Uncertainty in clinical data and stochastic model for in vitro fertilization

Kirti M. Yenkie a,b, Urmila Diwekar a,b,*

a Department of Bioengineering, University of Illinois, Chicago, IL 60607, USA
b Center for Uncertain Systems: Tools for Optimization and Management (CUSTOM), Vishwamitra Research Institute, Clarendon Hills, IL 60514, USA

HIGHLIGHTS
- First study to analyze the uncertainty in the ongoing superovulation cycle.
- Ito process models are capable of capturing time-dependent uncertainties in superovulation.
- Stochastic differential equation model predictions are better matched with the clinical data.
- A promising approach for efficient and robust modeling of biomedical processes

ABSTRACT
In vitro fertilization (IVF) is the most widely used technique in assisted reproductive technologies (ART). It has been divided into four stages: (i) superovulation, (ii) egg retrieval, (iii) insemination/fertilization and (iv) embryo transfer. The first stage of superovulation is a drug induced method to enable multiple ovulation, i.e., multiple follicle growth to oocytes or matured follicles in a single menstrual cycle. IVF being a medical procedure that aims at manipulating the biological functions in the human body is subjected to inherent sources of uncertainty and variability. Also, the interplay of hormones with the natural functioning of the ovaries to stimulate multiple ovulation as against single ovulation in a normal menstrual cycle makes the procedure dependent on several factors like the patient’s condition in terms of cause of infertility, actual ovarian function, responsiveness to the medication, etc. The treatment requires continuous monitoring and testing and this can give rise to errors in observations and reports. These uncertainties are present in the form of measurement noise in the clinical data. Thus, it becomes essential to look at the process noise and account for it to build better representative models for follicle growth. The purpose of this work is to come up with a robust model which can project the superovulation cycle outcome based on the hormonal doses and patient response in a better way in presence of uncertainty. The stochastic model results in better projection of the cycle outcomes for the patients where the deterministic model has some deviations from the clinical observations and the growth term value is not within the range of 0.3–0.6. It was found that the prediction accuracy was enhanced by more than 70% for two patients by using the stochastic model projections. Also, in patients where the prediction accuracy did not increase significantly, a better match with the trend of the clinical data was observed in case of the stochastic model projections as compared to their deterministic counterparts.

ARTICLE INFO
Article history:
Received 3 May 2014
Received in revised form 23 September 2014
Accepted 3 November 2014
Available online 4 December 2014
Keywords:
Infertility
Assisted reproduction
Multiple ovulation
Uncertainty
Stochastic model

1. Introduction
Infertility is the inability of a couple to achieve conception or to bring a pregnancy to term after a year or more of regular, unprotected intercourse. The word infertility literally translates to ‘non-fertile’ and therefore can be interpreted as sterile. However, this is not true, infertility is a medical condition with diminished ability or inability to produce offspring in either the male or female partner and they are better categorized as ‘subfertile’ (Habbema et al., 2004). According to worldwide statistics, around 80 million couples experience some kind of infertility problems (Ombelet and Campo, 2007). Hence, these patients have an option to seek medical aid and there are several assisted reproduction technologies (ARTs) like IVF which can help them in conceiving. Nearly 200,000 IVF cycles are performed worldwide annually and more than 1,000,000 children have been born

* Corresponding author at: Center for Uncertain Systems: Tools for Optimization and Management (CUSTOM), Vishwamitra Research Institute, Clarendon Hills, IL 60514, USA. Tel.: +1 630 886 3047.
E-mail address: urmila@vri-custom.org (U. Diwekar).

http://dx.doi.org/10.1016/j.jtbi.2014.11.004
0022-5193/© 2014 Elsevier Ltd. All rights reserved.
since the first successful birth of the ‘test-tube’ baby, Louise Joy Brown in 1978, as per the data reported by Bayer et al. (2007). In the year 2010, Robert G. Edwards was awarded the Nobel prize in the category of ‘Physiology or Medicine’ for the development of human IVF therapy (Nobel Media, 2010). This reveals the prevalence of infertility, the number of people sorting to IVF as a therapy and the overall significance of its development to benefit mankind.

IVF involves fertilization of oocytes by a sperm outside the body in a laboratory simulating similar conditions as in the human body and then implanting the fertilized eggs back in the uterus of the carrier mother for full term pregnancy. It is a four stage medical procedure (Speroff and Fritz (2005)) and the success of superovulation, its first stage, is critical to proceed with the next stages in the IVF cycle. Also it requires maximum medical attention and investment of time and money as compared to the other stages. The major hormonal drug involved in superovulation is the follicle stimulating hormone (FSH), which is responsible for follicle growth and development (Jánát-Amsbury et al., 2009). In our previous work (Yenkie et al., 2013a), a deterministic model for the prediction of multiple follicle growth, dependent on the amount of FSH dosage was built by drawing similarities with the particulate process of batch crystallization (Hill et al., 2006; Yenkie and Diwekar, 2013). This model was validated with clinical data from 50 IVF cycles, available from our collaborating hospital in India. The model fitted very well to most of the data and almost 85% predictions were within 0 to 30% range of the mean error (Yenkie et al., 2014). However, the fact that IVF is actually a medical procedure that involves manipulations in the biological functions of the human body and specifically to the functioning of the ovaries, is subjected to inherent sources of uncertainty and variability.

Uncertainties regarding the normal range of values for interpreting the fertility tests is a major concern and this can result in improper diagnosis and treatment specially in cases where the cause of infertility is unexplained (Bachus and Walmer, 1993). Johnson et al. (1987) discuss about the uncertainties experienced by the patients at each stage of the IVF procedure. Sometimes the patients end up overestimating the IVF success rate based on the current popularity of the procedure, while, underestimating the chances of failure in one of the less publicized stage in the overall procedure. The anxiety and stress of the overall procedure can have a huge impact on the patient’s responsiveness to the medications. The report on incorporating natural variations into IVF clinic tables by Lemmers et al. (2007) suggest that the result of any IVF cycle will lie within the best-case and worst-case scenario observed at that particular clinic.

The previous studies on uncertainty analysis in IVF are mostly based on anticipation and stress related to the psychology of the patient, emotional state and fears concerning the medical procedure and their ability to accept the fact that medical intervention can help in reproduction (Ardenti et al., 1999; Thiering et al., 1993). In most of their conclusions the results are reported in terms of relative percentage of success of the cycle in depressed and non-depressed patients. The work by Ardenti et al. (1999) suggests that the uncertainty of outcome generates maximum anxiety levels during the oocyte retrieval and embryo transfer. Both these procedures are highly dependent on the success of superovulation. If high number of follicles grow to mature oocytes, the excess oocytes can be saved for the next cycle if the first attempt fails and hence the anxiety levels of the patient can be reduced by providing them with a better outcome in superovulation. Prediction of better superovulation outcomes and enhancement in the success rate by application of optimal control for pre-determined hormonal drug dosage has already been addressed in our previous work (Yenkie and Diwekar, 2014a, 2014b).

However, the uncertainty in the clinical observations has not gained enough attention in the previous work. Since the existing protocols (Meniru and Craft, 1997; Louradis et al., 2007) for superovulation are largely dependent upon the clinical observations for deciding the hormonal dosage on the next day of the cycle, it is essential to look into the possible errors in the follicle growth measurements and reporting. The uncertainty in the IVF outcomes specific to the particular clinic has been addressed by Lemmers et al. (2007), but we intend to characterize the uncertainty specific to the patient and the ongoing cycle, which makes the procedure more acceptable. In the current work, we are proposing a method to characterize the uncertainty in the ongoing cycle. The clinical data on IVF cycles when used for fitting the deterministic model and validating the projected results had some deviations from the expected behavior (Yenkie et al., 2014). Thus, it can be used for evaluating the effects of uncertainty in the process.

Previously, the work done in our group on uncertainty characterization and modeling has revealed that the Ito processes are quite efficient in capturing the time-dependent uncertainties in batch processes (Ulás et al., 2005; Benavides and Diwekar, 2012; Yenkie and Diwekar, 2013). Also, they were equally efficient in capturing the time-dependent variations in the blood glucose levels of insulin-dependent diabetes patients (Ulás and Diwekar, 2010). IVF being a medical procedure and the superovulation model developed from the principles of batch crystallization, motivated us to look into Ito processes for modeling the associated uncertainties.

2. Methodology

The deterministic model for the superovulation stage in the form of ordinary differential equations (ODEs) is discussed briefly in Section 2.1. Then the occurrence of uncertainty in the clinical data, its characterization as suitable stochastic processes and the development of a more robust model for superovulation in terms of stochastic differential equations (SDEs) is discussed in Section 2.2.

2.1. The deterministic model

The similarities between the process of crystallization and superovulation was used for the modeling of multiple follicle growth under the influence of injected hormones. The concept of the moment based model for crystal growth in batch crystallization (Hu et al., 2005) was used as the basis for modeling superovulation in IVF (Yenkie et al., 2013b) because it had the advantage of evaluation of moments of different orders. From the literature by Randolph and Larson (1988) it is known that moments correspond to specific features of the particles like the zeroth moment ($\mu_0$) corresponds to the particle number, first moment ($\mu_1$) to their size and second ($\mu_2$) to their shape, etc. The growth term in cooling batch crystallization is temperature dependent and hence temperature is considered to be the most promising decision variable for eventually achieving a desired particle size distribution (PSD). On similar lines in IVF, the follicle growth is dependent upon the doses of hormones injected to the patient. Thus, the follicle growth term ($G$) is dependent on the amount of follicle stimulating hormone (FSH) injected ($\Delta C_{FSH}$) to the patient at the particular time ($t$) in the cycle and is represented as shown in Eq. (1).

$$G(t) = k\Delta C_{FSH}(t)^{\alpha}$$

(1)

Here, $k$ and $\alpha$ are kinetic constants of the growth term.

In the literature by Baird (1987), it was suggested that the number of follicles activated for growth during a particular superovulation cycle is relatively constant for a particular patient, hence the zeroth moment ($\mu_0$) is assumed to have a constant value at all times during the FSH dosage regime. The 0th to 6th order moments are used in the model, since they help in better prediction of
moment values as well as help in efficient recovery of the size distributions as against the lower order moments (Flood, 2002). The moment equations for the follicle dynamics can be written as in Eqs. (2) and (3).

\[ \mu_0 = \text{constant} \]  

\[ \frac{d\mu_i}{dt} = G(t)\mu_{i-1}(t); \quad (i = 1, 2, \ldots, 6) \]  

Here, \( \mu_i \) is the \( i \)th moment and \( G \) is the follicle growth term.

The data available from our collaborators is in the form of follicle size, their number and dosage of hormones on selected days; usually the first day, fifth day and then every alternate day till the follicles grow to the desired size in the superovulation cycle. The follicle size data is converted into mathematical moments by assuming the follicles to be spherical in shape. Eq. (4) is used for converting follicle size to moments (see Appendix A for more details).

\[ \mu_i = \sum_n n_i(r, t)|r\Delta r_j \]  

Here, \( \mu_i \) is the \( i \)th moment, \( n_i(r, t) \) are the number of follicles in \( j \)th bin with mean radius as \( r \) at time \( t \), \( r_j \) is the mean radius of the \( j \)th bin and \( \Delta r \) is the range of radii variation in each bin.

The data converted into the form of moments by Eq. (4) is then used for model fitting and validation. The validation method requires the conversion of predicted moments back to the follicle size and number and this is done by adapting the methodology proposed by Flood (2002) (see Appendix B). The method suggested is represented in Eq. (5)

\[ \overline{\Pi} = A \overline{\pi} \]  

Here, \( \overline{\Pi} \) is the vector of number of follicles in \( n \) bins on \( i \)th day, \( \overline{\pi} \) is the vector of moments on \( i \)th day and \( A \) is the inversion matrix derived from Eq. (4). The elements in matrix ‘\( A \)’ are evaluated using the formula shown in Eq. (6).

\[ a_{ij} = r_j^3 \Delta r_j \]  

Here, \( a_{ij} \) is the order of the moment (increment in rows), \( j \) is the representative bin (increment in columns), \( r_j \) is the mean radius of the bin and \( \Delta r \) is the range of the bin. The conversion of moments to follicle size distribution (FSD) is applicable to the deterministic as well as stochastic model.

2.2. Prediction accuracy of the deterministic model

The deterministic modeling approach was applied to the clinical data on 50 superovulation cycles. The predictions of the follicle size distribution (FSD) were compared with the actual follicle size and numbers reported in the clinical data. The mean square error (MSE) for the accuracy of the FSD projections on final FSH dosing day was evaluated using the formula shown in Eq. (7).

\[ \text{MSE} = \sum \left( \frac{n_{obs} - n_{sim}}{\text{Max}(n_{obs})} \right) \]  

Here, \( n_{obs} \) is the number of follicles observed and \( n_{sim} \) is the number of follicles predicted by the model in the \( i \)th bin on last day of FSH administration. This error value was evaluated twice (i) using the available data on all days and (ii) using the data for initial two days of observation. In MSE evaluation the follicles in the size bins corresponding to the desired size range of 9–12 mm were used. The error analysis from our previous work (Yenkie et al., 2014) as shown in Fig. 1 indicates that the model accuracy is lower when initial two day data is used as compared to all day data, which is intuitive. However, our purpose of using the all day data is to check the model feasibility as well as to develop a comparative standard in terms of the MSE value against the two day data model projections. The comparison of the MSE values between these two measures of accuracy show that the model predictions from the two day data are also quite good and for almost 60% cases. This suggests the sufficiency of two day data in those 60% cases for projecting the overall superovulation outcome. However, the deterministic model does not match very well in cases where the MSE values are more and we also found that the growth term \( G \) values for those patients were not within the usual range of 0.3–0.6, suggesting the need for additional monitoring and the requirement of additional data to predict the superovulation outcome with more accuracy.

This lack of accuracy in the deterministic model could also be due to the uncertainty in the clinical data. Thus, the modeling accuracy can also be enhanced by including these uncertainties and by keeping the requirement of monitoring restricted to initial two days. This shall reduce the risks of overdosing on the patients as well as the costs on monitoring. Thus, we analyze the source of time-dependent uncertainty. In this work, we are proposing a ‘personalized model’ for each patient and each cycle, so this model is not subjected to any parametric uncertainty as discussed in our previous work on batch crystallization (Yenkie and Diwekar, 2013a). However, there can be errors in the follicle size measurements as the time changes resulting in the uncertainty in the clinical data. This is also known as the measurement noise in the follicle sizes as well as their numbers in a particular superovulation cycle.

In the previous Section 2.1. on the deterministic model development and validation, we reported the use of the discrete moment formula (Eq. (4)) for conversion of the clinical data to the moment values. Thus, the uncertainty in the follicle size measurements will be transferred to the uncertainty in the moment values for the patients. The time-dependent nature of these uncertainties can be seen in the moment plots shown for the Patient-A (Fig. 2(a)) and Patient-D (Fig. 2(b)). The deterministic model predictions are shown by the black profile for the 2nd moment in the case of Patient-A and the 4th moment in case of Patient-D. The gray profile with the markers are the evaluated moments from the observed clinical data on the corresponding days. It can be seen that the model predictions are usually smooth curves, but the clinical data from these two patients provide the moment values (from Eq. (4)) which have some deviations from the smooth curve and also they have a sudden change in the slope of the curve at some points, which are difficult to capture in a deterministic model and hence are an indicator of the possible time-dependent uncertainties in the follicle size data. Thus, we propose the development of the stochastic model for the better prediction accuracy in these patients.

![Fig. 1. Deterministic model prediction accuracy for data on 50 superovulation cycles.](Image 307x70 to 547x221)
2.3. The stochastic model

The stochastic model for the superovulation stage is developed by using the Ito processes. Initially, these Ito processes were mostly applied in the areas of finance and stock price modeling to capture and represent the uncertainties for predicting the market behavior. Since Ito processes can characterize time-dependent uncertainties and can be integrated and differentiated using the rules of Ito’s stochastic calculus, they can prove beneficial in modeling biomedical processes too. Since we are trying to develop a stochastic model for a biomedical treatment process of IVF, the use of Ito processes as a modeling basis is reasonable. The first part of the stochastic modeling is to prove that the uncertainties in the superovulation stage of IVF can be modeled as an Ito process. Hence, we introduce the fundamental properties of the Wiener process, the simplest form of an Ito process, and then look at the superovulation process and perform a check whether it satisfies the properties of the Wiener process.

2.3.1. Wiener process

The simplest example of an Ito process is the Brownian motion which is also known as the Wiener process in the continuous form. For any stochastic process to be characterized as the Wiener process it must follow the following three important properties (Diwekar, 2008).

1. It should follow the Markov property: the probability distribution for future values of the process depends only on its current value.
2. It should have independent increments in time. The probability distribution for the changes in the process over any time interval is independent of any other time interval.
3. Changes in the process over a finite time interval should be normally distributed, with variance linearly dependent on the length of time interval, $\Delta t$ (i.e., $N(0, \sqrt{\Delta t})$ for all $t > 0$).

The Wiener process acts as a building block for modeling and representing different types of Ito processes. Depending upon the behavior of the system under random influences it can be modeled as one of the suitable forms of the Ito process. Some examples of Ito process are simple brownian motion, geometric brownian motion, mean reverting process and geometric mean reverting process. The sample paths for geometric brownian motion or brownian motion with drift and mean reverting process are shown in Fig. 3.

2.3.2. Superovulation: a Wiener process

We study the process of superovulation and check whether it classifies as a Wiener process. As per the current protocols, the hormonal dosage is decided based on daily monitoring of the patient’s response and the follicle size observed at the particular time. Thus, the next follicle size distribution (FSD) after injection of hormones is dependent on the FSD at that particular time,

---

![Image](https://example.com/image1.png)

**Fig. 2.** Uncertainty characterized in terms of deviations of the moment values evaluated from the clinical data in comparison to the moment values projected from deterministic model.

![Image](https://example.com/image2.png)

**Fig. 3.** Sample paths of different types of Ito processes (Diwekar, 2008).
hence it is reasonable to categorize the superovulation stage as a Markov process.

The follicle growth term developed in the deterministic model and which satisfies the data is dependent on the amount of FSH injected (ΔCFSH) to the patient at the particular day (t) in the cycle as shown earlier in Eq. (2). Thus it satisfies the second property, where the FSH dosage can have independent increments.

The follicle growth and number is represented as the FSD, where the follicle size is divided into equispaced bins and number of follicles are represented as the discrete markers in each of the size bins as shown in Fig. 4. When these markers are connected with a smooth curve, it can be seen that the data tends to follow the normal distribution and as the time progresses, we see that it continues to follow the same trend, with a change in the mean and variance values. Hence, the uncertainties associated with the superovulation process can be represented in terms of a Wiener processes.

### 2.3.3. Ito process model for superovulation

Depending upon the time-dependent behavior of the states of the system (the moments of the superovulation) due to uncertainties present in the clinical data, suitable type of Ito process can be selected to develop the stochastic model for superovulation. We checked the different types of Ito process models and found that the uncertainties could be best captured using the simple Brownian motion type of the Ito process, thus resulting in the following (see Eqs. (8)–(13)) set of stochastic differential equations (SDEs):

\[
d\mu_i(t) = G(t)\mu_i \, dt + \sigma_i \varepsilon_i \sqrt{dt}
\]

(8)

\[
d\mu_1(t) = 2G(t)\mu_1 \, dt + 3\sigma_1 \varepsilon_1 \sqrt{dt}
\]

(9)

\[
d\mu_2(t) = 3G(t)\mu_2 \, dt + 4\sigma_2 \varepsilon_2 \sqrt{dt}
\]

(10)

\[
d\mu_3(t) = 4G(t)\mu_3 \, dt + 5\sigma_3 \varepsilon_3 \sqrt{dt}
\]

(11)

\[
d\mu_4(t) = 5G(t)\mu_4 \, dt + 6\sigma_4 \varepsilon_4 \sqrt{dt}
\]

(12)

\[
d\mu_5(t) = 6G(t)\mu_5 \, dt + \sigma_5 \varepsilon_5 \sqrt{dt}
\]

(13)

Here, \( \mu_i \) is the \( i \)th moment, \( \sigma_i \)'s are the standard deviation terms for each moment equation corresponding to the noise in the clinical data, \( \varepsilon_i \) is the set of random numbers derived from unit normal distribution \( N(0,1) \) and \( dt \) is the time interval.

### 2.4. Calculations

#### 2.4.1. Parameter estimation and validation

The stochastic model requires parameter estimation methods which are capable of predicting the parameters in the deterministic as well as stochastic parts of the equations and hence we have used a novel methodology (Yenkie et al., 2013c) which combines the advantages of deterministic optimization methods and computes the stochastic parameters directly from the noise in the clinical data. The methodology is shown in Fig. 5 and it can be seen that it is a two level method in which the parameter in the stochastic part (\( \sigma \)) of the differential equation is computed from the initial guess for the deterministic parameter set (\( \theta \)), available model information and the clinical data. The stochastic parameter (\( \sigma \)) computed in the inner level of the method is then plugged into the modified objective function for the enhanced estimation of the parameters present in the deterministic part (\( \theta \)). The advantage of this method is that we can use deterministic optimization algorithms for determining the parameters of the SDEs. It should be noted that while performing

| Table 1 Parameter values obtained for Patient A. |
|---|---|---|
| Parameter | Deterministic | Stochastic |
| \( k \) | 171.6 | 1.048 |
| \( \alpha \) | –1.097 | –0.10 |
| \( \sigma_1 \) | 0.243235 | 0.088484 |
| \( \sigma_2 \) | 0.189031 | 0.088484 |
| \( \sigma_3 \) | 0.686597 | 0.088484 |
| \( \sigma_4 \) | 1.812936 | 0.088484 |
| \( \sigma_5 \) | 3.816483 | 0.088484 |

![Fig. 5. Two level approach for parameter estimation in SDEs (Yenkie et al. (2013a)).](image_url)
the estimation of the parameter set \( \theta \), the optimization is still subjected to the stochastic model equations. The methodology is explained with an example in Appendix C. The results from this new method are compared with the results from existing SDE estimation methods to show the enhanced accuracy of this method.

The results obtained for the parameters are then plugged into the model to project the moment values on the successive days of the cycle. Since the SDE model has the term of unit normal distribution \( \epsilon \) in their stochastic part, we obtain a range of moment values for the 1st–6th moments when we run the stochastic model for some finite number of times, as against the same moment value at a particular time in case of the deterministic model. These different moment trajectories are called as scenarios for the moments. The aim of this stochastic simulation is to account for the possible deviations in the moment values. To accomplish this objective, we generate about 100 scenarios for all the moments by running the model 100 times and then calculating the expected value of each of these moment trajectories. Thus, we get the estimate of the moment values which are inclusive of the possible variations and uncertainties and their values would resemble the characteristics of the growing follicles more closely as compared to their deterministic counterparts.

These expected values of the projected moments are converted back into FSDs by using the follicle number prediction algorithm (Appendix B) from our previous work on the development of deterministic model for superovulation stage (Yenkie et al., 2013a). These results are compared with the clinical observations (O) and deterministic model predictions (D) to check the performance of the stochastic model (S). The material from our previous work on the FSD backprediction from the moment values is explained in Appendix B.

### 2.4.2. Predictive improvements due to stochastic model

To provide more insight into the work, the model prediction error for the deterministic (D) and stochastic (S) model is evaluated using the Eq. (7) listed in Section 2.2. The purpose is to check the improvements in FSD predictions due to the inclusion of uncertainty in the clinical data for some patients. These patients had the growth term ‘G’ values beyond the desired range of ‘0.3–0.6’ and also the MSE values were more for the final day FSD prediction.

### 3. Results and discussions

The parameters for the system are evaluated using the two level approach (Yenkie et al., 2013b) as shown in Fig. 5, in which the parameters in the stochastic part, i.e., the values of the standard deviations \( \sigma_i \), are evaluated from the model information and the variance in the data. The follicle size measurements are converted to moments by the method discussed earlier in Yenkie et al. (2013c). The parameters in the deterministic part \( k \) and \( \alpha \) are evaluated using the modified form of the sum of least square error objective function involving the previously computed values...
of standard deviations. The parameter values predicted for Patient-A are shown in Table 1.

The results for the FSDs are shown in Fig. 6(a) and (b) for day 5 and day 9 respectively. The plots show the observed clinical data as discrete data points (O), the deterministic model (D) projections as dotted profiles and the stochastic model (S) projections as continuous profiles. We can see that the results projected by the stochastic model match the data better when compared to the deterministic results, specifically the final FSH dosing day projections (i.e. the 9th day in case of Patient-A). The results resemble the data more closely which is one of the expected results from this study, since it tends to include the possible sources of error in the measurements.

Some more results for patients B, C, D and E are shown in Figs. 7–10(a) and (b). The final day FSD projections for patients B (Fig. 7(b)) and C (Fig. 8(b)) are quite close to the actual observations. The FSD projections for patients D (Fig. 9(b)) and E (Fig. 10(b)) do not match very well to the clinical data, but they tend to follow the trend of the clinical observations closely when compared to the deterministic predictions. The comparison of the model projections for the deterministic and stochastic cases are shown in Table 2 in terms of the mean square error on the final FSH dosing day. For all these patients, the FSD projections on using the stochastic model are greatly improved as compared to the deterministic model projections. The comparison of the stochastic and the deterministic model predictions in terms of the percentage error shows that there is an improvement in the outcome projection for all the patients. In fact, the percentage improvements for patients A, B and E are quite significant and suggest that the uncertainties can have a huge impact on mathematical models developed for biomedical processes.

The aim of the study was to analyze the possible uncertainty in the clinical data and devise a methodology for including those variations in the superovulation model for enhanced predictability. The use of Ito process, a special kind of stochastic process to capture the uncertainty in IVF data proves promising and provides a reasonably better model predictivity as compared to the deterministic model. Another advantage of Ito process models are that they have been studied extensively and have a well developed Ito calculus for integration and differentiation along with the numerical methods (Kloeden and Platen (1999)) for solving equations which are unsolvable analytically.

The suggested approach for uncertainty characterization in clinical data and development of a stochastic model is independent of the clinic and previous success or failure rates. This methodology can be applied for modeling any superovulation cycle based on the current observations. Our future aim is to use the SDE model as a basis for predicting the optimal hormonal dosing profiles, which shall be more robust and would be able to provide even better outcomes than the deterministic optimal control methodology (Yenkie and Diwekar, 2014a). The methodology for stochastic control in Ito process models has already been developed by Ramirez and Diwekar (2004), and has been applied to existing problems of insulin dosing in the biomedical processes like diabetes control (Ulas and Diwekar, 2010).
4. Conclusions

The approach to develop a stochastic model for the superovulation stage in IVF in terms of Ito form of stochastic differential equations is a novel aspect of the study and the results look very promising. The approach used for stochastic model development involving a novel parameter estimation procedure is an added contribution in this work and can eliminate a lot of computation difficulties in development of SDE models. The model predictions obtained from the stochastic model are inclusive of the noise in the clinical data and hence match better as compared to deterministic results for the study done on five selected patients. The mean square error in modeling for the deterministic and stochastic cases provide a numerical estimate of the improvements in predictability and robustness of the SDE model. The predictions from the stochastic superovulation model can be used for predicting robust drug dosing policies using stochastic control methods for achieving the desired outcome in superovulation and thus contributing to the study on enhancement of the success rate of IVF cycles.

Acknowledgments

We thank Dr. Vibha Bhalerao and the Jijamata Hospital in Nanded, India for help with the data for superovulation. We also thank the Institutional Review Board at UIC for approving the current research to meet with the requirements of Protection of Human Subjects research under the Protocol #2012-0317 titled ‘Understanding the Superovulation Stage of IVF Process for Optimal Drug Delivery’.

Appendix A. Conversion of follicle size data into moments

The clinical data from a successful superovulation cycle in shown in Table A.1. It shows that during FSH dosage regime, as time progresses the size of the follicles increase. The data has been reorganized into six size bins so that they can be converted into the moment values of different orders (order 1–6) (Yenkie et al., 2014a).

The follicle size data shown in Table A.1 can be converted to moments using the general expression shown in Eq. (A.1) also listed as Eq. (4) in Section 2.1.

\[
\mu_i = \sum \eta_j(t_r) r_j^i \Delta r_j \quad (A.1)
\]

The follicles are assumed to be spherical in shape and from the data it is evident that the number of follicles which enter into growth on the first day tend to remain constant, we just see the increase in their size as the FSH dosing progresses on each day of the cycle. The collection of the first two day data is very important because it is used for the model parameter estimation in the deterministic as well as stochastic model. The moment values obtained from the dataset shown earlier in Table A.1 is shown in Table A.2.

Table A.2
Error comparison for deterministic and stochastic model predictions for final FSH day.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Growth term (G)</th>
<th>Mean square error (deterministic)</th>
<th>Mean square error (stochastic)</th>
<th>% Improvement (stochastic over deterministic)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.72</td>
<td>1.1239</td>
<td>0.0748</td>
<td>93.34</td>
</tr>
<tr>
<td>B</td>
<td>0.77</td>
<td>0.927</td>
<td>0.1999</td>
<td>78.4</td>
</tr>
<tr>
<td>C</td>
<td>1.55</td>
<td>1.4813</td>
<td>1.3740</td>
<td>7.24</td>
</tr>
<tr>
<td>D</td>
<td>0.82</td>
<td>1.3009</td>
<td>1.1865</td>
<td>8.79</td>
</tr>
<tr>
<td>E</td>
<td>0.70</td>
<td>1.6541</td>
<td>0.9351</td>
<td>43.47</td>
</tr>
</tbody>
</table>

Fig. 10. Comparison of the follicle size distribution from observed clinical data (O), deterministic (D) and stochastic (S) model predictions for Patient-E on day 7(A) and day 9(B).

Table A.1
Variation of follicle size (diameter) with time and FSH dose for a patient.

<table>
<thead>
<tr>
<th>Size bins (mm)/time</th>
<th>Number of follicles</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Day 1</td>
</tr>
<tr>
<td>0–4</td>
<td>4</td>
</tr>
<tr>
<td>4–8</td>
<td>12</td>
</tr>
<tr>
<td>8–12</td>
<td>8</td>
</tr>
<tr>
<td>12–16</td>
<td>2</td>
</tr>
<tr>
<td>16–20</td>
<td>0</td>
</tr>
<tr>
<td>20–24</td>
<td>0</td>
</tr>
<tr>
<td>FSH dose (IU/ml)</td>
<td>150</td>
</tr>
</tbody>
</table>

IU—International units used for hormonal dosage measurement listed as Eq. (4) in Section 2.1.

\[
\mu_i = \sum \eta_j(t_r) r_j^i \Delta r_j \quad (A.1)
\]
Table A.2
Experimental moments evaluated using Eq. (a.1).

<table>
<thead>
<tr>
<th>Sr. no</th>
<th>Time (day)</th>
<th>μ₀</th>
<th>μ₁</th>
<th>μ₂</th>
<th>μ₃</th>
<th>μ₄</th>
<th>μ₅</th>
<th>μ₆</th>
<th>FSH (IU/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>52</td>
<td>188</td>
<td>820</td>
<td>4028</td>
<td>21,556</td>
<td>123,068</td>
<td>738,100</td>
<td>150</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>52</td>
<td>308</td>
<td>1924</td>
<td>12,740</td>
<td>89,428</td>
<td>662,228</td>
<td>5,131,684</td>
<td>75</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>52</td>
<td>400</td>
<td>3140</td>
<td>25,120</td>
<td>204,500</td>
<td>1,691,440</td>
<td>14,189,540</td>
<td>75</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>52</td>
<td>488</td>
<td>4660</td>
<td>45,224</td>
<td>445,492</td>
<td>4,449,128</td>
<td>44,994,100</td>
<td>75</td>
</tr>
</tbody>
</table>

Appendix B. The backprediction of follicle size distribution (FSD) from the moment values

To predict the number of follicles in a particular size bin on a cycle day of FSH dosage regime the non-linear programming constrained optimization approach is used (Yenkie et al., 2013a).

The decision variables in the back prediction algorithm are the initial values for the SDE parameter set \( \theta \) of follicles per day in the cycle.

Step 1: Assign some initial values to \( n \) (number of follicles/day) within the different size bins used in the model

Step 2: Obtain the moment values by multiplying the matrix \( A \) (see Eqs. (5) and (6)) with the initially assumed \( n \) values.

Step 3: Introduce the constraint for total number of follicles in an IVF cycle for a particular patient to be \( \mu_0/2 \).

Step 4: Restrict the values of \( n \) to be either positive or zero since number of follicles can never be negative.

Step 5: Write the objective function to minimize the sum of square of errors between the model predicted moments and the moment values evaluated from the data using Eq. (4).

Step 6: Use a constrained non-linear optimization method to obtain the values of \( n \).

Step 7: Compare the optimum values of \( n \) obtained from this constrained optimization method to the actual data observed for the patient.

This algorithm remains the same for the deterministic as well as stochastic model validation.

Appendix C. Details of the new two level parameter estimation method used for the stochastic model development

The SDE consists of two parts; the drift term or the deterministic component and the diffusion term or the stochastic component (Eq. (C.1)). The parameters describing the dynamics of the system, which are also present in the deterministic model and are called as the SDE parameter set \( \theta \). The additional parameter in the SDE due to the randomness is addressed as the standard deviation \( \sigma \).

\[
dx_t = a_t(x_t, \theta) \, dt + \sigma b_t(x_t, t) \, dz_t
\]  
(C.1)

Here, \( x_t \) is the state variable, \( a_t \) and \( b_t \) are known functions, \( dt \)–time interval, \( dz_t \) is the Wiener process increment \( (\epsilon_t \sqrt{dt}) \).

The estimation procedure is divided into two levels:

**Outer level**—Procedure for evaluation of \( \theta \), the SDE parameter set, using the deterministic optimization methods

**Inner level**—Procedure for evaluation of the randomness estimate or standard deviation \( \sigma \), as an embedded function of \( \theta \), model and variance of the available data

The inner level evaluations can be performed by choosing some initial values for the SDE parameter set \( \theta \), the model information and the available data. The outer level involves the computation of better estimates for \( \theta \) and modifications in the deterministic objective function with an additional term involving the standard deviation \( \sigma \) estimated in the inner level.

The steps involved in this method are given below as follows:

**Step 1**: Specify initial values of \( \theta \)

**Step 2**: Obtain values of the standard deviation \( \sigma \) using the initial model information and data

**Step 3**: Find the value of the modified objective function for the specified value of \( \theta \).

**Step 4**: Find the derivative value and check if this value of objective function is optimum, if yes then stop, else find new values of \( \theta \) using deterministic nonlinear programming method and go to Step 2.

An example from the species population decline modeled using the stochastic logistic function (Eq. (C.2)) is shown. The parameter values are evaluated using the deterministic approach (without the second term in Eq. (C.2)), two conventional methods for SDE parameter estimation (simulated maximum likelihood estimation (MLE) and the generalized method of moments (GMM)) and are compared with the values estimated from the new two level method.

\[
dx_t = r x_t \left( 1 - \frac{x_t}{K} \right) dt + x_t \sigma dz_t
\]  
(C.2)

Here, \( x_t \) is the species population, \( r \) and \( K \) are the parameters corresponding to the combined birth and death rates (both are second order functions of \( x_t \)), \( \sigma \) is the standard deviation corresponding to the uncertainty in the system and \( dz_t \) is the Wiener process increment. The equation has a second order function with respect to \( x \) in the deterministic part and also has a dependency on \( x \) in the stochastic part. Thus, this closely resembles to the geometric Brownian form on the Ito process. The SDE parameter set \( \theta \) has two constants \( r \) and \( K \) to be estimated in the outer level of the proposed method. The results from this example, the stochastic logistic function is shown in Fig. C.1. The discrete markers are the data points, the dotted curve is the result from the deterministic model (i.e., considering only the first term in the RHS of Eq. (C.2))), the results from the simulated MLE method and the GMM approach are also shown along with the results from the proposed new method (solid black curve). The results for the three stochastic methods (MLE, GMM, new method) are evaluated by taking an average over 100 scenarios obtained on running the stochastic model using the estimated parameters from the respective methods. The parameter values estimated from the different methods are shown in Table C.1. It can be seen that the new method gives much better results as compared...
to the deterministic approach as well as the traditional MLE and GMM methods.

The advantages of the method are that it does not require sampling and hence is less time consuming as compared to other stochastic methods. However, it cannot be used for stochastic differential equations having parameters in addition to the standard deviation ($\sigma$) in the stochastic part.

## References


