Extending the Modified Bayesian Information Criterion (mBIC) to Dense Markers and Multiple Internal Mapping

by

M. Bogdan, P. Biecek
Wroclaw University of Technology

F. Frommlet
Vienna University

R. Cheng, J.K. Ghosh and R.W. Doerge
Purdue University

Technical Report #07-05

Department of Statistics
Purdue University

2007

Extending the Modified Bayesian Information Criterion (mBIC) to dense markers and multiple interval mapping

M. Bogdan,[1,*] F. Frommlet,[2] P. Biecek,[1] R. Cheng,[3] J. K. Ghosh[3] and R. W. Doerge[3]

[1]Institute of Mathematics and Computer Science, Wrocław University of Technology, Wrocław, Poland

[2] Department of Statistics, Vienna University, Vienna, Austria

[3]Department of Statistics, Purdue University, West Lafayette, IN 47907, USA

April 12, 2007

SUMMARY. The modified version of Bayesian Information Criterion (mBIC) is a relatively simple model selection procedure that can be used when locating multiple interacting quantitative trait loci (QTL). The statistical properties of mBIC have been demonstrated for situations where the average genetic map interval is at least 5 cM. In this work mBIC is adapted to genome searches based on a dense map and, more importantly, to the situtation where consecutive QTL and interactions are located by multiple interval mapping. Easy to use formulas for the extended BIC are given. A simulation study, as well as the analysis of real data, confirm the good properties of the extended mBIC.

* *email:* Malgorzata.Bogdan@pwr.wroc.pl

KEY WORDS: Model selection criterion, Multiple QTL, QTL interaction

1. Introduction

Many quantitative traits in plants, animals and humans are, to a certain extent, determined genetically. Regions of the genome that influence such traits are called quantitative trait loci (QTL), and typically molecular markers are employed to detect and locate QTL using statistical models. These molecular markers are polymorphic (exhibiting variation) at identifiable locations on chromosomes, and their genotypes can be identified experimentally. From a statistical point of view, marker genotypes are qualitative explanatory variables and the task of locating QTL relies on the associations between marker genotypes and the trait values.

The earliest methods for QTL mapping date back to Sax (1923), Thoday (1961) and Soller et al. (1976). These early approaches used standard statistical procedures such as t-tests or analysis of variance and were limited by the availability of markers. To address the accuracy of QTL location Lander and Botstein (1989) proposed interval mapping (IM), based on the EM algorithm (Dempster et al. (1977)), which searches for QTLs within known intervals defined by genetic markers. The main disadvantage of interval mapping is that it is based on a single QTL statistical model that leads to biased estimators of both QTL effect size and QTL location when the trait is influenced by more than one QTL. To account for multiple QTLs Zeng (1994) and Jansen and Stam (1994) proposed Composite Interval Mapping (CIM) and Multiple QTL Mapping (MQM), respectively, by including additional background markers into the model. Although CIM and MQM both have potential to increase the accuracy of detecting QTLs with additive effects,

they remain unable to detect epistatic effects (i.e., interactions). Epistasis is a common phenomenon that plays an important role in the genetic determination of complex traits (see e.g., Doerge (2002), Carlborg and Haley (2004) and references given there), as well as in evolution (see e.g., Wolf et al. (2000)). Neglecting these epistatic effects may lead to oversimplified models for inheritance of complex traits and, as noted by Carlborg and Haley (2004), may result in a relatively low economic gain if such models are used for marker-assisted selection.

The growing awareness of the importance of epistasis has lead to the development of new statistical methods for QTL mapping (see Kao et al. (1999), Carlborg et al. (2000), Bogdan et al. (2004), or Yi et al. (2005)). Although simple approach that acknowledges epistasis is multiple regression or ANOVA models with interactions, the most difficult part in fitting such a model lies in the identification of the nonzero coefficients of the model. The standard approach is to perform sequential statistical tests (Doerge and Churchill (1996), Kao et al. (1999), Carlborg and Andersson (2002)), but is limited to the comparison of nested models. An alternative is to employ a model selection criteria (e.g., Bayesian Information Criterion (BIC, Schwarz (1978)) or the Akaike Information Criterion (AIC, Akaike (1974)) when formulating the QTL model.

AIC and BIC belong to a class of so called penalized maximum likelihood methods that select the best model by maximizing the likelihood of the data minus a penalty for the model dimension. By comparison to other penalized likelihood methods BIC has a relatively large penalty and is generally considered to be one of the most conservative model selection criteria. How-

ever, in an exploration of BIC as applied to QTL mapping, Broman (1997) and Broman and Speed (2002) reported that it overestimates the number of QTLs, and proposed to increase the penalty for the model dimension. Simulations reported in Broman and Speed (2002) show that the resulting modified version of BIC behaves well and in some cases detects the correct model more often than composite interval mapping. Later, Bogdan et al. (2004) explained the same overestimation phenomenon by observing that, due to the large number of markers used in typical genome scans, the number of possible high dimensional models is disproportionately large compared to the number of low dimensional models. Therefore high dimensional models have a larger probability to be chosen by chance. To address this problem Bogdan et al. (2004) proposed a modification of BIC (mBIC), which takes into account the number of markers used in the analysis while allowing prior knowledge on the QTL number to be incorporated. When prior knowledge is lacking they proposed a standard version of mBIC which adjusts for multiple testing and controls the type I error. In Baierl et al. (2006) mBIC is further extended by a two-step procedure. In the second step the prior is adjusted according to the results of the initial step. Simulations reported in Bogdan et al. (2004) and Baierl et al. (2006) demonstrate that both mBIC, and the two-step version retain good power when distances between genetic markers are larger than 5cM. However, when distances between markers are smaller than 5cM the penalty for mBIC becomes too large, and results in an unnecessary decrease of statistical power. Motivated by the fact that current QTL mapping populations are relatively large, and dense genetic maps are common, mBIC is extended to situations where map distances are smaller

than 5 cM. This extension also sets the stage for applying mBIC to interval mapping and association mapping (discussed elsewhere).

The article is supplied with the web appendix, available at `http://www.im.pwr.wroc.pl/ mbogdan/mBIC/appendix.pdf` . The appendix contains some additional, theoretical material, as well as the example of the real data analysis, demonstrating the properties of our method.

2. Methods

Consider a backcross population, the QTL genotypes at each location are denoted by Q_{ij}: $Q_{ij} = -\frac{1}{2}$ if the ith individual is homozygous at the jth QTL, and $Q_{ij} = \frac{1}{2}$ if it is heterozygous. Assume the relationship between quantitative trait values and multiple QTL genotypes is given by a normal regression model

$$Y_i = \mu + \sum_{j=1}^{m} \beta_j Q_{ij} + \sum_{1 \le j \le l \le m} \gamma_{jl} Q_{ij} Q_{il} + \varepsilon_i, \tag{1}$$

where m is the number of QTL and $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ is environmental noise. The second summation corresponds to pairwise epistatic interactions. Since the coefficients β_j and γ_{ij} can both equal zero, denote the number of QTL with nonzero main effects as p, and the number of nonzero epistatic terms as q. Because locating so many interacting QTLs by multiple interval mapping poses a complex multidimensional computational problem, as a first step QTL location is restricted to marker positions. Specifically, the problem of detection and location of QTL is based on choosing the best model of the form

$$Y_i = \mu + \sum_{j \in I} \beta_j X_{ij} + \sum_{(u,v) \in U} \gamma_{uv} X_{iu} X_{iv} + \varepsilon_i, \tag{2}$$

where X_{ij} is the genotype of the ith individual at the jth marker, and I and U are sets of markers with significant main and epistatic effects, respectively. When the mBIC criterion (Bogdan et al. (2004)) is employed to select the best marker model, it chooses the model which minimizes

$$mBIC = n \log RSS + (p+q) \log n + 2p \log(l-1) + 2q \log(u-1) \, ,$$

where p is the number of main effects in the model, q is the number of epistatic effects, and RSS is the corresponding residual sum of squares. The penalty coefficients l and u depend on the prior distribution of the number of QTL effects and the number of epistatic effects. When the expected values of the number of main effects and epistatic effects are equal to c_1 and c_2 respectively, then the penalty coefficients can be calculated as $l = N_m/c_1$ and $u = N_e/c_2$. Here, $N_e = N_m(N_m - 1)/2$ is the number of possible pairwise interactions between N_m markers. For situations where the prior knowledge on the number of QTL is not available Bogdan et al. (2004) proposed the use of $c_1 = 2.2$ and $c_2 = 2.2$ which leads to a standard form of mBIC

$$mBIC = n \log RSS + (p+q) \log n + 2p \log(N_m/2.2 - 1) + 2q \log(N_e/2.2 - 1) \, . \tag{3}$$

2.0.1 Calibrating mBIC for dense markers. Committing a weak sense familywise Type I error occurs when at least one QTL is detected when there is none. In this context, mBIC might either wrongly detect a QTL with one of the simple regression models, including only one marker, or detect false QTL based on a multiple regression model when all simple regression models do not detect QTL. The extensive simulation studies reported in Bogdan

et al. (2004) and Baierl et al. (2006) demonstrate that the probability of the second event is extremely small. Thus, the familywise error rate (FWER) of mBIC is mainly determined by the results of the initial search over single markers, a fact that will be used when calibrating mBIC for densely spaced markers.

As mentioned previously increasing the marker density increases the penalty in the mBIC, and results in a significant loss of statistical power. Specifically for densely spaced markers (less than 5cM) the correlation between neighboring marker genotypes becomes very strong. Adding a new marker to an already dense genetic map does not provide much new information and is unlikely to increase the maximum of the likelihood ratio statistic over the entire genome, which suggests that the distance between markers when computing mBIC penalty has to be taken into account.

The challenge of computing the penalty for highly correlated markers is related to the issue of choosing a threshold for the test of QTL presence, when the QTL search is performed over many markers simultaneously. As is well known, any multiple testing problem can be dealt with by using a Bonferroni correction. While the Bonferroni correction provides a good control of FWER when individual tests are independent, it can be strongly conservative (i.e., give smaller FWER , which results in decreasing of power) when test statistics are correlated. Cheverud (2001) proposed a modification of the Bonferroni procedure. Instead of dividing the overall significance level by the number of markers N_m, it is divided it by an effective number of tests, N_m^{eff}, which should provide an accurate control of the probability of the overall type I error. N_m^{eff} is estimated using the eigenvalues of the empirical matrix of

correlations between marker genotypes.

Cheverud's method of estimating N_m^{eff} turned out to be inaccurate in many cases (see e.g. Salyakina et al. (2005)). In this article we propose an alternative method for computing N_m^{eff} by comparing the threshold value for the maximum of the likelihood ratio tests at N_m markers with the corresponding threshold for N_m^{eff} independent tests. The issue of computing the genome-wise threshold value for the single marker and interval QTL mapping has been intensively studied. In particular Lander and Botstein (1989), Feingold et al. (1993), Dupuis and Siegmund (1999) and Rebaï et al. (1994) addressed the problem by approximating the distribution of the likelihood ratios at neighboring marker locations by the square of the Gaussian process. The most common alternative approach for obtaining QTL thresholds is based on permutation tests (Churchill and Doerge (1994), Doerge and Churchill (1996)). In this article we use computer simulations to estimate the chromosome-wise thresholds for detecting main and epistatic effects and, in case of main effects, we compare our results with asymptotic theoretical values provided by Dupuis and Siegmund (1999) and Rebaï et al. (1994).

2.1 *Searching over markers.*

The standard likelihood ratio (LRT) statistic for testing of association between the trait data Y and the genotype of jth marker is given by

$$LRT(j) = 2\log\frac{L(Y|M_j)}{L(Y|M_0)} \ ,$$

where $L(Y|M_j)$ is the maximum likelihood of the data for the model $Y_i = \mu+\beta X_{ij}+\varepsilon_i$, with X_{ij} being the genotype of i^{th} individual at j^{th} marker, and $L(Y|M_0)$ is the corresponding likelihood for the null model (when $\beta = 0$).

When the null hypothesis is true $LRT(j)$ has approximately χ^2_1 distribution. To detect a QTL we perform a sequence of such tests at every marker.

Dupuis and Siegmund (1999) showed when a genome scan based on one chromosome with N markers spaced every δ cM is performed, the weak sense FWER can be approximated by

$$\begin{aligned} \alpha &= P_{H_0}\left(\max_{i\in\{1,\ldots,N\}} LRT(i) > c\right) \\ &\approx 1 - \exp\left(-2\left[1-\Phi(\sqrt{c})\right] - 0.04L\sqrt{c}\phi(\sqrt{c})\nu\left(\sqrt{0.04c\delta}\right)\right) , \quad (4) \end{aligned}$$

where c is the threshold for the likelihood ratio test, $L = (N-1)\delta$ is the length of the chromosome in cM, ϕ is the density and Φ is the distribution function of the standard normal distribution, and $\nu(t)$ is

$$\nu(t) = 2t^{-2}\exp\left[-2\sum_{n=1}^{\infty} n^{-1}\Phi(-\frac{1}{2}|t|n^{1/2})\right]. \quad (5)$$

Alternatively, the weak sense FWER resulting from performing N_m^{eff} tests at unlinked markers is

$$\begin{aligned} \alpha &= P_{H_0}\left(\max_{i\in\{1,\ldots,N_m^{eff}\}} LRT(i) > c\right) = 1 - \prod_{i=1}^{N_m^{eff}} (1 - Pr(LRT(i) > c)) \\ &\approx 1 - \left(1 - 2\left(1 - \Phi\left(\sqrt{c}\right)\right)\right)^{N_m^{eff}} . \quad (6) \end{aligned}$$

Comparing equations (4) and (6) the effective number of independent tests corresponding to N markers spaced every δcM is computed as

$$N_m^{eff} = \frac{\log(1-\alpha)}{\log(2\Phi(\sqrt{c}) - 1)} , \quad (7)$$

where c depends on α, δ and N according to equation (4). It turns out that the dependence of N_m^{eff} on α is comparably small, therefore only the case of $\alpha = 0.05$ is considered. Smaller levels of significance give rise to slightly larger effective numbers making the procedure slightly more conservative.

[Figure 1 about here.]

For the purpose of scaling the corresponding penalty in mBIC when searching over densely spaced markers we define a weight $w_{SM}^{add}(\delta, N)$, which is assigned to additive effects if the average distance between markers is equal to δcM. If we consider a single chromosome,

$$w_{SM}^{add}(\delta, N) = \frac{N_m^{eff}}{N} . \tag{8}$$

This weight (Equation 8) clearly depends on δ, but also on the number of markers on the chromosome. It is evident that for any fixed δ, $w_{SM}^{add}(\delta, 1) = 1$. Moreover in web appendix it is shown that $\lim_{N\to\infty} w_{SM}^{add}(\delta, N) = 1$. This result illustrates that the dependence between neighboring markers is of a "short range" and has a negligible influence on the behavior of the maximum of the likelihood ratio statistic when the length of the chromosome converges to infinity. However, Figure 1a demonstrates that the rate of increase of w_{SM}^{add} is very slow, and the weights remain relatively stable over a wide range of N.

Equation (4) from Dupuis and Siegmund (1999) is based on the assumption that the likelihood ratios at neighboring locations can be asymptotically described by the square of a Gaussian process. This fact was utilized when designing a simulation study to estimate the weights for the additive and epistatic effects. In the first step the correlation matrix S for genotypes of markers uniformly spaced on a chromosome was calculated, and 10,000 instances were generated from a multivariate normal distribution with mean 0 and the covariance matrix S. For each of these 10,000 random vectors the maximum over the squares of its coordinates was calculated and 0.95 quantile of the empirical distribution of these maxima was used as the estimate of the

chromosome-wise threshold value c for the likelihood ratio statistic to detect main effects. Finally weights w_{SM}^{add} were computed by applying equations (7) and (8). As seen in Figure 1b, the simulated values tend to be consistently slightly smaller than the theoretical counterparts, but the difference is so small that the effect on mBIC is negligible.

It is assumed that markers on different chromosomes are inherited independently. Therefore, when considering the whole genome N_m^{eff} is calculated separately for each chromosome. The effective numbers are then added and divided by the overall number of markers to achieve the appropriate weights. For chromosomes of equal length this procedure is the same as choosing the weight for a single chromosome. As a general rule, since the difference in weights is small in the range between 100 and 300 cM, we suggest using the weights computed for 150 cM for a standard adjustment of mBIC for correlated markers. When the distance between markers is larger than 30cM then w_{SM}^{add} is larger than 0.90 and the modification of the penalty in mBIC is not necessary. Otherwise, w_{SM}^{add} can be very well approximated by

$$\hat{w}_{SM}^{add}(\delta) = 1 - 0.9e^{(-10\delta/100+10(\delta/100)^2)} \quad . \tag{9}$$

The particular shape of $\hat{w}_{SM}^{add}$ was chosen in accordance with the observation that w_{SM}^{add} tends towards 1 for large δ. Figure 2a shows that $\hat{w}_{SM}^{add}$ serves well both for chromosomes of length 150cM as well as 300cM.

[Figure 2 about here.]

While dealing with individual markers is relatively straightforward, the structure of dependence between epistatic terms is much more complicated. In fact, when considering interactions there is no theoretical result that is

equivalent to equation (4). However, our simulation approach led to reasonably good approximations of the weights for additive effects, therefore the same strategy is applied to interactions. Thus, in the first step, the correlation matrix $\mathrm{Cov}(I_{i,j}, I_{k,l})$ between all pairwise interaction terms $I_{i,j} = X_i X_j$ is calculated and 10,000 instances from a related multivariate normal distribution are generated (the formulas for the covariances between different interaction terms are given in web appendix). For each of these 10,000 random vectors the maximum over the squares of its coordinates is computed and 0.95 quantile of the empirical distribution of these maxima is used as the estimate of the chromosome-wise threshold value c for the likelihood ratio statistic to detect epistatic effects. Finally, the equation (7) is employed to compute the effective number of interactions, N_e^{eff}, and the weight for epistatic effects is computed according to the formula $w_{SM}^{epi} = \frac{N_e^{eff}}{N_e}$, where $N_e = \binom{N}{2}$ is the total number of interactions. The dependence of these simulated weights on the number of markers and on the distance δ is illustrated in Figure 1c. Note that $w_{SM}^{epi}(\delta, N)$ appears to have the same qualitative behavior as $w_{SM}^{add}(\delta, N)$ in the sense that for $N = 2$, $w_{SM}^{epi}(\delta, 2) = 1$, and the simulation results suggest that $\lim_{N\to\infty} w_{SM}^{epi}(\delta, N) = 1$. Figure 2b illustrates that the weights of chromosomes of length 150cM and 300cM for different δ can be well approximated by

$$\hat{w}_{SM}^{epi}(\delta) = e^{(-10.7\delta/100 + 8.7(\delta/100)^2)} \quad . \tag{10}$$

For situations where the markers are not equally spaced the average intermarker distance $\bar{\delta}$ can be used with weights $w_{SM}^{add} = \hat{w}_{SM}^{add}(\bar{\delta})$ and $w_{SM}^{epi} = \hat{w}_{SM}^{epi}(\bar{\delta})$ according to equations (9) and (10). Based on these results the standard version of mBIC for dense and unequally spaced markers recommends

choosing the model for which

$$mBIC_{SM} = n \log RSS + (p+q) \log n + 2p \log \left(\frac{w_{SM}^{add} N_m}{2.2} \right) + 2q \log \left(\frac{w_{SM}^{epi} N_e}{2.2} \right), \tag{11}$$

obtains a minimal value.

Multiple Interval Mapping. Interval mapping (IM; Lander and Botstein (1989)), while originally based on a single QTL model, was extended to multiple interval mapping (MIM; Kao et al. (1999)) for the purpose of locating multiple interacting QTL. In the present paper we apply a simplified version of multiple interval mapping based on the approach proposed by Haley and Knott (1992). The method relies on replacing missing genotypes of putative QTL by their expected values conditioned on the genotypes of neighboring markers. The procedure is very simple and quick and is particularly advantageous when there are multiple interacting QTL and many competing models need to be searched to estimate QTL number and their location. A comparison between the EM algorithm, applied in Kao et al. (1999), and Haley and Knott regression did not provide significant differences in the performance of mBIC in an interval mapping setting.

MIM based on Haley and Knott method relies on fitting the multiple regression model

$$Y_i = \mu + \sum_{j \in I} \beta_j G_{ij} + \sum_{(u,v) \in U} \gamma_{uv} G_{iu} G_{iv} + \varepsilon_i, \tag{12}$$

on a dense net of possible QTL locations. Here G_{ij} is the expected value of the genotype of ith individual at jth position on the genome (the formulas for G_{ij} are provided e.g. in Kao (2000)), and I and U are sets of

locations corresponding to QTL with significant main effects and epistatic effects, respectively. To estimate the number of QTL and their locations the appropriate version of mBIC is used.

The main difference between multiple interval mapping (MIM) and searching over individual markers is that a much larger set of possible QTL locations is investigated with MIM. To accommodate the increased number of investigations the penalty of mBIC needs to be increased. Because the predictor variables corresponding to the given inter-marker locations are completely specified by genotypes of neighboring markers the correlations between likelihood ratio statistics at neighboring locations are stronger than when searching over a dense map of markers. To adapt mBIC for MIM the effective number of tests N_{IM}^{eff} corresponding to the genome search based on one dimensional interval mapping is calculated. As mentioned previously one can make use of the threshold value of the likelihood ratio statistic for a one dimensional interval mapping, because under the null hypothesis of no QTL, the probability is very small for obtaining at least one false positive QTL in MIM, where there are no significant effects found in one dimensional IM. To find this threshold one option is to use the theoretical results from Rebaï et al. (1994), which state that the significance level α of the genome search based on interval mapping can be approximated by

$$\begin{aligned} \alpha &= Pr\left(\sup_{0\leq x\leq\sum_{i=1}^{k}\delta_i} LTR(x) > c^2\right) \\ &\approx 2\Phi(-c) + \frac{2}{\pi}\exp\left(-\frac{1}{2}c^2\right)\sum_{i=1}^{k}\arctan\left(\sqrt{\frac{r_i}{1-r_i}}\right) \end{aligned} \tag{13}$$

where c^2 is the threshold, k is the number of intervals, δ_i is the length of the ith interval and r_i is the probability of recombination for the ith interval. In

this case markers can be either equally or unequally spaced.

To find the related number of independent tests N_{IM}^{eff} equation (7) is used, with dependence between α and c provided by equation (13). When searching intervals (IM) the additive effect weight is $w_{IM}^{add} = \frac{N_{IM}^{eff}}{N}$. As might be expected, the behavior of $w_{IM}^{add}(\delta, N)$ differs from that of $w_{SM}^{add}(\delta, N)$. Specifically, in web appendix we prove that $\lim_{N\to\infty} w_{IM}^{add}(\delta, N) = \infty$, or more precisely for every fixed δ $w_{IM}^{add}(\delta, N) = O(\sqrt{\log(N)})$ when $N \to \infty$. Therefore, w_{IM}^{add} increases slowly with N, and in particular it is not expected to be bounded by 1 (Figure 1d). This result is very appealing since it demonstrates that strong correlation between likelihood ratio statistics at neighboring locations has a negligible influence on the distribution of the maximum of the likelihood ratio statistic when the length of the genome tends to infinity. Note however that this result no longer holds in the situation when the interval mapping is performed by maximizing the likelihood function at a finite grid of locations spaced every 1 or 2 cM. In this case the effective number of tests is always bounded by the number of positions at which tests are performed. To deal with this more realistic situation, simulations are used to estimate the corresponding weights. Imputations that are equally spaced within each interval at distances equal to 2cM are considered. When calculating covariance matrices for genotypes of imputed positions the Haley and Knott imputations are approximated by linear combinations of the flanking marker genotypes,

$$\hat{G}_i = \frac{\delta_1 X_{2i} + \delta_2 X_{1i}}{\delta_1 + \delta_2} \quad , i = 1, \ldots, n \quad ,$$

where δ_1 and δ_2 are the genetic distances between the putative QTL and the flanking markers, and X_{1i} and X_{2i} are the corresponding genotypes of flanking markers for ith individual. This approximation, neglecting the pos-

sibility of a double crossover within a given interval, is very accurate when the distance between neighboring markers is small. In cases where the distance between flanking markers does not exceed 30cM the relative error in approximating G_i is never larger than 4.53% with the largest value obtained in the middle of a 30cM interval with both flanking markers having the same genotype (if $X_{1i} = X_{2i} = 0.5$ our approximation leads to $\hat{G}_i = 0.5$ while the value of G_i computed upon the assumption of no interference is equal to $G_i = 0.4783$). In a subsequent simulation study interval mapping based on exact Haley and Knott regression demonstrates that the weights computed according to the presented simulation strategy perform very well. In Figures 1d and 1e simulation results are compared to theoretical values of Rebaï et al. (1994). As expected, the simulated weights are systematically smaller, which is due to the fact that they model the more realistic situation of a discrete set of imputations.

Simulations were also used to estimate the weights corresponding to interaction terms (Figure 1f and 2). For the purpose of reducing the size of the covariance matrix required to simulate weights for interactions corresponding to a chromosome of the length of 300cM, a "loose" grid of imputations, separated by approximately 4cM was used. For a 150cM chromosome both "medium" (approximately 2cM) and "loose grid of imputations" were implemented. The difference in weights resulting from the "medium" and "loose" grid turned out to be negligible, while the difference between 150 and 300 cM chromosomes is substantial only when the distance between markers exceeds 20 cM (Figure 2). For the purpose of aiding the reader in applying the mBIC_{IM} to real data the following empirical approximations for the additive

weights can be used

$$\hat{w}_{IM}^{add}(\delta) = -0.15 + 3.1\sqrt{\delta/100} - 1.3\delta/100 \ . \tag{14}$$

The corresponding interval mapping weights for interactions (for chromosome length of approximately 150cM) can be well approximated by

$$\hat{w}_{IM}^{epi}(\delta) = -0.53 + 5.4\sqrt{\delta/100} - 2.7\delta/100 \ . \tag{15}$$

Approximations (14) and (15) were obtained after examining several classes of simple functions (including logarithm) for approximating the simulated weights. The coefficients were calculated using the method of least squares, and the accuracy of these approximations is demonstrated in Figure 2. In situations where markers are not equally spaced the average distance between markers $\bar{\delta}$ can be computed to yield weights $\hat{w}_{IM}^{add}(\bar{\delta})$ and $\hat{w}_{IM}^{epi}(\bar{\delta})$ (equations (14) and (15) respectively). The adjusted version of mBIC for MIM recommends choosing the model for which

$$mBIC_{IM} = n\log RSS + (p+q)\log n + 2p\log\left(\frac{w_{IM}^{add}N_m}{2.2}\right) + 2q\log\left(\frac{w_{IM}^{epi}N_e}{2.2}\right) \ , \tag{16}$$

obtains a minimal value.

3. Simulations

Both dense markers and multiple interval mapping adjustments to mBIC are studied via simulation. Marker and QTL genotypes for a backcross with 3 unique 100cM chromosomes are considered under two sample sizes: $n = 200$ and $n = 500$. All simulations are based on 1000 replicates. Due to the computational complexity of large scale simulations a simple forward selection with the standard version of mBIC is used.

3.1 *Type I error*

For purpose of examining whether the proposed modifications to mBIC allows the control of weak sense FWER at the desired level quantitative trait data were simulated under a null hypothesis represented by a normal distribution with $\mu = 0$ and $\sigma = 1$. Both a dense marker and multiple interval mapping setting is investigated. Specifically genotypes of markers densely spaced every 2, 5, 10 and 20cM were simulated. Furthermore MIM based on simulated markers spaced every 5, 10, 20 and 25cM was considered. To adjust mBIC weights based on equations (9) and (10) for densely spaced markers and equations (14) and (15) for MIM were implemented. Finally, 30 unequally spaced markers (Figure 3) were simulated such that the average width of the intermarker distance was 11.1 cM and the corresponding weights were equal to $w_{SM}^{add} = w_{SM}^{epi} = 0.66$ for single marker analysis and $w_{IM}^{add} = 0.74$, $w_{IM}^{epi} = 0.97$ for MIM.

[Figure 3 about here.]

Table 1 gives the estimated probabilities of wrongly detecting both the main and the epistatic effects, as well as the total Type I error. These results demonstrate that the weak sense FWER both for single marker analysis and MIM are comparable, and that both are at the assumed level of 8% for $n = 200$ and 5% for $n = 500$. Since mBIC is a consistent model selection procedure FWER will further diminish as the sample size increases. Furthermore even though the penalty for including epistatic effects is much larger than the penalty for main effects, the probability of falsely detecting the epistatic term is comparable to the probability of falsely detecting the main

effect. This is due to an increasingly large number of potential epistatic effects investigated by mBIC_{IM}.

[Table 1 about here.]

3.2 *Power and accuracy of QTL localization*

The statistical power and accuracy of QTL localization based on the adjusted mBIC is considered under three different search strategies: single marker analysis over a relatively sparsely spaced set of markers (scenario 1), multiple interval mapping (scenario 2), and single marker analysis over markers spaced every 2cM (scenario 3).

QTL locations, as well as locations of markers used in scenario 1 are presented in Figure 3. Six QTL (Q1-Q6) were simulated, two on each of three chromosomes. Three of these QTL: Q1, Q4 and Q6 have main effects. The corresponding effect sizes, according to the model (1), are: $\beta_{Q1} = 0.6$, $\beta_{Q4} = 0.7$, $\beta_{Q6} = 0.5$. Additionally, three interaction effects are simulated: interaction 1 involving Q1 and Q6, interaction 2 between Q2 and Q4, and interaction 3 between Q3 and Q5. The corresponding effect sizes are $\gamma_{Q1Q6} = 1.2$, $\gamma_{Q2Q4} = 1.4$ and $\gamma_{Q3Q5} = 1$. The standard deviation of the error term ε is 1. Define a single effect heritability as $h^2 = \sigma^2_{eff}/\sigma^2_Y$, where σ^2_{eff} is the variance of the trait due to a particular effect, and σ^2_Y is the total trait variance. The heritabilities corresponding to the simulated effects are equal to $h^2_{Q1} = h^2_{Q1Q6} = 0.058$, $h^2_{Q4} = h^2_{Q2Q4} = 0.08$, $h^2_{Q6} = h^2_{Q3Q5} = 0.04$. The overall broad sense trait heritability $H^2 = 0.355$.

Table 2 illustrates both the power of detection for each of the simulated effects, and the standard error of the estimate of QTL location. When $n = 200$ the standard errors of the estimates of location reach the level of 15cM,

therefore an effect identified by mBIC is qualified a true positive if it is within 30cM of the true QTL. Epistatic effects are classified as true positives if both detected positions are within 30 cM of the true epistatic QTL. For the sample size $n = 500$ this detection window was decreased to within ± 15 cM of the true QTL. If more than one effect was found in a detection window only one was classified as a true positive. All other effects were classified as false positives. We also report the estimate of the proportion of false positives, $pfp = \frac{\mathbf{S}fp}{\mathbf{S}ap}$, where $\mathbf{S}fp$ and $\mathbf{S}ap$ are the sums of false positives (fp) and all positives (ap) over all 1000 replicates.

[Table 2 about here.]

Table 2 summarizes both the power and precision of QTL location over increasing sample size. As expected, the power of mBIC increases with an increase of the sample size and for $n = 500$ the power of detecting the weakest main effect with $h^2 = 0.04$ reaches 93% and the corresponding, weakest epistatic effect is detected with the power of 84%. Note that lower power for epistatic effects results from using a larger penalty than for main effects.

Results demonstrated in table 2 show that using dense genetic maps can greatly increase the power of detecting QTL. For $n = 200$ the power of detecting the weakest main effect, Q6, which was 11cM from the closest marker in the sparse map, increased from 35% for the sparse map to 53% for a dense map. Furthermore, a dense genetic map also allowed a greater power of detection of the weakest interaction. For $n = 500$ this power increased from 40% for a sparse map to 84% for a dense map. Note that one of the QTL

involved in this interaction, Q6, is 15cM from the closest marker in the sparse map and thus it is difficult to detect. The advantage of using a dense map is however not so obvious when the QTL effects are large and located close to markers from a sparse map. In fact in this situation the search over a dense map may even yield slightly lower power due to the necessity of adjusting the detection thresholds to the increased multiple testing problem. However, even in the situation where the dense map does not yield the highest power it usually allows for a more precise QTL localization. The only exception for these simulations is Q4, which is located outside the last marker on chromosome 2 (Figure 3). The dense map was constructed only within the markers on this chromosome and did not allow for a more precise localization of Q4.

The use of MIM had a small influence on the power of QTL detection when compared to the sparse map case. However, as anticipated it substantially increased the precision of QTL localization. This difference is clearly visible for $n = 500$ where the standard deviation of QTL localization estimates obtained by interval mapping was smaller than the distance from the closest flanking marker. Again, Q4 is an exception to this, where interval mapping actually increases the standard deviation.

Our simulations demonstrate that the precision of QTL location increases significantly with the sample size. In case of multiple interval mapping the standard deviation of the estimate of the location of the weakest QTL, Q6, decreases from 13.2 cM for $n = 200$ to 6.3 cM for $n = 500$. In case of the dense marker map the standard deviation of the estimate of Q6 location changes from 9.8 cM for $n = 200$ to 4.3 cM for $n = 500$. However, in case when the search is performed over markers, then the standard error of QTL

location is always larger than the distance between the QTL and the closest marker. Therefore, in the sparse map case, the improvement of the precision with increasing sample size is limited.

4. Discussion

The mBIC of Bogdan et al. (2004) is adapted to two unique and applicable situations, namely the search over markers from the dense marker maps and multiple interval mapping. Results based on simulations demonstrate that the proposed methods of relaxing the penalty in the standard version of mBIC allows for the control of the weak sense FWER and the proportion of false positives at the assumed level. The need to relax the penalty in the standard version of mBIC is apparent when markers are spaced closer than by 5 cM. In fact, using mBIC with such dense marker maps may help to increase the power and precision of QTL location. For multiple interval mapping the implemented penalty keeps the FWER error rate at the assumed level and only slightly depends on the number of tests performed within intermarker intervals. Based on our results multiple interval mapping only slightly increases the power of QTL detection when compared to the search over relatively distant markers, but can substantially increase the precision of QTL localization.

The particular weights to calibrate the penalty as presented in Figures 1 and 2 are specific for the backcross design. However, a general method for computing the "efficient" number of markers to be used in mBIC criterion and provided by the equation (7) can be used for other experimental designs. In situations where the structure of correlations between the regressor variables is known, the critical value c can be simulated using the approach presented

in this paper. When the structure of correlations is not known the critical value c can be approximated using the permutation approach described in Churchill and Doerge (1994).

Under complicated genetic scenarios (Zeng et al. (2000)) there are usually many comparable statistical models that provide a good fit to the data. The comparison between such models can be made within a Bayesian framework. Computing posterior probabilities of different models allows Bayesian averaging (Hoeting et al. (1999)) to estimate the genetic parameters, (e.g., heritability). In fact, mBIC allows an approximation to the posterior probabilities of different models

$$P(M_i|Y) \approx \frac{\exp(-mBIC(i)/2)}{\sum_j \exp(-mBIC(j)/2)} \ , \tag{17}$$

where the sum in the denominator is over all possible statistical models. It is worth mentioning that the calibration of the proposed penalty corresponds to the choice of the prior distribution on the QTL number, such that the resulting Bayesian model selection procedure has good frequentist properties. Furthermore, when estimating the posterior probability of a given model by mBIC it is enough to visit each of the plausible models just once. This allows to substantially reduce the computational burden as compared to standard MCMC methods often used in QTL mapping (see Yi et al. (2005) and references given there), which require multiple visits of each model. However, even for mBIC the estimates of $P(M_i|Y)$ are accurate only if the majority of plausible models are represented in the denominator of (17). Therefore, to use mBIC in a Bayesian context, a suitable, computationally efficient search strategy still needs to be developed.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automat. Control.* **19**, 716–723.

Baierl, A., Bogdan, M., Frommlet, F. and Futschik, A. (2006). On locating multiple interacting quantitative trait loci in intercross designs. *Genetics* **173**, 1693–1703.

Bogdan, M., Doerge, R. W. and Ghosh, J. (2004). Modifying the schwarz bayesian information criterion to locate multiple interacting quantitive trait loci. *Genetics* **167**, 989–99.

Broman, K. W. (1997). Identifying quantitative trait loci in experimental crosses. PhD dissertation, Department of Statistics, University of California, Berkeley.

Broman, K. W. and Speed, T. P. (2002). A model selection approach for the identification of quantitative trait loci in experimental crosses. *J. R. Stat. Soc. B* **64**, 641–656.

Carlborg, O. and Andersson, L. (2002). The use of randomisation testing for detection of multiple epistatic qtl. *Genet. Res.* **79**, 175–184.

Carlborg, O., Andersson, L. and Kinghorn, B. (2000). The use of a genetic algorithm for simultaneous mapping of multiple interacting quantitative trait loci. *Genetics* **155**, 2003–2010.

Carlborg, O. and Haley, C. (2004). Epistasis:too often neglected in complex trait studies ? *Nature Rev. Genet.* **5**, 618 – 625.

Cheverud, J. (2001). A simple correction for multiple comparsions in interval mapping genome scans. *Heredity* **87**, 52–58.

Churchill, G. A. and Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics* **138**, 963–971.

Dempster, A. P., Laird, N. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via em algorithm. *J. Roy. Statist. Soc. B* **39**, 1–38.

Doerge, R. W. (2002). Mapping and analysis of quantitative trait loci in experimental populations. *Nature Reviews Genetics* **3**, 43–52.

Doerge, R. W. and Churchill, G. A. (1996). Permutation tests for multiple loci affecting a quantitative character. *Genetics* **142**, 285–294.

Dupuis, J. and Siegmund, D. (1999). Statistical methods for mapping quantitative trait loci from a dense set of markers. *Genetics* **151**, 373–386.

Feingold, E., Brown, P. O. and Siegmund, D. (1993). Gaussian models for genetic linkage analysis using complete high-resolution maps of identity by descent. *Am. J. Hum. Genet.* **53**, 234–251.

Haley, C. and Knott, S. (1992). A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**, 315–324.

Hoeting, J., Madigan, D., Raftery, A. E. and Volinsky, C. (1999). Bayesian model averaging: a tutorial. *Statistical Science* **14**, 382–401.

Jansen, R. and Stam, P. (1994). High resolution of quantitative traits into multiple loci via interval mapping. *Genetics* **136**, 1447–1455.

Kao, C. (2000). On the differences between maximum likelihood and regression interval mapping in the analysis of quantitative trait loci. *Genetics* **156**, 855–865.

Kao, C., Zeng, Z. and Teasdale, R. (1999). Multiple interval mapping for quantitative trait loci. *Genetics* **152**, 1203–16.

Lander, E. and Botstein, D. (1989). Mapping mendelian factors underlying quantitative traits using rflp linkage maps. *Genetics* **121**, 185–199.

Rebaï, A., Goffinet, B. and Mangin, B. (1994). Approximate thresholds of interval mapping test for qtl detection. *Genetics* **138**, 235–240.

Salyakina, D., Seaman, S. R., Browning, B. L., Dudbridge, F. and M uller-Myhsoka, B. (2005). Evaluation of nyholt's procedure for multiple testing correction. *Human Heredity* **60**, 19–25.

Sax, K. (1923). The association of size difference with seed-coat pattern and pigmentation in phaseolus vulgaris. *Genetics* **8**, 552–560.

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Stat.* **6**, 461464.

Soller, M., Brody, T. and Genizi, A. (1976). On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. *Theor. Appl. Genet.* **47**, 35–39.

Thoday, J. M. (1961). Location of polygenes. *Nature* **191**, 368–370.

Wolf, J. B., Brodie, E. D. and Wade, M. J. (2000). Epistasis and the evolutionary process. Oxford University Press, New York.

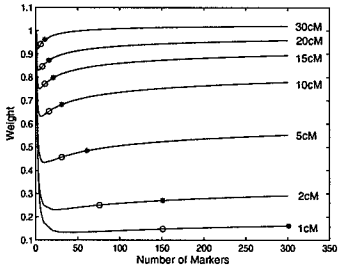
Yi, N., Yandell, B. S., Churchill, G., Allison, D., Eisen, E. and Pomp, D. (2005). Bayesian model selection for genome-wide epistatic qtl analysis. *Genetics* **170**, 1333–1344.

Zeng, Z. B. (1994). Precision mapping of quantitative trait loci. *Genetics* **136**, 1457–1468.

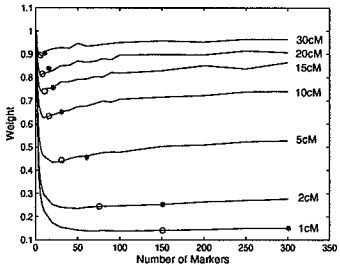
Zeng, Z. B., Liu, J., Stam, L. F., Kao, C.-H., Mercer, J. M. and Laurie, C. C. (2000). Genetic architecture of a morphological shape difference between two *drosophila* species. *Genetics* **154**, 299–310.

Figure 1. Dependence of the weights on the number of markers (N) and the genetic distance between markers (δ) for both additive (add) and epistatic (epi) effects when searching over markers (SM) and for multiple interval mapping (MIM), respectively. For additive effects both theoretical and simulation results are reported. Circles and stars denote chromosomes of length 150cM and 300cM, respectively.

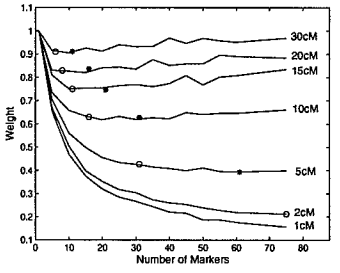
(a) SM: w_{SM}^{add} theoretical



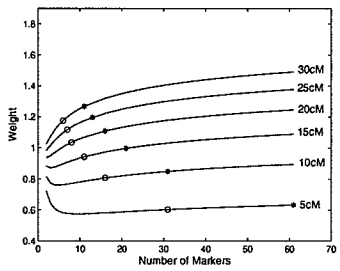
(b) SM: w_{SM}^{add} simulated



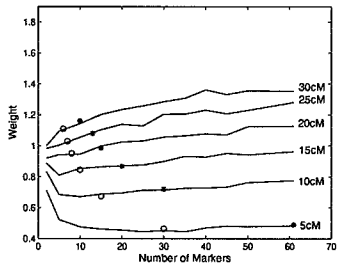
(c) SM: w_{SM}^{epi} simulated



(d) MIM: w_{IM}^{add} theoretical



(e) MIM: w_{IM}^{add} simulated



(f) MIM: w_{IM}^{epi} simulated

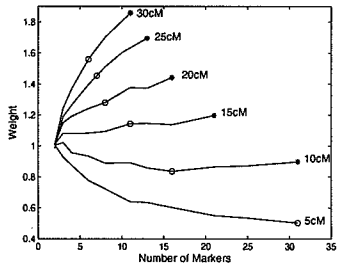
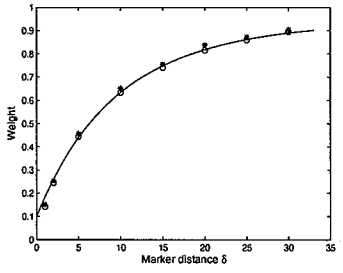
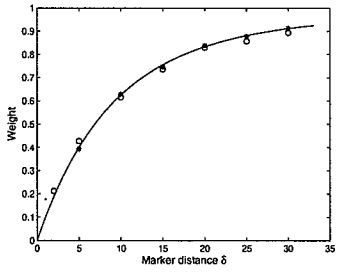


Figure 2. Investigating the accuracy of the approximation for simulated weights. Circles and stars denote chromosomes of length 150cM and 300cM, respectively. For multiple interval mapping the weights for main effects are obtained based on 2cM grid of imputations. For epistasis circles mark the observations for 150 cM and 2 cM grid, and dots for 150 cM and 4cM grid of imputations. The weights for epistasis for a 300 cM chromosome are denoted by stars and were obtained using 4cM grid of imputations.

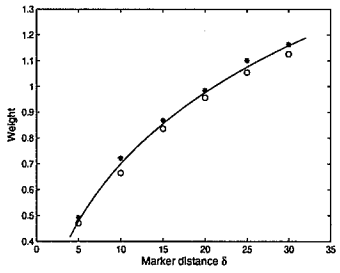
(a) $\hat{w}_{SM}^{add} = 1 - 0.9e^{(-10\delta/100+10(\delta/100)^2)}$



(b) $\hat{w}_{SM}^{epi} = e^{(-10.7\delta/100+8.7(\delta/100)^2)}$



(c) $\hat{w}_{IM}^{add} = -0.15 + 3.1\sqrt{\delta/100} - 1.3\delta/100$



(d) $\hat{w}_{IM}^{epi} = -0.53 + 5.4\sqrt{\delta/100} - 2.7\delta/100$

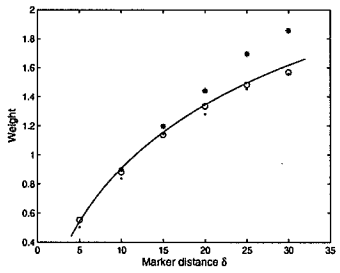


Figure 3. Marker and QTL locations for a scenario with 30 unequally spaced markers. Markers are denoted by black vertical lines. QTL locations are represented as circles. The distance between markers is specified, the average interval distance is 11.1 cM

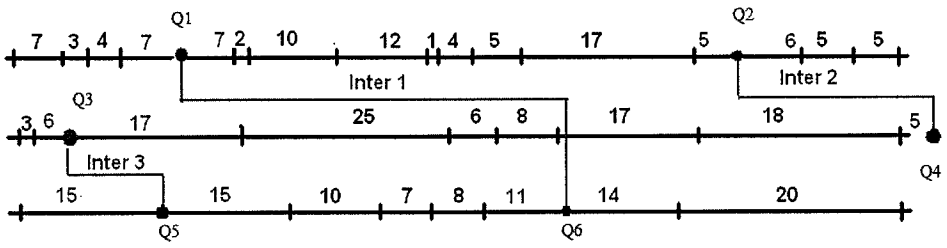


Table 1

Probabilities of false detections (in %). e_m and e_e denote the percentage of simulations for which main and epistatic effects were falsely detected. $e_t = e_m + e_e$ is the percentage of simulations for which at least one false signal was detected (i.e. weak sense FWER). SM denotes the search over markers and IM denotes our multple interval mapping approach. The significance levels are assumed to be 8% for $n = 200$ and 5% for $n = 500$.

n	method	δ	e_m	e_e	e_t
200	SM	2cM	4.4	2.4	6.9
	SM	5cM	4.8	2.9	7.8
	SM	10cM	3.7	3.6	7.3
	SM	20cM	4.7	3.4	8.1
	SM	Fig. 3	4.8	3.0	7.8
200	IM	5 cM	4.9	2.9	7.8
	IM	10 cM	4.1	4.0	8.1
	IM	20 cM	4.8	4.2	9.0
	IM	25 cM	3.9	4.3	8.2
	IM	Fig. 3	4.8	3.1	7.9
500	SM	2cM	1.7	1.5	3.2
	SM	5cM	2.9	2.4	5.3
	SM	10cM	2.8	2	4.8
	SM	20cM	3.4	2.4	5.8
	SM	Fig. 3	3.3	2.4	5.7
500	IM	5 cM	3	2.6	5.6
	IM	10 cM	2.4	2.6	5.0
	IM	20 cM	2.6	2.4	5.0
	IM	25 cM	3.1	2.2	5.3
	IM	Fig. 3	3.2	2.1	5.3

Table 2

Estimates of power and the precision for QTL location (in cM) for the search over markers in a sparse map (scenario 1), multiple interval mapping (scenario 2), and the search over markers spaced every 2cM (scenario 3). pfp denotes the estimated proportion of false positives as defined in the text.

n	scenario	Q1 *power* *std.loc.*	Q4 *power* *std.loc.*	Q6 *power* *std.loc.*	Int 1 *power* *std.loc.Q1* *std.loc.Q6*	Int 2 *power* *std.loc.Q2* *std.loc.Q4*	Int 3 *power* *std.loc.Q3* *std.loc.Q5*	*pfp*
200	1	0.67 10.2	0.76 6.9	0.35 14.8	0.24 10.9 14.5	0.47 9.2 7.8	0.07 10.1 16.6	0.06
200	2	0.69 9.8	0.76 8.1	0.39 13.2	0.28 9.7 12.1	0.47 8.0 8.6	0.10 9.0 11.4	0.07
200	3	0.71 7.8	0.75 9.6	0.53 9.8	0.40 8.3 8.2	0.45 7.3 9.7	0.19 7.1 8.2	0.06
500	1	0.96 8.0	0.98 5.0	0.86 12.9	0.79 8.2 13.0	0.94 5.8 5.0	0.40 9.2 15.0	0.07
500	2	0.97 5.6	0.997 7.0	0.83 6.3	0.80 6.2 6.1	0.97 4.7 6.9	0.48 5.7 6.8	0.07
500	3	0.99 3.6	0.985 6.3	0.93 4.3	0.955 3.8 3.4	0.97 3.2 6.2	0.84 3.8 4.4	0.04