The development of genetic maps of markers based upon DNA polymorphisms is beginning to provide the experimental geneticist and the plant and animal breeder with powerful tools for the study of quantitative genetic variation. The use of markers to detect individual loci responsible for quantitative genetic variation (quantitative trait loci or QTL) provides greater power than segregation analysis without marker information. The use of flanking marker methods has proved to be a powerful tool for the mapping of quantitative trait loci (QTL) in the segregating generations derived from crosses between inbred lines (Haley and Knott 1992). The Haley-Knott regression method is based on multiple regressions which can be applied using any general statistical package, developed by Haley and Knott (1992). They used the example of mapping in an F2 population and showed that these regression methods produce very similar results to those obtained using maximum likelihood (Haley and Knott 1992). The Haley–Knott (HK) regression method continues to be a popular approximation to standard interval mapping (IM) of quantitative trait loci (QTL) in experimental crosses (Feenstra et al. 2006). Currently, the HK regression method is preferred as a fast approximation to the IM method for estimating model parameters (Feenstra et al. 2006).

In addition to the correctly estimated QTL position, the size of the CI (95%) serves as an important criterion in QTL mapping. Bootstrapping method had more per cent of accurate intervals but average size of the intervals was very high in more scenarios. 1 LOD support interval and bayesian credible interval resulted to be preferable with high per cent of accurate and small confidence intervals, moreover they weekly affected by parameters such as population size and SD of QTL. This study investigated that we can predict CIs for more combination of simulated parameters using best trained ANN. By this study it is sugestive to consideration of more combinations of simulated parameters using the model obtained by best structured ANN to expanding of original study.

**Key words:** Artificial Neural Network, F2 Design, QTL mapping, Regression method
to determine the effect of marker spacing, population size, standard deviation of QTL effect, ratio of additive to dominance effect of QTL and QTL location relative to flanking markers on the percent of accurate confidence intervals (PACI) and Mean of ACI obtained using HK regression method. Then an optimal ANN was designed to predict PACI and Mean of ACI. Finally the designed adequate ANN model were used for prediction of PACI and mean of ACI for more combination of simulated parameters (scenarioes).

MATERIAL AND METHODS

Haley and Knott regression method

We assume that $y_i|g_i ~ N(\mu_{g_i}, \sigma^2)$, where $y_i$ is the phenotype of individual $i$ and $g_i$ is its (unobserved) QTL genotype. Therefore we assumed that the phenotypic data ($y_i$) in every group of animals (animals with same QTL genotype) has a mean depending on QTL genotype and a residual variance. Then we calculate conditional QTL genotype given marker genotype, $p_{ij} = pr(g_i | M_i)$ where $M_i$ is marker genotype data for individual $i$. Phenotype of individual $i$ for a given marker data follows a mixture of normal distribution, $E(y_i| M_i) = \sigma p_{ij} \mu_j$ where $\mu_j$ is the mean of individual’s phenotype with $j$th QTL genotype so the conditional phenotype average given marker length was simulated. The genotypes of markers and QTL population the trait value had a normal distribution. E(yi|Mi)= $\sigma p_{ij} \mu_j$ for individual $i$. Phenotype of individual $i$ and $g_i$ is its (unobserved) QTL genotype.

Data simulation

An F$_2$ population derived from crossing between two inbred line each with alternate homozygote genotype in marker loci and QTL, with different population size was simulated. 11 markers on one chromosome with different equal spaces of 5 and 10 CM were simulated. Chromosome length was different corresponding to marker spaces from 50 and 100 respectively with presenting one QTL (between 6th and 7th marker). An F$_2$ populations with different combination of population sizes ($P_s$) of 300, 600 and 900, Standard deviation of QTL effect (SDQ) of 0.2, 0.5, 0.8 and with different portion of additive to dominance effects (Rad) of 0.25, 0.5 and 0.75 were simulated. In each combination of simulated parameters (scenario) QTL located between sixth and seventh markers relatively with 0, 0.25, 0.5 of the interval separated from sixth marker. Therefor 162 scenarioes with different combinations of 2 levels of marker spacing (Ms), 3 level of Population size (Ps), 3 level of QTL effect (SDQ) each with 3 different levels of proportion of additive dominance effect (Rad) and 3 level of QTL location relative to adjacent flanking markers (rpQ) were considered. Each scenario was replicated 100 times. In each simulated population the trait value had a normal distribution.

For each individual one chromosome with corresponding length was simulated. The genotypes of markers and QTL was sampled from binomial distribution using haldane mapping function, thus Crossing over between markers and between markers and QTL was simulated using haldane mapping function. Trait value was sampled from normal distribution with corresponding mean according to genotype of QTL and with unexplained standard deviation (unexplained by QTL) equal 1 SD($\sigma=1$).

QTL mapping analysis

Analysis was carried out using Haley-Knott regression method using R/qtl (Broman et al. 2003) package. At first carried out a genome scan with a single QTL model for estimating probable QTL position with higher LOD score on the chromosome. The LOD scores calculated as LOD=$(n/2)\log_{10}(RSS_0/RSS_1)$ where $n$ is sample size, RSS$_0$ is the null residual sum of squares and RSS$_1$ is model residual sum of squares (the model defined as regression of phenotypes on the conditional QTL genotypes depending on markers genotypes). For estimating intercept (mean), additive effect and dominance effect and corresponding standard error of them we fitted a single QTL model using HK regression method. To determine the significant treshold of LOD score ($\alpha=0.05$), a permutation with 1000 replicate using defined model were done for each replication of a scenario. 1 and 1.5 LOD support intervals calculated as interval in wich LOD score is within 1 and 1.5 units of its maximum on the chromosome. With a priori of being QTL existing any where of chromosome is equal, posterior distribution obtained by rescaled LOD to $10\log_{10}(RSS_0/RSS_1)$ were calculated.

Comparision statistics

Here Accurate Confidence Intervals (ACI) is the estimated confidence interval in the replication with significant ($\alpha=0.05$) QTL that covers real QTL position on the chromosome. Per cent of Acurate Confidence Intervals (PACI) for each scenario was calculated as,

$$PACI = \frac{\text{Number of ACIs}}{\text{Number of significant replicates in the scenario}}$$

We calculated mean of ACIs size as,
Artificial neural network modeling and evaluation

For fit an ANN model to predict PACI and mean of ACI for different parameters using Ps, Ms, SDQ, Rad and rQp as inputs, we developed an ANN model by backpropagation algorithms. The data set obtained from simulation study were divided into learning and testing data set, then learning data set used to train the ANN and testing data set were used for validation of the trained ANN model. The best of number of input, output, learning rate, momentum coefficient, number of hidden layers, number of hidden neurons, and number of training cycles or epochs were chosen to obtain the optimal ANN. To determine adequacy of the ANN model we used three statistics containing $R^2$, $T$ and root MSE (RMSE). The $T$ statistics measures the scattering around fitted line using the ANN. It’s better when close to 1. The formula of calculating $T$ is as below (Khazaei et al. 2008),

\[ T = \frac{\sum_{i=1}^{n} (X_m - X_p)^2}{\sum_{i=1}^{n} (X_m - \bar{X})^2} \]

where $n$ is the number of data set, $\bar{X}$ is the average of X over the $n$ samples, and $X_m$ and $X_p$ are the actual and by ANN model predicted HK efficiency parameters, respectively.

Expanding of original simulation study using the ANN models

For appropraite methods of calculating CI accourding to results from the simulation study, we considered more scenarios. Mean of ACI and PACI for these scenarios were predicted using the trained ANN models. Thus More different levels of simulated parameters were used to create large-scale scenarios. Then the adequate ANN models were used to predict outputs. In the original study the space from different levels of simulated parameters was equal thus it can support to proper the use of the adequate ANN model to expanding the original study.

RESULTS

Confidence intervals

PACI using all methods positively affected by increasing Ps and decreasing rQp. Table 1 shows average effect of other parameters than Ps on PACIs obtained from different methods. Per cent of accurate 1 LOD support interval affected slightly by SDQ and affected by Ms specially in low level of SDQ. SDQ increased the PACI obtained from 1.5 LOD interval the increase was higher when SDQ increased from 0.2 to 0.5. Ms in high and medium SDQ did not affect PACI of 1.5 LOD interval. Increase of SDQ decreased PACIs obtained from bayesian credible interval, other 2 parameters. PACI in the bootstrap 95% and 99% was not affected by Ms, SDQ and Rad.

As presented in Table 2 mean of ACIs in all methods affected by rQp and sharply by Ps. CIs was small in larg Ps and when the QTL located on the marker. CIs in 1 LOD interval became high with increase of Rad and Ms, size of the CI was narrowed by increas in SDQ, the trend was same in other methods.

In total CIs of bootstrap method was very large than others, size of CIs in different methods. In the SDQ of 0.8 all methids had small CI, and in the SDQ of 0.2 efficiency of bootstraping method decreased. 1 LOD interval was small in SDQ of 0.2, but BCI was better on the context of 0.5 and 0.8.

ANN models and evaluations

To fit an ANN model for prediction of mean of ACIs we used ANN structure of 4–10–5–5, containing number of

Table 1. Average effect of SDQ, Rad and Ms on PACI in different methods

<table>
<thead>
<tr>
<th>SDQ=0.2</th>
<th>SDQ=0.5</th>
<th>SDQ=0.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rad=0.25</td>
<td>Rad=0.5</td>
<td>Rad=0.75</td>
</tr>
<tr>
<td>† Ms</td>
<td>91.15</td>
<td>91.05</td>
</tr>
<tr>
<td>5</td>
<td>90.87</td>
<td>91.93</td>
</tr>
<tr>
<td>10</td>
<td>97.30</td>
<td>97.00</td>
</tr>
<tr>
<td>‡ Ms</td>
<td>95.63</td>
<td>96.90</td>
</tr>
<tr>
<td>5</td>
<td>93.47</td>
<td>94.03</td>
</tr>
<tr>
<td>10</td>
<td>92.36</td>
<td>94.18</td>
</tr>
<tr>
<td>ª Ms</td>
<td>98.04</td>
<td>99.17</td>
</tr>
<tr>
<td>5</td>
<td>99.60</td>
<td>98.46</td>
</tr>
<tr>
<td>10</td>
<td>99.89</td>
<td>99.76</td>
</tr>
<tr>
<td>§ Ms</td>
<td>100.00</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Ms, Marker spece, Rad, ratio of additive to dominance Standard Deviation, SDQ, Standard Deviation of QTL effect, PACI, per cent of accurate confidence interval (per cent of confidence intervals that containing QTL), † PACIs that obtained using 1 LOD support interval, ‡ PACIs that obtained using 1.5 LOD support interval, † PACIs that obtained using baysian credible interval with 95%, *PACIs that obtained using bootstrap method with 95%, § PACIs that obtained using bootstrap method with 99%.
inputs \((Ps, Ms, SDQ\text{and } rpQ)\), number of neurons in first and second hidden layer and number of outputs, respectively. Results from 162 scenarios partitioned to 110 scenarios for training data set and 52 scenarios for testing data set. Adequacy statistics for every predicted output (mean of ACI resulted from different methods) using the ANN model were presented in Table 3. The ANN trained with Learn Rule of Delta, TanH transfer function, with learning rate and momentum equal to 0.3 and 0.4, respectively. The results show that the ANN model is able to learn the relationship between the inputs and mean of ACIs for different methods.

Adequate ANN structure for prediction of PACI resulted from 1 LOD and BCI methods were 5–10–9–5–2 (number of input, number of neurons for first, second and third hidden layer and number of output, respectively). Other parameters containing learning rule, transfer function and etc were same with the ANN that fitted for mean of ACI. Adequacy statistic of the ANN model to predict of PACI for 1 LOD and BCI method using the ANN were presented in Table 4. The results show that the ANN can’t predict PACI properly, in other word there are other parameter or parameters rather than considered inputs that affects the PACIs.

Expanded study using the ANN models

A total of 3,276 different scenarios using different simulated parameters that presented in Table 5 were used to predict mean of ACIs by the trained ANN model. Fig. 1 presents mean of ACIs resulted from 1 LOD support interval in different combinations of Ps and Ms. As presented in the Fig. 1 increase of Ps and decrease of Ms decreases the mean of ACIs. The results show that mean of ACI resulted from BCI affected by Ps and Ms with same trend as 1 LOD support method. Moreover effect of SDQ and Rad on mean of ACIs obtained from 1 LOD support interval were presented in Fig. 2. Effect of the parameters on mean of ACI resulted from BCI followed the same trend as presented in Fig. 2.

Consideration of PACI using the ANN trained for this regard carried out using 36036 different scenarios that...
ACI, accurate confidence interval (confidence intervals containing QTL), Ps, Population size, Ms, Marker space

Fig. 1 presents mean of ACIs resulted from 1 LOD support interval in different combinations of Ps and Ms.

SDQ, standard deviation of QTL effect, ACI, accurate confidence interval (confidence intervals containing QTL)

Fig. 2. Effect of SDQ and Rad on mean of ACIs for 1 LOD support interval.

Ps, Population size, Ms, Marker space, PACI, per cent of accurate confidence interval (per cent of confidence intervals containing QTL)

Fig. 3. Effect of Ps and Ms on different combinations of Ps and Ms on PACI using 1 LOD method.

Table 4. Adequacy statistics of predicted PACI for 1 LOD support interval and BCI using the ANN model

<table>
<thead>
<tr>
<th>Data</th>
<th>Statistics</th>
<th>1 LOD</th>
<th>BCI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Testing data set</td>
<td>$R^2$</td>
<td>0.637163</td>
<td>0.788005</td>
</tr>
<tr>
<td></td>
<td>$T$</td>
<td>0.313621</td>
<td>0.774682</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.137845</td>
<td>0.087068</td>
</tr>
<tr>
<td>Training data set</td>
<td>$R^2$</td>
<td>0.919026</td>
<td>0.903801</td>
</tr>
<tr>
<td></td>
<td>$T$</td>
<td>0.918762</td>
<td>0.90253</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.048167</td>
<td>0.054305</td>
</tr>
<tr>
<td>All data set(train+test)</td>
<td>$R^2$</td>
<td>0.777918</td>
<td>0.861341</td>
</tr>
<tr>
<td></td>
<td>$T$</td>
<td>0.737094</td>
<td>0.860303</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.086376</td>
<td>0.066155</td>
</tr>
</tbody>
</table>

PACI, percent of accurate confidence interval (percent of confidence intervals containing QTL), 1 LOD, confidence interval resulted from 1 LOD support interval, BCI, confidence interval resulted from Bayesian Credible Interval.

P increasing the per cent of accurate confidence intervals for 1 LOD support and BCI methods. Because of low adequacy of the ANN model trained for PACI no more consideration carried out.

DISCUSSION

As presented in Table 1 Bootstrapping method for confidence interval has a high per cent of accurate confidence interval in all scenarios but mean of confidence interval wide
in all scenarios and only fairly preferable when QTL effect is high. Effect of marker space on size of confidence interval was very small, either Wisscher et al. (1996) reported that the Marker spacing had only a small effect on the average empirical confidence interval obtained by Bootstrapping method. Although bayesian credible interval has appropriate per cent of accurate confidence specially in combination of low effect of QTL comparing other methods but this method in the level of effect has higher size of confidence interval comparing 1 lod support interval. Percent of accurate interval of LOD support intervals low affected by parameters, 1 LOD support interval is preferable specially in high marker space because of pretty per cent of accuracy (Table 1) and size of confidence interval. “Resolving power” defined by Darvasi and Soller (1997) as the 95% of confidence interval for QTL map location. Their study show that resolving power is inversely proportional to sample size and to square of the QTL gene effect.

The results of this study implies that 1 LOD support interval is better for it’s low affection from simulated parameters. This study resulted that ANN can properly learn relationship between simulated parameters and mean of ACI, but it’s not useful for prediction of PACI.

ACKNOWLEDGEMENT

The authors are grateful to Ferdowsi University of Mashhad for supporting the research.

REFERENCES