_i = 1$. Typically, Sobol indices with values greater than 0.05 are considered significant. Unlike the PRCC results, however, Sobol coefficients do not indicate positive or negative influence but merely significance.

2.4. Single-cell model of hepatitis B viral dynamics in the liver

We performed global sensitivity analysis of a HBV model (Murray & Goyal, 2015) representative of single-cell viral infection dynamics in the liver (see Fig. 2). HBV invades hepatocytes via the NTCP and establishes the stable covalently-closed circular viral DNA (cccDNA) in the nucleus. Subsequent transcriptions are two intermediate forms such as the single-stranded DNA (ssDNA) and the dual-strand DNA (dsDNA). Of the milieu of viral proteins produced by HBV, the model tracks *p36* whose action determines whether the dsDNA intermediate continues to a complete HBV and is released, or instead reinforces the nucleic cccDNA pool. In the model, the intracellular numbers of ssDNA, dsDNA, and infecting rcDNA are respectively denoted by S, D, and R. The number of cccDNA copies in the cell is denoted by C; the number of protein (p36) molecules inside and outside of the cell are respectively denoted by P and P_E, and the number of virions in serum originated from the cell is denoted by V. The intra-cellular viral replication dynamics in a single cell, shown in Fig. 2, is modelled by the following set of time-delayed differential equations:
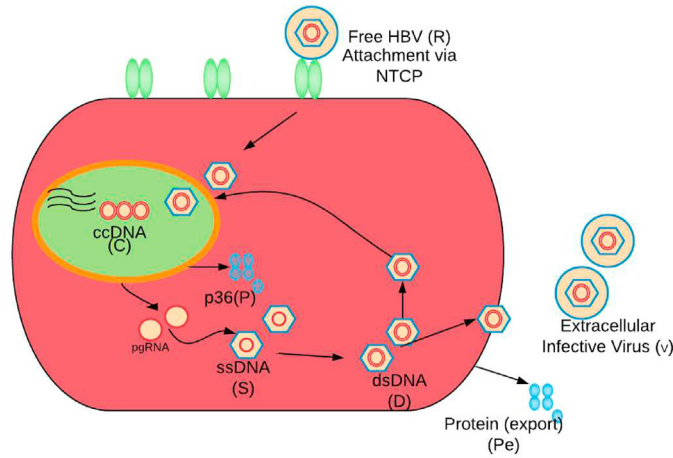
Fig. 2. The schematic shows the HBV replication cycle in a single cell. HBV enters the hepatocyte from outside of the cell through NTCP, denoted here as variable R, and the genome (cccDNA) is transferred to the nucleus, labeled as C. By the transcription of cccDNA, protein (p36) and pre-genomic RNA (pgRNA) are produced. These are cytoplasmic pgRNA and are packaged with polymerase and envelope proteins into nucleocapsids, and by the reverse transcription, RNA is converted into DNA. In this process, first single-stranded DNA (ssDNA), labeled S, and then double-stranded DNA (dsDNA), denoted D, are produced. Depending upon the level of p36 proteins in the cytoplasm, labeled P, dsDNA will either return to the nucleus or will be released as complete infective HBV virions labeled V outside of the cell to infect other hepatocytes in the liver.

$$
\begin{cases}
\dfrac{dR}{dt} = kV(t-\tau) - \mu_R R - b_R e^{(-\lambda P)} R \\[2mm]
\dfrac{dC}{dt} = b_R e^{-\lambda P(t-\tau)} R(t-\tau) + b_D e^{-\lambda P(t-\tau)} D(t-\tau) - \mu C \\[2mm]
\dfrac{dS}{dt} = aC(t-\tau) - b_S S \\[2mm]
\dfrac{dD}{dt} = b_S S(t-\tau) - b_D D \\[2mm]
\dfrac{dP}{dt} = a_P C(t-\tau) - b_P P \\[2mm]
\dfrac{dP_E}{dt} = b_P P(t-\tau) - c_{P_E} P_E \\[2mm]
\dfrac{dV}{dt} = b_D(1 - e^{-\lambda P(t-\tau)}) D(t-\tau) - c_V V
\end{cases}
\tag{1}
$$

Here, k is a rate at which virions infect a cell and produce rcDNA which is lost at a rate μ_R; depending on the level of p36, rcDNA is transported to the nucleus at a rate b_R and there at the same rate rcDNA is converted to cccDNA which is lost at a rate μ; cccDNA is converted to ssDNA at a rate a and is lost at a rate b_S; ssDNA is converted to dsDNA at a rate b_S and is also lost at a rate b_D; a_P is the rate of synthesis of protein p36 which is lost at a rate b_p; protein (p36) is exported outside of the cell at a rate b_P and lost at a rate c_{PE}, and virions are released from the cell at a rate b_D and lost in serum at a rate c_V. λ maintains the average level of p36 that directs R and D to nucleus, and V to the outside of the cell. The time delay, τ, is considered as 30 min in a day for all simulations. The parameter names with some symbolic modifications without changing their meaning and values are adopted from the original paper published by Murray and Goyal (Murray & Goyal, 2015). The list of parameters with their mean values and ranges used for the generation of parameter samples for SA is given in Table 1.

The model in Eq. (1) is completed by adding initial conditions. For all simulations, we use initial conditions for all populations as: $(R, C, S, D, P, P_E, V) = (0, 0, 0, 0, 0, 0, 1)$. Solving the model and performing the sensitivity analysis is completed as described in Appendix E "Computational Aspects".

Table 1

Model parameters used in the simulation are taken predominantly from those used by Murray and Goyal (Murray & Goyal, 2015) with a slight symbolic modification for the visualization of outputs corresponding to the respective parameters with a clear distinguish-ability. As we do not have realistic ranges for all parameters, instead of varying some in realistic ranges, and others over wide ranges, we choose to vary all in wide ranges from 50 to 150%. Using a wide range for each parameter, we have a big parameter space to compare Sobol and LHS-PRCC methods.

| Parameter	Symbol (in (Murray & Goyal, 2015))	Symbol (in our study)	Value ($\frac{1}{day}$)	Reference
Cell infection rate	k	k	0.3	(Cangelosi et al., 2017; Murray & Goyal, 2015)
cccDNA synthesis rate	b	b_R	$log(2)$	(Cangelosi et al., 2017; Murray & Goyal, 2015)
Conversion rate of cccDNA to ssDNA	a	a	50	(Cangelosi et al., 2017; Murray & Goyal, 2015)
Conversion rate of ssDNA to dsDNA	b	b_S	$log(2)$	(Cangelosi et al., 2017; Murray & Goyal, 2015)
$P36$ synthesis rate	a_P	a_P	$1000 \times a$	(Cangelosi et al., 2017; Murray & Goyal, 2015)
[1ex] $P36$ exporting rate	b_P	b_P	$log(2)$	(Cangelosi et al., 2017; Murray & Goyal, 2015)
dsDNA transportation rate to the nucleus/Virions release rate	b	b_D	$log(2)$	(Cangelosi et al., 2017; Murray & Goyal, 2015)
rcDNA degradation rate	μ_R	μ_R	$log(2)$	(Cangelosi et al., 2017; Murray & Goyal, 2015)
cccDNA degradation rate	μ	μ	$log(2)/50$	(Cangelosi et al., 2017; Murray & Goyal, 2015)
p36 influence on R and $dsDNA \rightarrow C$, and on R to export V	λ	λ	1/100000	(Cangelosi et al., 2017; Murray & Goyal, 2015)
p36 degradation rate	c	c_{P_E}	$24 \times log(2)/4$	(Cangelosi et al., 2017; Murray & Goyal, 2015)
Virions degradation rate	c	c_V	$24 \times log(2)/4$	(Cangelosi et al., 2017; Murray & Goyal, 2015)
Time delay	τ	τ	30/1440	(Cangelosi et al., 2017; Murray & Goyal, 2015)

3. Results

3.1. Time-evolution of cccDNA population

Typical cccDNA replication dynamics for each sample of parameters show a disease equilibrium state, as shown in Fig. 3(a). Copies of cccDNA appear to be in steady-state after about 12 days of infection for each set of parameters, though some cccDNA curves appear in steady-state afterwards. All cccDNA curves show a steady state at the end of the simulation (at $t = 300$ days) with substantial variation in magnitudes as shown by the histograms in Fig. 3(b); non-zero particle counts at the steady-state equilibrium of cccDNA indicates the chronic HBV infection with the immune-tolerant status of patients (Wu & Chang, 2015). Variations in the steady-state cccDNA copy levels illustrate the sensitivity of this particular model output to the range of inputs tested. Precisely which parameters hold the greatest influence we consider next.

3.2. Scatter plots: monotonic relationship between input and output variables

Simulation results of the model (Murray & Goyal, 2015) based on LHS samples of size 5200 are visualized by scatter plots (shown in Fig. 4) and they demonstrate a monotonic relationship between all output variables and input variables. These apparent monotonic relationships suggest the PRCC analysis is thus suitable for application to the HBV model utilized here. We see a wide range of sensitivities varying from strong sensitivity to negligible. In fact, all outputs except $p36$ increase monotonically with the increase of b_P parameter values, all outputs except $p36$ released from the cell decrease monotonically with the increase of a_P values, and the increase of λ contributes to a monotonic decrease of all outputs. The values of parameters k, a, b_S and C_{PE} contribute to the nominal change of respective outputs (results not shown). Thus, the parameters k, a,
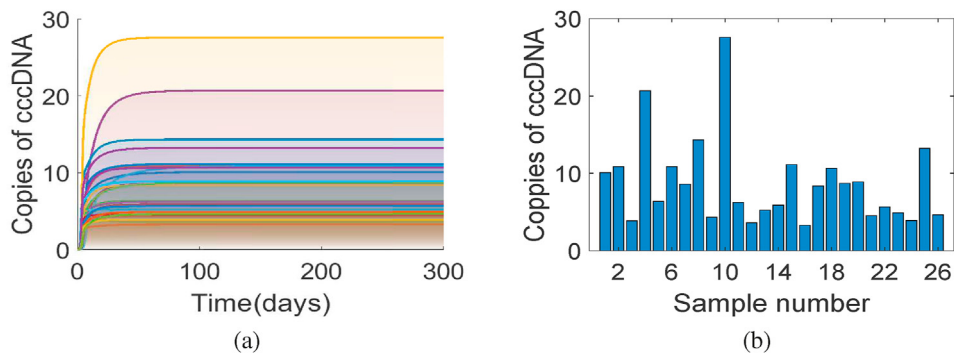
Fig. 3. Simulation runs are carried out for all samples of size $N = 5200$ and the time evolution of all cccDNA curves is shown in Appendix F. For visibility, a few cccDNA curves are plotted showing the evolution of cccDNA copies over time in Figure (a). In Figure (b), for each of these cccDNA curves, the number of cccDNA at $t = 300(days)$ is plotted to show the variation of cccDNA copies for different sets of input data.

b_S and C_{PE} are nominally sensitive or insensitive (results not shown), and b_P, a_P and λ are strongly sensitive to the outputs, but the sensitivity of outputs to other parameters are not observed (results not shown).

cccDNA dictates infection intensity and duration of HBV, so it is instrumental to know which parameter contributes most to the cccDNA replication. The scatter plots as shown in Fig. 4 indicate that three parameters - protein synthesis rate (a_P), protein export rate (b_P) and rcDNA transportation rate (λ) to nucleus are highly correlated with cccDNA replication within the cell. We observe here a positive correlation to protein export rate (b_P) but a negative correlation to protein synthesis rate (a_P) and rcDNA transportation rate (λ). The positive correlation of cccDNA replication to b_P ensures that if the rate increases, i.e. if protein p36 exports enough to maintain the protein level in the cell, then cccDNA levels double (Murray & Goyal, 2015). On the other hand, a negative correlation of cccDNA to the parameter a_P shows that if the values of a_P increase, the protein level in the cell is balanced at a constant level by the loss of it at a rate b_P leading to reduced transport of rcDNA to the nucleus; thus the values of cccDNA fall. The parameter λ exhibits similar behaviour to a_P. We see similar results with the PRCC and Sobol index analysis discussed in the following section. Overall, we see the level of protein *p36* produced by the infected cell has an immense effect on the replication of cccDNA − and this is reflected by the model structure (Murray & Goyal, 2015). *p36* concentrations are modulated by both feedback regulation or simple export to the extracellular space; moreover, cccDNA replication is inversely related to *p36* levels and hence key to the HBV infection dynamic. For other parameters, such as a, b_R, b_S, b_V, μ_R, μ, c_{PE}, and c_V, the synthesis of cccDNA within the infected cell has a nominal (or no) effect (results not shown).

3.3. Time-varying sensitivity analysis

Parameter influence may vary over the temporal evolution of the model, and sensitivity analysis allows us to assess how their significance varies over the time interval (Marino et al., 2008). For both acute and chronic HBV infections, parameter influence on the model output may vary over time. Consider our results in Fig. 3 for case (a), where the curves of cccDNA copies initially increase exponentially and after a course of time they reach a steady-state for all parameter samples. After cccDNA replication starts, HBV infection turns into an acute phase during early time points, when the autoimmune system remains activated. This subsequently turns into the persistent mode of varying degrees over time for all parameter samples. cccDNA persistence after six months in the cell is considered a chronic infection: the immune system failed to clear the virus (Ganem & Prince, 2004; Smolders et al., 2020). Thus the time points at early days of infection (roughly in the acute phase) and the time points at which cccDNA levels equilibrate at the chronic phase are crucial for pathological, physiological and pharmaceutical aspects. The fortunate case of acute infection and clearance we set aside, and instead focus on parameter values leading to such chronic infection outcomes. In this regard, where cccDNA copies increase, propagate, and persist at steady state for further replication is of interest. When specific time points are not provided, temporal sensitivity analysis may identify a significant time-dependent relationship between inputs and outputs for the whole course of simulation time.

In order to better capture the natural variability of HBV infection processes over time, for all parameters, we have calculated both PRCC and Sobol indices considering model outcomes at some time points chosen over the simulation time. We present results here for the parameters a_P, b_P and λ in Fig. 5. A general trend of piece-wise linear and exponential progressions occur for the PRCC (measuring monotonic sensitivity) and Sobol (measuring contribution to variability) indices, respectively. Both indices exhibit strong sensitivity to the parameters b_P − the protein export rate, a_P − the protein (p36) synthesis rate, and λ − the rate of influence of protein on rcDNA to direct it to nucleus. A notable exception is that p36 is not sensitive to the parameters a_P and b_P, and export of p36 is not sensitive to the parameter a_P.

Because of strong positive correlation of b_P to cccDNA, as seen in Fig. 4, changes in parameter b_p induce a linear incremental effect on cccDNA synthesis, which means that with a linear increase of the rate of removal of protein molecules p36 created in the nucleus by cccDNA, the replication of cccDNA occurs with an overall linear increase. On the other hand, the parameters a_p
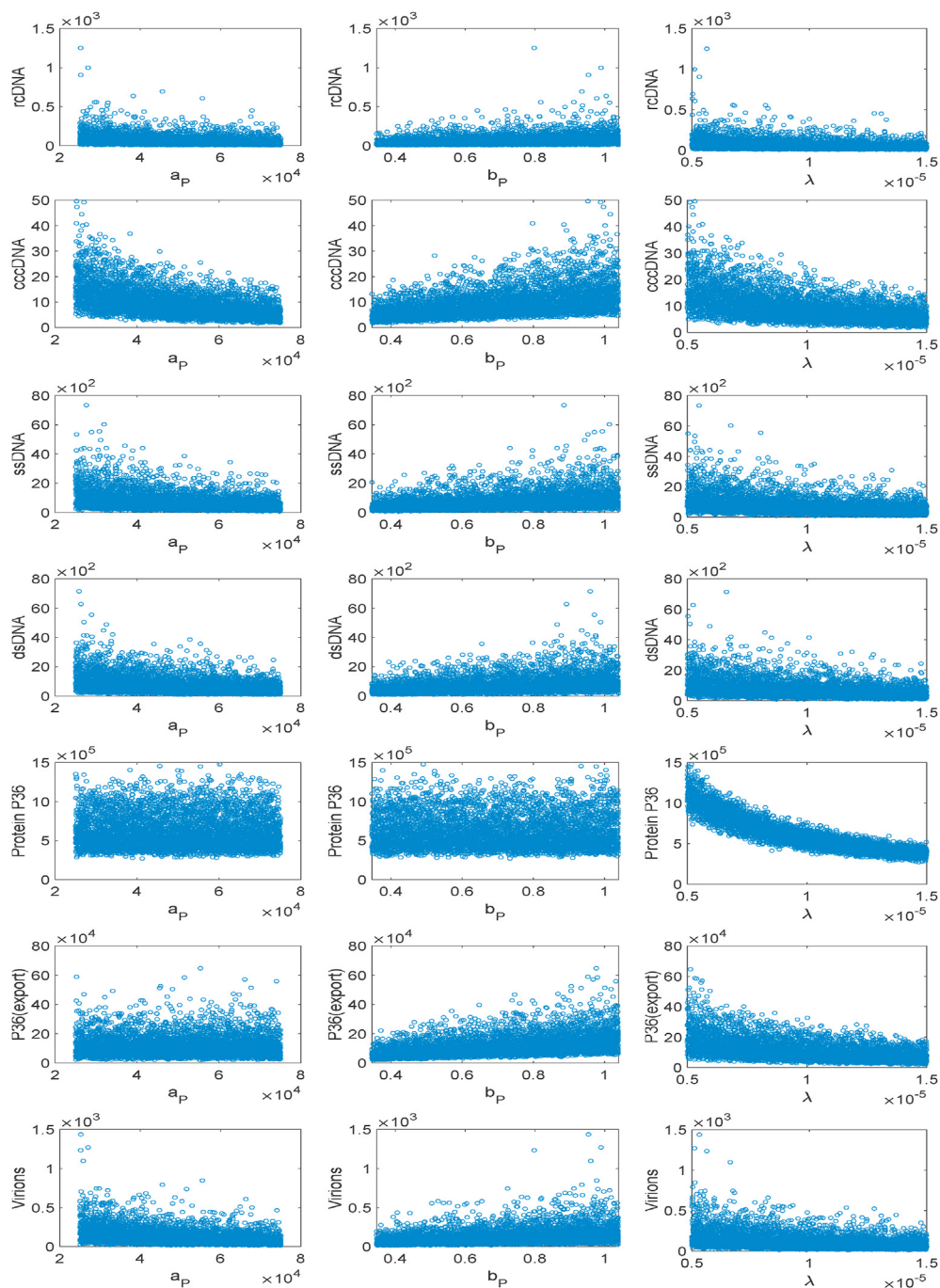
Fig. 4. Model is simulated using Latin hypercube samples of size $N = 5200$ and the scatter plots are depicted for each population with respect to different interesting parameters to show the monotonicity of the model outputs over the parameter regime.

and λ are strongly and negatively correlated (see Fig. 4) to cccDNA replication while the infection progresses monotonically to its steady-state (see Fig. 3).

The correlation between parameters and outputs is further annotated in the PRCC and Sobol indices analysis. The time-varying PRCCs, as seen in Panel (A) in Fig. 5, show a strong positive sensitivity of ccDNA to the parameter b_p, after a short time of infection (early time points) and overall, progress constantly over time. Considering the impact of b_p – the rate at which $p36$ is released into the extracellular space – we see that the positive sign of PRCC for b_p indicates that if parameter b_p is increased, cccDNA increases (and vice versa) over time (Murray & Goyal, 2015). On the other hand, the negative sign of PRCC

(A)   (B)   (C)
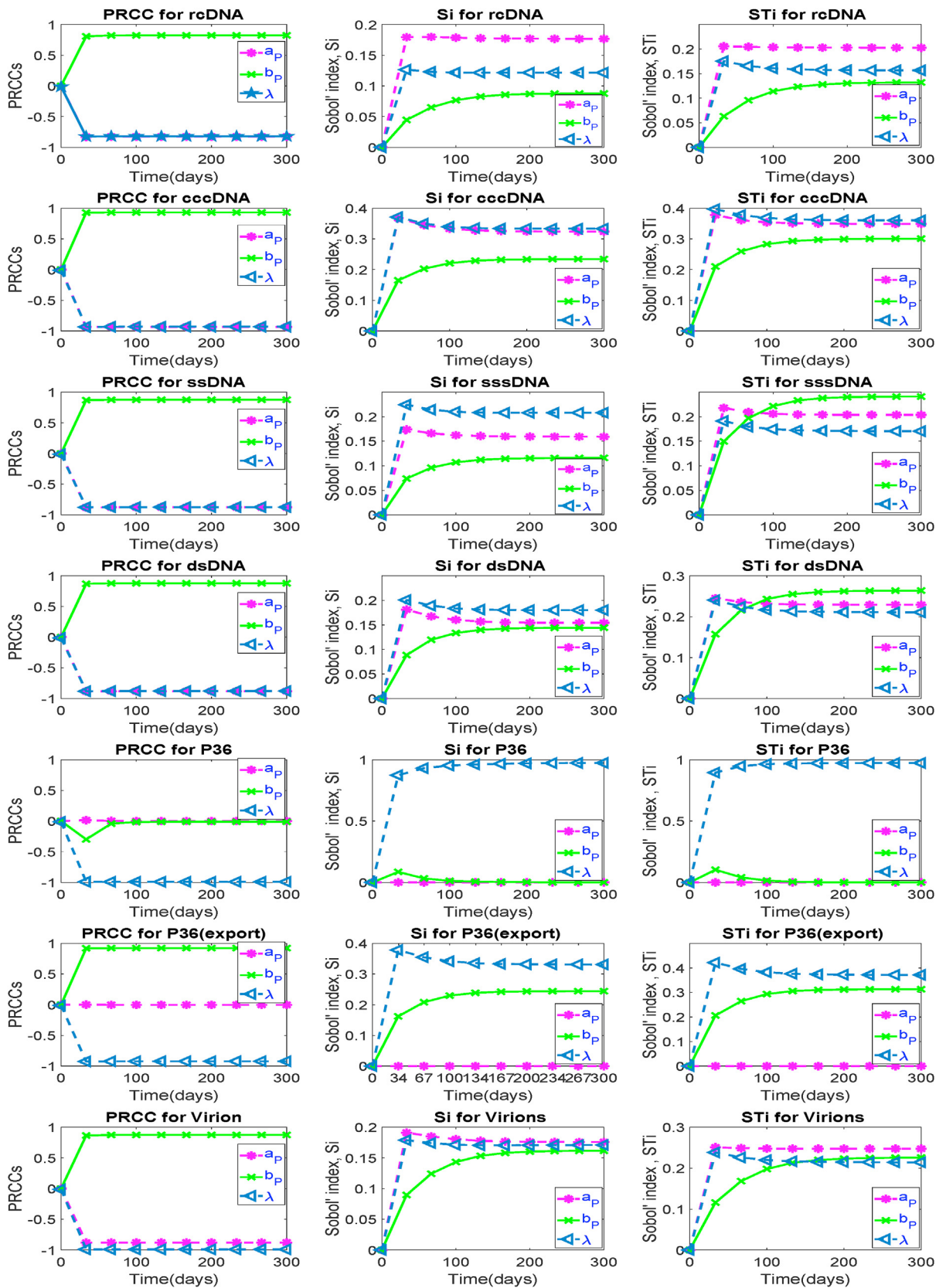
for a_p and λ (see in Panel (A) in Fig. 5) suggests that if parameter a_p and λ decrease, cccDNA synthesis decreases (and vice versa), i.e., the level of infection caused by HBV decreases over time. Sobol indices are positive and they progress exponentially over time, indicating that the parameters - a_p, b_p, and λ are sensitive to the respective outputs with some exemptions over time.

Other parameters, such as k, b_R, a, b_S, b_R, μ_R, μ, c_{PE}, and c_V remain insignificant over time with regards to the equilibrium state of cccDNA according to the PRCC method (Appendix H). Consistent with this PRCC result, these parameters according to the Sobol indices remain insignificant as cccDNA replication progresses to the equilibrium state (Appendix H).

4. Discussion and conclusion

Uncertainty and sensitivity analyses are capable of evaluating a model's effectiveness, and determining what factors affect model outputs. We have here investigated the sensitivity and interaction of parameters of a single-cell HBV model, illuminating model behaviour and HBV infection phenomena in a single cell. Two specific types of sensitivity analysis methods considered reliable and efficient, namely, a method based on LHS sampling (Partial rank correlation coefficient-PRCC) and a method based on variance (Sobol method) are compared. We apply them to the HBV model chosen, identify the critical parameters characterizing their influence on the model outputs, and compare the sensitivity indices for both methods. The relative merits of both approaches are further considered: each identifies similar parameter influences, yet complement with insights into their respective importance.

Both methods, for our model, are very reliable and accurate in determining the most important parameter that has the greatest effect on a specific output. Based on the PRCC indices and the Sobol first-order and total sensitivity indices, Si and STi, of the twelve parameters of interest, three parameters-a_P, λ and b_P are found to be the leading contributors to the variance in population size of cccDNA, and nine are found to be poorly significant or even insignificant and have very low effects on the HBV model's outcomes as well as on the intensity of HBV infection in the liver. The three significant parameters are characterised by both methods as the most sensitive (treated as crucial parameters) to the outcomes of the HBV model.

Besides these three parameters-a_P, λ and b_P, PRCC shows that some other parameters are significant with a moderate effect on outputs, whereas the Sobol method is very strict to predict that only the parameters a_P, λ and b_P are the most sensitive to the model outcomes. Therefore, the Sobol method is more robust in determining the most critical parameters compared to the PRCC method. Moreover, the Sobol indices show overall exponential progression over time, which may provide more insight on the time dependency of outputs on the parameters than PRCCs as they progress constantly over time. In the case of screening parameters during model building and simulation for a complex system, thus sensitivity analysis via the Sobol method would be precise, contributory and reliable. However, our SA results overall indicate that the PRCC and Sobol methods agree with each other.

PRCC is able to find both positive and negative correlations on model outputs. In our case, PRCC analysis reveals that the parameters, a_P and λ have a strong negative correlation to cccDNA replication, but b_P has a strong positive correlation with cccDNA outcome; both pieces of information are very important to take control measures against HBV viral infection and propagation in the liver. On the other hand, the Sobol indices measure just the significance of parameters with- and - without taking the effect of parameter interactions, thus it may provide reliable information for a complex system.

PRCC is efficient and brings useful insights on global sensitivity. However, PRCC requires monotonic relationships between parameters and model outputs. The Sobol method can deal with non-linear models efficiently and rigorously even though the model outputs are not monotonically related to the model parameters. In our case, scatter plots ensure that outputs are monotonic with input variables; so the PRCC method functions similarly in determining the crucial model parameters, as the Sobol method does. So, for the sensitivity analysis of this particular model, the Sobol method does not play a superior role to PRCC.

The sensitivity analysis of the single-cell HBV model using both SA methods suggests that the protein level in the infected hepatocyte (which can be controlled by adjusting the production of p36 molecules and exporting them outside the cell) is crucial for cccDNA replication. With this, three parameters a_p —protein synthesis rate, λ —P36 influence and b_p — protein exporting rate are termed as crucial to cccDNA replication and establishment in the nucleus, leading to chronic infection. To control chronic HBV infection, a pharmaceutical protocol for the development of HBV drug and selection of an appropriate dose of any anti-viral medicines may be built up through the adjustment of the crucial parameters.

The temporal variability of state variable response to parameters is rarely addressed for research into infectious disease modeling (Wu et al., 2013). Modelers should investigate this temporal effect as their assumptions about the value of specific parameters may be wrong, or they may misinterpret the robustness of their results if sensitivity is only tested at a single point in time. For example, if sensitivity is measured at a point after the epidemic peak, it may seem a meaningless consideration for parameters that are instrumental to the increasing epidemic curve over time. For our model, if we calculate sensitivity indices at the time point where acute HBV infection occurs, we might miss measuring the sensitivity of parameters to the chronic

Fig. 5. Model is simulated using samples of size $N = 5200$ and the time-varying PRCC and Sobol indices (both 1st order, Si and total effect, STi) are calculated for each population (mean value) at different time points and plotted for key parameters in the model identified to show the monotonicity of sensitivity indices over time. Recall, the PRCCs in the interval [− 0.5 0.5] are not significant, otherwise significant. For Sobol method, parameters are significant when indices are in the interval [0.05 1.0].

HBV infection (i.e. the persistence of HBV for a long time in the nuclei). The temporal sensitivity analysis of parameters to the model outputs resolves this efficiently.

HBV infection and propagation may depend on spatial aspects of liver geometry. We, therefore, want to build a spatial model of HBV (Cangelosi et al., 2017; Wang & Wang, 2007) incorporating spatial factors such as cell-cell infection and sinusoidal structures. We also suggest that the LHS-PRCC and Sobol methods are valuable for investigating parameter sensitivity in a spatial model depending on its linearity/non-linearity.

Acknowledgements

Declaration of competing interest

The authors declare that they have no conflict of interest.

Appendices

Appendix A. Latin hypercube sampling

Latin hypercube sampling is a class of stratified Monte Carlo sampling methods without replacement, and was introduced by Mckay et al. (Helton & Davis, 2003; Iman & Conover, 1982; Iman et al., 1981; McKay et al., 1979). It aims to generate samples of model inputs for investigating the sensitivity of their influence on the model outputs via PRCC analysis (see below). LHS generates a sample of size N for each of the p variables $X = [X_1, X_2, ..., X_p]^T$ with the probability density function $f(X)$ in the following manner. The range of each parameter is divided uniformly into N disjoint intervals of equal probability $1/N$. From each interval, one value is selected randomly according to the probability density in the interval. The N values of X_1 are thus obtained and paired in a random manner with the N values of X_2. These N pairs are then combined with the N values of X_3 to form N triplets, and continued until a set of N p-tuples is formed. Thus, for a given sample size N and the number of parameters p, $(N!)^{p-1}$ possible interval combinations for Latin hypercube sample are formed into a matrix sized $N \times p$. Simulations of the mathematical model are then performed iterating over the N parameter suites.

Appendix B. Partial rank correlation coefficient (PRCC)

Partial rank correlation coefficient (PRCC) method is a sampling-based sensitivity analysis method that calculates the partial rank correlation coefficients for the model inputs and outputs. For two variables, X_i (input variable) and Y (output variable), a normal correlation coefficient C is determined as follows:

$$C(X_i, Y) = \frac{Cov(X_i, Y)}{\sqrt{Var(X_i)Var(Y)}} = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_i (X_i - \bar{X})^2 \sum_i (Y_i - \bar{Y})^2},$$

where $Cov(X_i, Y)$ is the covariance between X_i and Y, $Var(X_i)$ and $Var(Y)$ are respectively the variance of X_i and the variance of Y; and X and Y are the sample means of X and Y, respectively. The coefficient C is called the Pearson correlation coefficient (*PCC*). If the data are rank-transformed, the result is called a rank correlation coefficient.

LHS-PRCC is able to draw an important conclusion regarding the importance and significance of unknown parameters of interest in contributing to the outputs of a specific model (Blower & Dowlatabadi, 1994; Zheng & Rundell, 2006). The details of the PRCC method can be found in several publications (Blower & Dowlatabadi, 1994; Marino et al., 2008; Zheng & Rundell, 2006).

Appendix C. Sobol method

The Sobol method is a variance-based GSA approach that is capable of estimating the influence of individual parameters, or a group of parameters, on the output variables of a nonlinear model (Sobol', 1990; Sobol', 2001). This method assembles a global index based on an ANOVA-like decomposition; the term ANOVA refers to the Analysis Of Variances. This decomposition

takes place within the boundaries of the n-dimensional unit hypercube I^n, where I is the unit interval $[0, 1]$ and n is the number of input factors.

We describe the model under investigation as $y = f(x)$, where y is the model output variable and $x = (x_1, x_2, x_3,..., x_n)$ is a vector of input factors. We can decompose $f(x)$ into summands of increasing dimensionality as follows:

$$f(x) = f_0 + \sum_{i=1}^{n} f_i(x_i) + \sum_{1 \le i < j \le n} f_{ij}(x_i, x_j) + ... + f_{1,2...,n}(x_1, ..., x_n) \tag{2}$$

The decomposition presented in Eq. (2) is called ANOVA-representation of $f(x)$ if f_0 is constant, and the integrals of every summand over any of its own variables must be zero, i.e.,

$$\int_0^1 f_{i_1,...,i_s}(x_{i_1}, ..., x_{i_s}) dx_k = 0, \quad \text{for} \quad k = i_1, ..., i_s \tag{3}$$

Equation (3) uniquely defines that the terms in Eq. (2) are orthogonal and can be expressed as integrals of $f(x)$. In fact, the terms of the decomposition are constructed as

$$f_0 = \int f(x) dx, \tag{4}$$

$$f_i(x_i) = \int f(x) \prod_{k \ne i} dx_k - f_0, \tag{5}$$

$$f_{ij}(x_i, x_j) = \int f(x) \prod_{k \ne i,j} dx_k - f_0 - f_i(x_i), \tag{6}$$

and so on.

We assume that $f(x)$ is square integrable, so, $f_{i_1...i_s}$ are also square integrable. Squaring Eq. (2) and integrating over I^n yields:

$$\int f^2(x) dx - f_0^2 = \sum_{s=1}^{n} \sum_{i_1 < ... < i_s}^{n} \int f_{i_1,...,i_s}^2(x_{i_1}, ..., x_{i_s}) dx_{i_1}, ..., dx_{i_s} \tag{7}$$

As x are chosen as a random point uniformly distributed within the unit hypercube I^n, then $f(x)$ and $f_{i_1,...,i_s}(x_{i_1,...,x_{i_s}})$ are also termed as random variables and defined in the unit hypercube I^n. From Eq. (7), we can write the total and partial variances, respectively, as follows:

$$D = \int f^2(x) dx - f_0^2 \quad \text{and} \quad D_{i_1,...,i_s} = \int f_{i_1,...,i_s}^2 dx_{i_1}, ..., dx_{i_s}. \tag{8}$$

By squaring and integrating Eq. (2) over I^n, the total variance is obtained as follows:

$$D = \sum_{i=1}^{n} D_i + \sum_{1 \le i < j \le n} D_{ij} + ... + D_{1,2...,n}, ... \tag{9}$$

D_i is the fraction of variance in $f(x)$ solely due to the variability of the ith parameter x_i, and used to measure the sensitivity of $f(x)$ to x_i; D_{ij} is the fraction of output variance if both parameters x_i and x_j are varied, and used to measure the sensitivity of $f(x)$ to the interaction between x_i and x_j.

Now, the main effect of each factor on the model output, called the Sobol global sensitivity indices, is given as follows:

$$S_{i_1,...,i_s} = \frac{D_{i_1,...,i_s}}{D}, \tag{10}$$

where

$$\sum_{s=1}^{n} \sum_{i_1,\ldots,i_s}^{n} S_{i_1,\ldots,i_s} = 1 \tag{11}$$

Two Sobol indices-the main effects, called the first order index - and the total effects of a parameter (Sobol', 1990; Sobol', 2001), are given respectively, as follows:

$$S_i = \frac{D_i}{D} \tag{12}$$

and

$$S_{T_i} = 1 - \frac{D_{\sim i}}{D}, \tag{13}$$

where $D_{\sim i}$ is the variance contributed by all parameters except ith parameters x_i.

In summary, the first-order model sensitivity to each parameter x_i quantifies the effect if only ith parameter x_i is varied while all other parameters are fixed. The total order index is one minus the fraction of total variance attributed to $D_{\sim i}$, which represents all parameters except ith parameter x_i. The total order index effectively removes parameter x_i from the analysis and attributes the resulting reduction in variance to that parameter. The difference between the first and total order indices is that a total-order index represents the effects of its interactions with other parameters, but a first-order index represents the effect of a single parameter variance in output. Generally, parameters with sensitivity indices greater than 0.05 are considered to be significant. The total-order sensitivity indices are greater than the first-order sensitivity indices.

Appendix D. An optimal choice for sample size

There is no *a priori* known rule determining the exact sample size for either the LHS-PRCC or Sobol method. Trial-and-error, however, proves its utility during simulations and processing of the model outputs. We do obtain a known minimum for both methods where $N = m + 1$ for LHS (Marino et al., 2008), and $N = D(m + 1)$ and $N = D(2m + 1)$ for the Sobol method. Here m is the number of unknown parameters of interest and D is the user-defined number. To obtain the correct sample size for first and second order accuracy of sensitivity indices, systematic incremental increases of sample size are made, and whether sensitivity indices will reliably capture and provide a similar set of the most significant effects is observed. If that holds for two consecutive studies of SA, there is no obvious advantage in increasing the sample size, as the conclusions will be the same. It is further noted that the sample size N increases if the number of model parameters increases. Usually, the higher number of samples provides better outcomes of sensitivity analysis, though the computational cost is a matter. By observation, we consider sample size $N = 5200$ for both methods in our simulation as a suitable compromise between accuracy and computational cost.

Appendix E. Computational aspects

The coupled delay different equations (DDEs) in Eq. (1) are simultaneously solved using DDE23 solver implemented in MATLAB. The solutions of the model are integrated to $t = 300$ days, where the outputs of the model remain at steady-state equilibrium. The SA technique PRCC algorithm is executed in MATLAB and an efficient SA tool SALib, a Python library, is used for calculating Sobol indices.

For SA, we sample twelve parameters- (k, b_R, a, b_S, a_P, b_P, b_R, λ, μ_R, μ, c_{PE}, and c_V) simultaneously, using Latin hypercube sampling and Sobol sequence techniques. The mean values of parameters, adopted from (Cangelosi et al., 2017; Murray & Goyal, 2015), provide the reasonably accurate mean values of outputs. The description of parameters and their mean values (see Value column) are given in Table 1.

The Latin hypercube sampling technique, using the uniform distribution, generates samples of parameters using parameter mean values and the desired parameter ranges, described in Table 1, as follows:

$$k = Unif(k_{min}, k_{mean}, k_{max}), \quad \ldots, \quad c_V = Unif(c_{V_{min}}, c_{V_{mean}}, c_{V_{max}})$$

The Sobol sequence technique also generates parameter samples using the same parameter ranges inside SALib.

Appendix F. Time-evolution of cccDNA population

We have simulated our model for samples of size $N = 5200$. For all samples, the time evolution of cccDNA copies is shown in Fig. 6. Though all cccDNA copies, initially, increases exponentially, they, eventually, show steady-state evolution as time progress.
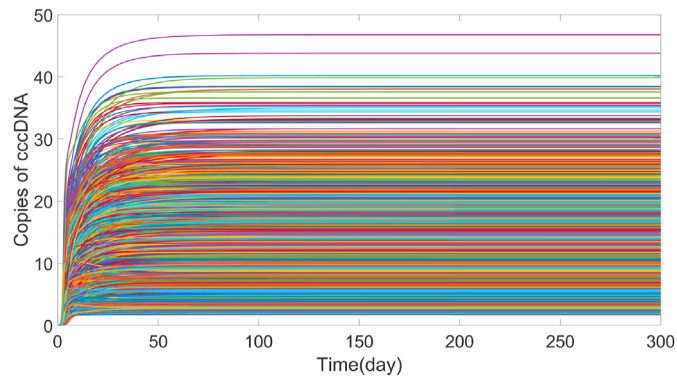
Fig. 6. For parameter samples of size $N = 5200$ simulation is carried and corresponding cccDNA copies are plotted to show their evolution over time.

*Appendix G. Monotonic relationship between inputs and outputs*

To show the monotonic relationship between input and output variables, we depict scatter plots of all outputs with respect to nine parameters − k, b_R, a, b_S, a_P, b_P, b_R, λ, μ_R, μ, c_{PE}, and c_V in Fig. 7, which shows that these nine parameters do not have significant influence on cccDNA replication.
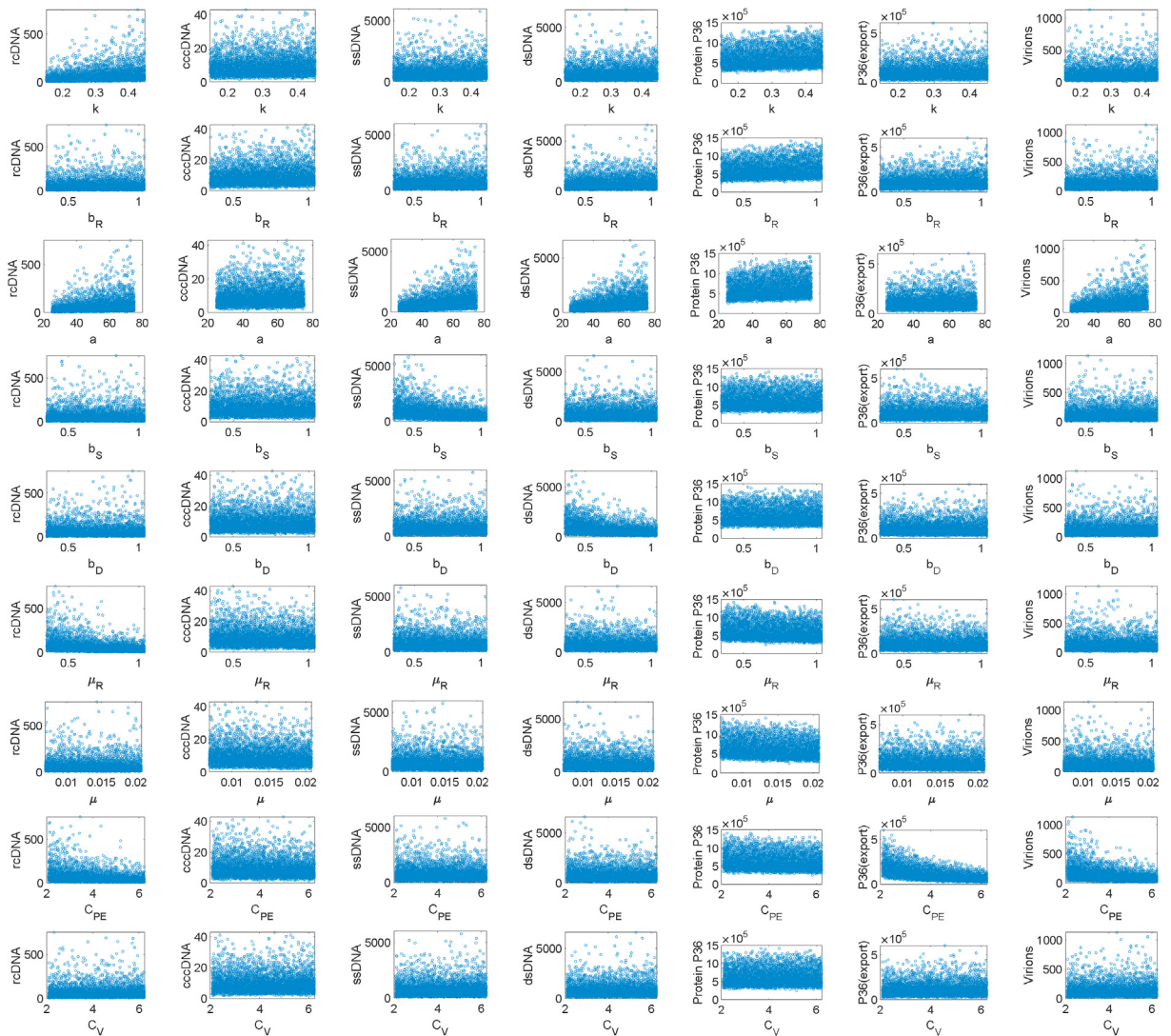
Fig. 7. Model is simulated using Latin hypercube samples of size $N = 5200$ and the scatter plots are depicted for each population with respect to nine parameters- k, b_R, a, b_S, b_D, μ_R, μ, c_{PE}, and c_V to show the monotonicity of the model outputs over the parameter regime.

Appendix H. Time-varying PRCC and Sobol indices analysis

Time-varying PRCC and Sobol indices are plotted for all parameters, where all parameters except a_P, b_P and λ show no influence on cccDNA replication over time, as shown in fig. 8.
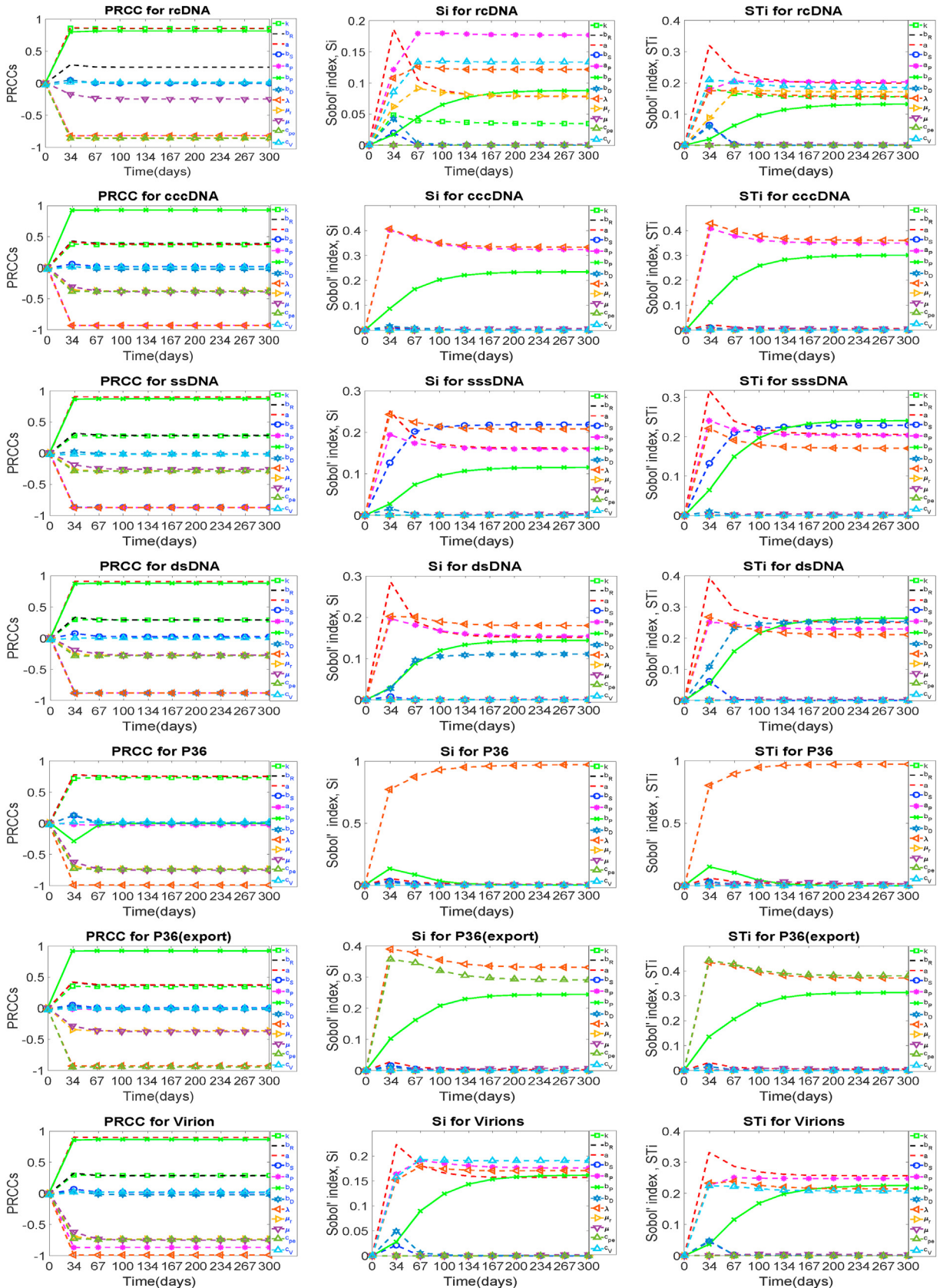
Fig. 8. Model is simulated using samples of size $N = 5200$ and the time-varying PRCC and Sobol indices (both 1st order, Si and total effect, STi) are calculated for each population (mean value) at different time points and plotted for different parameters to show the monotonicity of sensitivity indices over time.

References

Aparna, S., Johannes, H., Mikolajczyk, R. T., Gerard, K., & Ott, J. J. (2015). Estimations of worldwide prevalence of chronic hepatitis B virus infection: A systematic review of data published between 1965 and 2013. *Lancet, 386*, 1546–1555.

Blower, S., & Dowlatabadi, H. (1994). Sensitivity and uncertainty analysis of complex models of disease transmission: An HIV model, as an example. *International Statistical Review/Revue Internationale De Statistique, 62*(2), 229–243. https://doi.org/10.2307/1403510

Cangelosi, Q., Means, S. A., & Ho, H. (2017). A multi-scale spatial model of hepatitis-B viral dynamics. *PLoS One, 12*(12), Article e0188209. https://doi.org/10.1371/journal.pone.0188209

Elgouhari, H. M., Abu-Rajab Tamimi, T. I., & Carey, W. D. (2008). Hepatitis B virus infection: Understanding its epidemiology, course, and diagnosis. *Cleveland Clinic Journal of Medicine, 75*(12), 881–889. https://doi.org/10.3949/ccjm.75a.07019

Fausto, N., & Campbell, J. S. (2003). The role of hepatocytes and oval cells in liver regeneration and repopulation. *Mechanisms of Development, 120*(1), 117–130. https://doi.org/10.1016/s0925-4773(02)00338-6

Ganem, D., & Prince, A. M. (2004). Hepatitis B virus infection—natural history and clinical consequences. *New England Journal of Medicine, 350*(11), 1118–1129. https://doi.org/10.1056/NEJMra031087

Goyal, A., & Murray, J. M. (2016). Modelling the impact of cell-to-cell transmission in hepatitis B virus. *PLoS One, 11*(8), Article e0161978. https://doi.org/10.1371/journal.pone.0161978

Guo, J. T., & Guo, H. (2015). Metabolism and function of hepatitis B virus cccDNA: Implications for the development of cccDNA-targeting antiviral therapeutics. *Antiviral Research, 122*, 91–100. https://doi.org/10.1016/j.antiviral.2015.08.005

Helton, J. C., & Davis, F. J. (2002). Illustration of sampling-based methods for uncertainty and sensitivity analysis. *Risk Analysis : An Official Publication of the Society for Risk Analysis, 22*(3), 591–622. https://doi.org/10.1111/0272-4332.00041

Helton, J. C., & Davis, F. J. (2003). Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering & System Safety, 81*, 23–69. https://doi.org/10.1016/S0951-8320(03)00058-9

Helton, J. C., & Oberkampf, W. L. (2004). Alternative representations of epistemic uncertainty. *Reliability Engineering & System Safety, 85*. https://doi.org/10.1016/j.ress.2004.03.001, 1-3, 1-10.

Iman, R. L., & Conover, W. J. (1982). Sensitivity-analysis techniques: Self-teaching curriculum. *United States*. https://doi.org/10.2172/5388062

Iman, R. L., Helton, J. C., & Campbell, J. E. (1981). An approach to sensitivity analysis of computer models: Part I-introduction, input variable selection and preliminary variable assessment. *Journal of Quality Technology, 13*(3), 174–183. https://doi.org/10.1080/00224065.1981.11978748

Liu, S., & Xin, Y. (2019). HBV cccDNA: The stumbling block for treatment of HBV infection. *Journal of clinical and translational hepatology, 7*(3), 195–196. https://doi.org/10.14218/JCTH.2019.00047

Marino, S., Hogue, I. B., Ray, C. J., & Kirschner, D. E. (2008). A methodology for performing global uncertainty and sensitivity analysis in systems biology. *Journal of Theoretical Biology, 254*(1), 178–196. https://doi.org/10.1016/j.jtbi.2008.04.011

McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics, 21*(2), 239–245. https://doi.org/10.1080/00401706.1979.10489755

Means, S., Ali, M. A., Ho, H., & Heffernan, J. (2020). Mathematical modeling for hepatitis B virus: Would spatial effects play a role and how to model it? *Frontiers in Physiology, 11*, 146. https://doi.org/10.3389/fphys.2020.00146

Murray, J. M., & Goyal, A. (2015). In silico single cell dynamics of hepatitis B virus infection and clearance. *Journal of Theoretical Biology, 366*, 91–102. https://doi.org/10.1016/j.jtbi.2014.11.020

Shi, Y. H., & Shi, C. H. (2009). Molecular characteristics and stages of chronic hepatitis B virus infection. *World Journal of Gastroenterology, 15*(25), 3099–3105. https://doi.org/10.3748/wjg.15.3099

Smolders, E. J., Burger, D. M., Feld, J. J., & Kiser, J. J. (2020). Review article: Clinical pharmacology of current and investigational hepatitis B virus therapies. *Alimentary pharmacology & therapeutics, 51*(2), 231–243. https://doi.org/10.1111/apt.15581

Sobol', I. M. (1990). On sensitivity estimates for nonlinear mathematical models. *Mathematics and Computers in Simulation, 2*(1), 112–118. http://mi.mathnet.ru/mm2320.

Sobol', I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation, 55*(1–3), 271–280. https://doi.org/10.1016/S0378-4754(00)00270-6

Tarao, K., Nozaki, A., Ikeda, T., Sato, A., Komatsu, H., Komatsu, T., Taguri, M., & Tanaka, K. (2019). Real impact of liver cirrhosis on the development of hepatocellular carcinoma in various liver diseases-meta-analytic assessment. *Cancer medicine, 8*(3), 1054–1065. https://doi.org/10.1002/cam4.1998

Van Griensven, A., Meixner, T., Grunwald, S., Bishop, T., Diluzio, M., & Srinivasan, R. A. (2005). Global sensitivity analysis tool for the parameters of multi-variable catchment models. *Journal of Hydrology, 324*(1–4), 10–23. https://doi.org/10.1016/j.jhydrol.2005.09.008

Wang, K., & Wang, W. (2007). Propagation of HBV with spatial dependence. *Mathematical Biosciences, 210*(1), 78–95. https://doi.org/10.1016/j.mbs.2007.05.004

World Health Organization. (2015). Hepatitis B fact sheet. http://www.who.int/mediacentre/factsheet/fs204/en/, 204.

Wu, J. F., & Chang, M. H. (2015). Natural history of chronic hepatitis B virus infection from infancy to adult life - the mechanism of inflammation triggering and long-term impacts. *Journal of Biomedical Science, 22*, 92. https://doi.org/10.1186/s12929-015-0199-y

Wu, J., Dhingra, R., Gambhir, M., & Remais, J. V. (2013). Sensitivity analysis of infectious disease models: Methods, advances and their application. *Journal of The Royal Society Interface, 10*(86), 20121018.

Zhan, C., Song, X., Xia, J., & Tong, C. (2013). An efficient approach for global sensitivity analysis of hydrological model parameters. *Environmental Modelling & Software, 41*, 39–52. https://doi.org/10.1016/j.envsoft.2012.10.009

Zheng, Y., & Rundell, A. (2006). Comparative study of parameter sensitivity analyses of the TCR-activated erk-MAPK signalling pathway. *IEE Proceedings - Systems Biology, 153*(4), 201–211. https://doi.org/10.1049/ip-syb:20050088. July 2006.

Zi, Z. (2011). Sensitivity analysis approaches applied to systems biology models. *IET Systems Biology, 5*(6). https://doi.org/10.1049/iet-syb.2011.0015, 336-6.