Uncertainty Quantification driven Predictive multi-scale model for synthesis of mycotoxins

Sourav Banerjee¹, Gabriel Terejanu², Anindya Chanda³

¹Dept. of Mechanical Engineering, University of South Carolina, Columbia, South Carolina, USA
²Dept. of Computer Science and Engineering, University of South Carolina, Columbia, South Carolina, USA
³Dept. of Environmental Health Sciences, University of South Carolina, Columbia, South Carolina, USA

Email address:
banerjes@cec.sc.edu (S. Banerjee), terejanu@cse.sc.edu (G. Terejanu), achnada@mailbox.sc.edu (A. Chanda)

To cite this article:

Abstract: Many toxic molds synthesize and release an array of poisons, termed mycotoxins that have an enormous impact on human health, agriculture and economy [1]. These molds contaminate our buildings, indoor air and crops, cause life threatening human and animal diseases and reduce agricultural output [2]. In order to design appropriate approaches to minimize the detrimental effects of these fungi, it is essential to develop diagnostic methodologies that can rapidly and accurately determine based on fungal strains and their growth patterns, the extent of mycotoxin mediated damage caused to the environment. Here we developed a novel multi-scale predictive mathematical model that could reliably estimate aflatoxin synthesis from growth features extracted from Aspergillus parasiticus, a well-characterized model for studying mycotoxin biosynthesis. We conducted acoustic imaging experiments to observe and extract the growth features from the biomass profiles of the growing Aspergillus colony growing on an aflatoxin-inducing solid growth medium. We employed the probability-based representation of uncertainty and used Bayes’ theorem to infer the uncertain parameters in our mathematical model using biomass observations of the colony at 24h (aflatoxin is not synthesized yet at this time-point) and 48 hours (aflatoxin synthesis occurs at peak levels). We demonstrate that our model could successfully predict with quantified uncertainties the levels of aflatoxin secreted to the environment by the fungus.

Keywords: Predictive Multi-scale Model, Aflatoxin synthesis, Fungi, Aspergillus, Scanning Acoustic Microscopy, Uncertainty Quantification

1. Introduction

A fungal mycelium is composed of tubular structure termed hyphae [3], which continuously branch out and find new source of nutrients to ensure growth and survival of the cellular colony. By creating hydrostatic pressure differences, hyphae advance their frontier [4]. Using the process of dichotomous and lateral branching [5], hyphae create the mycelial network, in which the older branches eventually collides with existing braches and annihilated by the process of anastomosis [6]. The nutrients are internalized by the hyphae [7-10] and are sent through the mycelial network to the area of high nutrient demand, using a process called translocation [8,11,12]. As hyphae propagate, they leave behind a stationary biomass [13] that consists of older cells, which are structurally stable and promote further growth of the hyphae. Earlier studies with mycotoxin producers suggest that mycotoxin production promotes hyphal growth and vice-versa. Biomass accumulation is positively correlated with mycotoxin synthesized and released by the mold[11] because hyphae propagate opposite to the increasing gradient of the mycotoxins and away from the toxigenic cells in the colony, a phenomenon described as negative chemotropism [14, 15]. Hence, we hypothesized that an accurate mathematical model for describe hyphal growth patterns in mycotoxin producing mold could successfully predict the levels of mycotoxin produced by the colony. In order to be accurate, such a model should be able to describe both the hyphal growth features (biological scale) and the negative chemotropism-driven colony expansion (behavioral scale). Hence, we reasoned that only a multi-scale model could satisfy such requirements.

Currently, two distinct mathematical models are popularly used to describe fungal growth: 1) discrete model [15-17], where a random model of branching structure is used for
hyphal network and is applicable to hyphal scale and 2) continuum model, where an extremely complex structure of mycelial network is simplified to a biomass density profile [18–20] and is applicable to a colony scale. A discrete model relies on the stochastic processes of hyphal growth and can only describe a small area of interest (hyphal scale). On the other hand, a continuum model aims to understand the macroscopic behavior of the mycelium; however, they do not consider either the multi-scale hyphal interaction or the reverse chemotropism occurring at the meso scale. To overcome these limitations, we developed a predictive growth model for mycotoxigenic fungi that incorporates the co-dependence of their multi-scale nature of growth and mycotoxin production. For the current work we used *Aspergillus parasiticus* our biological model. A radially growing *A. parasiticus* colony on an agar growth medium is shown in Figure 1. The fungus is a well-characterized model for studying mycotoxin biosynthesis because it synthesizes the hepatocarcinogen, aflatoxin that contaminates our crops, resulting in billions of dollars (~$2.5 billion per year) of loss to agriculture and economy [1,2].

To develop the predictive model, we have considered two continuum models at two different scales and coupled them with a systematic uncertainty quantification processes and designed the necessary experiments for model driven data collection. To maximize the number of observables during data collection for growth, we employed an acoustic microscopy based approach that has been developed by us recently (manuscript submitted elsewhere) and is especially suited for the multi-scale hyphal growth patterns in the most minimally invasive way.

It is obvious that several unknown parameters in the model cannot be measured by any means. The values of those parameters are also uncertain. Such uncertainties could lead to complete erroneous interpretations. To circumvent this, uncertainty quantification (UQ) of such model parameters was performed with Bayesian inference. We demonstrate here that after calibration with the feedback from real time acoustic imaging data our model successfully predicted the spatial and temporal profiles of aflatoxin concentrations in an *A. parasiticus* colony.

### 2. Mathematical Formulation

With uniform nutrient concentration available in all directions, *A. parasiticus* colony grows radially almost uniformly in all directions (Figure 1). Hence, we considered an axisymmetric problem and analyses were performed only along the radial direction [18, 24]. A stationary radial biomass with radius $R$ at time $t$ resulting from continuous hyphal growth was denoted as $B_m(R, t)$. As mentioned earlier, the biomass also supports further hyphal growth.

We denoted $(\partial T_c / \partial R)$ as the toxin gradient along the radius, which motivates the hyphal propagation against the aflatoxin gradient and the annihilation of the hyphal tips at the meso scale resulting in the mycelial network. We denoted the rate of change of biomass equal to the gradient of advective flux ($\Psi = -v \partial T_c / \partial R$) created by hyphal front where $v$ is the hyphal growth velocity. Considering the physical phenomena described above, we used three coupled differential equations. The equation that simulated the local and meso scale behaviors of nutrient uptake by the fungal colony is written below:

$$U(r, t) = S_0 - S_e(r, t)$$

where, $U(r, t)$ denoted nutrient uptake, $S_0$ denoted the initial uniform nutrient concentration, $S_e(r, t)$ denoted the remaining nutrient in the colony and $r$ and $t$, respectively denoted the special and temporal independent variables. The scale $r$ was considered 100 times smaller than the special variable $R$ in the macro scale. Since nutrient absorption occurs at the hyphal scale (by the individual hyphae), we considered the nutrient uptake to be a meso scale phenomenon, a scale smaller than colony scale by the order of 100 (an experimental observation from our imaging experiments with *Aspergillus* colonies). In the colony scale the formation of stationary biomass is a dominant phenomenon coupled with the synthesis of aflatoxin. Since the events at the two scales (nutrient uptake, formation of...
biomass and aflatoxin synthesis) influence each other, a set of coupled differential equations was introduced to capture the complete physics. For two different scales the following equations were used to describe the synthesis of the aflatoxin in an A. parasiticus colony.

Lower scale or meso scale equation
\[
\frac{\partial u}{\partial t} = D_e \frac{\partial^2 u}{\partial x^2} - k_u \frac{\partial B_m}{\partial x} \tag{1}
\]

Higher scale or macro scale equations
\[
\frac{\partial B_m}{\partial t} = k_1 \frac{\partial^2 B_m}{\partial x^2} - v \frac{\partial \tau_c}{\partial x} + a_1 U \tag{2}
\]
\[
\frac{\partial \tau_c}{\partial t} = D_c \frac{\partial^2 \tau_c}{\partial x^2} + a_2 B_m \tag{3}
\]

All the relevant parameters in the equations were coupled with their respective proportionality constants. The coupling was performed to consider the respective influence of the dependent variables in the comprehensive phenomena. These parameters cannot be measured by any means, using any possible experimental setup. However, the biomass can be measured using acoustic imaging. The acoustic microscope with varying frequency was capable of measuring the ultrasound wave velocity in the fungal specimens and from the travel time of the wave between the top and the bottom of the colony, the thickness of the colony could be determined. The thickness of the colony, a radial function of (i.e. the function of \( R \), is directly proportional to the biomass at any special position in the colony. From an infinitesimal pixel size at any point (e.g. acoustic image pixel) inside the colony in 2D, the local thickness of the colony (gives volume) and density of the colony could be calculated to get the mass of the colony at that pixel point. For our imaging experiment set specific frequency the pixel size is well defined and hence constant.

Assuming the density of the colony (which is non measurable) to be constant, we obtained the thickness of the colony could be calculated to get the mass of the colony at any special position in the colony. From an infinitesimal pixel size at any point (e.g. acoustic image pixel) inside the colony in 2D, the local thickness of the colony (gives volume) and density of the colony could be calculated to get the mass of the colony at that pixel point. For our imaging experiment set specific frequency the pixel size is well defined and hence constant. For the current work, the term “thickness of the colony” is used synonymously as “biomass” (\( B_m(R, t) \)), the dependent variable. \( B_m(R, t) \) was obtained from acoustic imaging experiments and the values (see Figure 1) have been used in the current work to calibrate the developed model.

In the equations 1, 2 and 3, two spatial scales, \( r \) (meso scale) and \( R \) (colony scale) and one time scale \( t \) are introduced. Scales were further normalized with respect to their relevant parameters, \( \lambda_{48} \) and \( R_{48} \), where \( \lambda_{48} \) is the wave length of hyphal wave front of the Aspergillus colony after 48 hours and is an observable parameter from imaging at meso scale and \( R_{48} \) is the radius of the colony after 48 hours, which is also a measurable quantity from imaging at the colony scale. Let \( x = r/\lambda_{48} \) and \( X = R/R_{48} \), considering the numerical values of these normalized scales we can say \( x/X = 10^{-2} \). All the other parameters were subsequently normalized as follows: \( B_m^* = B_m/M_{48} \), \( T_c^* = T_c/T_{48} \), \( U^* = U/S_0 \) and \( t^* = t/48 \). Here, \( M_{48} \) is the total biomass of the colony after 48 hours and \( T_{48} \) is the total aflatoxin concentration produced by the colony after 48 hours. Time was normalized by 48 hours, since aflatoxin production at this time point reaches peak levels.

Normalizing all the equations (1) – (3) and dropping the * from the variables we get following three dimensionless coupled differential equations.

\[
\frac{\partial u}{\partial t} = D_e \frac{\partial^2 u}{\partial X^2} - K_u \frac{\partial B_{m^*}}{\partial X} \tag{4}
\]
\[
\frac{\partial B_{m^*}}{\partial t} = K_1 \frac{\partial^2 B_{m^*}}{\partial X^2} - V \frac{\partial \tau_{c^*}}{\partial X} + a_1 U \tag{5}
\]
\[
\frac{\partial \tau_{c^*}}{\partial t} = D_c \frac{\partial^2 \tau_{c^*}}{\partial X^2} + a_2 B_{m^*} \tag{6}
\]

Where,
\[
D_e = 48 \times 10^3 D_e/\lambda_{48}^2
\]
\[
D_c = 48 D_c/R_{48}^2
\]
\[
K_2 = 48 k_2 M_{48}/R_{48} S_0
\]
\[
K_1 = 48 k_1/R_{48}^2
\]
\[
V = 48 T_{48} V/R_{48} M_{48}
\]
\[
\alpha_1 = a_1 S_0 \text{ and } \alpha_2 = a_2 M_{48}
\]

It can be easily identified that except \( R_{48} \) and \( \lambda_{48} \), the parameters in the equations are too complicated and cannot be measured by any possible experimental methods. Thus in this foregoing discussion we classified these parameters as uncertain. We considered that the only possible measurable quantity is the biomass density (described before). Please note that all the parameters are dimensionless (varies between 0-1) and the biomass observable shown in Figure 1 was normalized with respect to the maximum thickness of the colony at the end of 72 hours, which is 1256.2 \( \mu \)m. The biomass profile at 24 hours was used to fit the initial condition and partially inform the prior distribution of the parameters and the 24 hours profile was used for calibration. The following boundary and initial conditions were used in solving the partial differential equations previously introduced.

<table>
<thead>
<tr>
<th>Left end</th>
<th>Right end</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \frac{\partial B_m}{\partial X} = 0 )</td>
<td>( B_m = 0 )</td>
</tr>
<tr>
<td>( \frac{\partial \tau_c}{\partial X} = 0 )</td>
<td>( T_c = 0 )</td>
</tr>
<tr>
<td>( \frac{\partial U}{\partial X} = 0 )</td>
<td>( U = 0 )</td>
</tr>
</tbody>
</table>

Initial Conditions
\[
B_m(x, 0) = H e^{-400 x^2}, \text{ for } x \leq 0.1 \text{ and 0 otherwise }
\]
\[
U(x, 0) = 0
\]
\[
T_c(x, 0) = 0
\]
3. Model Calibration and Numerical Results

The inherent uncertainties in the previously derived model are the result of both parametric uncertainty and model structure inadequacies. Given that the normalized biomass is experimentally observable, the following measurement model will be used to calibrate our model.

\[ \bar{B}_m(x) = B_m(x, \theta, t) + \epsilon_g(x) \]  \hspace{1cm} (7)

Here, \( B_m(x, \theta, t) \) represents the normalized prediction of the model at a specific time \( t \) and distance from the center of the colony \( x \), given a set of parameters \( \theta = [V, K_1, \alpha_1, D_\alpha, K_2, D_\epsilon, \alpha_2] \) and \( \bar{B}_m(x) \) is the corresponding experimental data obtained using SAM. Possible missing interaction terms in the partial differential equations cause structural errors in the model that need to be quantified. To address this, a very common approach in the Bayesian literature initially suggested by Kennedy and O’Hagan [25], is to introduce a statistical model, \( \epsilon_g \), that can explain the discrepancy between model predictions and observations. This statistical model is essential to ensure a correct model calibration as well as to provide additional insights regarding the source and magnitude of additional errors and the measurement errors. At this initial stage there is no prior information regarding the magnitude of any of the two error types, thus the variance of the normal distribution will be inferred from the data along with the parameters defining the growth model. Given our reliance on probability to represent uncertainty, we formulate this inference problem as a Bayesian update,

\[ p(\theta, \sigma^2 | \bar{B}_m) = \frac{p(\bar{B}_m | \theta, \sigma^2)p(\theta, \sigma^2)}{p(\bar{B}_m)} \]  \hspace{1cm} (9)

The discrepancy model has to explain both the modeling errors and the measurement errors. At this initial development stage there is no prior information regarding the magnitude of any of the two error types, thus the variance of the normal distribution will be inferred from the data along with the parameters defining the growth model. Given our assumption that the biomass observations are spatially independent, the likelihood function is given by the product of individual likelihood functions

\[ p(\bar{B}_m | \theta, \sigma^2) = \prod_{i=1}^{N} p(\bar{B}_m(x_i) | \theta, \sigma^2) \]
\[ = \prod_{i=1}^{N} N(\bar{B}_m(x_i) - B_m(x_i, \theta, t), \sigma^2) \]  \hspace{1cm} (10)

The denominator in the Bayesian formula, \( p(\bar{B}_m) \) is a normalization constant, and the posterior pdf \( p(\theta, \sigma^2 | \bar{B}_m) \) is the Bayesian solution of the inverse problem, representing the desired estimate of the parameters with associated uncertainties. It is assumed that a priori the model parameters \( \theta \) are independent of the variance \( \sigma^2 \) of the discrepancy model.

\[ p(\theta, \sigma^2) = p(\theta)p(\sigma^2) \]  \hspace{1cm} (11)

In the current study we use a non-informative prior for the parameters, \( p(\theta) \propto \text{const.} \), and we use the biomass profile at 24h to fit the initial condition and inform the prior pdf of the variance, and the data at 48h to infer the parameters of our model. The following model form defines the initial condition:

\[ B_m(x, 24h) = ae^{-bx^2} \]  \hspace{1cm} (12)

The two parameters \( a \) and \( b \), are estimated by minimizing the mean square error (MSE) between the biomass observations at 24h and model predictions. The cost function is minimized using the function fminunc in the optimization package of Octave [26]. The results of this optimization are shown in Figure 1(i) where the initial condition is given by \( a = 0.26 \) and \( b = 344.33 \). The minimum MSE found is also used to inform the prior pdf of the variance that takes the form of an inverse gamma pdf.

\[ p(\sigma^2) = \frac{\beta^\alpha}{\Gamma(\alpha)} \sigma^{\alpha-1} e^{-\frac{\beta}{\sigma^2}} \]  \hspace{1cm} (13)

The inherent uncertainties in the previously derived model are the result of both parametric uncertainty and model structure inadequacies. Given that the normalized biomass is experimentally observable, the following measurement model will be used to calibrate our model.

\[ \bar{B}_m(x) = B_m(x, \theta, t) + \epsilon_g(x) \]  \hspace{1cm} (7)

Here, \( B_m(x, \theta, t) \) represents the normalized prediction of the model at a specific time \( t \) and distance from the center of the colony \( x \), given a set of parameters \( \theta = [V, K_1, \alpha_1, D_\alpha, K_2, D_\epsilon, \alpha_2] \) and \( \bar{B}_m(x) \) is the corresponding experimental data obtained using SAM. Possible missing interaction terms in the partial differential equations cause structural errors in the model that need to be quantified. To address this, a very common approach in the Bayesian literature initially suggested by Kennedy and O’Hagan [25], is to introduce a statistical model, \( \epsilon_g \), that can explain the discrepancy between model predictions and observations. This statistical model is essential to ensure a correct model calibration as well as to provide additional insights regarding the source and magnitude of additional model errors that we might have overlooked. The spatial errors in the biomass profile are assumed to be independent and identically distributed with a zero mean normal distribution.

\[ \epsilon_g(x) \sim N(0, \sigma^2) \]  \hspace{1cm} (8)

The discrepancy model has to explain both the modeling errors and the measurement errors. At this initial development stage there is no prior information regarding the magnitude of any of the two error types, thus the variance of the normal distribution will be inferred from the data along with the parameters defining the growth model. Given our assumption that the biomass observations are spatially independent, the likelihood function is given by the product of individual likelihood functions

\[ p(\bar{B}_m | \theta, \sigma^2) = \prod_{i=1}^{N} p(\bar{B}_m(x_i) | \theta, \sigma^2) \]
\[ = \prod_{i=1}^{N} N(\bar{B}_m(x_i) - B_m(x_i, \theta, t), \sigma^2) \]  \hspace{1cm} (10)

The denominator in the Bayesian formula, \( p(\bar{B}_m) \) is a normalization constant, and the posterior pdf \( p(\theta, \sigma^2 | \bar{B}_m) \) is the Bayesian solution of the inverse problem, representing the desired estimate of the parameters with associated uncertainties. It is assumed that a priori the model parameters \( \theta \) are independent of the variance \( \sigma^2 \) of the discrepancy model.

\[ p(\theta, \sigma^2) = p(\theta)p(\sigma^2) \]  \hspace{1cm} (11)

In the current study we use a non-informative prior for the parameters, \( p(\theta) \propto \text{const.} \), and we use the biomass profile at 24h to fit the initial condition and inform the prior pdf of the variance, and the data at 48h to infer the parameters of our model. The following model form defines the initial condition:

\[ B_m(x, 24h) = ae^{-bx^2} \]  \hspace{1cm} (12)

The two parameters \( a \) and \( b \), are estimated by minimizing the mean square error (MSE) between the biomass observations at 24h and model predictions. The cost function is minimized using the function fminunc in the optimization package of Octave [26]. The results of this optimization are shown in Figure 1(i) where the initial condition is given by \( a = 0.26 \) and \( b = 344.33 \). The minimum MSE found is also used to inform the prior pdf of the variance that takes the form of an inverse gamma pdf.

\[ p(\sigma^2) = \frac{\beta^\alpha}{\Gamma(\alpha)} \sigma^{\alpha-1} e^{-\frac{\beta}{\sigma^2}} \]  \hspace{1cm} (13)

Here, \( \Gamma(\cdot) \) is the gamma function, the shape parameter \( \alpha = 10 \) and the scale parameter is given by the minimum MSE, namely \( \beta = 3.0553e-05 \). Finally, having defined the likelihood function and the prior pdf, we can obtain samples from the posterior pdf using Markov Chain Monte Carlo (MCMC) methods. In this paper we have adopted the MCMC Delayed Rejection and Adaptive Metropolis [27], which adapts the proposal distribution both locally after...
each rejection as well as globally using the past chain. The histogram corresponding to the posterior pdf of all eight parameters are given in the Figure 2(a)-(h).

Calibrating the model does not guarantee that its outputs will be consistent with the data. It is therefore necessary that consistency of model outputs with the experimental observations be explicitly checked. The discrepancy between the data and model outputs was quantified in this study using posterior predictive checks [28] such as determining whether the data lies inside the 90% credible interval of the posterior predictive pdf of $B_m$. Based on the literature, we determined that appropriate for our use are highest posterior density (HPD) confidence intervals [29]. The 90% HPD interval is an interval or a set of intervals for which the probability of belonging to it is 90% and the density of all the points in the interval are greater than the density of the points outside the interval. If more than 10% of the data is not in the 90% HPD region, then the data would be considered an implausible outcome of the model. This is the plausibility criterion we used here and the concept was clearly similar to p-values and confidence intervals in classical statistics, though they differ in detail and interpretation. Finally, note that for multi-modal distributions, an HPD region may consist of multiple disjoint intervals [30]. In Figure 3 all three profiles, biomass, toxin and uptake were predicted at 48h with their associated uncertainties. The majority of biomass readings were within the 90% HPD region.

Future work will focus on further strengthening the predictive capability of our current mathematical model. We are already conducting experiments that are designed to simultaneously measure both the biomass and the nutrient uptake under various experimental settings to provide additional insight on the correlation between nutrient distribution, fungal biomass profiles and the spatiotemporal profiles of aflatoxin synthesis with a colony. The knowledge will significantly help in improving the predictive capability of our model and further reduce uncertainties. The same posterior predictive checks for consistencies between model predictions and predictive experimental data (as described in the current work) will be conducted in those future studies.

4. Conclusion

The main goal of the current work was to develop a mathematical model that can be used to effectively predict mycotoxin profiles in mold colonies. We used one of the most well-studied aflatoxin (a hepatocarcinogenic mycotoxin) producers, Aspergillus parasiticus to develop the predictive model. Based on a data set of just biomass profiles of the colonies growing in a medium with known nutrient concentration, we successfully quantified the profile of aflatoxin concentration in the fungal colonies. Obtaining such profiles with established ELISA or chromatography-based methods is tedious, time consuming, costly and has occupational hazards of mycotoxin exposure. In addition the mathematical model can also predict several non-measurable parameters without altering or perturbing the specimens.

The use of uncertainty quantification should also allow the adaptability of our model for other mycotoxin producers. Except biomass, all other parameters in the model are assumed to be uncertain. Hence, we anticipate that irrespective of species, only the measurement of mycelial biomass profiles could lead to successful quantification of their mycotoxin profiles without any tedious wet laboratory experiments.

Acknowledgements

Authors would like to thank the Office of Vice President of Research at the University of South Carolina and their respective departments for providing financial support and the infrastructure for this collaborative research venture. Authors would also like to thank Dr. Robin Brigmon with Savannah River National Laboratory, Aiken, South Carolina, for his valuable suggestions in this joint collaborative effort.

References


