Johns Hopkins University, Dept. of Biostatistics Working Papers

1-18-2008

STATISTICAL METHODS FOR AUTOMATED DRUG SUSCEPTIBILITY TESTING: BAYESIAN MINIMUM INHIBITORY CONCENTRATION PREDICTION FROM GROWTH CURVES

Xi Zhou
Division of Biostatistics and Epidemiology, Department of Public Health, Weill Medical College of Cornell University

Merlise A. Clyde
Department of Statistical Science, Duke University

James Garrett
Manager, Bioinformatics, Algorithms and Non-Clinical Statistics Group, Becton-Dickinson Diagnostic Systems

Viridiana Lourdes
Vice-President, Investment Management, Morgan Stanley

Michael O'Connell
Director of Life Sciences, Insightful Corporation

See next page for additional authors

Authors
Xi Zhou, Merlise A. Clyde, James Garrett, Viridiana Lourdes, Michael O'Connell, Giovanni Parmigiani, David J. Turner, and Tim Wiles

Statistical Methods for Automated Drug Susceptibility Testing: Bayesian Minimum Inhibitory Concentration Prediction from Growth Curves

Xi Zhou, Merlise A. Clyde, James Garrett, Viridiana Lourdes, Michael O'Connell, Giovanni Parmigiani, David J. Turner, and Tim Wiles *

Abstract

Determination of the minimum inhibitory concentration (MIC) of a drug that prevents microbial growth is an important step for managing patients with infections. In this paper, we present a novel probabilistic approach that accurately estimates MICs based on a panel of multiple curves reflecting features of bacterial growth. We develop a probabilistic model for determining whether a given dilution of an antimicrobial agent is the MIC given features of the growth curves over time. Because of the potentially large collection of features, we utilize Bayesian model selection to narrow the collection of predictors to the most important variables. In addition to point estimates of MICs, we are able to provide posterior probabilities that each dilution is the MIC based on the observed growth curves. The methods are easily automated and have been incorporated into the Becton-Dickinson PHOENIX automated susceptibility system that rapidly and accurately classifies the resistance of a large number of micro-organisms in clinical samples. Over seventy-five studies to date have shown this new method provides improved estimation of MICs over existing approaches.

Key Words: Bayes, BIC, decision theory, logistic regression, model selection, model uncertainty

*Xi Kathy Zhou is Assistant Professor, Division of Biostatistics and Epidemiology Department of Public Health Weill Medical College of Cornell University 411 East 69th Street, KB-321 New York, NY 10021 (email: kaz2004@med.cornell.edu), Merlise Clyde is Associate Professor, Department of Statistical Science, Duke University, Durham, NC 27708-0251 (email: clyde@stat.duke.edu), James Garrett is Manager, Bioinformatics, Algorithms, and Non-Clinical Statistics Group, Becton-Dickinson Diagnostic Systems, Sparks, MD (email: Jim_Garrett@bd.com), Viridiana Lourdes is Vice-President, Investment Management, Morgan Stanley, NY,NY (email viridiana.lourdes@morganstanley.com), Michael O'Connell is Director Life Sciences, Insightful Corporation (email: moconnell@insightful.com), Giovanni Parmigiani is Professor, School of Medicine, Johns Hopkins University, Baltimore, MD 21205 (email: gp@jhu.edu), David Turner, Senior Scientist, ID/AST Research and Development, Becton-Dickinson Diagnostic Systems, Sparks, MD (email: David_J_Turner@bd.com), Tim Wiles, Senior Scientist, ID/AST Research and Development, Becton-Dickinson Diagnostic Systems, Sparks, MD (email: Tim_Wiles@bd.com). This research was partially supported by Becton-Dickenson and the National Science Foundation grant DMS–0406115. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of Becton-Dickinson or the National Science Foundation.

1 Introduction

Since the discovery of penicillin in the late 19th century, microbiology has undergone rapid development. Large number of antibiotics have been identified and have greatly improved the management of patients with infectious diseases. Identification and susceptibility testing of clinically obtained isolates are performed daily in clinical laboratories across the world. Rapid and accurate identification of micro-organisms and especially their drug resistance is critical for the management of patients. Speedy reporting of these testing results have been shown to have clinical and financial benefits not only to patients but also to clinical laboratories (Barenfanger et al. 1999).

Despite such rapid progress, routine laboratory susceptibility testing methods still take 24 hours to provide results. Current methods for antimicrobial susceptibility testing (AST) are largely phenotypic approaches based on the growth patterns of the micro-organisms in antimicrobial agents (Wheat 2001). The minimum inhibitory concentration (MIC) is defined as the lowest concentration above which bacterial growth is inhibited in vivo testing. The MIC value is then utilized to determine the clinical effectiveness of a given antibiotic for the isolate being tested. The interpretation of the MIC is reported as susceptible (S) if the combination is likely to be effective (a low MIC value) and resistant (R) if treatment failure is likely (a high MIC value). In some cases an intermediate result is reported when the MIC value falls between the S and R values.

Efforts for speeding up the process of susceptibility testing have resulted in the development of automated AST systems. With automation, a large number of tests can be performed at the same time and the amount of the data to be analyzed increases. There is a crucial need, however, for developing statistical methods that produce reliable MIC estimates from such systems. In this paper, we develop a novel real-time statistical method that allows more accurate and rapid determination of MICs based on a panel of growth curves for a

given isolate exposed to various dilutions of an anti-microbial agent. This methodology is currently implemented in Becton Dickinson's (BD) newly developed PHOENIX AST system (BD Diagnostic Systems, Sparks, MD). This system provides rapid and reliable susceptibility testing of a majority of clinically encountered bacterial strains and is currently utilized by a large number of hospitals across the globe. To date, more than seventy-five recent studies have shown good reliability of this AST system in MIC estimation and breakpoint determination for various micro-organisms and drugs (see Fahr et al. 2003; Donay et al. 2004; Horstkotte et al. 2004, for example).

In the next section, we describe the structure of the panel growth data obtained during the development of the BD PHOENIX AST system. Our goal is to identify and use features of the growth curves to accurately predict the MIC for each set of panel data. In order to train the statistical algorithms, in section 3 we develop a model that predicts for each concentration a growth or no growth response, using available features of the curves from each panel. Features in the growth data that are crucial for predicting the probability of growth are selected using Bayesian model selection. For each isolate and drug dilution in a panel, we construct an estimate of the probability of growth using these selected features of the growth curves. In Section 4, we present a novel method to combine the estimated growth probabilities for the sequence of drug dilutions in a panel to construct a probability distribution for the MIC. A decision theoretic approach for estimating the MIC is presented which balances the different types of errors in making an MIC call. In Section 5, we validate the method and illustrate its performance using two selected antibiotics, piperacillin and cefoxitin. We conclude with discussion in Section 6.

2 The PHOENIX AST System

PHOENIX is an instrumented antimicrobial susceptibility testing (AST) system for rapid identification and susceptibility determination of clinical samples. In this system, a biological sample is placed in each of eight wells on a test panel, where each panel is a single PHOENIX diagnostic disposable cartridge. Multiple panels are placed in a revolving carousel and at a sequence of twenty minute intervals the system moves the panels past a detector where red, green, and blue wavelengths are directed at each well. The system measures the resulting wavelengths that emanate from the wells; these optical measurements are used to generate two measures of microbial growth: the redox state and turbidity characteristics of the sample wells which are directly correlated with microbial growth. The redox state of a sample in a well is measured by utilizing the change in red, green and blue readings that occurs over time as a result of the reduction of a growth indicator, such as resazurin, by the antimicrobial material in the well. As the resazurin is reduced, the color of the sample in the well changes from blue to red to clear. This change in redox is represented numerically as a continuum, with the value 0 indicating an unreduced growth indicator (blue=resazurin), the value 0.5 indicating that the indicator is 50% reduced (red=resazurin), and the value 1.0 indicating that the indicator has been completely reduced (clear=dihydroresorufin). The turbidity is also estimated by using the red, green and blue readings. The initial signal has a value of 0 and a maximum of 2.25 units can be estimated. The units for turbidity correspond to McFarland units.

The system monitors growth over time to decide if the samples have incubated long enough to estimate the MIC. If the processor determines that the maximum redox state for the growth control well of the test panel is greater than 0.07 but less than a predetermined value of 0.2, the panel continues to be incubated, unless the elapsed incubation period has exceeded 16 hours. In this case, the test panel is failed as having insufficient sample

growth, and no results are reported for that test panel. If the processor determines that the maximum redox state for the growth control well is indeed greater than 0.2, the processor then determines whether the redox curve for the growth control well indicates that the sample is a slow or fast growing sample or not yet classifiable as either a slow or fast growing sample. In the latter case, panel incubation continues and the panel is retested twenty minutes later as discussed above. However, if the processor determines the curve classification is either fast or slow growing, the time at which this occurs is labeled as the "time-to-result" and the turbidity and redox data for each of the wells in the panel are extracted. A second-degree polynomial local regression algorithm (LOESS) is used to smooth the time series data for both redox and turbidity values calculated for each well over the elapsed period of time. From the LOESS fit any reading at any time point can be estimated. Moreover, first- and second-derivatives can be estimated at any point as a function of fitted local regression coefficients.

Figure 1 illustrates these interpolated growth curves for three PHOENIX panels in which three gram-negative bacterial strains were combined with eight dilutions of the drug piperacillin (PIP). Each curve corresponds to a different dilution of the drug PIP. The vertical line represents the time-to-result, the time at which features are extracted for determining the MIC. Independently from the PHOENIX AST system, a reference MIC is determined for each strain. In Figure 1, all curves corresponding to dilutions greater than the reference MIC are colored red, while all curves with dilutions less than or equal to the reference MIC are colored blue. The dilution corresponding to the reference MIC is illustrated with a blue dashed line.

3 Modelling Probability of Growth

Theoretically, we should see no growth in a well with a concentration higher than the true MIC. For our purposes, we will treat the reference MIC as the truth as this is the best available information. Using this external reference value of the MIC, we create a binary response variable Y_{ij} for dilution (curve) j in panel i defined as

$$Y_{ij} = \begin{cases} 1 \text{ if dilution } j \text{ is } < \text{ Reference MIC} \\ 0 \text{ if dilution } j \text{ is } \geq \text{ Reference MIC.} \end{cases}$$

This definition treats the reference MIC data as if it were measured without error and ignores any possible miss-classifications of curves. Inspection of the panels in the second row in Figure 1 indicates that several curves with dilutions below the reference MIC exhibit growth. Upon further verification, in this case the problem is not an error in the reference MIC data, but is conjectured to be related to the formulation of the panel where the actual concentrations of drugs in the wells are less than their nominal values. The predictive accuracy of the training model may possibly be improved through an iterative procedure by omitting such "outlier" panels that exhibited strong growth for dilutions below the MIC when the reference MIC could be verified as being correct.

3.1 Models

Conditional on the unknown probabilities (and MIC), we assume that indicators of growth in different wells and across panels are independent of each other,

$$Y_{ij} \mid \pi_{ij} \overset{\text{ind}}{\sim} \mathsf{Ber}(\pi_{ij}) \quad i = 1, \ldots, n, \quad j = 1, \ldots, J \tag{1}$$

where π_{ij} is the probability of growth or equivalently that dilution j in panel i is less than the MIC, n is the number of panels, and J is the number of wells in a panel ($J = 8$). While it is conceivable that observations in the same panel may be dependent, this assumption implies that such dependence may be captured through a model for the probabilities π_{ij}, for example, a panel specific random effect.

The indicator of GROWTH/NO GROWTH may be predicted as a function of the growth curves by modelling the probabilities of growth, π_{ij}, as a function of the time series for dilution j in panel i. While of course the entire time series could be used for modeling the GROWTH indicators, certain features of the curves may be sufficient for differentiating the pattern of GROWTH/NO GROWTH. In an experiment with no noise in the growth curves, we would actually need only the difference in growth measurements from the beginning to the time-to-result to separate the dilutions corresponding to GROWTH or NO GROWTH. However with noisy curves, it is not that simple and other features may be better predictors of growth. In Table 3.1, we list the features of the curves that were viewed as being scientifically relevant for predicting growth. These include the difference in growth, the area under the growth curve, the first derivative, the second derivative at the time-to-result and time points where maxima occurred. Most features are defined relative to the growth control well to standardize them across different isolate/drug combinations.

We use a logistic regression model, a generalized linear model (GLM), to relate the GROWTH indicator to the collection of features, where the linear predictor $\eta_{ij} \equiv \text{logit}(\pi_{ij}) \equiv \log(\pi_{ij}/(1-\pi_{ij}))$ is expressed as a linear function of selected features. As we are uncertain that the relationship is actually linear in the features in Table 3.1, we may try more flexible GLM models where we replace the linear functions by up to a third order polynomial in each feature. As the linear, quadratic, and cubic terms are typically highly correlated with each other, orthogonal polynomials are preferable from a computational perspective for model selection, but give equivalent predictions. More generally, the GLM can be represented in

FEATURE LABEL	DESCRIPTION
T.FD	Turbidity, 1st derivative of the Test Well
T.SD	Turbidity, 2nd derivative of the Test Well
T.IN	Turbidity, Integral of the Test Well
T.AB.M	Turbidity, Maximum Absolute of the Test Well
T.FD.M	Turbody, Maximum 1st Derivative of the Test Well
T.SD.M	Turbidity, Maximum 2nd Derivative of the Test Well
T.AB.M.R	T.AB.M(Test Well)/T.AB.M(Control Well)
T.FD.M.R	T.FD.M(Test Well)/T.FD.M(Control Well)
T.SD.M.R	T.SD.M(Tets Well)/T.SD.M(Control Well)
T.IN.R	T.IN(Test Well)/T.AB(Control Well)
T.FD.T	$t_{T.FD.M}$(Test Well) - $t_{T.FD.M}$(Control Well)
T.SD.T	$t_{T.SD.M}$(Test Well) - $t_{T.SD.M}$(Control Well)
R.FD	Redox, 1st derivative of the Test Well
R.SD	Redox, 2nd derivative of the Test Well
R.IN	Redox, Integral of the Test Well
R.AB.M	Redox, Maximum Absolute of the Test Well
R.FD.M	Redox, Maximum 1st Derivative of the Test Well
R.SD.M	Redox, Maximum 2nd Derivative of the Test Well
R.AB.M.R	R.AB.M(Test Well)/R.AB.M(Control Well)
R.FD.M.R	R.FD.M(Test Well)/R.FD.M(Control Well)
R.SD.M.R	R.SD.M(Test Well)/R.SD.M(Control Well)
R.IN.R	R.IN(Test Well)/R.AB Control Well)
R.FD.T	$t_{R.FD.M}$(Test Well) - $t_{R.FD.M}$(Control Well)
R.SD.T	$t_{R.SD.M}$(Test Well) - $t_{R.SD.M}$(Control Well)

Table 1: Characteristic features of growth curves used as covariates in the GROWTH or NO GROWTH model.

matrix notation as $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$, where $\mathbf{X}$ represents the $n \times p$ matrix of feature variables with columns for the linear, quadratic and cubic terms, $\boldsymbol{\beta}$ is the p dimensional vector of the unknown regression coefficients and $\boldsymbol{\eta}$ is the n-dimensional vector of the linear predictor. Once we have estimates of $\boldsymbol{\beta}$, these are used to calculate $\hat{\eta}_{ij}$, the estimate of the linear predictor, which in turn is used to obtain the estimates of the probabilities of GROWTH,

$$\hat{\pi}_{ij} = \frac{\exp(\hat{\eta}_{ij})}{1 + \exp(\hat{\eta}_{ij})}.$$

We may find that not all features are needed to model the probability of GROWTH; the variables represent features that could potentially explain the growth patterns and in some cases may be redundant. Using all features (and the higher polynomial terms) may lead to over-fitting of the model training data, and poor predictions on the out of sample validation data. Reducing the set of features is also important because of limited storage and computing capacity in the system for making predictions. Bayesian variable selection (Hoeting et al. 1999; Clyde and George 2004) can be used to reduce the set of features to prevent over-fitting while still providing excellent out-of-sample properties.

3.2 Bayesian Model Selection

We used Bayesian model selection based on the Bayes Information Criterion (Schwarz 1978; Kass and Raftery 1995) to reduce the set of features used in the prediction of GROWTH. This approach is equivalent to using a penalized deviance criterion for model selection, however the penalty depends on the sample size, which ensures that in large samples that the probability of the true model goes to one (under modest conditions), and provides consistent model selection (Kass and Raftery 1995). This choice typically results in more parsimonious models than selection based on, for example, the often-used Akaike's Information Criterion (AIC) (Akaike 1973, 1983a,b).

Models $\mathcal{M}$ correspond to different choices of features and polynomials in the features used to capture the smooth functions. Using BIC, the probability of a model, $\mathcal{M}_m$, given the collection of panel data $\mathbf{Y} = \{Y_{ij}\}_{i=1,\ldots,n,j=1,\ldots 8}$ is approximated by

$$p(\mathcal{M}_m|\mathbf{Y}) = \frac{\exp(\mathcal{L}(\mathcal{M}_m) - \frac{d_m}{2}\log(n))}{\sum_{m'}\exp(\mathcal{L}(\mathcal{M}_{m'}) - \frac{d_{m'}}{2}\log(n))}$$

where $\mathcal{L}(\mathcal{M}_m)$ is the log likelihood under model $\mathcal{M}_m$ evaluated at the maximum likelihood estimates $\hat{\boldsymbol{\beta}}$ and d_m is the number of terms in the model. Because of the large number of features and higher order terms, we cannot enumerate all models. To identify the highest posterior probability models, we used the deterministic search algorithm `bic.glm()` in S-PLUS to identify the best BIC models (Hoeting et al. 1999).

Because of the large number of potential features and limitations in the leaps and bounds algorithm used in `bic.glm()`, it was necessary to use a two stage procedure to identify which features would be incorporated in the selected model. The first stage was a screening stage where we identified the highest probability models that included only linear terms in the features. Using the subset of features that were included in the top model, we then added quadratic and cubic terms in these features, and repeated the calculations of the posterior model probabilities. While such a sequential approach had the potential not to identify the highest probability model overall, this scheme corresponds to a hierarchical structure that incorporates second order (and higher) terms only if there are important main (linear) terms. Secondly, because the search algorithm needed to be run for hundreds of other drugs, this led to a reasonably efficient computational strategy that could be run on computers at BD. After this two stage selection approach, the highest posterior probability model was used to determine the distribution of the MIC, as described in Section 4.

3.3 Model Fitting Results

The above procedures are applied to study the turbidity and redox growth curves for gram-negative bacterial strains exposed to dilutions of cefoxitin (FOX) (n = 1647 panels) and piperacillin (PIP) (n = 1599 panels). These two antibiotics were selected to provide a testbed for method development. FOX was considered to be an easier system to model, with more clearly distinguished cases of GROWTH/NO GROWTH, while PIP was viewed as being a more challenging case. For modelling purposes here, both fast and slow growing strains were combined.

In modelling GROWTH/NO GROWTH, we have made no effort to constrain the estimated probabilities π_{ij} in a panel to be ordered based on dilution. We chose not to constrain the π_{ij} as the actual dilution in a panel may differ from its nominal level leading to observed growth curves that may cross. Rather than explicitly building an order constraint into the model for GROWTH/NO GROWTH, we take the theoretical ordering into account in developing the prediction of the MIC, as described in the next section.

4 MIC Estimation

4.1 Overview of MIC Estimation

The modeling described in Section 3 operates at the level of a single well, and is aimed at determining the probability that each well is displaying growth. Note, however, that prediction of the MIC must be done at the level of a panel. In this section we describe our strategy for combining the well-level predictions and obtaining a MIC estimate for each panel. For panel i, our prediction algorithm consists of the following:

Estimate growth probabilities. For each well j in panel i estimate the probability π_{ij} that there is growth in well j.

Combine growth probabilities. For each well j in panel i estimate the probability ρ_{ij} that the dilution present in well j is the MIC. This distribution of the MIC is constructed by combining all the $\hat{\pi}_{ij}$ for panel i and is described in detail in Section 4.2.

Estimate the MIC. Using the distribution of the MIC, ρ_{ij}, derive an estimate for the MIC in panel i. Two estimates are discussed in detail: the "modal MIC" and a "decision theoretic MIC". The probability distribution of the MIC may also be used to delay the call, when there is a high degree of uncertainty about which dilution is the MIC.

Statistically, this overall procedure achieves the important practical goals of a) estimating growth based on simple physical features representing growth in a well; b) training the growth probability model on large datasets; c) adjusting the probability of growth in a well depending on the observations made on other wells in the same panel; and d) providing the basis for sequential estimation of the MIC, and decisions about delaying the call.

4.2 Probability Distribution of the MIC

We now describe how to derive the probabilities $\rho_{i1}, \ldots, \rho_{iJ}$ that the dilution present in each well is the MIC, using the set of probabilities $\hat{\pi}_{i1}, \ldots, \hat{\pi}_{iJ}$ that there is growth in each well. Suppose that the dilution for well j (denoted as D_j) is the MIC for panel i. If this were true, theoretically we would have the following sequence of curves: all curves with dilutions less than D_j would exhibit GROWTH, and all dilutions greater than or equal to D_j would have NO GROWTH. How do we compute the probability of this sequence given the probability of GROWTH/NO GROWTH for each well?

First we need to make additional assumptions of order (O) and independence (I):

O: Suppose that the dilution (concentration) in well j, D_j, is greater than the dilution in well k, D_k. Then, if D_k is inhibitory, so is D_j.

I: The outcome in well j, Y_{ij}, is independent of the outcome Y_{ik} in any other well $k \neq j$.

We previously utilized Assumption I in constructing the model for GROWTH/NO GROWTH. Within the nested restrictions of Assumption O, the probability of the sequence of curves that can leads to D_j being the MIC, can be computed under independence, Assumption I.

Consider the set of all sequences of growth curves that satisfy the ordering assumption O. We call these "valid" sequences. For example, if a panel had only three wells, the set of valid sequences of GROWTH/NO GROWTH results would be

D_1	$<$	D_2	$<$	D_3	MIC
0		0		0	$< D_1$
1		0		0	D_2
1		1		0	D_3
1		1		1	$>D_3$

By restricting attention to the reduced set of ordered sequences, we are now in a position to compute the probability ρ_{ij} of the sequence of curves that leads to D_j being the MIC in panel i. This ordered sequence is

D_1	$\ldots$	D_{j-1}	D_j	$\ldots$	D_J
1	1	1	0	0	0.

By comparing the likelihood of the sequence above to the likelihood of the other valid sequences, we can compute MIC probabilities that are consistent with the ordering assumption O,

$$\rho_{ij} = P(\text{MIC}_i = D_j) = \frac{\prod_{k<j} \pi_{ik} \prod_{k \geq j} (1 - \pi_{ik})}{\sum_{j=1}^{J+1} \prod_{k<j} \pi_{ik} \prod_{k \geq j} (1 - \pi_{ik})} \tag{2}$$

where the denominator is a sum over all $J + 1$ valid sequences.

For a particular model $\mathcal{M}_m$, this can be estimated by plugging in the $\hat{\pi}_{ij}$. To take into account model uncertainty, one can average the above expression over several models using the posterior probabilities of models as weights. While such a procedure may give better estimates, we have not pursued this direction because of the limited capacity for the PHOENIX system, and instead use the highest probability model.

The probabilities ρ_{ij} can be used in several ways to determine an estimate of the MIC. We experimented extensively with two approaches: the "modal MIC" and a "decision theoretic MIC". The modal MIC consists simply of choosing the dilution D_j for panel i with the largest probability ρ_{ij}. This is the optimal estimator under a loss function that is one if the estimate is not the true MIC and zero if the estimate is exactly the MIC. Our decision theoretic MIC takes into account that under-estimation and over-estimation have different costs and that errors of plus or minus one dilution are unimportant.

4.3 Decision Theoretic MIC Estimation

In evaluating performance in practice, a number of error types are evaluated. The essential agreement between estimated MICs and reference MICs are defined as agreement in MIC to within $\pm$ 1 dilution, and reflect that errors in estimation of $\pm$ one dilution are unimportant. According to the Food and Drug Administration (FDA 2007, page 21), the essential agreement should be greater than or equal to 90% . In general, underestimation of the MIC by more than one dilution is considered to be worse than overestimating the MIC by more than one dilution.

A second approach for evaluating performance utilizes the categorical classification of isolates as susceptible, resistant or intermediate. The S/I/R interpretations are compared and categorical agreement is evaluated. Based on FDA guidelines, the overall categorical agreement should be greater than or equal to 90%. Acceptable error rates are based on the clinical significance of the error. A very major error (VME) occurs when an isolate is called

sensitive when the isolate is in fact resistant. A VME rate greater than 1.5% of the resistant strains would be unacceptable. The major error threshold is more forgiving, less than or equal to 3%, and occurs when an isolate is called resistant when in actuality the isolate is sensitive. While we do not have the S/I/R classification for the isolates in our study, we can devise a decision theoretic estimator that captures the ideas behind the categorical errors, where underestimating the MIC leads to more severe consequences than overestimating it. One way to capture this is to attach a weight, or a loss, to each dilution error size. We use the following notation:

Error	*Loss*
Exact prediction (no error)	0
Predicted MIC within 1 dilution of reference MIC	w_3
Predicted MIC > reference MIC by more than 1 dilution	w_2
Predicted MIC < reference MIC by more than 1 dilution	w_1

The w's need to be chosen to reflect the relative importance of the errors and will typically be such that $w_1 > w_2 > w_3$.

To obtain the MIC prediction we choose the dilution that minimizes the expected loss. For dilution D_j this is

$$\mathcal{L}_j = w_1 \sum_{k>j+1} \rho_{ik} + w_2 \sum_{k<j-1} \rho_{ik} + w_3 \left(\rho_{i(j-1)} + \rho_{i(j+1)}\right).$$

The decision theoretic MIC prediction is the D_j associated with the j that minimizes $\mathcal{L}_j$. If one chooses $w_1 = w_2 = w_3$ the modal MIC and the decision theoretic MIC coincide. For the analyses reported later we used $w_1 = 5, w_2 = 1$ and $w_3 = 0$, reflecting that under estimating the MIC by more than one dilution is 5 times worse than over estimating it by more than one dilution. There is no penalty for being within one dilution of the reference MIC.

4.4 Uncertainty and Making a Call

The distribution of the MIC may be useful in formulating guidelines for deciding when to make a call on the MIC prediction or when to continue incubating a panel. One could determine the modal MIC or decision theoretic MIC and how much probability they receive, as well as how much probability mass the estimated MIC plus or minus 1 dilution receives. Sampling could continue until these reach a specified level of confidence. We expect that as sampling continues, the estimated π_{ij} will become closer to 1 or 0, so that the probability of the model MIC (the dilution with the largest probability) should increase to 1. However, early on the probabilities will not be as concentrated.

4.5 Illustration

Figure 2 shows the MIC distribution for the three bacterial strains that were presented in the growth curves in Figure 1, while Table 2 lists the estimates of the MIC and the probability mass that they receive. For strain 144882, the distribution of the MIC is uni-modal with

TEST ID	REF MIC P(REF ±1)	Modal MIC P(Modal MIC)	Decision Theoretic MIC P(DT MIC)	P(Valid Sequence)
14482	2	2	2	
	(0.96)	(0.47)	(0.47)	0.63
14598	2	1	2	
	(0.90)	(0.52)	(0.29)	0.74
14617	4	128	128	
	(0.00)	(0.58)	(0.58)	0.50

Table 2: MIC prediction results for the three strains of interest. P(REF ±1) represents the probability that the mass assigned to dilutions within 1 dilution of the reference MIC. P(Modal MIC) and P(DT MIC) are the probabilities that the modal and the decision theoretic MICs estimates are the reference MIC, respectively. P(Valid Sequence) is the sum of probabilities of all possible valid sequences.

a peak at two, which corresponds to the modal MIC, the decision theoretic MIC and the reference MIC. For strain 14598, we also have a uni-modal distribution, but underestimate

the MIC by one dilution using the modal MIC (reference MIC = 2, modal MIC = 1), while the decision theoretic MIC agrees with the reference MIC. Because the estimate is within one dilution of the reference MIC, this error is not important. Finally for strain 14617, the modal MIC is 128, while the Reference MIC is 4. While overestimation of the MIC is better than underestimation, this results in an unacceptably large error. The reference MIC and dilutions within one of it receive virtually no support under the estimated distribution of the MIC. Also, the probability of all valid sequences is 0.5 indicating that curves are less likely to be ordered by dilution than with the other strains, where the probability is 0.63 and 0.74. Recall, however, that in this case there are several curves with dilutions above the reference MIC that exhibited strong growth. Verification of the reference data indicated that the response variable is in fact not in error. One possible explanation for the growth is that the amount of antimicrobial agent in the panel may be less than its nominal level. In the case of this type of outlier, we feel that the data for the panel may be deleted and the model refit using the remaining data.

5 Validation

To carry out model validation for each drug, we divide the data into a training and validation group. We used a random sample of 65% of the panel data for training and the remaining 35% for validation. In the training stage we identify the best model using BIC as described previously. This model is then used to predict the MICs for the validation group.

The MIC prediction for the validation group proceeds as follows.

1. The features from each dilution j of each panel i in the validation group are used along with the estimated coefficients from the training data to predict the probability of growth in each well, $\hat{\pi}_{ij}$.
2. The probability that a dilution is the MIC, $\hat{\rho}_{ij}$ is calculated (see equation (2)). These

ESTIMATE	EST - REF < -1	-1 ≤ EST - REF ≤ 1	EST - REF > 1
Modal MIC	3.99	91.84	4.17
Decision Theoretic MIC	1.57	90.91	7.52

Table 3: Essential agreements from the FOX validation data.

ESTIMATE	EST - REF < -1	-1 ≤ EST - REF ≤ 1	EST - REF > 1
Modal MIC	24.46	71.07	4.46
Decision Theoretic MIC	11.61	80.36	8.04

Table 4: Essential agreements from the PIP validation data.

probabilities are estimated by plugging in the estimates of π_{ij} obtained in Step (1) for the panels in the validation set.

3. The estimated MIC probabilities ($\hat{\rho}_{ij}$) for each dilution in each panel are then used to estimate the modal or decision theoretic MIC.

For each panel in the validation set we thus obtain a modal and decision theoretic MIC estimator.

5.1 Essential Agreements

The essential agreements are summarized graphically in Figures 3 and 4 and in Tables 3 and 4 for the FOX and PIP data. In Figure 3 note that the choice of MIC estimator markedly effects the error type rates. For the modal MIC the error rate associated with under estimation by more than one dilution was 3.99% whereas that for the decision theoretic MIC, the error rate was 1.57%. Essential agreements for the two MIC estimates were similar, 91.84% for the modal MIC and 90.91% for the decision theoretic MIC. The error rates for PIP are noticeably higher as anticipated based on initial perceptions of PIP being a more challenging system. The decision theoretic MIC has a much higher essential agreement than the modal MIC, with a reduction by half in the underestimation error rate (11.6% versus 24.5%).

To further explore the errors, a plot of $\log_2$(estimated MIC) - $\log_2$(reference MIC) versus the $P(\text{MIC} = D_j)$ is given in Figures 3 and 4, where the estimated MIC is either the modal MIC or decision theoretic MIC. Note that for the decision theoretic MIC in the FOX data, a probability of less than about 0.4 produces a wider scatter in $\log_2$(estimated MIC) - $\log_2$(reference MIC), corresponding to more cases of poor agreement between the estimated and reference MIC. It is possible that by delaying the call because of the substantial uncertainty in the MIC and allowing the samples to incubate longer, that we may reduce such errors.

6 Discussion

In this paper we have illustrated a probabilistic approach for estimating MICs based on panels of microbial growth curves. Given the necessity to fit a large number of models to hundreds of antibiotics, we had to make several compromises in our modelling approach due to the computational environments available at the time and the current computing/storage constraints of the PHOENIX device. Given advances in computing environments today, more flexible models could be obtained by replacing the cubic polynomials with piecewise cubic splines as in generalized additive models (GAM). While a fully Bayesian analysis might be preferable, the use of approximate model probabilities using BIC and plug-in estimates provided a solution that could be implemented in real-time in a device with limited computing and storage capacity.

Despite the potential shortcomings given above, various studies have tested the accuracy of the PHOENIX AST system and have shown that this system provides accurate estimates of MICs. Currently, 85 drug indications are cleared by the Food and Drug Administration, with an additional 20 to 25 in clinical trials. In a European collaborative two-center trial (Fahr et al. 2003), this system was tested for 469 clinically obtained bacterial isolates with

64 antimicrobial drugs. The results were compared to those of frozen standard broth microdilution (SBM) panels according to National Committee for Clinical Laboratory Standards' guidelines. It was found out that the PHOENIX AST performance was equivalent to that of the SBM reference method. In addition to accuracy in prediction Donay et al. (2004) found that the PHOENIX AST system required significantly less time to obtain results than by the disk diffusion method.

Predicting the emergence of antibiotic resistance is a challenging problem that many automated AST systems fail to adequately address (Tenover et al. 2006). In some cases resistance to higher levels of an antibiotic is virtually unknown. Because the models developed in this paper predict inhibition of growth based on features of the curves, and not dilution, the PHOENIX AST system may predict high MICs indicative of emerging resistance, even though the models have not necessarily been trained on resistant strains. The PHOENIX system has already proved to be effective in detecting the emergence of *Staphylococcal* resistance to vancomycin (Deal et al. 2002).

References

Akaike, H. (1973), "Information theory and an extension of the maximum likelihood principle," in *Second International Symposium on Information Theory*, eds. B. Petrox and F. Caski, Akademia Kiado, Budapest, pp. 267–281.

Akaike, H. (1983a), "Information measures and model selection (STMA V25 937)," *Bulletin of the International Statistical Institute*, 50, 277–290.

Akaike, H. (1983b), "On minimum information prior distributions," *Annals of the Institute of Statistical Mathematics*, 35, 139–149.

Barenfanger, J., Drake, C., and Kacich, G. (1999), "Clinical and Financial Benefits of Rapid Bacterial Identification and Antimicrobial Susceptibility Testing," *Journal of Clinical Microbiology*, 37(5), 1415–1418.

Clyde, M. and George, E. I. (2004), "Model Uncertainty," *Statistical Science*, 19, 81–94.

Deal, M., Votta, M., Turng, S. H. B., Wiles, T., and Reuben, J. (2002), "Detection of Glycopeptide Intermediate or Resistant *Staphylococcus aureus* Strains Using BD Phoenix^{TM}

Automated Microbiology System," in *101st General Meeting of the American Society for Microbiology*, Salt Lake City, Utah, Poster C–119.

Donay, J.-L., Mathieu, D., Fernandes, P., Prgermain, C., Bruel, P., Wargnier, A., Casin, I., Weill, F. X., Lagrange, P. H., and Herrmann, J. L. (2004), "Evaluation of the Automated Phoenix System for Potential Routine Use in the Clinical Microbiology Laboratory," *Journal of Clinical Microbiology*, 42(4), 1542–1546.

Fahr, A.-M., Eigner, U., Armbrust, M., Caganic, A., Dettori, G., Chezzi, C., Bertoncini, L., Benecchi, M., and Menozzi, M. G. (2003), "Two-Center Collaborative Evaluation of the Performance of the BD Phoenix Automated Microbiology System for Identification and Antimicrobial Susceptibility Testing of Enterococcus spp. and Staphylococcus spp." *Journal of Clinical Microbiology*, 41(3), 1135–1142.

FDA (2007), *Class II Special Controls Guidance Document: Antimicrobial Susceptibility Test (AST) System; Guidance for Industry and FDA*, Center for Devices and Radiological Health, Food and Drug Administration, U.S. Department of Health and Human Services.

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999), "Bayesian model averaging: a tutorial (with discussion)," *Statistical Science*, 14, 382–401, corrected version at http://www.stat.washington.edu/www/research/online/hoeting1999.pdf.

Horstkotte, M. A., Knobloch, J. K.-M., Rohde, H., Dobinsky, S., and Mack1, D. (2004), "Evaluation of the BD PHOENIX Automated Microbiology System for Detection of Methicillin Resistance in Coagulase-Negative Staphylococci," *Journal of Clinical Microbiology*, 42(11), 5041–5046.

Kass, R. E. and Raftery, A. E. (1995), "Bayes factors," *J. Amer. Statist. Assoc.*, 90, 773–795.

Schwarz, G. (1978), "Estimating the Dimension of a Model," *Annals of Statistics*, 6, 461–464.

Tenover, F. C., Kalsi, R. K., Williams, P. P., Carey, R. B., Stocker, S., Lonsway, D., Rasheed, J. K., Biddle, J. W., Jr., J. E. M., , and Hanna, B. (2006), "Carbapenem Resistance in *Klebsiella pneumoniae* Not Detected by Automated Susceptibility Testing," *Emerging Infectious Diseases*, 12, 1209–1213.

Wheat, P. F. (2001), "History and development of antimicrobial susceptibility testing methodology," *Journal of Antimicrobial Chemotherapy*, 48 Suppl. S1, 1–4.

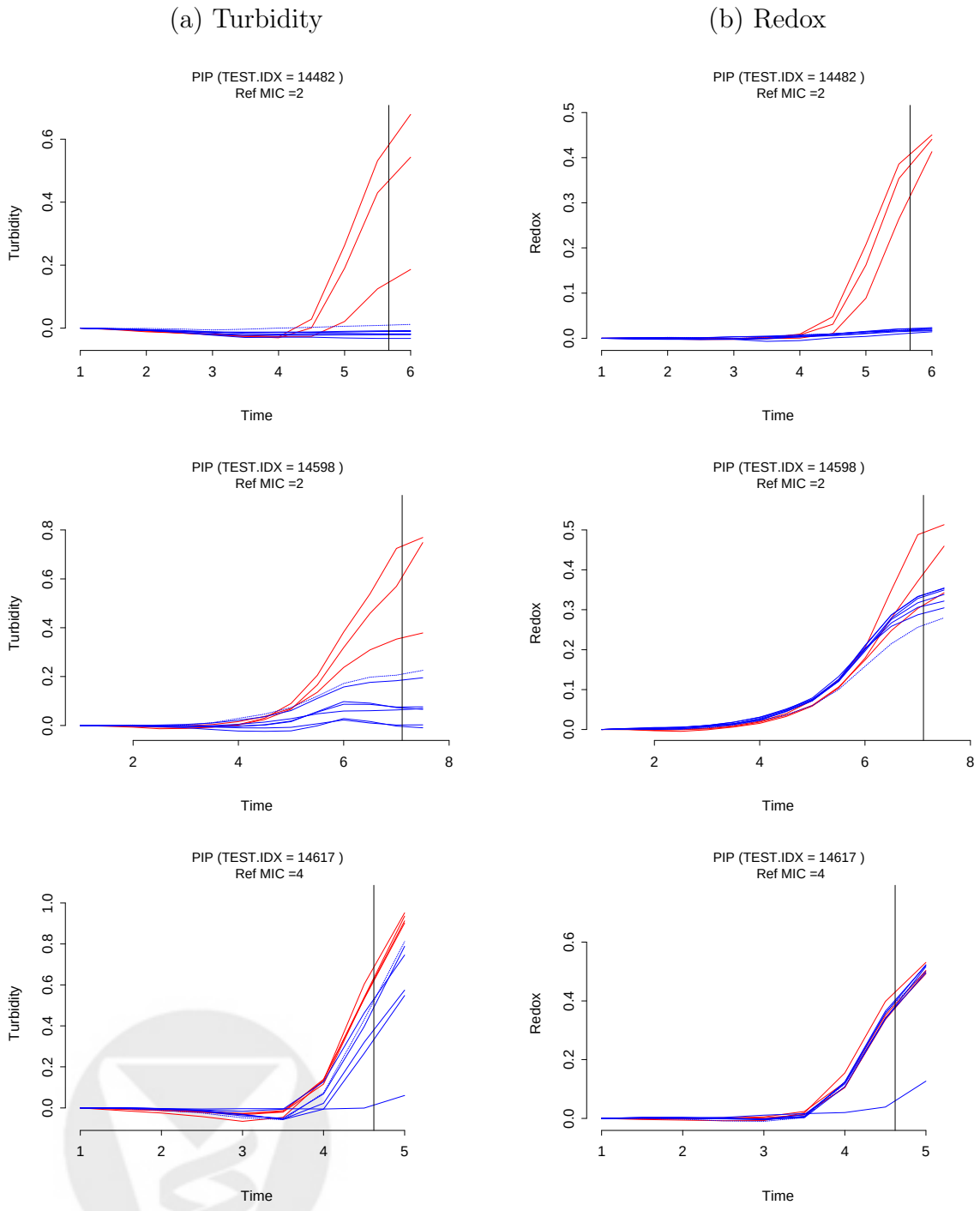

Figure 1: Panel data for three gram-negative bacterial strains exposed to dilutions of piperacillin showing turbidity (a) and redox (b) over time in hours. The vertical line is at the time-to-result when features are extracted for estimating the MIC. All curves in red have concentrations less than the reference MIC, while those in blue have concentrations greater than or equal to the reference MIC.

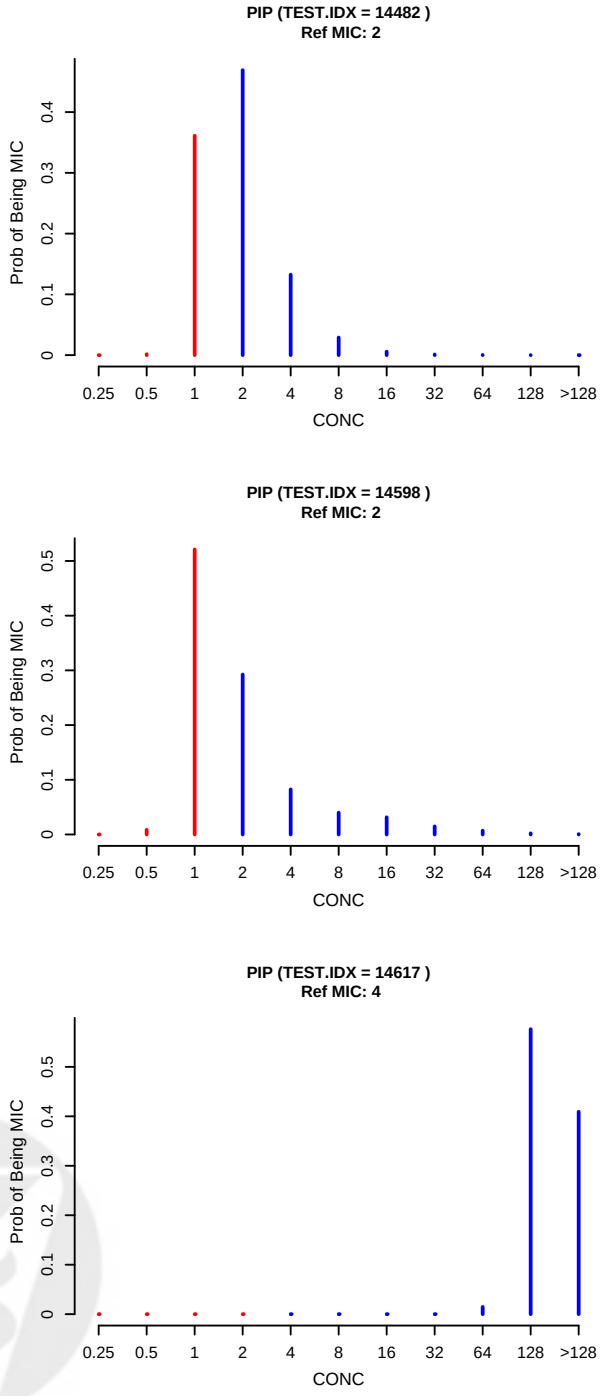

Figure 2: Probability distribution of the MIC for the three strains presented in Figure 1. The bars in red represent dilutions below the reference MIC, while the bars in blue correspond to dilutions greater than or equal to the reference MIC.

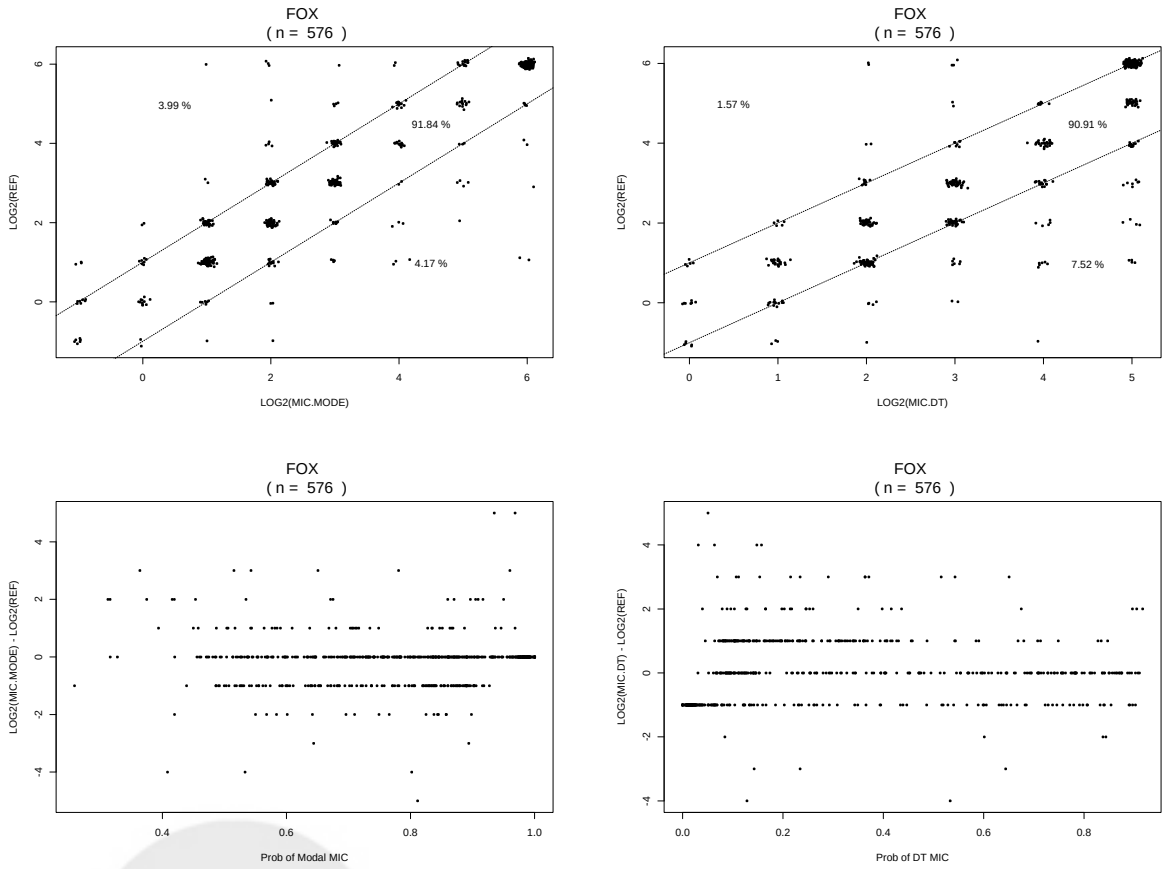

Figure 3: Essential agreement (top) and residual plots (bottom) for FOX validation set using the Modal MIC (left) and the Decision Theoretic MIC (right).

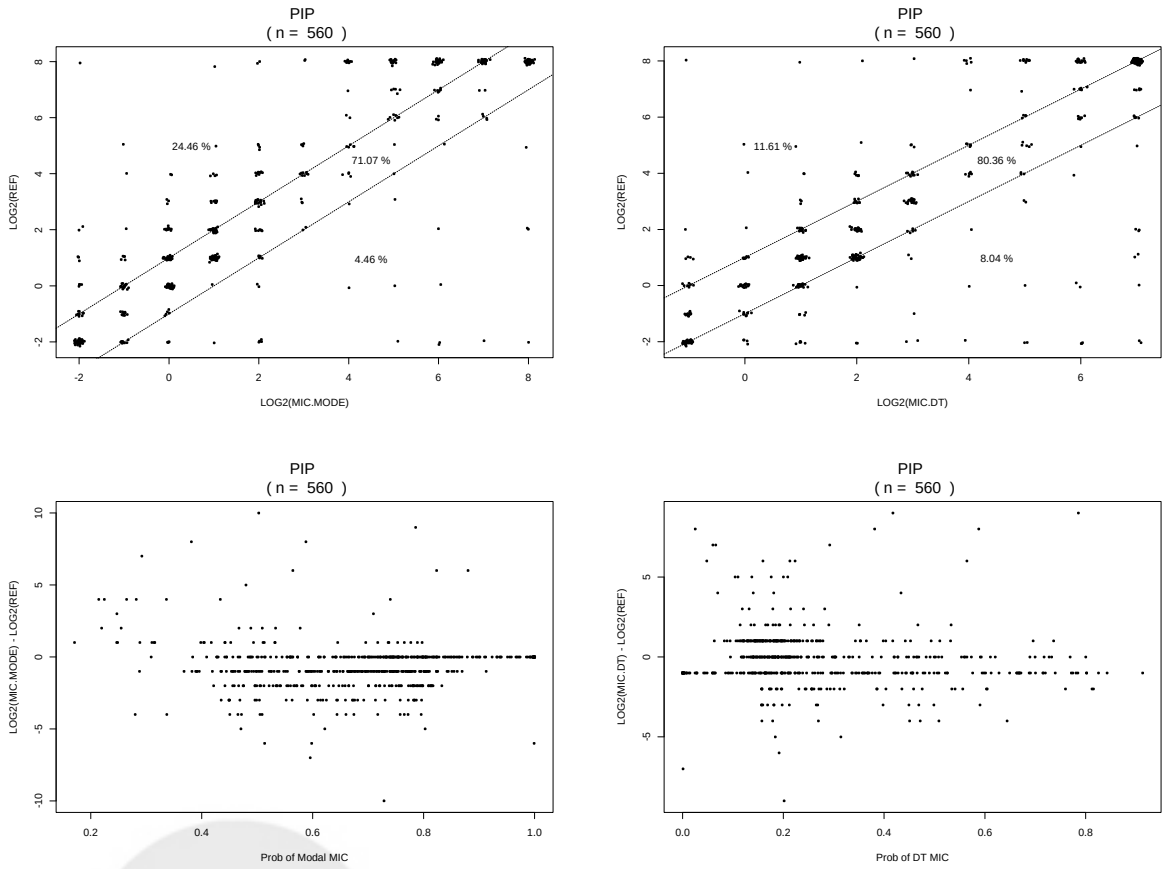

Figure 4: Essential agreement (top) and residual plots (bottom) for PIP validation set using the Modal MIC (left) and the Decision Theoretic MIC (right).