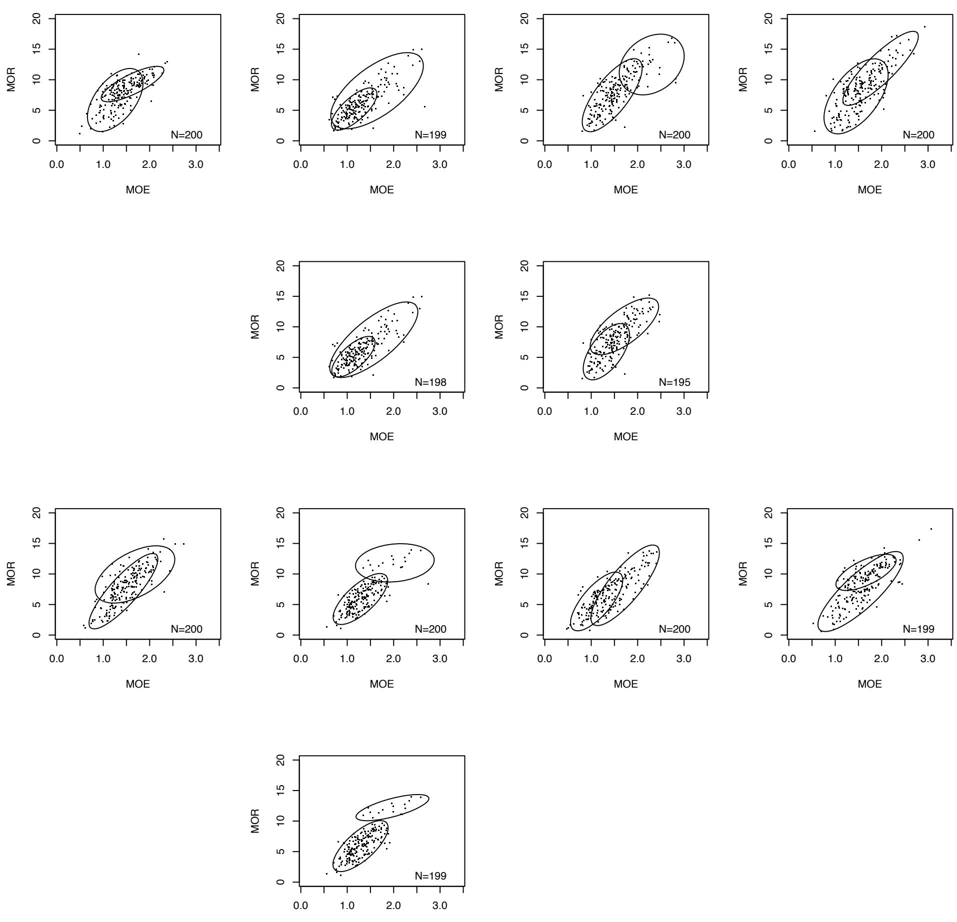
United States Department of Agriculture

Estimated Probability of Breakage of Lumber of a Fixed “Grade” Can Vary Greatly from Mill to Mill and Time to Time

Steve P. Verrill
Frank C. Owens
Matthew A. Arvanitis
David E. Kretschmann
Rubin Shmulsky
Robert J. Ross
Patricia Lebow



Forest Service
Forest Products Laboratory
Research Paper FPL–RP–705
August 2020

／
Abstract

To evaluate the reliability of lumber structures, we need (among many other things) good models for the strength and stiffness distributions of visual and machine-stress-rated (MSR) grades of lumber. Verrill et al. established theoretically and empirically that the strength properties of visual and MSR grades of lumber are not distributed as two-parameter Weibulls. Instead, strength properties of grades of lumber must have (at least to a first approximation) “pseudo-truncated” distributions. To properly implement Verrill et al.’s pseudo-truncation theory, we must know the true mill run modulus of elasticity (MOE) and modulus of rupture (MOR) distributions.

Owens et al. investigated the mill run distributions of MOE and MOR at two times for each of four mills. They found that univariate mill run MOE and MOR distributions are well-modeled by skewnormal distributions or mixtures of normal distributions, but not so well-modeled by normal, lognormal, two-parameter Weibull, or three-parameter Weibull distributions.

Verrill et al. investigated a mixture of two bivariate normals model for the mill run bivariate MOE–MOR population at a single time at a single mill. (Some possible sources of two-component mixture relationships include a mixture of trees from a fast-grown plantation stand and a suppressed stand, trees of two separate species, small-diameter trees and large-diameter trees, and lumber from the pith region versus lumber from the bark region.) They found that a mixture of two bivariate normals model performed well.

In this paper, we apply this model to all eight of the Owens et al. lumber samples. We find that the model continues to yield good fits. However, we also find that the fits differ from mill to mill and time to time. Some variability is, of course, to be expected. However, we find that the fitted models differ to such an extent that the calculated probability that a piece of lumber randomly drawn from a fixed “grade” breaks at a fixed load can vary by a factor as large as 35 when we permit both season and mill to vary, and as large as 15 when we permit only mill to vary. Similar factors were found when we replaced fixed loads with loads randomly drawn from fixed load distributions.

Keywords: Machine-stress-rated lumber, MSR lumber, MOE binned lumber, visually graded lumber, lumber property distribution, lumber reliability, bivariate normal, mixture of bivariate normals

August 2020

Verrill, Steve P.; Owens, Frank C.; Arvanitis, Matthew A.; Kretschmann, David E.; Shmulsky, Rubin; Ross, Robert J.; Lebow, Patricia. 2020. Estimated probability of breakage of lumber of a fixed “grade” can vary greatly from mill to mill and time to time. Research Paper FPL-RP-705. Madison, WI: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory. 47 p.

A limited number of free copies of this publication are available to the public from the Forest Products Laboratory, One Gifford Pinchot Drive, Madison, WI 53726-2398. This publication is also available online at www.fpl.fs.fed.us. Laboratory publications are sent to hundreds of libraries in the United States and elsewhere.

The Forest Products Laboratory is maintained in cooperation with the University of Wisconsin.

Estimated Probability of Breakage of Lumber of a Fixed "Grade" Can Vary Greatly from Mill to Mill and Time to Time

Steve P. Verrill, Mathematical Statistician
USDA Forest Service, Forest Products Laboratory, Madison, WI

Frank C. Owens, Assistant Professor
Department of Sustainable Bioproducts, Mississippi State University, MS

Matthew A. Arvanitis, Mathematical Statistician
USDA Forest Service, Forest Products Laboratory, Madison, WI

David E. Kretschmann, Research General Engineer (retired[1])
USDA Forest Service, Forest Products Laboratory, Madison, WI

Rubin Shmulsky, Professor and Department Head
Department of Sustainable Bioproducts, Mississippi State University, MS

Robert J. Ross, Supervisory Research General Engineer
USDA Forest Service, Forest Products Laboratory, Madison, WI

Patricia Lebow, Mathematical Statistician
USDA Forest Service, Forest Products Laboratory, Madison, WI

1 Introduction

To evaluate the reliability of lumber structures, we need (among many other things) good models for the strength and stiffness distributions of visual and machine-stress-rated (MSR) grades of lumber. Verrill et al. (2012, 2013, 2014, 2015, 2019, 2020) established theoretically and empirically that the strength properties of visual and MSR grades of lumber are not distributed as two-parameter Weibulls. Instead, strength properties of grades of lumber must have (at least to a first approximation) "pseudo-truncated" distributions.

"Pseudo-truncation" has a technical meaning. The concept, at least, of pseudo-truncation was recognized in an ASCE pre-standard report (ASCE 1988). Section B3 of that standard notes that "an improved strength distribution can be obtained by . . . thinning the lower tail by sorting on a correlated variable." For example, if the full ("mill run") bivariate modulus of elasticity–modulus of rupture (MOE–MOR) distribution were a bivariate Gaussian (normal)–Weibull, then truncating or "binning" on the basis of MOE values (as in MSR lumber) would lead to a pseudo-truncated

[1]now President, American Lumber Standard Committee, Inc., Frederick, MD

MOR distribution. That is, because MOE and MOR are not perfectly correlated, truncating on the basis of lower and upper MOE limits does not lead to perfect truncation of the MOR distribution, but it does, of course, lead to a MOR distribution whose tails are thinned. For the case in which the mill run joint MOE–MOR distribution is a bivariate Gaussian–Weibull, Verrill et al. (2012, 2015) derived the exact form of this “pseudo-truncated Weibull” distribution. (They obtained its probability density function.) They also showed that it cannot have tail behavior that matches that of a Weibull distribution.

To properly implement Verrill et al.'s (2012, 2015) pseudo-truncation theory, we must know the true mill run MOE and MOR distributions.

Verrill et al. (2017) and Owens et al. (2018, 2019, 2020) investigated the mill run distributions of Southern Pine 2x4 MOE and MOR at two times for each of four mills. They found that univariate mill run MOE and MOR distributions are well-modeled by skewnormal distributions or mixtures of normal distributions, but not so well-modeled by normal, lognormal, two-parameter Weibull, or three-parameter Weibull distributions.

Verrill et al. (2018) investigated a mixture of two bivariate normals model for the mill run bivariate MOE–MOR population at a single time at a single mill. (Some possible sources of two-component mixture relationships include a mixture of trees from a fast-grown plantation stand and a suppressed stand, trees of two separate species, small-diameter trees and large-diameter trees, and lumber from the pith region versus lumber from the bark region.) They found that a mixture of two bivariate normals model performed well.

In this paper, we apply this model to all eight of the Owens et al. (2018, 2019, 2020) lumber samples. We find that the model continues to yield good fits.

However, we also find that the fits differ from mill to mill and from time to time. This finding is in accord with Anderson et al. (2019), who focused on changes in mean and standard deviation within a mill over time. Some variability in parameter estimates is, of course, to be expected. However, we find that the fitted models differ to such an extent that the calculated probability that a piece of lumber randomly drawn from a fixed “grade” breaks at a *fixed load* can vary by a factor as large as 35 when we permit both season and mill to vary, and as large as 15 when we only permit mill to vary.

Similar factors were found when we replaced fixed loads with loads randomly drawn from *fixed load distributions.*

2 A Note on “Grade”

In this paper we place the word “grade” in quotation marks. We are not working with visual grades. We are not working with MSR grades. Instead we are working with a “grade” that is based on MOE boundaries. The boundaries, 1.311 and 1.772, are approximate 40th and 80th percentiles of the 1591 MOE mill run samples that we obtained. For reference, these values lie above the Southern Pine Inspection Bureau (SPIB) median value for No. 2 non-dense 2x4 lumber and below the SPIB median value for No. 1 dense 2x4 lumber.

We are not working with visual grades because our pseudo-truncation work is not yet capable of handling visual grades. We are not working with MSR grades because, in addition to MOE limits, MSR grades include both visual grade requirements and quality control requirements.

Instead, we are focusing on addressing the question of whether it is reasonable to believe that the probabilities of breakage associated with a "grade" of lumber *loosely* associated with actual grades of lumber might vary somewhat widely depending on mill and season.

If so, we expect that potentially large efficiencies may be gained through the development of data-based models that better explain and predict the performance of individual pieces of lumber.

3 A Note on "Probability of Breakage"

Formally, in our notation,

$$\text{Probability of Breakage} = \text{Prob}(\text{Load} > \text{MOR})$$

In this paper, the MOR distribution is modeled as a pseudo-truncated mixture of normals. We have theoretical and empirical reasons (discussed in Verrill *et al.* (2015, 2018), and in the current paper) for believing that this is a reasonable model, and that our strength distribution fits may have real-world implications.

On the other hand, in this paper, loads are fixed at (strength distribution 5th percentiles)/2.1 or are modeled as lognormal distributions that exceed 5th/2.1 values with probability 0.02 and have coefficients of variation equal to 0.30.

Thus, in this paper, *strength* distributions are based on theory, on data, and on tests of goodness of fit, but *load* distributions are not. Thus, we cannot and do not claim that our estimates of "probability of breakage" have direct real-world meaning.

However, we also argue that the *variability* in estimates of "probability of breakage" across mills and times observed in fits to our data is worthy of note and deserves further investigation.

In particular, we see probability of breakage ratios between mills or times or mill-times that have similarly high values across a range of fixed "5th/2.1" loads (Section 9) and for a lognormal load distribution (Section 10). This suggests that if we replace these load values with, for example, values randomly drawn from an appropriate data-based 50 year max load distribution, we might continue to see similarly high breakage ratios between mills, times, and mill-time combinations.

4 Data

The data collection process is described in detail in Owens et al. (2018, 2019, 2020). It yielded eight samples (four mills at two times) of southern pine (*Pinus* spp.) 2x4 lumber, each containing approximately 200 pieces. The MOE values used in our analyses are in millions of pounds per square inch. The MOR values are in thousands of pounds per square inch.

We observed 7 "visual outliers" among the 1598 data points (2 of the 1600 pieces of lumber were broken and untestable). We found that 2 of these visual outliers were statistically significant outliers. The visual outliers are listed in Table 1. Statistical significance results are listed in column 6 of Table 1. The manner in which the statistical significance values were obtained is discussed in Appendix A.

By removing visual outliers we formed 3 modified data sets. The resulting 11 data sets (8 original, 3 modified) are identified in column 2 of Table 2. In the tables, the word "all" in a data set description indicates that no outliers were omitted from the data set.

5 Graphical Evidence for a Mixture of Bivariate Normals

The models discussed in section 4.1 of Verrill et al. (2017) and Owens et al. (2018, 2019, 2020) were "univariate" models. That is, they were models for the distributions of single variables. This univariate work suggested that mill run MOE and MOR distributions might be modeled as mixtures of two univariate normals, or as skew normal distributions, but not, in general, as normal, lognormal, two-parameter Weibull, or three-parameter Weibull distributions.

Because strength and stiffness are correlated, we gain extra information by modeling the joint behavior of MOE and MOR. In this section, we discuss a bivariate model — a mixture of two bivariate normals — that is suggested by the data. We note that if a bivariate stiffness–strength distribution is a mixture of *bivariate* normals, then the corresponding univariate (or "marginal") stiffness and strength distributions will be mixtures of *univariate* normals. (This is well known to statisticians. For the convenience of readers, Verrill et al. (2018) provided an elementary proof in their appendix A.) Thus, a MOE–MOR distribution that is well modeled by a mixture of bivariate normals would explain the fact that MOE and MOR are individually well modeled by mixtures of univariate normals.

Some possible sources of two-component mixture relationships include a mixture of trees from a fast-grown plantation stand and a suppressed stand, trees of two separate species (for example, *Pinus taeda* and *Pinus palustris*), small-diameter trees and large-diameter trees, and lumber from the pith region versus lumber from the bark region.

We used maximum likelihood methods to fit mixtures of two bivariate normal distributions to each of the 11 data sets. The probability density functions of a bivariate normal distribution, of a mixture of two bivariate normal distributions, and of a pseudo-truncated mixture of two univariate normals are given in or derived in appendices A and B of Verrill et al. (2018). For the reader's convenience, we also provide the probability density functions of a bivariate normal distribution and of a mixture of two bivariate normal distributions in Appendix B of the current paper. The programs that we wrote to perform the maximum likelihood fits can be found at
`http://www1.fpl.fs.fed.us/mixbivn.2019.html`. In Table 2, we provide the results of these fits.

The fitted $\hat{\mu}_{\text{MOE},1}$, $\hat{\sigma}_{\text{MOE},1}$, $\hat{\rho}_1$, $\hat{\mu}_{\text{MOR},1}$, $\hat{\sigma}_{\text{MOR},1}$, $\hat{\mu}_{\text{MOE},2}$, $\hat{\sigma}_{\text{MOE},2}$, $\hat{\rho}_2$, $\hat{\mu}_{\text{MOR},2}$, $\hat{\sigma}_{\text{MOR},2}$, and $\hat{p}$ in Table 2 correspond to the parameters μ_{X1}, σ_{X1}, ρ_1, μ_{Y1}, σ_{Y1}, μ_{X2}, σ_{X2}, ρ_2, μ_{Y2}, σ_{Y2}, and p of equation (1) in Appendix B. In particular, $\hat{p}$ is our estimate of the proportion of specimens that come from the leftmost bivariate normal population.

Given these fits, what is the graphical evidence for a mixture of bivariate normals?

Statisticians know that if variables X and Y have a bivariate normal distribution, then a plot of Y versus X values will be an approximately elliptical cloud of data points. In Figures 1-12 we provide scatter plots of MOR versus MOE for the eight populations. (There is an initial "summary plot," and there are two plots for Mill 2 "summer," Mill 3 "summer," and Mill 2 "winter" because these data sets contain visual outliers and we provide plots with and without outliers.) Each of the scatter plots has been overlaid with the (elliptical) 0.90 probability content contours (calculated from the maximum likelihood fits — again see `http://www1.fpl.fs.fed.us/mixbivn.2019.html`) corresponding to the two bivariate normal components of the fitted mixture. The leftmost cloud in a scatter plot corresponds to the "Leftmost" fit in Table 2. The rightmost cloud corresponds to the "Rightmost" fit in Table 2. All 11 of the scatter plots suggest that the bivariate MOE–MOR distributions might be well approximated by mixtures of two bivariate normal distributions.

The mixture of bivariate normal fits can also be used to calculate probability content contours for the full distribution (as opposed to contours for the two components). (See `http://www1.fpl.fs.fed.us/mixbivn.2019.html` for listings of the Fortran programs that were used to perform the fits and calculate the contours.) In Figures 13–24 we plot 0.90 probability content contours for the full mixed bivariate normal distributions. (Figure 13 is a "summary plot.") We obtain one of these "full distribution" contours by finding a line along which the fitted mixed bivariate normal probability density function has a constant value and within which 0.90 of the data is expected to fall (from samples randomly drawn from the mixed bivariate normal distribution). Again these plots suggest that a mixture of two bivariate normals model might be reasonable for most of the 11 data sets.

Of course, the plots do not *prove* anything. However, they do lend support to the intuition that mill run lumber stiffness–strength distributions can often be mixtures of bivariate distributions. For our four mill, two times data, the MOE–MOR distributions appear to be generally well-fit by mixtures of two bivariate normal distributions.

In the next two sections we supplement this "graphical" evidence with results from formal tests.

6 Likelihood Ratio Tests

For each of the 11 data sets, we performed a likelihood ratio test that compared a single bivariate normal distribution to a mixture of two bivariate normals. As noted by a reviewer (we thank the reviewer), for nested mixture models, the likelihood ratio statistic will have a non-standard distribution, not the chi-squared distribution with 6 degrees of freedom that one might expect. (See, for example, McLachlan and Rathnayake (2014).) In this case, one can perform a parametric bootstrap. (See Efron and Hastie (2016) or Davison and Hinkley (1997) for a general discussion of parametric bootstraps, and see, for example, the "Resampling Approach" section in McLachlan and Rathnayake (2014) for a discussion of parametric bootstraps when comparing mixed normal distributions). We performed 1,000 trial parametric bootstraps for each of the 11 data sets. The resulting p-values are reported in column 3 of Table 3. Low p-values reject a single bivariate normal distribution.

In column 4 of Table 3, we provide the p-value from a test of normality for the MOR values of the data set. In column 5 of Table 3 we provide the p-value from a test of normality for the MOE values of the data set. These p-values were calculated via the shapiro.test function of the R statistical package (R Core Team, 2018).

The p-values reported in columns 3-5 of Table 3 suggest that, in general, a model that differs from a single bivariate normal is required. Given the p-values in columns 4 and 5, the only case in which a single bivariate normal might be appropriate for the MOE–MOR distribution appears to be the Mill 1, winter case. Listings of the programs that were used to perform the likelihood ratio tests can be found at `https://www1.fpl.fs.fed.us/mixbivn.2019.html`.

7 Chi-Squared Goodness-of-Fit Tests

Having rejected a single bivariate normal model, we next wanted to test whether a mixture of two bivariate normals might yield a good fit. To do so we implemented a "chi-squared" goodness-of-fit test for a mixture of bivariate normals. Chi-squared goodness-of-fit tests are described in many statistical textbooks. Moore (1986) provides a detailed description of such tests and suggests that an appropriate number of "cells" lies between $1.88 \times n^{2/5}$ and $3.76 \times n^{2/5}$ where n is the sample size. For a sample of size 200, this suggests a number of cells between 16 and 32. We chose to work with 20 cells, each of which (under the null hypothesis of a mixture of two bivariate normal distributions) contained about 0.05 of the probability. That is, the expected number of observations in each cell was approximately $200 \times 0.05 = 10$. (Moore (1986) recommends that cells should be of approximately equal probability.)

We followed several steps to perform the test:

1. Obtain a maximum likelihood (ML) fit of a mixture of two bivariate normals to the set of 200 stiffness–strength data pairs.

2. Choose a rectangular region (e.g., 0 to 3 by 0 to 15 for a stiffness–strength region [MOE values divided by 1,000,000, MOR values divided by 1,000]) that contains essentially all the probability (e.g., 0.997 and above).

3. Divide this region into 1,000,000 (1,000 by 1,000) rectangles.

4. Take as the estimate of the probability of one of these small rectangles,

   $$\text{(pdf at center of small rectangle)} \times \text{(area of small rectangle)}$$

   (This amounts to numerical integration.)

5. Use the small rectangles to divide the large rectangle into J columns, each of which contains approximately $1/J$ of the probability.

6. Use the small rectangles to divide each of the J columns into I rows (where $J \times I = 20$), each of which contains approximately 0.05 of the probability. The exact probability associated with a cell will be the sum of all the probabilities associated with the small rectangles contained in the cell. For the ijth cell, $j \in \{1, \ldots, J\}$, $i \in \{1, \ldots, I\}$, denote this probability by p_{ij}.

7. Take the chi-squared statistic to be

   $$\chi^2 = \sum_{j=1}^{J} \sum_{i=1}^{I} (O_{ij} - E_{ij})^2 / E_{ij}$$

where O_{ij} is the observed number of stiffness–strength pairs in the ijth cell and $E_{ij} = p_{ij} \times N$ where N is the number of observations in the data set.

8. Because we used full data maximum likelihood techniques rather than grouped data maximum likelihood techniques to estimate the parameters of the mixture of bivariate normals, the "chi-squared" statistic is not actually distributed as a chi-squared random variable (see section 3.2.2.1 in Moore (1986)). Thus, we needed to do a "parametric bootstrap" (a kind of simulation) to obtain estimates of the p-values associated with the statistic. For each of the chi-squared tests, we performed 1000 simulations. In each of these simulations, we generated a sample of size N from the fitted (to the original stiffness–strength data) mixture of bivariate normals. For each of these simulated data sets, we calculated a chi-squared statistic by the same method used to calculate the chi-squared statistic for the original data. This resulted in the original chi-squared statistic based on the original data, and 1000 chi-squared statistics based on 1000 samples of size N drawn from the mixed bivariate normal distribution that was fit to the original data.

9. From these simulations it is possible to obtain an approximate p-value. If we order the 1000 simulated chi-squared values from smallest to largest, and the original chi-squared value calculated from the real data lies between the mth and $(m + 1)$th of the 1000 ordered chi-squared values, then the approximate p-value is $(1000 - m)/1000$. Thus, for example, if the original chi-squared statistic lies below only 100 of the 1000 simulated chi-squared values, we would say that the approximate p-value is 0.10. If the original chi-squared statistic lies below only 50 of the 1000 simulated chi-squared values, we would say that the approximate p-value is 0.05.

Verrill et al. (2018) considered four J, I (column, row) cases — $J = 5, I = 4$; $J = 4, I = 5$; $J = 10, I = 2$; and $J = 2, I = 10$. They found that their conclusions did not depend on the I, J that were used. We use $J = 5, I = 4$ here.

Listings of the programs that we wrote to perform the chi-squared tests and simulations can be found at `http://www1.fpl.fs.fed.us/mixbivn.2019.html`.

Column 6 of Table 3 contains the approximate p-values for the 11 chi-squared goodness-of-fit tests. The p-values ranged from 0.057 to 0.763.

Taken together, the plots and the chi-squared tests suggest that we cannot reject the hypothesis that the MOE–MOR values measured for the 11 data sets are drawn from mixtures of two bivariate normals. Having said this, we realize that a mixture of bivariate normals model *cannot* be completely correct because it predicts non-zero probabilities for negative stiffness and strength values. Also, of course, we do not expect that mixtures will always have exactly two components (as opposed to one or three or . . .).

8 Tests of the Equality of the Parameter Fits

We are interested in whether the fitted mixture of two bivariate normals changes with mill and/or time. We performed 26 comparisons. These comparisons are listed in Tables 4a and 4b.

The statistical test that we used to perform the comparisons is discussed in Appendix C. Listings of the programs that we wrote to perform the comparisons are provided at `https://www1.fpl.fs.fed.us/mixbivn.2019.html`.

The test is simulation-based and results will depend on which of the two time-mill combinations being compared provides the parameter estimates being used as the basis for the simulation. (See Appendix C.) Taking the p-value for a comparison — a row in Table 4a or 4b — to be the larger of the two values in columns 4 and 5 of the row, we find that 10 of the 16 comparisons in Table 4a (complete data) are statistically significant at a 0.05 level, while 7 of the 10 comparisons in Table 4b (visual outliers removed) are statistically significant at a 0.05 level.

A careful look at appropriate pairs of plots from among Figures 1-24 suggest that this statistical significance is unsurprising. To facilitate the comparisons of the figures we have produced a web page, `https://www1.fpl.fs.fed.us/4x2.mix.plots.html`, that permits one to readily compare relevant figures.

9 The Variability of Probability of Breakage at a Fixed Load as a Function of Mill and Time

Given the discussions and analyses provided above, we claim that a mixture of two bivariate normals model for MOE–MOR makes physical sense (at least in some cases) and does a good job of fitting our eight mill run populations.

However, as noted in Section 8, we have also found that although a mixture of two bivariate normals *model* is not rejected visually or via chi-sqared tests, the particular *fits* (parameter estimates) differ (at least statistically) from mill to mill and time to time.

In this section we discuss the fact that these statistically detectable differences in parameter estimates have practical implications. In particular, we have found that because the fitted model parameter values (see Table 2) vary from mill to mill and time to time, the estimated probabilities of breakage for pieces of lumber randomly drawn from a fixed "grade" when the load is at a fixed value can vary widely from mill to mill and time to time.

Given a fit of a mixture of two bivariate normals to a mill run MOE–MOR sample, one can calculate the MOR distribution obtained when we restrict ourselves to a truncated range of MOE values. For this study, the truncating MOE values, 1.311 and 1.772, were approximate 40th and 80th percentiles of the 1591 MOE mill run samples. This excludes two missing values and seven "outliers." The results should not change much if we were to include the "outliers." As noted in the Introduction, we refer to the resulting MOR distribution as a pseudo-truncated distribution. The density function of a pseudo-truncated mixture of two bivariate normals is derived in appendix B of Verrill et al. (2018).

In Table 5 we report estimated probabilities of breakage for the 11 data sets (columns 4 through 14) at 11 fixed loads (column 3). The fixed load in row i of column 3 corresponds to the calculated 5th percentile/2.1 for the pseudo-truncated MOR distribution fit to the data set in row i, column 2 of the table.

The probabilities of breakage in columns 4 through 14 of Table 5 were obtained from the pseudo-truncated MOR distributions calculated from the eleven fitted mixture models. (The estimated probabilities of breakage were calculated via numerical integrations of the fitted pseudo-truncated probability density functions — see equation 18 in Verrill et al. (2018) and the listing of test.allow3.11.f, which is available at `http://www1.fpl.fs.fed.us/mixbivn.2019.html`.)

Table 5 indicates that the breakage probability (reported in one of columns 4 through 14 of the table) associated with a fixed load (loads are specified in column 3 of the table) and one of the 11 fitted pseudo-truncated MOR distributions (corresponding to one of columns 4 through 14 of the table) can vary considerably with the fit.

The breakage probabilities indicate that mills can be different for a fixed season. Compare, for example, summer mills 2 and 4 or winter mills 1 and 3. Comparing winter mill 3 with winter mill 1 yields breakage probability ratios as large as 15.

The breakage probabilities also indicate that seasons can be different for a fixed mill. Compare summer and winter breakage probabilities for mill 1, or summer and winter breakage probabilities for mill 4.

Combining mill and season changes can yield even larger differences. Comparing summer mill 4 with winter mill 1 yields breakage probability ratios as large as 35.

These differences suggest that for a fixed load, probabilities of breakage can vary significantly from mill to mill and time to time.

To obtain estimates of the uncertainty in the "35" and "15" probability of breakage ratios mentioned above, we performed bootstraps (a type of simulation).

In the "35" case, we drew (with replacement) 100 samples of size 200 from the summer mill 4 data set and 100 independent samples of size 200 from the winter mill 1 data set. For each of the 100 pairs of simulated samples, we fit separate mixtures of two bivariate normals to the summer and winter samples. From each of these 100 pairs of fits, we then obtained estimates at shared loads (column 3 of Table 5) of probabilities of breakage corresponding to the summer mill 4 fits and of probabilities of breakage corresponding to the winter mill 1 fits. This permitted us to calculate 100 probability of breakage ratios at the various loads (column 3 of Table 5). In column 4 of Table 6 we provide the probability of breakage ratios for the original data sets. In column 5 we provide the 5th order statistic of the 100 simulated ratios. In column 6 we provide the median of the 100 simulated ratios. In column 7 we provide the 95th order statistic of the 100 simulated ratios. These results suggest that in the "35" case, the probability of breakage ratios are both statistically and practically different from 1.

Our corresponding "15" case results appear in Table 7 and again suggest that the probability of breakage ratios in this case are statistically and practically different from 1.

Listings of the computer programs that were used to obtain Tables 5 through 7 are available at `https://www1.fpl.fs.fed.us/mixbivn.2019.html`.

10 Variability of Probability of Breakage Under a Lognormal Load Distribution as a Function of Mill and Time

We extended the work discussed in Section 9 by replacing fixed loads with load distributions. A reviewer of Verrill *et al.* (2013) suggested that it would be reasonable to model the load as a lognormal with coefficient of variation 0.30 and probability equal to 0.02 of exceeding a 5th/2.1

strength distribution value (e.g., the values in column 3 of Tables 5 through 10). (Obviously, one could consider other load distribution forms and other values for the load distribution parameters.)

Tables 8 through 10 correspond to Tables 5 through 7 with lognormal load distributions replacing fixed loads.

Probability of breakage values clearly decline in Tables 8 through 10 in comparison to Tables 5 through 7 as we would expect when we go from loads *fixed* at 5th/2.1 values to lognormally distributed loads that can approach zero and that only exceed 5th/2.1 values with probability 0.02. However, Tables 8 through 10 demonstrate that, again, probability of breakage values can vary widely from mill to mill and time to time.

In Table 8 we see summer mill 4 to winter mill 1 ratios on the order of 50, and winter mill 3 to winter mill 1 ratios on the order of 20.

Tables 9 and 10, like Tables 6 and 7, are based on bootstrapping. They demonstrate that in the lognormal load case, as in the fixed load case, probability of breakage ratios for different mills and times are statistically and practically different from 1.

Listings of the computer programs that were used to obtain Tables 8 through 10 are available at `https://www1.fpl.fs.fed.us/mixbivn.2019.html`.

11 Summary

We obtained 11 mill run MOE–MOR lumber data sets by collecting mill run samples of lumber from four mills at each of two times (roughly "summer" and "winter"). The extra three data sets correspond to cases in which we have removed "outliers."

We found via plots and statistical tests that these data sets appear to be well fit by mixtures of bivariate normals.

For each of these data sets, we calculated the MOR distribution of the corresponding "grade" consisting of only those specimens for which the MOE lies between bounds that approximate the 40th and 80th percentiles of our combined sample of 1591 specimens. The calculated MOR distributions for these "grades" are pseudo-truncated mixtures of two normals.

We used these fitted pseudo-truncated MOR distributions to estimate the probabilities of breakage associated with the 11 data sets under fixed loads and under lognormal load distributions. We found that the probabilities of breakage at a fixed load can vary by a factor as large as 35 when we permit both season and mill to vary, and as large as 15 when we only permit mill to vary. Under a lognormally distributed load (rather than a fixed load), factors as large as 50 were found when we permitted both mill and time to vary, and as large as 20 when we permitted only mill to vary. Very roughly, lower 90% confidence bounds on the 35, 15, 50, and 20 factors are 7, 3.2, 10, and 3.6 — see Tables 6, 7, 9, and 10. These estimates also depend on the mixture of two bivariate normals model.

This suggests that a purely MOE-based grading system might yield lumber that is more variable than we would desire. Of course, this does *not* directly apply to MSR lumber, which also includes visual grade restrictions and quality control procedures. We also emphasize that we are *not* suggesting that lumber that has been visually graded is less variable than lumber that has been graded via an MSR process. We have not yet developed the statistical tools needed to obtain theoretical estimates of the mill to mill and time to time variabilities of breakage probabilities for either MSR or visual grades of lumber.

We do argue that our variability results for a set of purely MOE-based "grades" suggest that there may be significant efficiencies that can be obtained through the development of computer models that yield real-time in-line estimates of lumber properties based on measurements of stiffness, specific gravity, knot size and location, slope of grain, and other strength predictors.

12 References

Anderson, G.C., Owens, F.C., Shmulsky, R., and Ross, R.J. (2019), "Within-mill Variation in the Means and Variance of MOE and MOR of Mill-run Lumber Over Time," *Wood and Fiber Science*, **51**(4), pp. 387-401.

ASCE (1988). Load and Resistance Factor Design for Engineered Wood Construction: A Pre-Standard Report, prepared by the Task Committee on Load and Resistance Factor Design for Engineered Wood Construction, American Society of Civil Engineers, New York, NY, 177 pages.

Davison, A.C. and Hinkley, D.V. (1997), *Bootstrap Methods and Their Application*, Cambridge University Press, Cambridge, UK, 582 pages.

Efron, B. and Hastie, T. (2016), *Computer Age Statistical Inference*, Cambridge University Press, Cambridge, UK, 475 pages.

McLachlan, G.J. and Rathnayake, S. (2014), "On the number of components in a Gaussian mixture model," *WIREs Data Mining Knowl Discov*, **4**, pp. 341-355.

Moore, D.S. (1986). Tests of Chi-Squared Type. In: *Goodness-of-Fit Techniques.* Edited by R.B. D'Agostino and M.A. Stephens. Marcel Dekker, Inc. New York, pages 63-96.

Owens, F.C, Verrill, S.P., Shmulsky, R., and Kretschmann, D.E. (2018), "Distributions of MOE and MOR in a Full Lumber Population," *Wood and Fiber Science*, **50**(3), pp. 265-279.

Owens, F.C, Verrill, S.P., Shmulsky, R., and Ross, R.J. (2019), "Distributions of Modulus of Elasticity and Modulus of Rupture in Four Mill-Run Lumber Populations," *Wood and Fiber Science*, **51**(2), pp. 183-192.

Owens, F.C, Verrill, S.P., Shmulsky, R., and Ross, R.J. (2020), "Distributions of Moduli of Elasticity and Rupture in Eight Mill-Run Lumber Populations (Four Mills at Two Times)," to appear in *Wood and Fiber Science.*

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Verrill, S.P., Evans, J.W., Kretschmann, D.E., and Hatfield, C.A. (2012). “Asymptotically Efficient Estimation of a Bivariate Gaussian–Weibull Distribution and an Introduction to the Pseudo-truncated Weibull.” Research Paper FPL-RP-666. Madison, WI: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory. 76 pages.

Verrill, S.P., Evans, J.W., Kretschmann, D.E., and Hatfield, C.A. (2013). “An Evaluation of a Proposed Revision of the ASTM D 1990 Grouping Procedure.” Research Paper FPL-RP-671. Madison, WI: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory. 34 pages.

Verrill, S.P., Evans, J.W., Kretschmann, D.E., and Hatfield, C.A. (2014), “Reliability Implications in Wood Systems of a Bivariate Gaussian–Weibull Distribution and the Associated Univariate Pseudo-truncated Weibull,” *ASTM Journal of Testing and Evaluation*, **42**, Number 2, pp. 412-419.

Verrill, S.P., Evans, J.W., Kretschmann, D.E., and Hatfield, C.A. (2015), “Asymptotically Efficient Estimation of a Bivariate Gaussian–Weibull Distribution and an Introduction to the Pseudo-truncated Weibull,” *Communications in Statistics – Theory and Methods*, **44**, pp. 2957-2975.

Verrill, S.P., Owens, F.C., Kretschmann, D.E., and Shmulsky, R. (2017). “Statistical Models for the Distribution of Modulus of Elasticity and Modulus of Rupture in Lumber with Implications for Reliability Calculations.” Research Paper FPL-RP-692. Madison, WI: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory. 51 pages.

Verrill, S.P., Owens, F.C., Kretschmann, D.E., and Shmulsky, R. (2018). “A Fit of a Mixture of Bivariate Normals to Lumber Stiffness–Strength Data.” Research Paper FPL-RP-696. Madison, WI: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory. 44 pages.

Verrill, S.P., Owens, F.C., Kretschmann, D.E., Shmulsky, R., and Brown, L.S. (2019). “Visual and MSR Grades of Lumber Are Not 2-Parameter Weibulls and Why It Matters (With a Discussion of Censored Data Fitting).” Research Paper FPL-RP-703. Madison, WI: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory. 40 pages.

Verrill, S.P., Owens, F.C., Kretschmann, D.E., Shmulsky, R., and Brown, L.S. (2020). “Visual and MSR Grades of Lumber Are Not 2-Parameter Weibulls and Why This May Matter,” *ASTM Journal of Testing and Evaluation*. In press. Web. 24 April 2019. `https://www.astm.org/DIGITAL_LIBRARY/JOURNALS/TESTEVAL/PAGES/JTE20180472.htm`

13 Appendix A — Checking for "Outliers"

We identified seven subjective visual outliers. They are listed in Table 1. (Also, see Figures 13, 15, and 19.) We checked these subjective judgments with quantitative tests. A test (see `http://www1.fpl.fs.fed.us/mixbivn.2019.html` for a listing of the programs used to perform outlier tests) consisted of the following steps:

1. Use maximum likelihood methods to fit a mixture of two bivariate normals to the data set (of size n).
2. Use the fit to estimate the probability p of obtaining (on any given draw) a data point with probability density function (pdf) as small as or smaller than the observed pdf at the visual outlier.
3. Then, if there were n observations in the original data set, calculate the probability of seeing a data point with pdf value as small as or smaller than the pdf at the visual outlier as $1-(1-p)^n$. If this value is 0.10 or smaller, report this "p-value" in column six of Table 1. If it is greater than 0.10, then report the value as "NS" (for "not significant") in Table 1.

The results reported in Table 1 suggest that we have no statistical justification for deleting any data points from the summer mill 3 data set. However, it might be reasonable to consider the summer mill 2 and winter mill 2 visual outliers as true outliers. In Tables 2 – 5 we provide results with and without the (possible) outliers for all three cases (the summer mill 2, summer mill 3, and winter mill 2 cases). The summer mill 2 and winter mill 2 outlier results should be granted more credibility than the summer mill 3 outlier results.

14 Appendix B — Probability Density Functions of a Mixture of Two Bivariate Normal Distributions

The probability density function of a mixture of two bivariate normal distributions has the form

$$\begin{aligned} &f_{\mathrm{M}}(x, y; \mu_{X1}, \sigma_{X1}, \rho_1, \mu_{Y1}, \sigma_{Y1}, \mu_{X2}, \sigma_{X2}, \rho_2, \mu_{Y2}, \sigma_{Y2}, p) \\ &= p \times f(x, y; \mu_{X1}, \sigma_{X1}, \rho_1, \mu_{Y1}, \sigma_{Y1}) + (1-p) \times f(x, y; \mu_{X2}, \sigma_{X2}, \rho_2, \mu_{Y2}, \sigma_{Y2}) \end{aligned} \tag{1}$$

where p is the mixing fraction (at any given "draw", we draw from population 1 with probability p and from population 2 with probability $1-p$); $\mu_{X1}, \sigma_{X1}, \rho_1, \mu_{Y1}, \sigma_{Y1}$ are the parameters of the population 1 bivariate normal; $\mu_{X2}, \sigma_{X2}, \rho_2, \mu_{Y2}, \sigma_{Y2}$ are the parameters of the population 2 bivariate normal; and $f(w, z; \mu_W, \sigma_W, \rho, \mu_Z, \sigma_Z)$ is the pdf of a single bivariate normal given by

$$f(w, z; \mu_W, \sigma_W, \rho, \mu_W, \sigma_Z) = \frac{1}{2\pi} \times \frac{1}{\sigma_W \sigma_Z \sqrt{1-\rho^2}} \times \exp\left(-\mathrm{arg}\right) \tag{2}$$

where

$$\mathrm{arg} = \left[\left(\frac{w-\mu_W}{\sigma_W}\right)^2 - 2\rho\left(\frac{w-\mu_W}{\sigma_W}\right)\left(\frac{z-\mu_Z}{\sigma_Z}\right) + \left(\frac{z-\mu_Z}{\sigma_Z}\right)^2\right] \div \left(2(1-\rho^2)\right)$$

and μ_W, σ_W are the mean and standard deviation of W; ρ is the correlation between W and Z; and μ_Z, σ_Z are the mean and standard deviation of Z.

15 Appendix C — Test of the Equality of Two Mixtures of Bivariate Normals

In Section 6 we discuss the results of tests that compared mixtures of two bivariate normals. Here we briefly outline the nature of these tests. Details can be found in the FORTRAN programs that implemented the tests. These can be found at `http://www1.fpl.fs.fed.us/mixbivn.2019.html`.

The tests are based on maximum likelihood fits of a mixture of two bivariate normals to MOE–MOR data. From maximum likelihood theory we know that (in the material below, a non-statistician can read $\stackrel{D}{\to}$ as $\approx$)

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \stackrel{D}{\to} N(\mathbf{0}, \boldsymbol{\Sigma}) \tag{3}$$

where, in the notation of Appendix B, $N(\mathbf{0}, \boldsymbol{\Sigma})$ denotes a multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$,

$$\boldsymbol{\beta}^T = (\mu_{X1}, \sigma_{X1}, \rho_1, \mu_{Y1}, \sigma_{Y1}, \mu_{X2}, \sigma_{X2}, \rho_2, \mu_{Y2}, \sigma_{Y2}, p),$$

$\hat{\boldsymbol{\beta}}$ denotes the maximum likelihood estimate of $\boldsymbol{\beta}$, the covariance matrix $\boldsymbol{\Sigma}$ is equal to the inverse of the information matrix associated with the maximum likelihood estimation, and the information matrix is approximated by

$$\sum_{i=1}^{n} - \left[\frac{\partial^2 \log(f_M(x_i, y_i; \hat{\boldsymbol{\beta}}))}{\partial \beta_j \partial \beta_k} \right] \Big/ n \tag{4}$$

Now, suppose that we want to compare mill–season combination 1 with mill–season combination 2.

We have

$$\sqrt{n_1}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) \stackrel{D}{\to} N(\mathbf{0}, \boldsymbol{\Sigma}_1) \tag{5}$$

and

$$\sqrt{n_2}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2) \stackrel{D}{\to} N(\mathbf{0}, \boldsymbol{\Sigma}_2) \tag{6}$$

Thus,

$$\sqrt{n_2}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) \stackrel{D}{\to} N(\mathbf{0}, (n_2/n_1)\boldsymbol{\Sigma}_1) \tag{7}$$

and combining results (6) and (7) we have

$$\sqrt{n_2}(\hat{\boldsymbol{\beta}}_2 - \hat{\boldsymbol{\beta}}_1 - (\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1)) \stackrel{D}{\to} N(\mathbf{0}, (n_2/n_1)\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2) \tag{8}$$

Under the null hypothesis that $\boldsymbol{\beta}_2 = \boldsymbol{\beta}_1$, this yields

$$\sqrt{n_2}(\hat{\boldsymbol{\beta}}_2 - \hat{\boldsymbol{\beta}}_1) \stackrel{D}{\to} N(\mathbf{0}, (n_2/n_1)\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2) \tag{9}$$

or

$$\sqrt{n_2}((n_2/n_1)\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2)^{-1/2}(\hat{\boldsymbol{\beta}}_2 - \hat{\boldsymbol{\beta}}_1) \xrightarrow{D} N(\mathbf{0}, \mathbf{I}_{11}) \tag{10}$$

or

$$S = n_2 \times (\hat{\boldsymbol{\beta}}_2 - \hat{\boldsymbol{\beta}}_1)^T((n_2/n_1)\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2)^{-1}(\hat{\boldsymbol{\beta}}_2 - \hat{\boldsymbol{\beta}}_1) \xrightarrow{D} \chi^2_{11} \tag{11}$$

That is, under the null hypothesis of no difference between mill–season combinations, the statistic S on the left of equation (11) should have approximately the distribution of a chi-squared random variable with 11 degrees of freedom. One can show that if $\boldsymbol{\beta}_2 \neq \boldsymbol{\beta}_1$, S will tend to be inflated.

The p-values that were initially reported by our software are the probabilities that a chi-squared random variable with 11 degrees of freedom would be as large as or larger than the observed S (from equation (11) with the $\mathbf{\Sigma}$ values replaced by approximations based on equation (4)). However, these p-values were so low that we suspected that we might have made a programming error or errors.

We found instead that the "asymptotics" that justify equation (11) had not yet kicked in. That is, for the sample sizes that were available to us ($n \approx 200$) the theoretical estimates of $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ were too "small" and the resulting estimates of S in equation (11) were too large. (In a simulation this problem did disappear as sample sizes became sufficiently large (between 1000 and 2000 in our case).)

Thus, rather than relying on asymptotic theory for the distribution of S in equation (1), we had to perform a "parametric bootstrap" (a form of simulation) to obtain correct p-values. These simulation-based p-values are reported in columns 4 and 5 of Tables 4a and 4b. The programs that were used to perform the simulations are available at
`http://www1.fpl.fs.fed.us/mixbivn.2019.html`. (See the summ*f, wint*f, and mill_compare*f files.) The simulation results differ somewhat (as we would expect) depending on whether the "generating" parameter vector comes from the data set corresponding to the second column in Tables 4a and 4b or the data set corresponding to the the third column in Tables 4a and 4b. Thus two distinct p-values are reported in columns 4 and 5 of Tables 4a and 4b. A more conservative conclusion about mill and time differences can be obtained by taking the larger of the columns 4 and 5 p-values as the estimated p-value.

Season	Mill	n	MOE	MOR	p-value
Summer	2	199	2.669	5.551	0.04
			2.818	10.148	NS
			2.817	9.511	NS
Summer	3	200	2.733	16.783	NS
			2.662	16.107	NS
			2.800	16.098	NS
Winter	2	200	2.746	8.404	0.10

Table 1: Visual outliers. The MOE, MOR values identify the visual outlier data points. The p-values are the probabilities that a probability density function (pdf) value as low as or lower than the observed pdf value at the visual outlier would occur in a sample of size n. “NS” indicates a value larger than 0.10. The MOE values used in the analyses are in millions of pounds per square inch. The MOR values used in the analyses are in thousands of pounds per square inch.

		“Leftmost” population					“Rightmost” population					
	Mill	$\hat{\mu}_{\mathrm{MOE},1}$	$\hat{\sigma}_{\mathrm{MOE},1}$	$\hat{\mu}_{\mathrm{MOR},1}$	$\hat{\sigma}_{\mathrm{MOR},1}$	$\hat{\rho}_1$	$\hat{\mu}_{\mathrm{MOE},2}$	$\hat{\sigma}_{\mathrm{MOE},2}$	$\hat{\mu}_{\mathrm{MOR},2}$	$\hat{\sigma}_{\mathrm{MOR},2}$	$\hat{\rho}_2$	$\hat{p}$
	1, all 200	1.26	.28	6.68	2.40	.54	1.63	.32	9.29	1.35	.79	.554
	2, all 199	1.15	.23	5.20	1.61	.72	1.64	.46	8.20	2.90	.66	.687
S	2, 198	1.13	.22	5.11	1.55	.70	1.57	.44	7.89	2.88	.75	.629
	3, all 200	1.45	.30	7.47	2.79	.74	2.30	.33	12.49	2.32	.27	.901
	3, 195	1.33	.23	5.93	2.15	.61	1.72	.34	10.07	2.16	.70	.543
	4, all 200	1.45	.32	7.28	2.86	.57	1.98	.38	11.86	2.83	.86	.755
	1, all 200	1.44	.35	7.18	2.88	.85	1.68	.40	9.92	2.18	.53	.766
	2, all 200	1.28	.27	5.90	1.94	.72	2.03	.39	11.82	1.46	.15	.921
W	2, 199	1.29	.28	5.91	1.95	.72	1.97	.37	12.20	1.00	.70	.929
	3, all 200	1.12	.26	5.50	2.24	.73	1.73	.35	8.15	3.08	.81	.572
	4, all 199	1.55	.43	7.13	3.05	.83	1.66	.30	10.26	1.38	.67	.665

Table 2: Maximimum likelihood parameter estimates for MOE–MOR data sets fitted as mixtures of two bivariate normals. “S” denotes summer. “W” denotes winter. The summer, Mill 2, all 199 data set had one missing data point. The summer, Mill2, 198 data set had one missing data point and one “outlier.” The summer, Mill 3, 195 data set had five “outliers.” The winter, Mill 2, 199 data set had one “outlier.” The winter, Mill 4, all 199 data set had one missing data point. $\hat{p}$ is our estimate of the proportion of specimens that come from the leftmost bivariate normal population. The MOE values used in the analyses are in millions of pounds per square inch. The MOR values used in the analyses are in thousands of pounds per square inch.

		Likelihood ratio test p-values	MOR S-W test p-values	MOE S-W test p-values	Chi-squared test p-values
Summer	Mill 1, all 200	.005	.001	.371	.140
	Mill 2, all 199	.004	.001	.001	.763
	Mill 2, 198	.006	.001	.001	.736
	Mill 3, all 200	.004	.104	.001	.469
	Mill 3, 195	.213	.004	.126	.057
	Mill 4, all 200	.291	.064	.024	.194
Winter	Mill 1, all 200	.325	.258	.298	.248
	Mill 2, all 200	.006	.001	.001	.316
	Mill 2, 199	.012	.001	.001	.656
	Mill 3, all 200	.137	.001	.001	.519
	Mill 4, all 199	.072	.027	.194	.467

Table 3: Likelihood ratio test — small p-values reject single bivariate normal distributions. S-W test (Shapiro-Wilk test of normality) — small p-values reject univariate normality (for MOR or MOE). Chi-squared test — small p-values reject a mixture of two bivariate normal distributions.

Comparison	First time–mill in the comparison	Second time–mill in the comparison	Simulation-based p-values: The generating $\hat{\boldsymbol{\beta}}$ is from the fit of the first time–mill	Simulation-based p-values: The generating $\hat{\boldsymbol{\beta}}$ is from the fit of the second time–mill
S mill versus W mill	summer, mill 1, all 200	winter, mill 1, all 200	.005	.155
	summer, mill 2, all 199	winter, mill 2, all 200	.009	.015
	summer, mill 3, all 200	winter, mill 3, all 200	.009	.040
	summer, mill 4, all 200	winter, mill 4, all 199	.112	.018
S mill versus S mill	summer, mill 1, all 200	summer, mill 2, all 199	<.001	.001
	summer, mill 1, all 200	summer, mill 3, all 200	.001	.023
	summer, mill 1, all 200	summer, mill 4, all 200	<.001	.050
	summer, mill 2, all 199	summer, mill 3, all 200	<.001	<.001
	summer, mill 2, all 199	summer, mill 4, all 200	<.001	.011
	summer, mill 3, all 200	summer, mill 4, all 200	.237	.352
W mill versus W mill	winter, mill 1, all 200	winter, mill 2, all 200	.221	.029
	winter, mill 1, all 200	winter, mill 3, all 200	.097	.049
	winter, mill 1, all 200	winter, mill 4, all 199	.246	.060
	winter, mill 2, all 200	winter, mill 3, all 200	.007	.046
	winter, mill 2, all 200	winter, mill 4, all 199	.001	.003
	winter, mill 3, all 200	winter, mill 4, all 199	.009	<.001

Table 4a: Simulation-based fit comparisons. "Outliers" present in some cases. Low p-values indicate a statistical difference between the fits of the two time–mill cases being compared. S – summer. W – winter.

Comparison	First time–mill in the comparison	Second time–mill in the comparison	Simulation-based p-values: The generating $\hat{\boldsymbol{\beta}}$ is from the fit of the first time–mill	Simulation-based p-values: The generating $\hat{\boldsymbol{\beta}}$ is from the fit of the second time–mill
S mill versus W mill	summer, mill 2, 198	winter, mill 2, 199	<.001	<.001
	summer, mill 3, 195	winter, mill 3, all 200	<.001	.055
S mill versus S mill	summer, mill 1, all 200	summer, mill 2, 198	<.001	.001
	summer, mill 1, all 200	summer, mill 3, 195	<.001	<.001
	summer, mill 2, 198	summer, mill 3, 195	.004	<.001
	summer, mill 2, 198	summer, mill 4, all 200	<.001	.007
	summer, mill 3, 195	summer, mill 4, all 200	.081	.276
W mill versus W mill	winter, mill 1, all 200	winter, mill 2, 199	.160	.002
	winter, mill 2, 199	winter, mill 3, all 200	<.001	.006
	winter, mill 2, 199	winter, mill 4, all 199	<.001	.005

Table 4b: Simulation-based fit comparisons. "Outliers" removed. Low p-values indicate a statistical difference between the fits of the two time–mill cases being compared. S – summer. W – winter.

Source of 5th/2.1			Probability that MOR lies below 5th/2.1										
			Summer						Winter				
			M1	M2	M2	M3	M3	M4	M1	M2	M2	M3	M4
Season	Mill	5th/2.1	all	all	198	all	195	all	all	all	199	all	all
S	1, all 200	2.413	.0027	.0035	.0022	.0032	.0044	.0127	.0006	.0009	.0009	.0076	.0036
	2, all 199	2.138	.0018	.0024	.0015	.0021	.0028	.0094	.0004	.0005	.0004	.0052	.0023
	2, 198	2.168	.0019	.0025	.0015	.0022	.0030	.0097	.0004	.0005	.0005	.0054	.0024
	3, all 200	2.188	.0020	.0026	.0016	.0023	.0031	.0099	.0004	.0005	.0005	.0056	.0025
	3, 195	2.059	.0016	.0022	.0013	.0019	.0025	.0086	.0003	.0004	.0004	.0047	.0020
	4, all 200	1.841	.0012	.0016	.0009	.0013	.0017	.0067	.0002	.0002	.0002	.0034	.0014
W	1, all 200	2.431	.0028	.0036	.0023	.0033	.0045	.0129	.0007	.0010	.0009	.0078	.0037
	2, all 200	2.158	.0019	.0025	.0015	.0022	.0029	.0096	.0004	.0005	.0005	.0054	.0024
	2, 199	2.167	.0019	.0025	.0015	.0022	.0030	.0097	.0004	.0005	.0005	.0054	.0024
	3, all 200	1.915	.0013	.0018	.0010	.0015	.0020	.0073	.0002	.0003	.0002	.0038	.0016
	4, all 199	2.140	.0019	.0025	.0015	.0021	.0028	.0094	.0004	.0005	.0004	.0052	.0023

Table 5: Probabilities of breakage at fixed loads (the 5th/2.1 values in column 3) for the 11 cases (4 mills $\times$ 2 seasons, all data cases and three cases in which “outliers” have been removed). S – Summer, W – Winter. M1 – Mill 1, M2 – Mill 2, M3 – Mill 3, M4 – Mill 4.

Source of 5th/2.1			p_{Br}(S,M4) / p_{Br}(W,M1)			
				Bootstrap		
Season	Mill	5th/2.1	Original	5%	median	95%
S	1, all 200	2.413	21	6.3	21	93
	2, all 199	2.138	23	7.6	28	141
	2, 198	2.168	24	7.4	28	135
	3, all 200	2.188	24	7.3	27	131
	3, 195	2.059	28	8.0	31	160
	4, all 200	1.841	33	9.3	40	228
W	1, all 200	2.431	18	6.2	21	90
	2, all 200	2.158	24	7.5	28	137
	2, 199	2.167	24	7.4	28	135
	3, all 200	1.915	36	8.8	37	202
	4, all 199	2.140	23	7.6	28	141

Table 6: Bootstrap results for probability of breakage ratios at fixed loads (the 5th/2.1 values in column 3) for the 11 cases (4 mills × 2 seasons, all data cases and three cases in which "outliers" have been removed). S – Summer, W – Winter. M1 – Mill 1, M4 – Mill 4. 100 trials.

Source of 5th/2.1			p_{Br}(W,M3) / p_{Br}(W,M1)			
				Bootstrap		
Season	Mill	5th/2.1	Original	5%	median	95%
S	1, all 200	2.413	12.7	3.2	13.4	66
	2, all 199	2.138	13.0	3.4	16.7	94
	2, 198	2.168	13.5	3.4	16.3	90
	3, all 200	2.188	14.0	3.4	16.0	88
	3, 195	2.059	15.7	3.4	17.6	104
	4, all 200	1.841	17.0	3.6	20.9	135
W	1, all 200	2.431	11.1	3.1	13.3	65
	2, all 200	2.158	13.5	3.4	16.4	91
	2, 199	2.167	13.5	3.4	16.3	90
	3, all 200	1.915	19.0	3.5	19.8	124
	4, all 199	2.140	13.0	3.4	16.6	93

Table 7: Bootstrap results for probability of breakage ratios at fixed loads (the 5th/2.1 values in column 3) for the 11 cases (4 mills × 2 seasons, all data cases and three cases in which "outliers" have been removed). W – Winter. M1 – Mill 1, M3 – Mill 3. 100 trials.

Source of 5th/2.1			$10^6 \times$ Probability that the MOR lies below the lognormal load										
			Summer						Winter				
			M1	M2	M2	M3	M3	M4	M1	M2	M2	M3	M4
Season	Mill	5th/2.1	all	all	198	all	195	all	all	all	199	all	all
S	1, all 200	2.413	800	1148	613	875	1132	4792	120	140	127	2297	935
	2, all 199	2.138	604	820	417	615	806	3557	75	81	71	1572	623
	2, 198	2.168	624	852	436	641	839	3679	79	86	75	1642	653
	3, all 200	2.188	639	875	450	660	862	3768	82	90	79	1693	675
	3, 195	2.059	549	735	367	546	718	3240	64	70	60	1390	543
	4, all 200	1.841	409	542	252	379	502	2519	39	46	39	986	362
W	1, all 200	2.431	814	1172	627	894	1155	4880	123	146	132	2350	958
	2, all 200	2.158	618	842	430	633	828	3641	77	85	74	1620	643
	2, 199	2.167	624	851	436	640	838	3676	79	86	75	1640	652
	3, all 200	1.915	455	601	288	432	571	2739	47	53	45	1108	417
	4, all 199	2.140	605	822	418	616	808	3565	75	82	71	1576	624

Table 8: 10^6 times the probability that a value randomly drawn from a lognormal load distribution exceeds a value randomly drawn from a fitted MOR distribution. The load distribution is constant for a row of the table. The fitted MOR distribution is constant for a column of the table. All of the lognormal distributions have coefficients of variation equal to 0.30. For a row, the lognormal distribution exceeds the value in column 3 with probability 0.02. The value in column 3 of row i is the estimated 5th percentile/2.1 of the pseudo-truncated mixed normal fit to the ith data set.

The 11 cases correspond to the 4 mills $\times$ 2 seasons, all data cases and the three cases in which "outliers" have been removed. S – Summer, W – Winter. M1 – Mill 1, M2 – Mill 2, M3 – Mill 3, M4 – Mill 4.

Source of 5th/2.1			p_{Br}(S,M4) / p_{Br}(W,M1)			
				Bootstrap		
Season	Mill	5th/2.1	Original	5%	median	95%
S	1, all 200	2.413	40	10	40	224
	2, all 199	2.138	47	11	48	307
	2, 198	2.168	47	10	47	298
	3, all 200	2.188	46	10	46	292
	3, 195	2.059	51	11	51	334
	4, all 200	1.841	65	13	62	443
W	1, all 200	2.431	40	10	40	212
	2, all 200	2.158	47	10	47	301
	2, 199	2.167	47	10	47	298
	3, all 200	1.915	58	12	58	397
	4, all 199	2.140	48	11	48	307

Table 9: Bootstrap results for ratios of the probability that a value randomly drawn from a lognormal load distribution exceeds a value randomly drawn from a fitted MOR distribution. The load distribution is constant for a row of the table. All of the lognormal distributions have coefficients of variation equal to 0.30. For a row, the lognormal distribution exceeds the value in column 3 with probability 0.02. The value in column 3 of row i is the estimated 5th percentile/2.1 of the pseudo-truncated mixed normal fit to the ith data set.

The 11 cases correspond to the 4 mills $\times$ 2 seasons, all data cases and the three cases in which "outliers" have been removed. S – Summer, W – Winter. M1 – Mill 1, M4 – Mill 4. 100 trials.

Source of 5th/2.1			p_{Br}(W,M3) / p_{Br}(W,M1)			
				Bootstrap		
Season	Mill	5th/2.1	Original	5%	median	95%
S	1, all 200	2.413	19	3.6	20	107
	2, all 199	2.138	21	3.7	22	153
	2, 198	2.168	21	3.7	22	149
	3, all 200	2.188	21	3.7	22	147
	3, 195	2.059	22	3.8	23	163
	4, all 200	1.841	25	3.9	25	196
W	1, all 200	2.431	19	3.6	20	102
	2, all 200	2.158	21	3.7	22	151
	2, 199	2.167	21	3.7	22	149
	3, all 200	1.915	24	3.8	24	183
	4, all 199	2.140	21	3.7	22	153

Table 10: Bootstrap results for ratios of the probability that a value randomly drawn from a lognormal load distribution exceeds a value randomly drawn from a fitted MOR distribution. The load distribution is constant for a row of the table. All of the lognormal distributions have coefficients of variation equal to 0.30. For a row, the lognormal distribution exceeds the value in column 3 with probability 0.02. The value in column 3 of row i is the estimated 5th percentile/2.1 of the pseudo-truncated mixed normal fit to the ith data set.

The 11 cases correspond to the 4 mills $\times$ 2 seasons, all data cases and the three cases in which “outliers” have been removed. W – Winter. M1 – Mill 1, M3 – Mill 3. 100 trials.

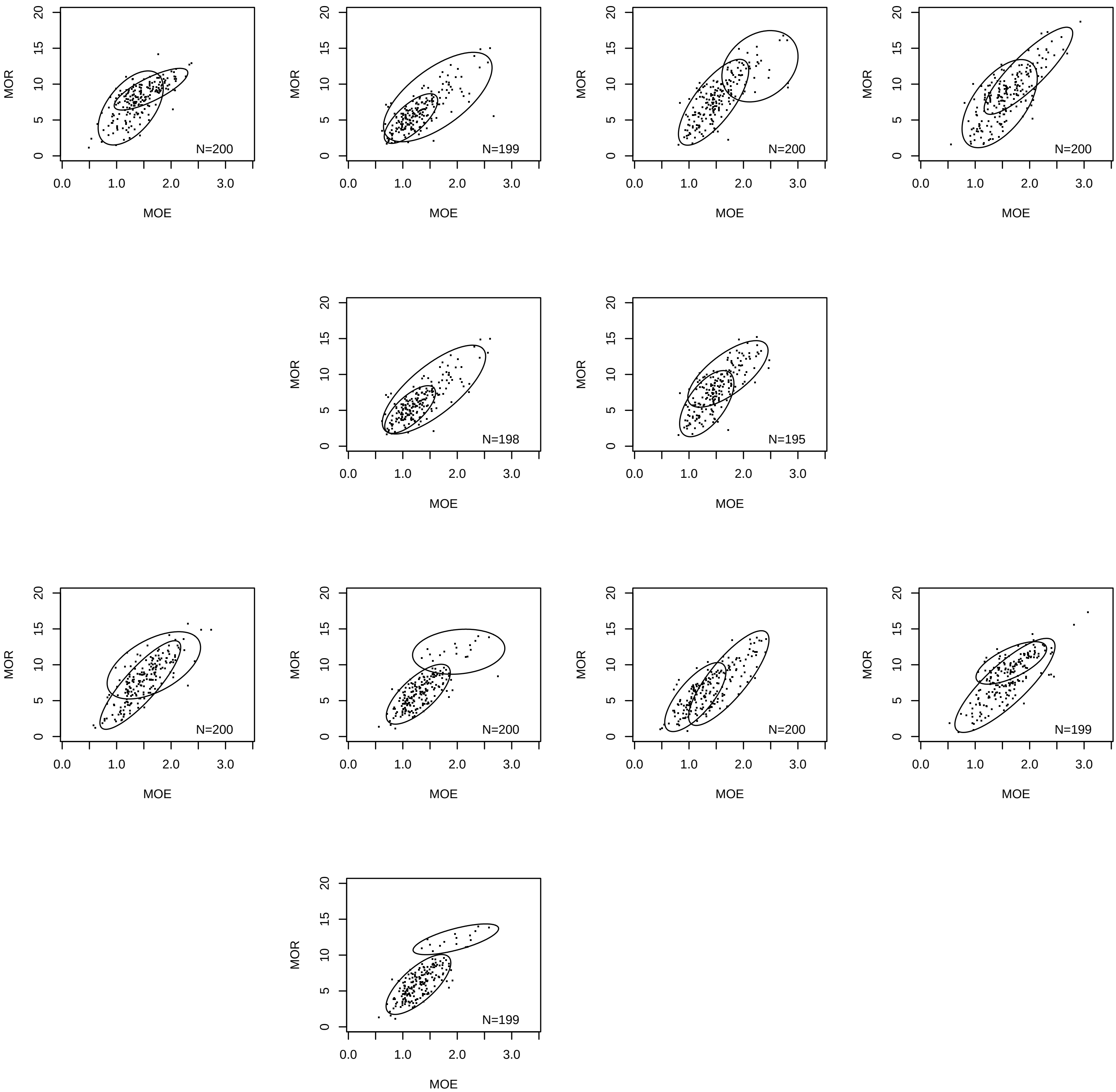


Figure 1: 0.90 content ellipses for the two bivariate normals in the fitted mixture distribution. Column 1 — Mill 1. Column 2 — Mill 2. Column 3 — Mill 3. Column 4 — Mill 4. Row 1 — Summer, full. Row 2 — Summer, "outliers" out. Row 3 — Winter, full. Row 4 — Winter, "outliers" out. The N values in the lower right corners of the plots are sample sizes.

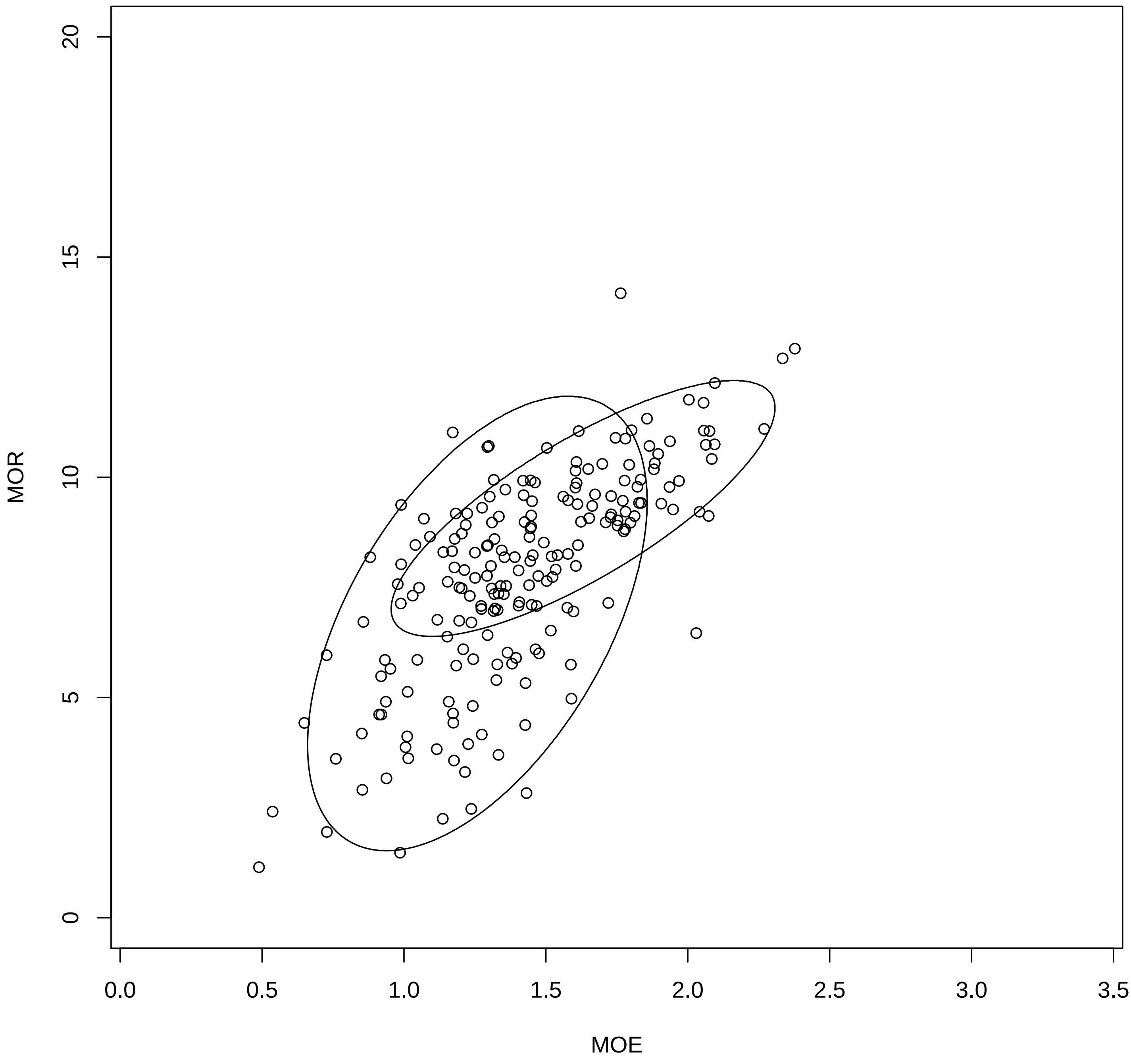


Figure 2: Summer, mill 1, 0.90 ellipses, all 200 data points

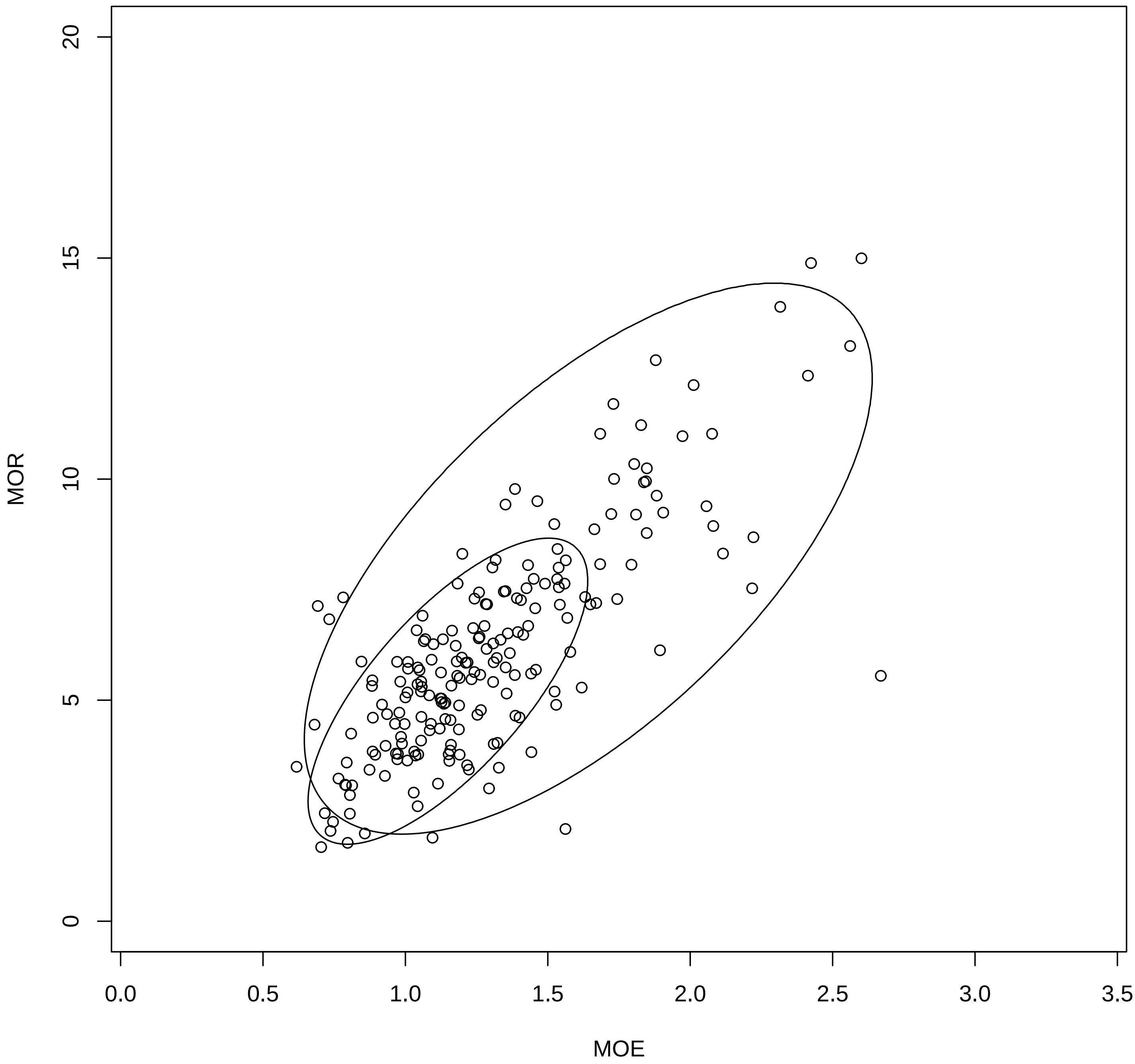


Figure 3: Summer, mill 2, 0.90 ellipses, all 199 data points

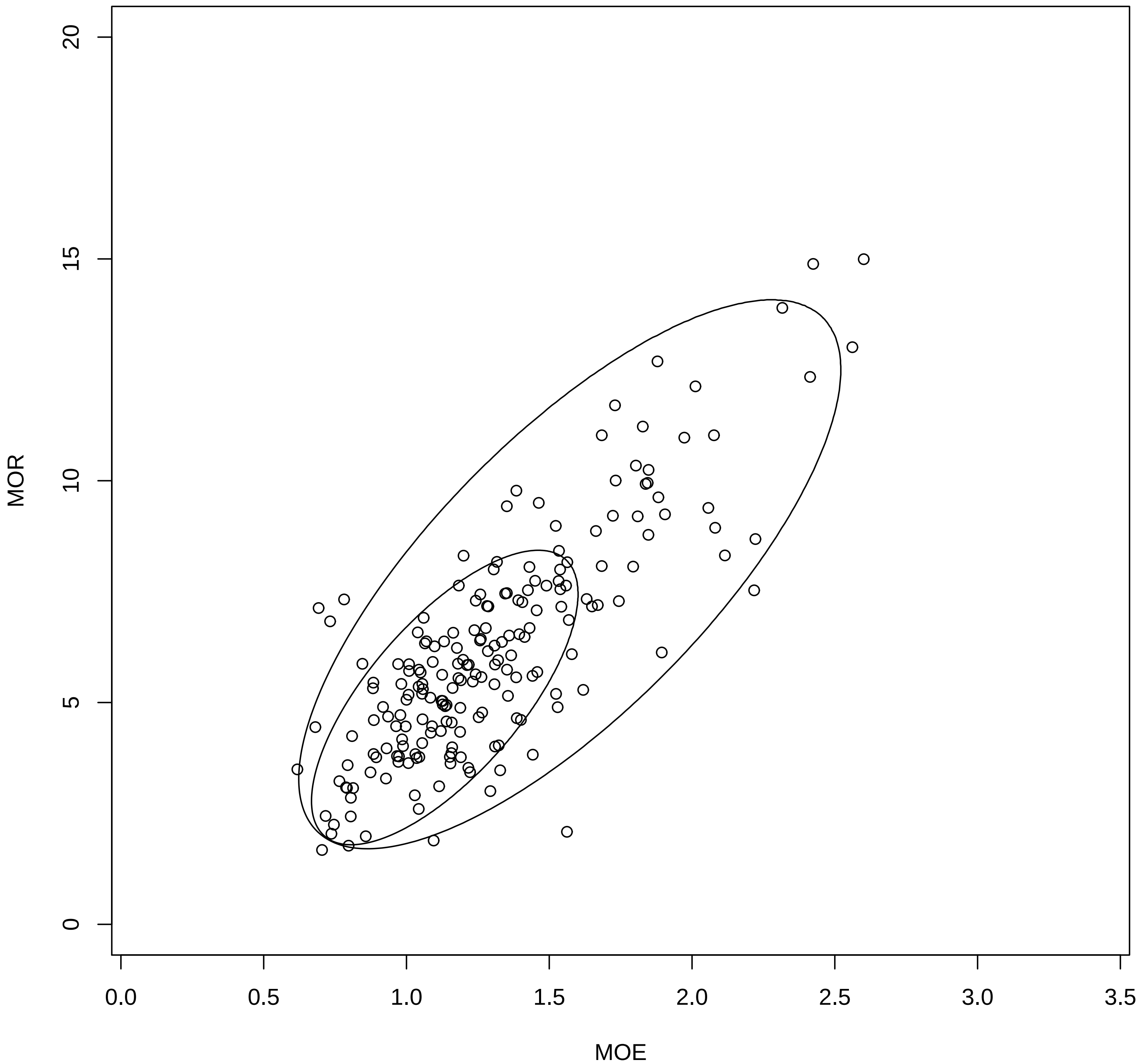


Figure 4: Summer, mill 2, 0.90 ellipses, 198 data points

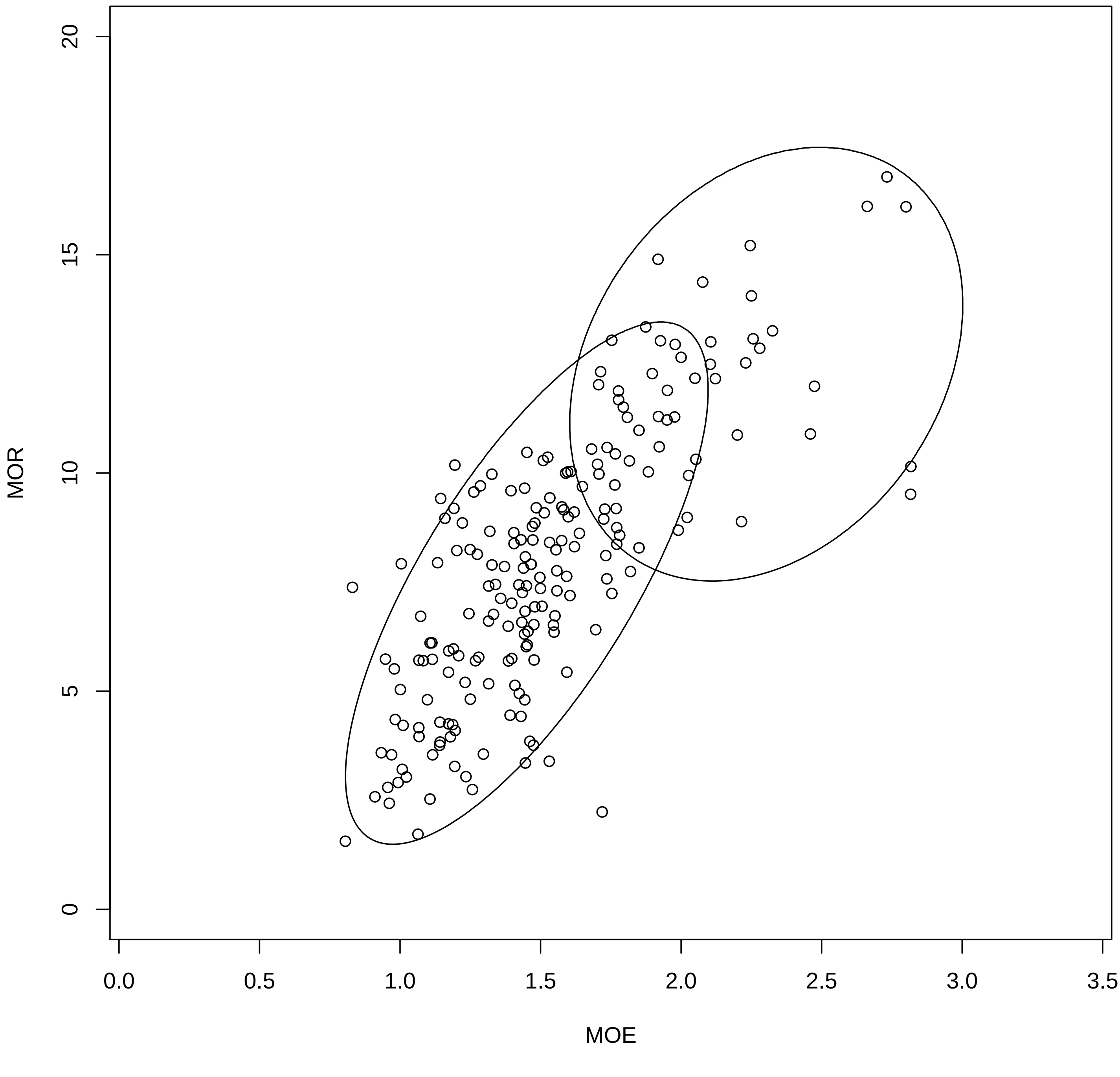


Figure 5: Summer, mill 3, 0.90 ellipses, all 200 data points

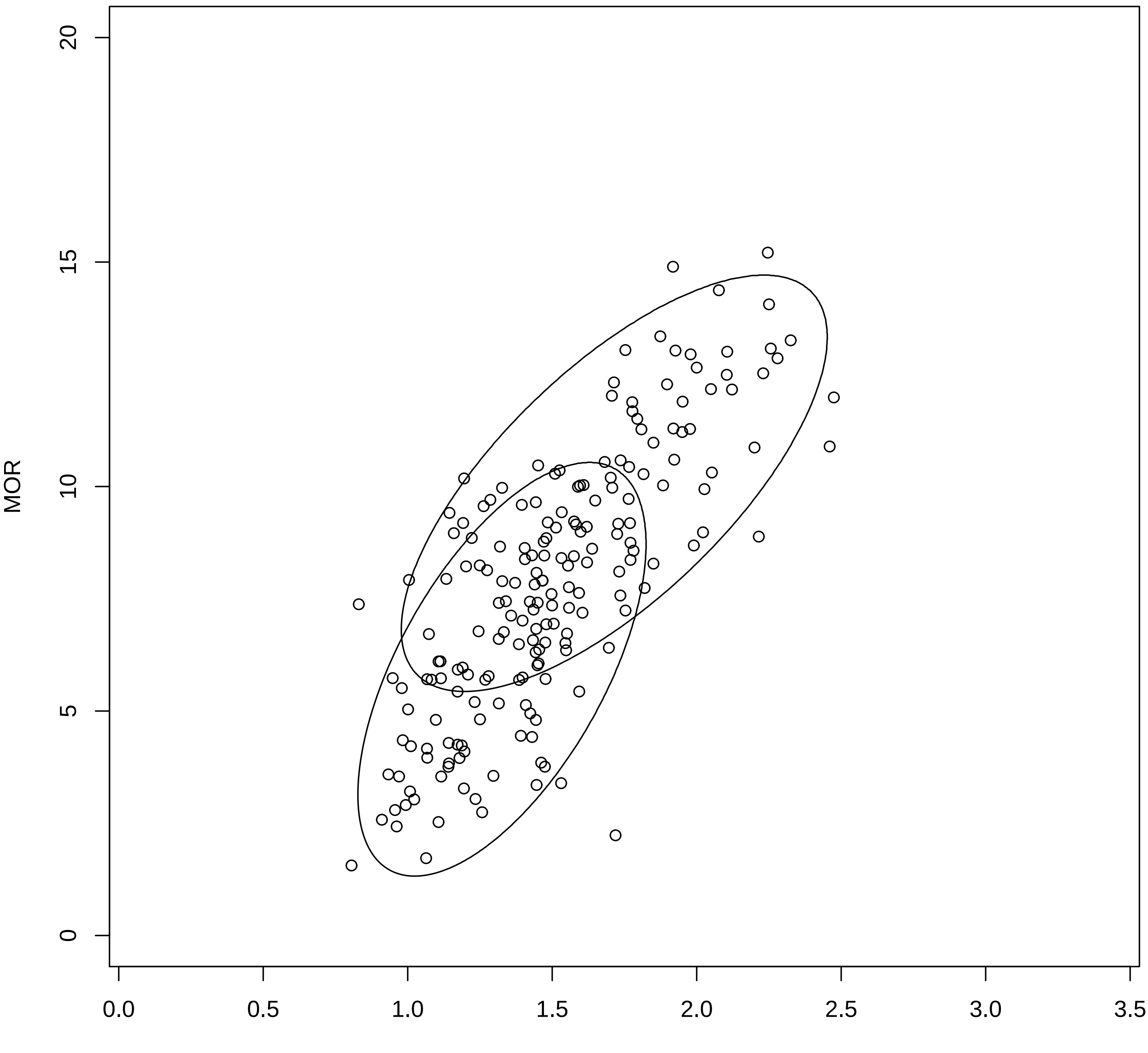


Figure 6: Summer, mill 3, 0.90 ellipses, 195 data points

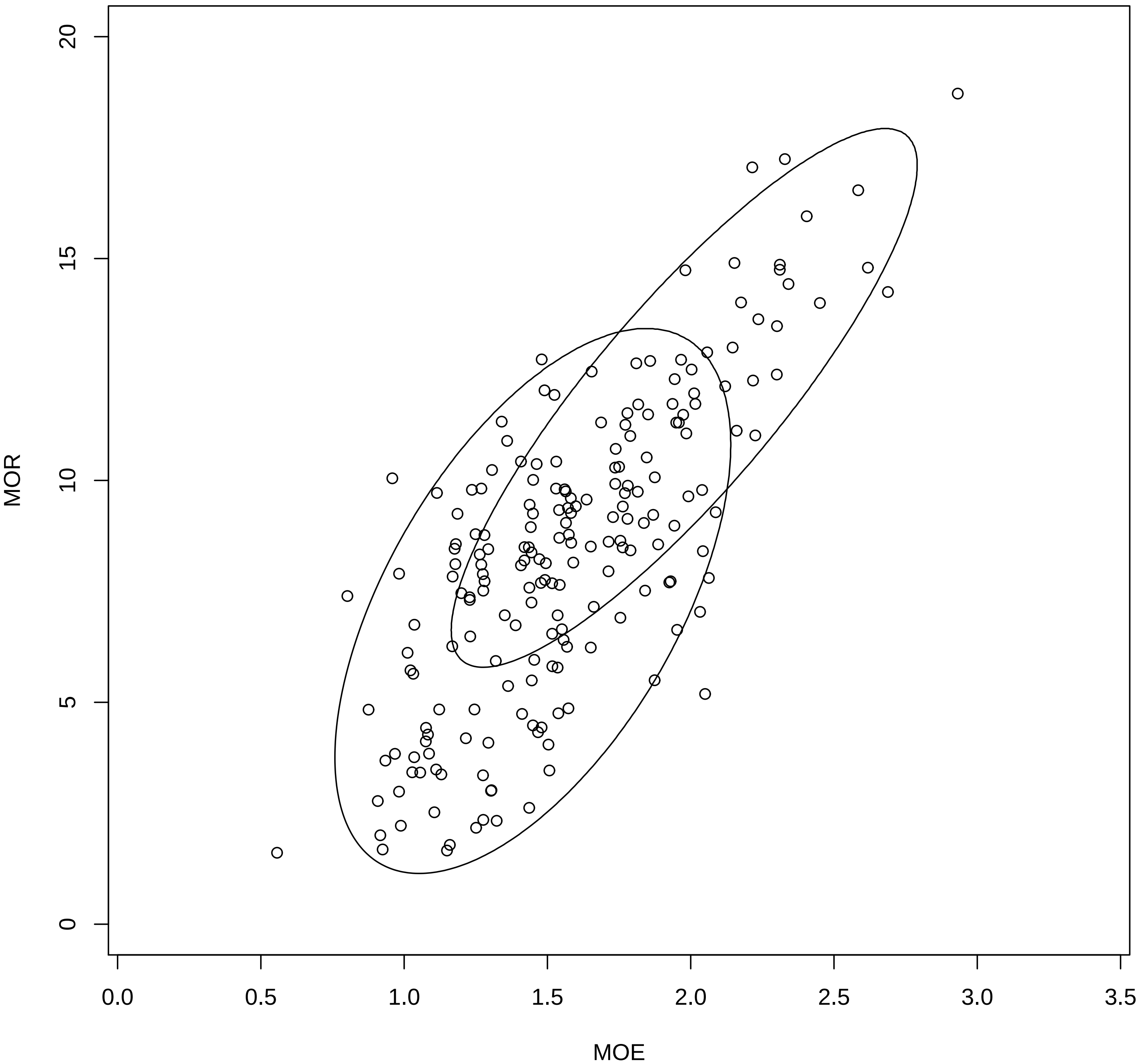


Figure 7: Summer, mill 4, 0.90 ellipses, all 200 data points

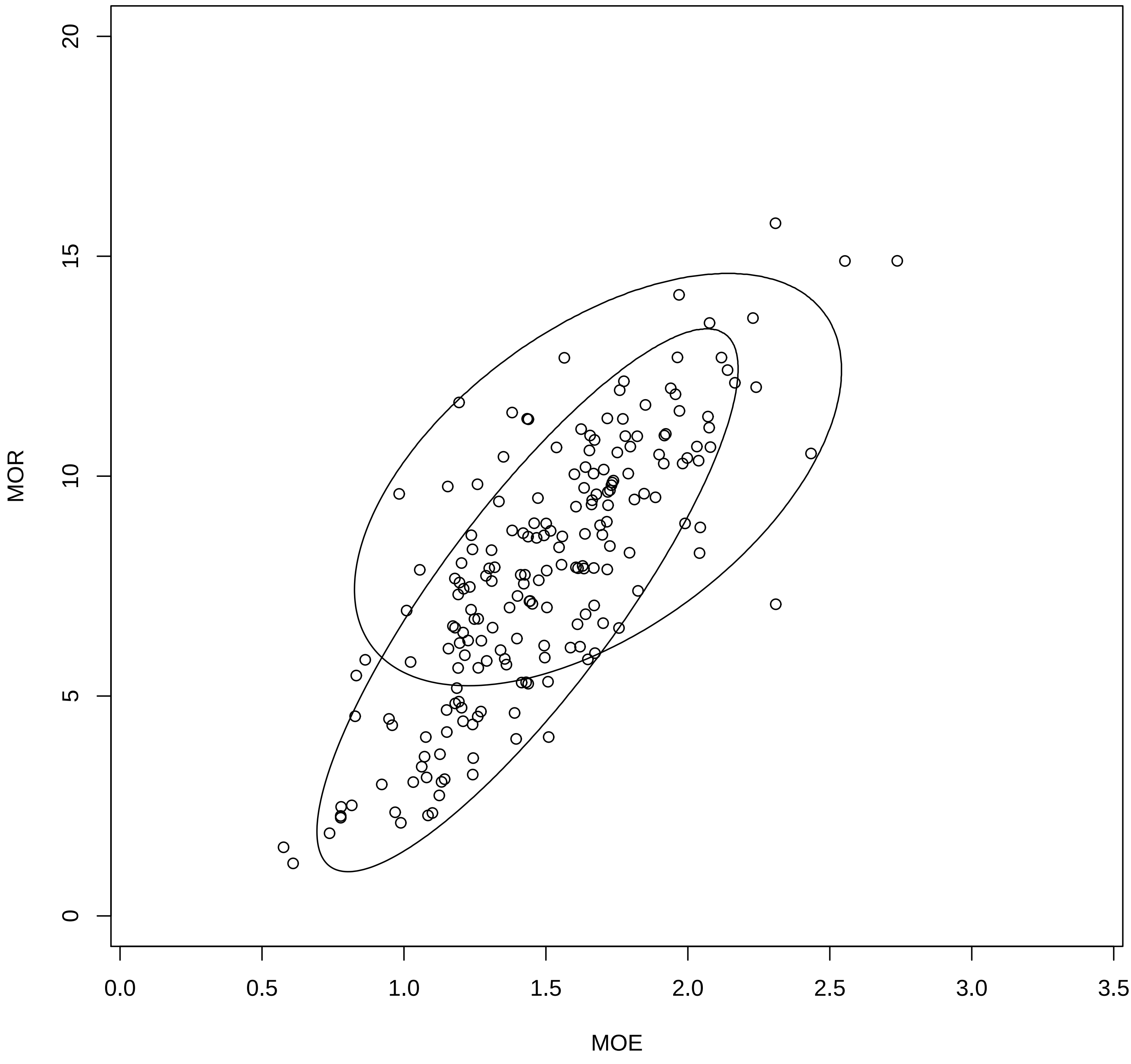


Figure 8: Winter, mill 1, 0.90 ellipses, all 200 data points

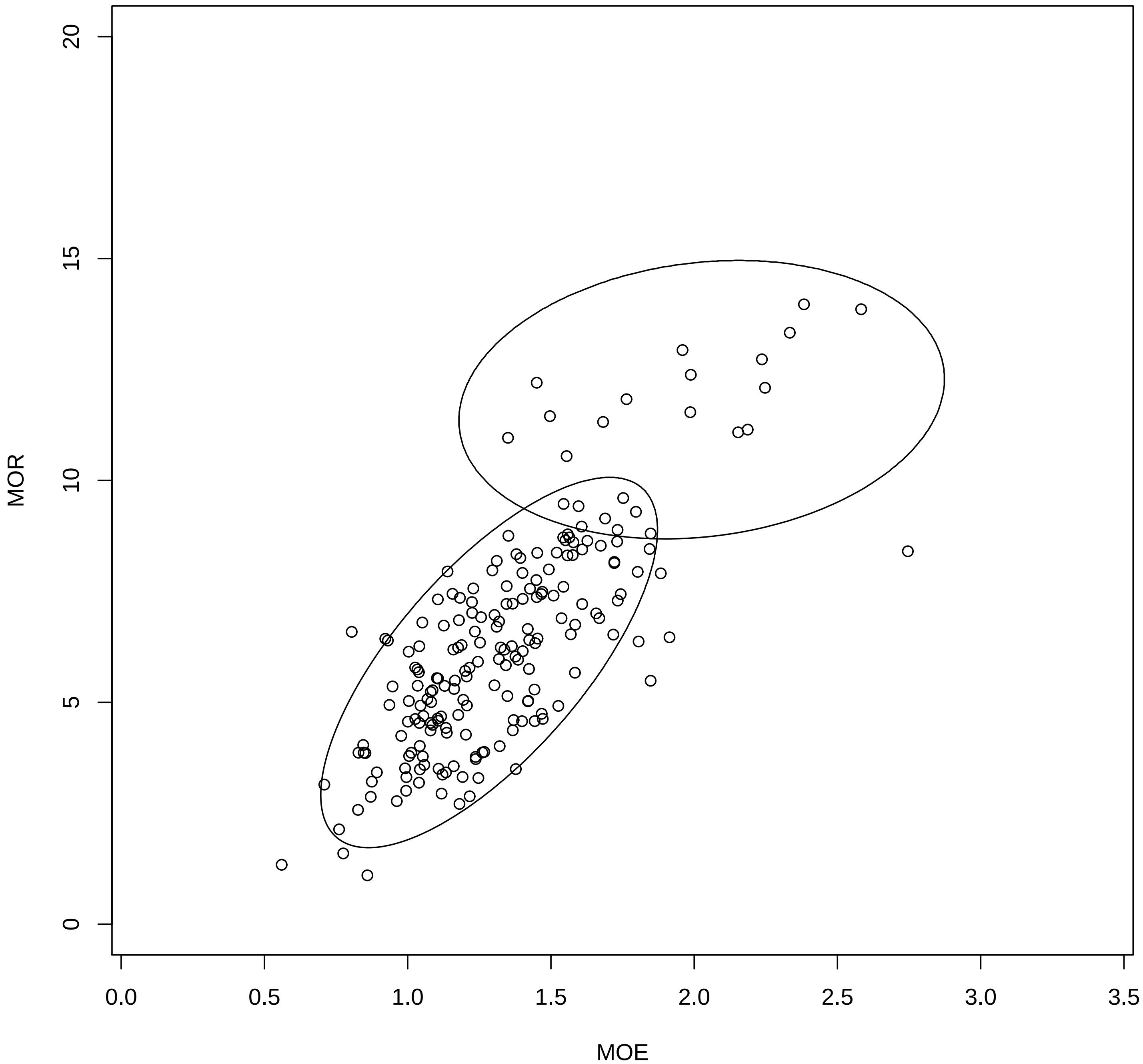


Figure 9: Winter, mill 2, 0.90 ellipses, all 200 data points

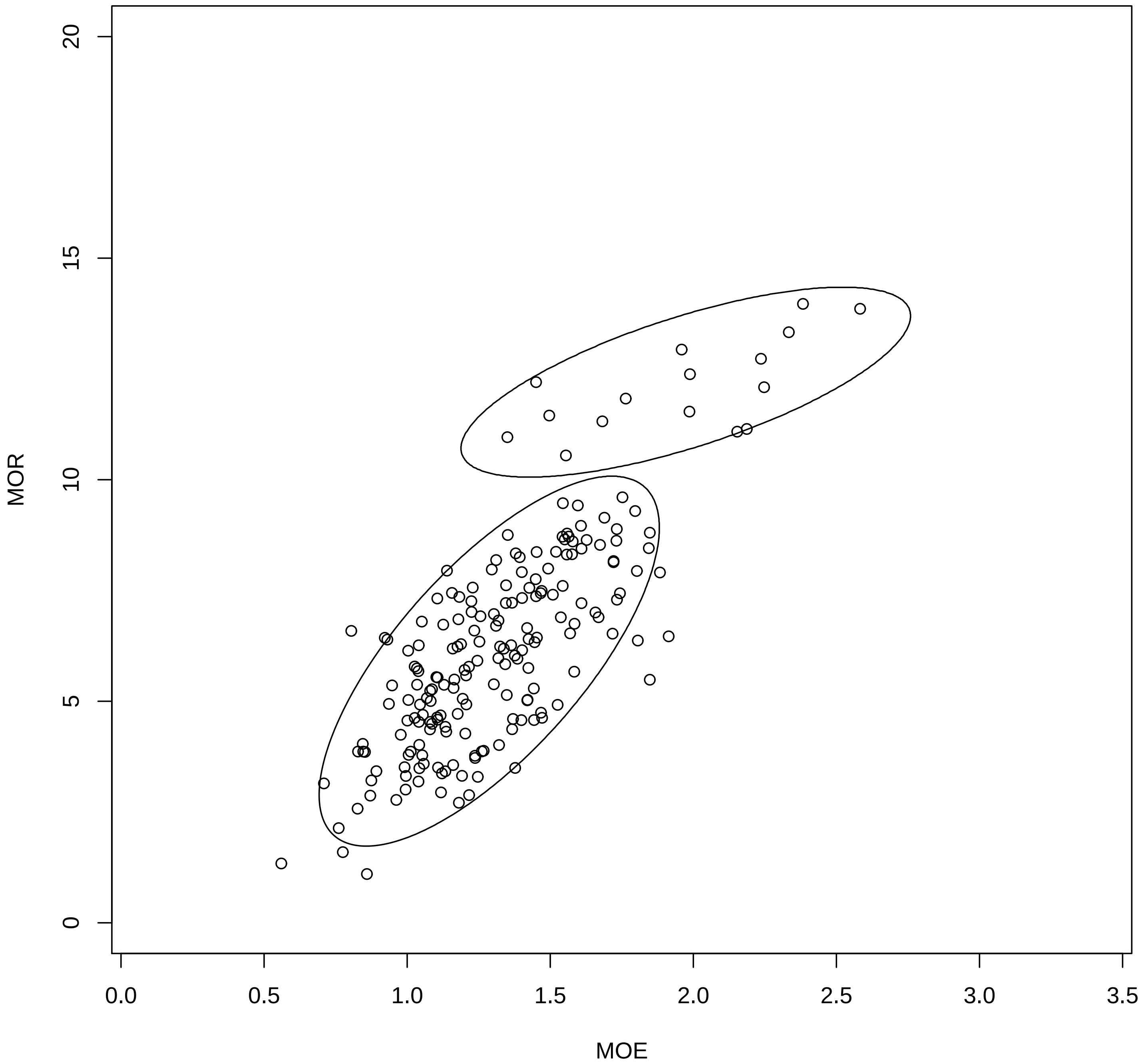


Figure 10: Winter, mill 2, 0.90 ellipses, 199 data points

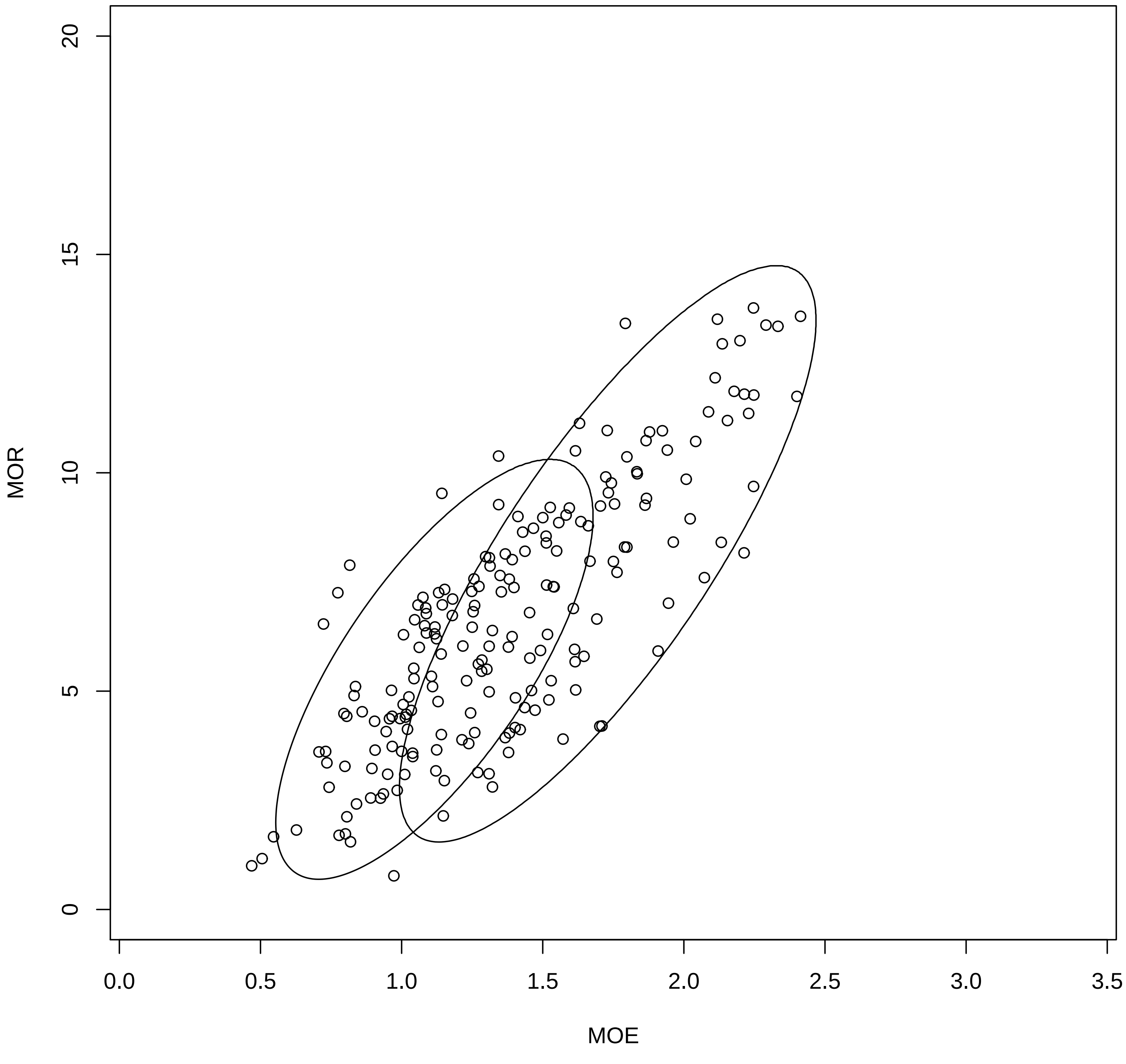


Figure 11: Winter, mill 3, 0.90 ellipses, all 200 data points

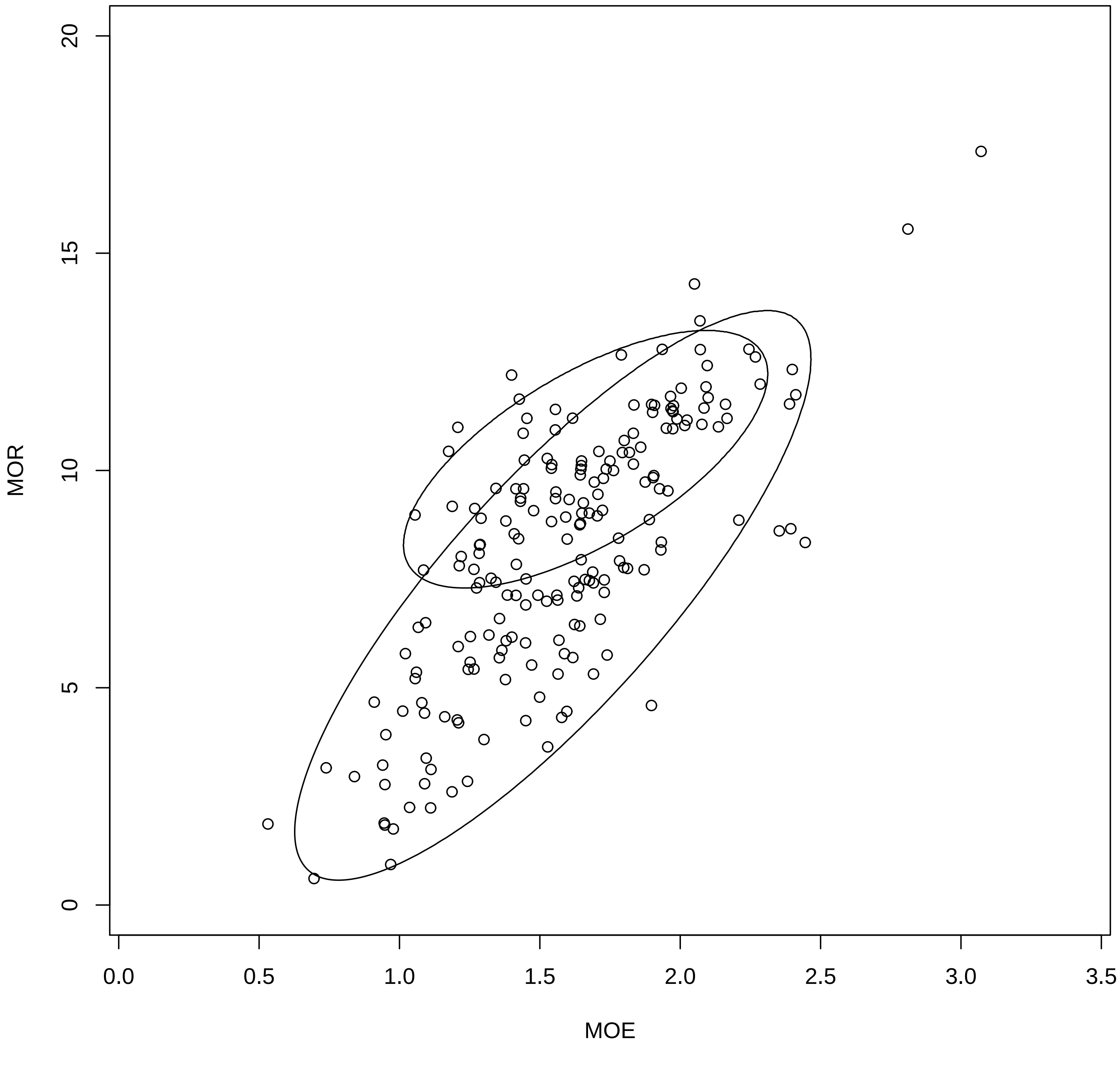


Figure 12: Winter, mill 4, 0.90 ellipses, all 199 data points

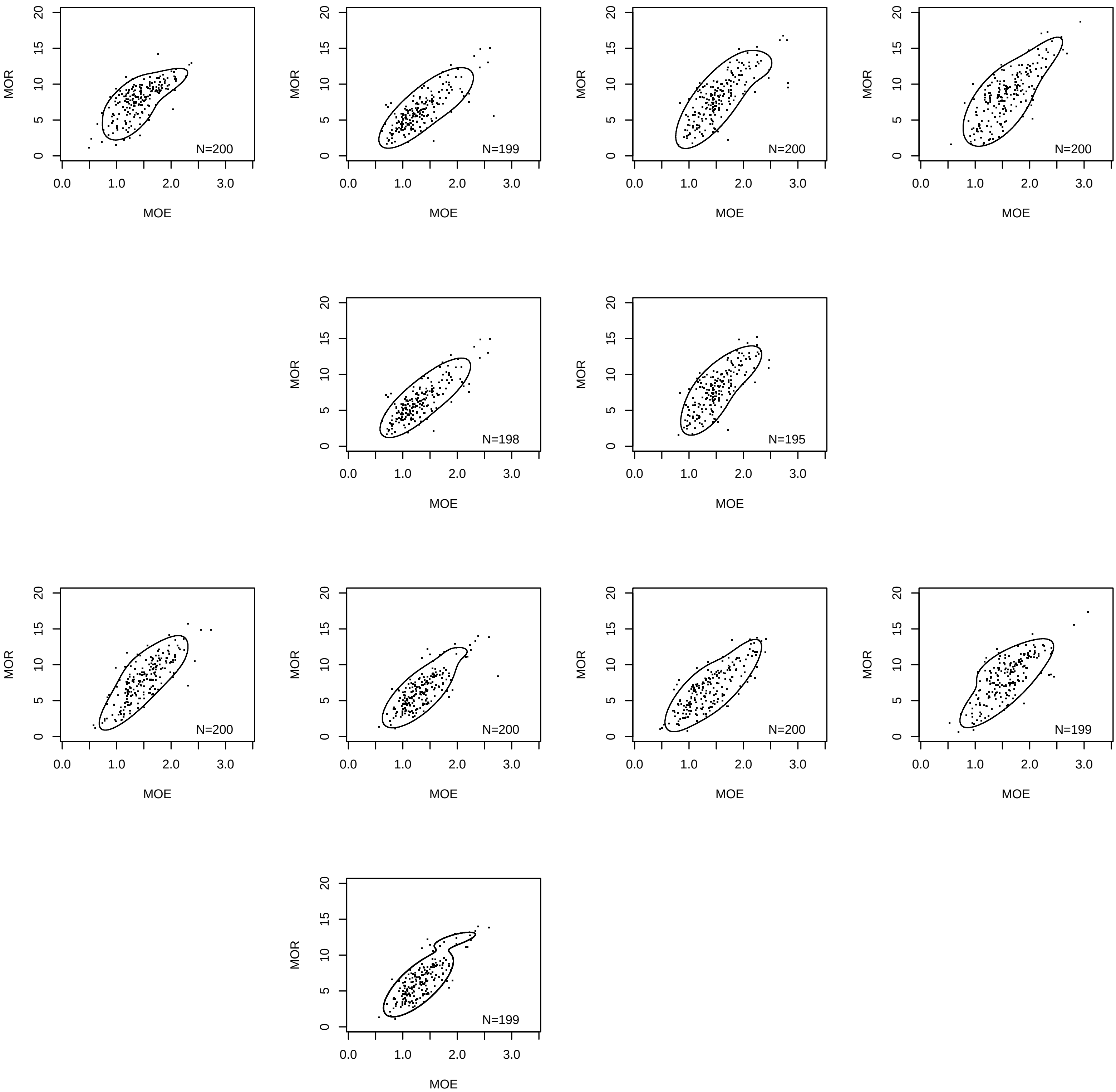


Figure 13: 0.90 content contours for the mixture distributions. Column 1 — Mill 1. Column 2 — Mill 2. Column 3 — Mill 3. Column 4 — Mill 4. Row 1 — Summer, full. Row 2 — Summer, "outliers" out. Row 3 — Winter, full. Row 4 — Winter, "outliers" out. The N values in the lower right corners of the plots are sample sizes.

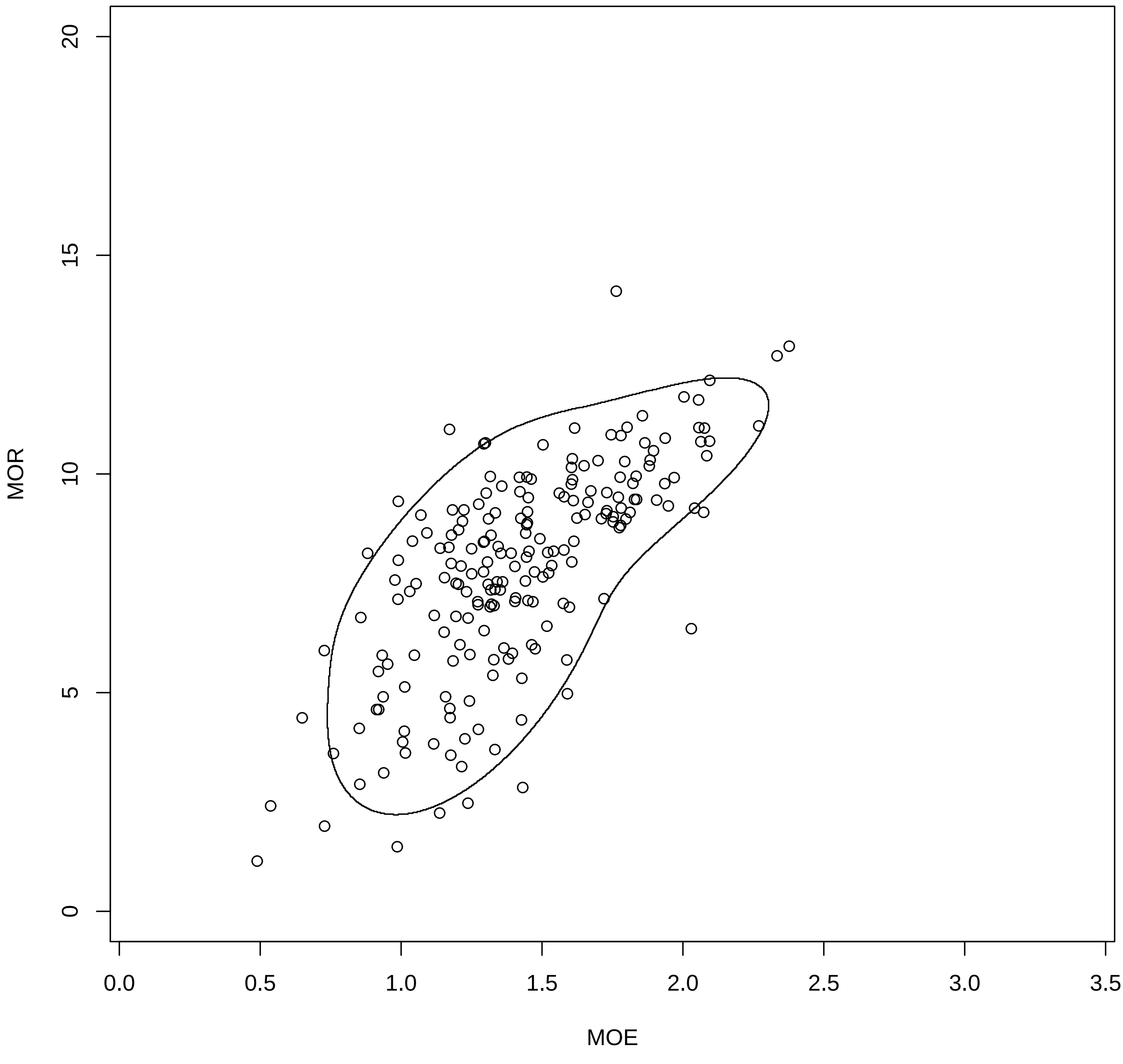


Figure 14: Summer, mill 1, 0.90 contour, all 200 data points

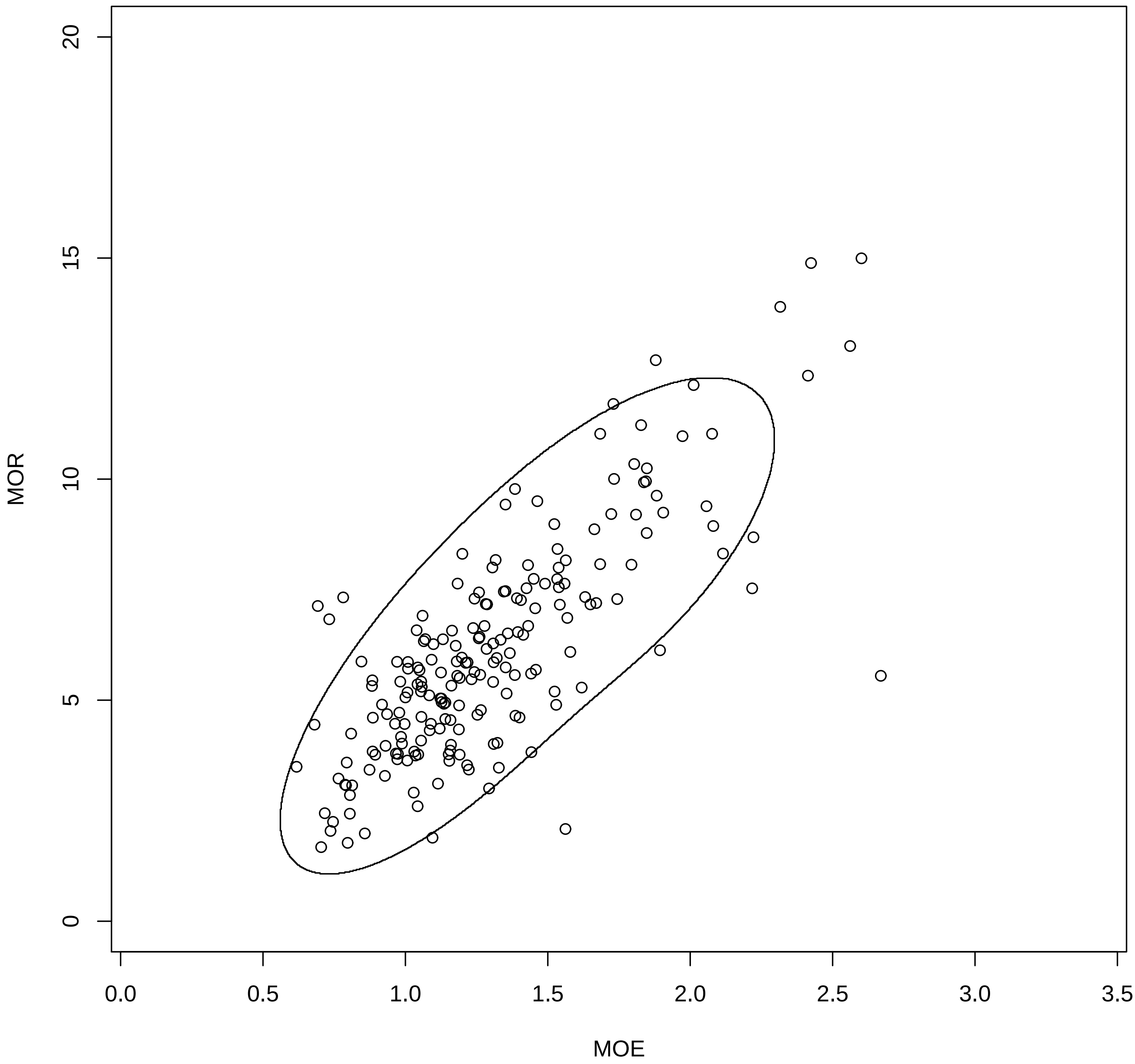


Figure 15: Summer, mill 2, 0.90 contour, all 199 data points

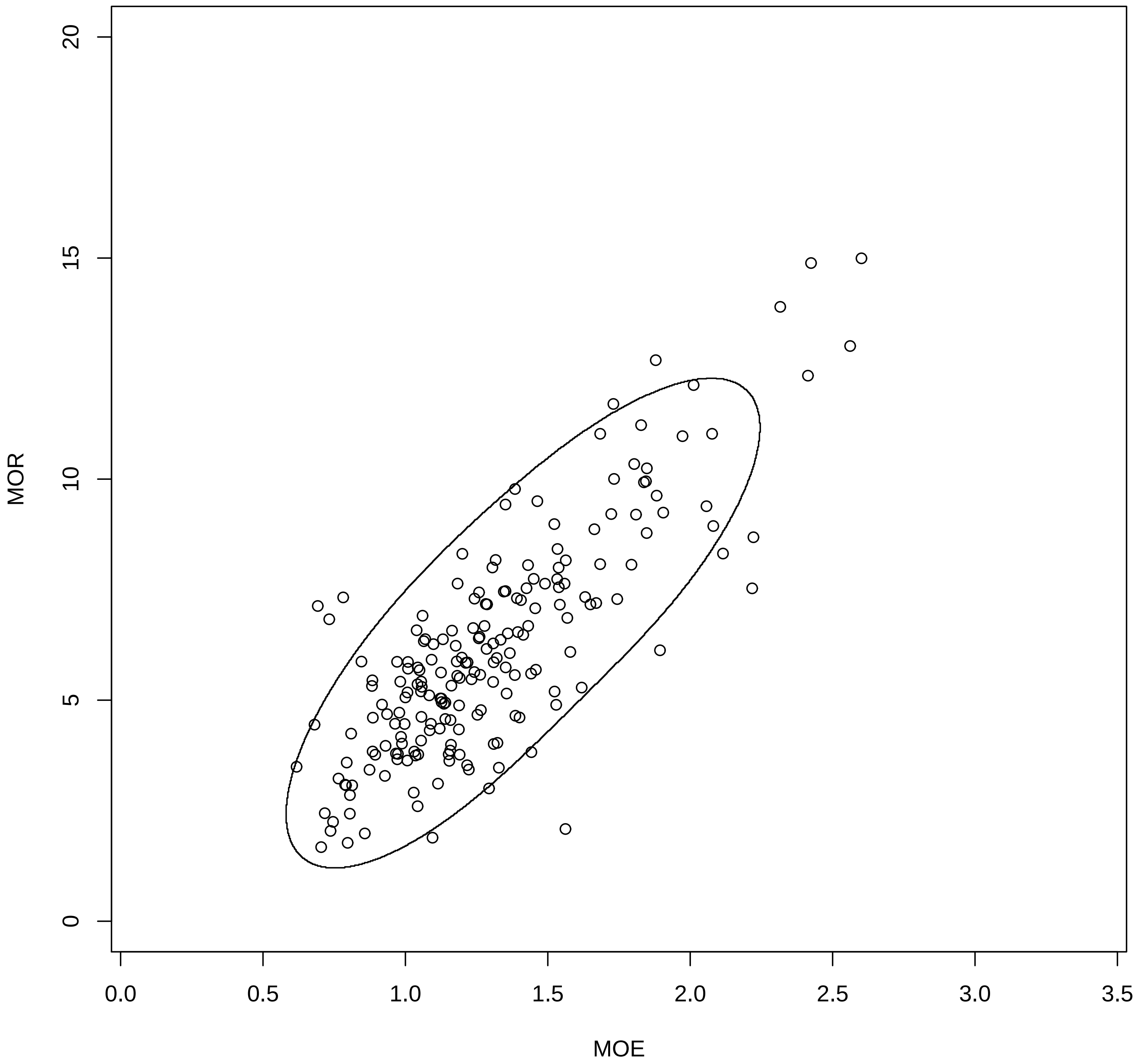


Figure 16: Summer, mill 2, 0.90 contour, 198 data points

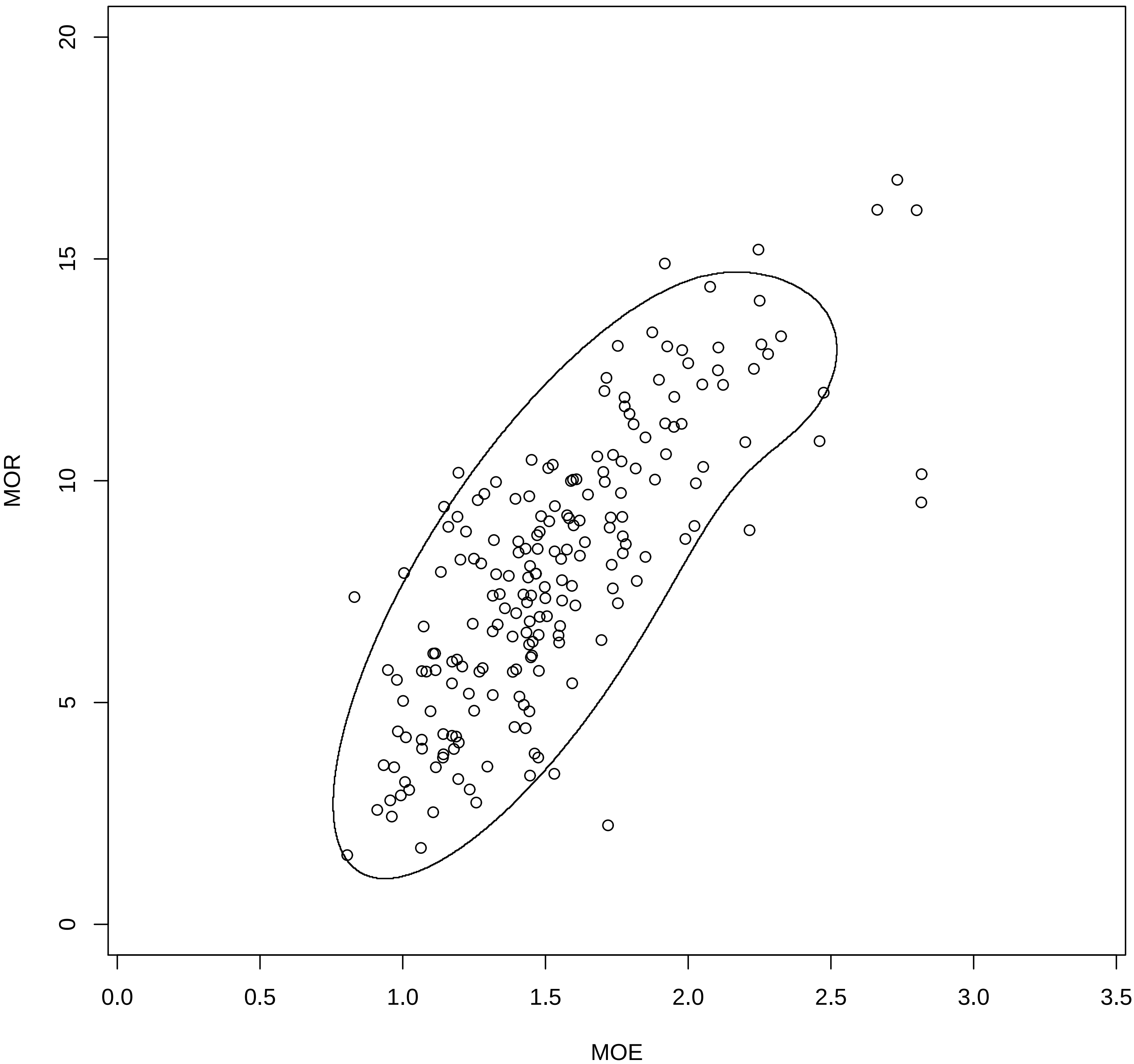


Figure 17: Summer, mill 3, 0.90 contour, all 200 data points

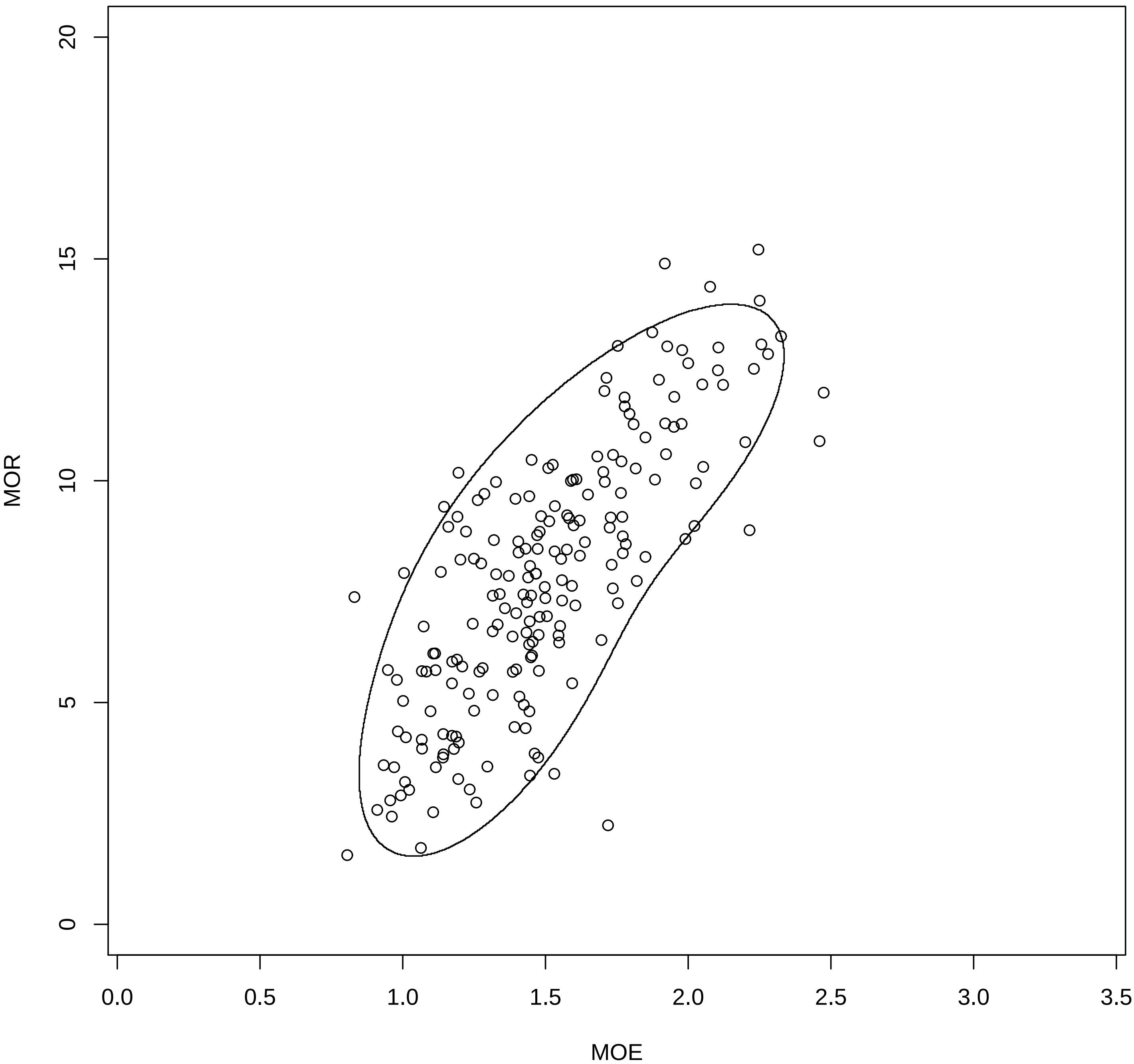


Figure 18: Summer, mill 3, 0.90 contour, 195 data points

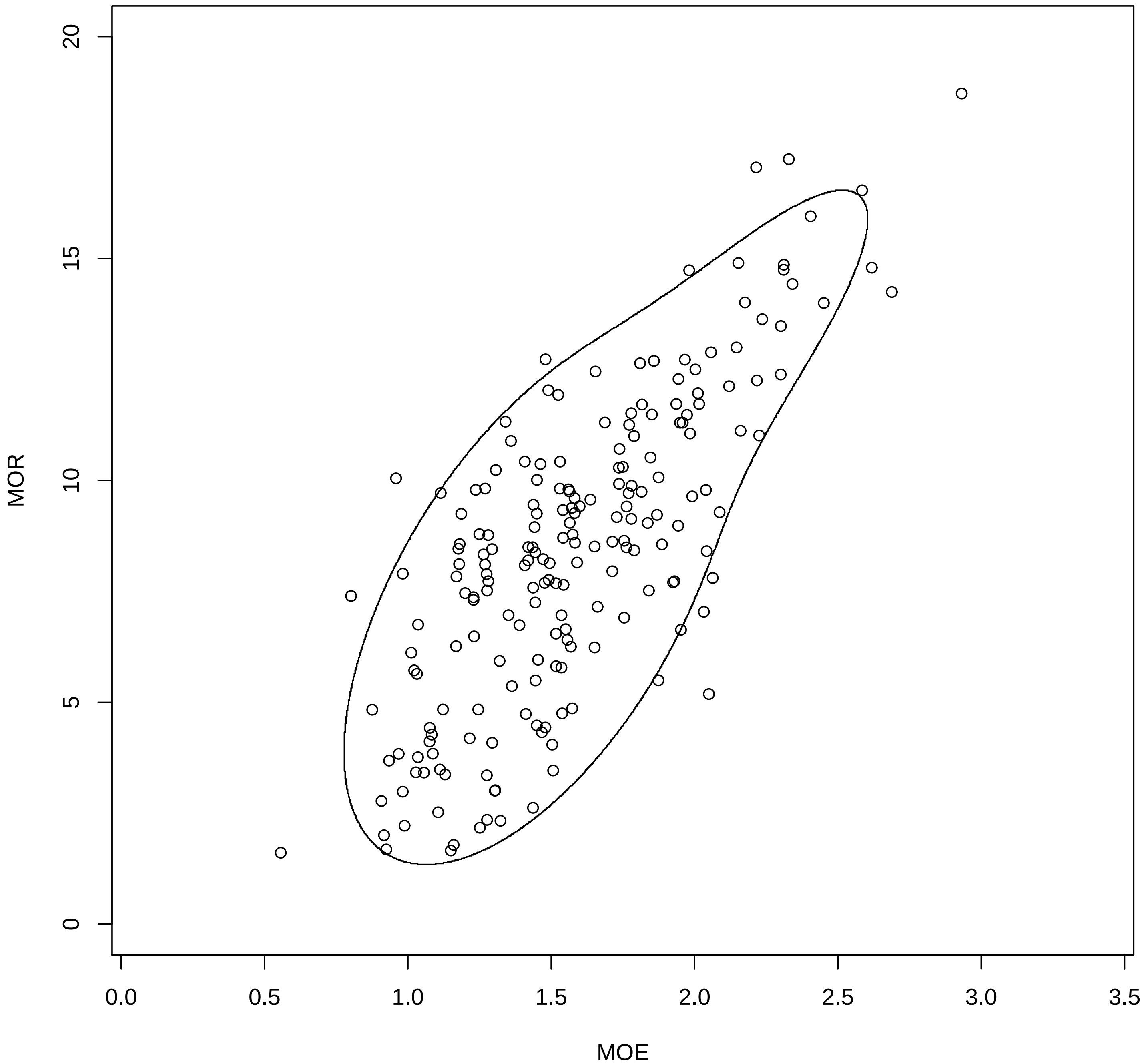


Figure 19: Summer, mill 4, 0.90 contour, all 200 data points

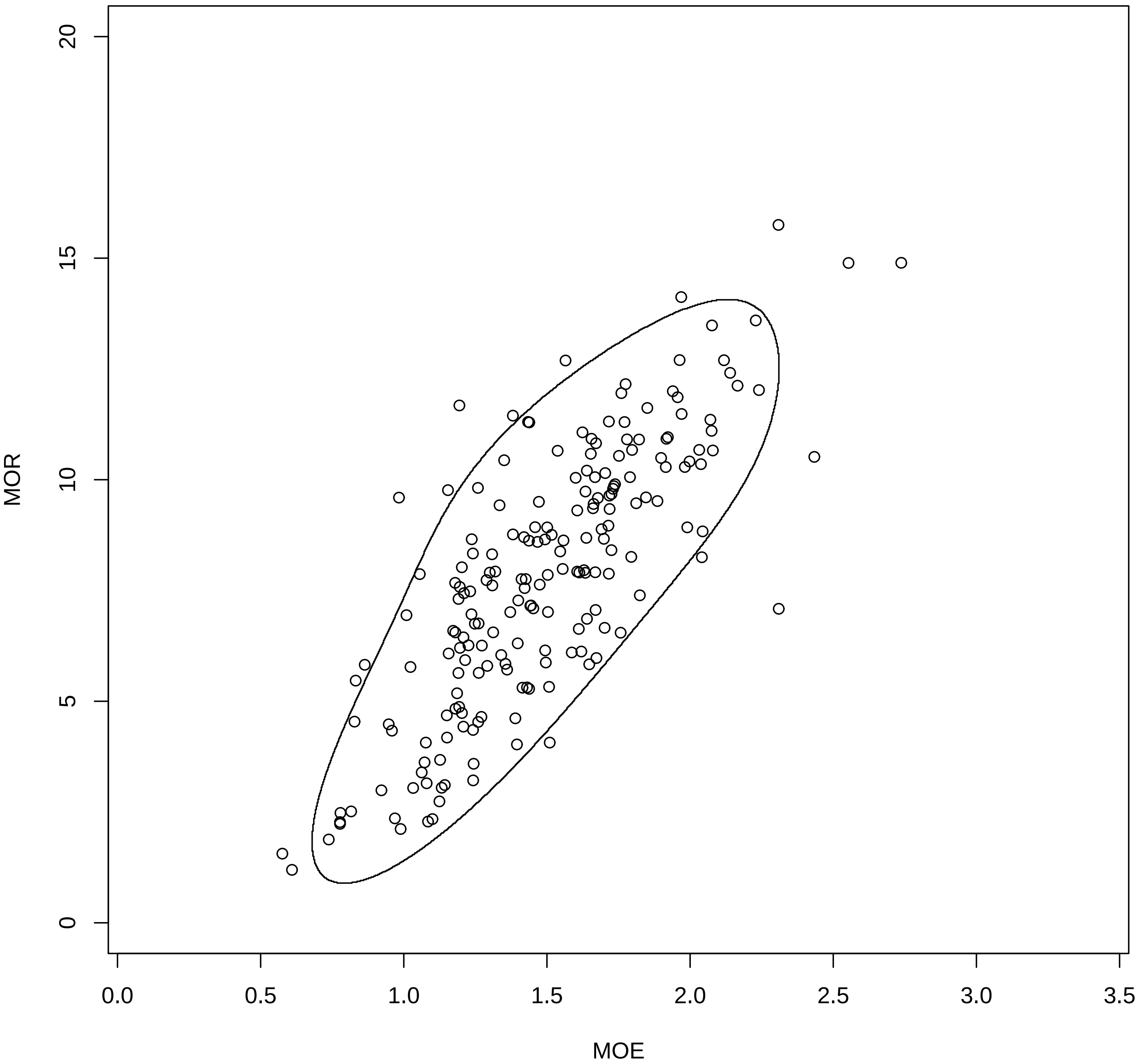


Figure 20: Winter, mill 1, 0.90 contour, all 200 data points

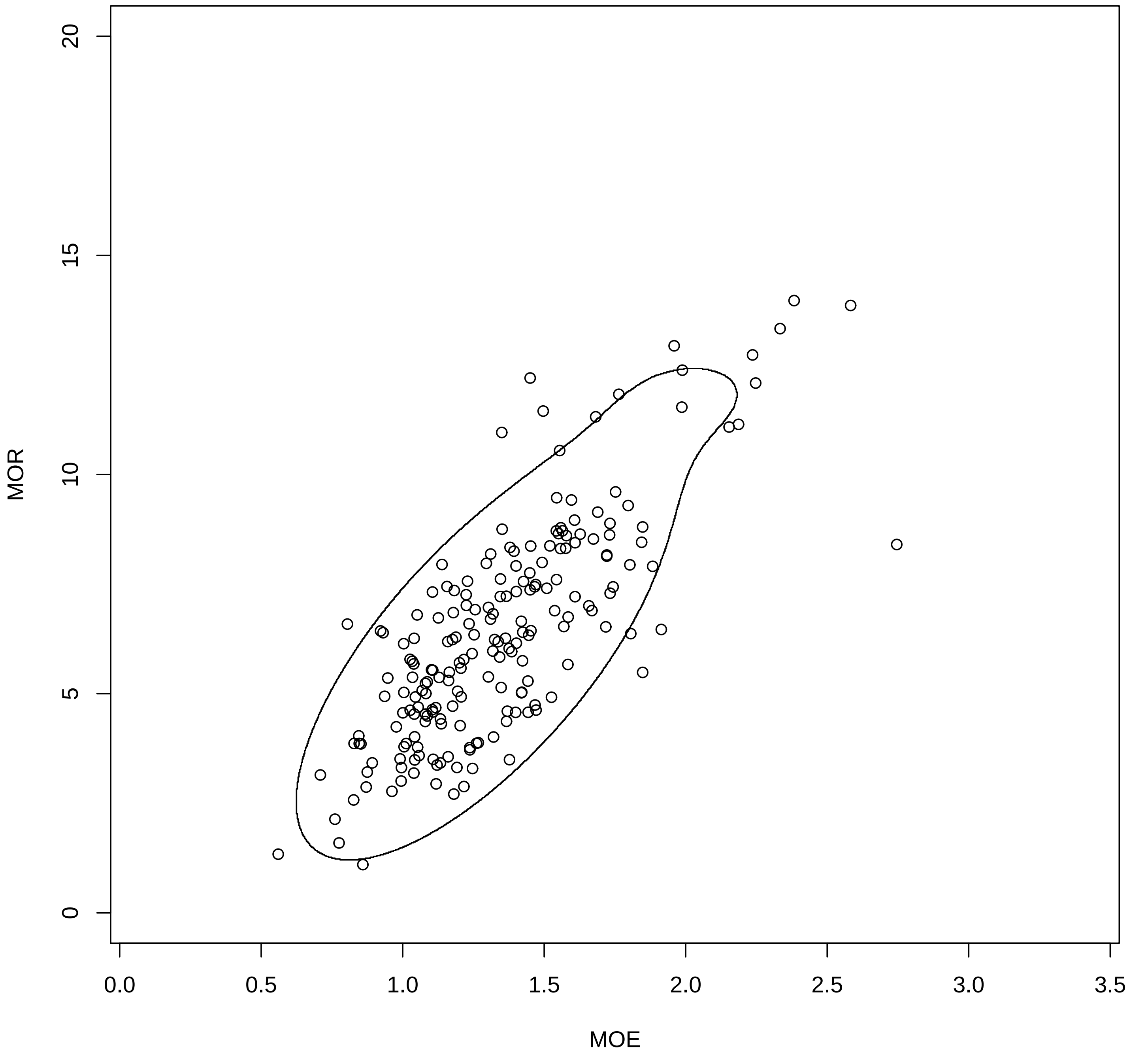


Figure 21: Winter, mill 2, 0.90 contour, all 200 data points

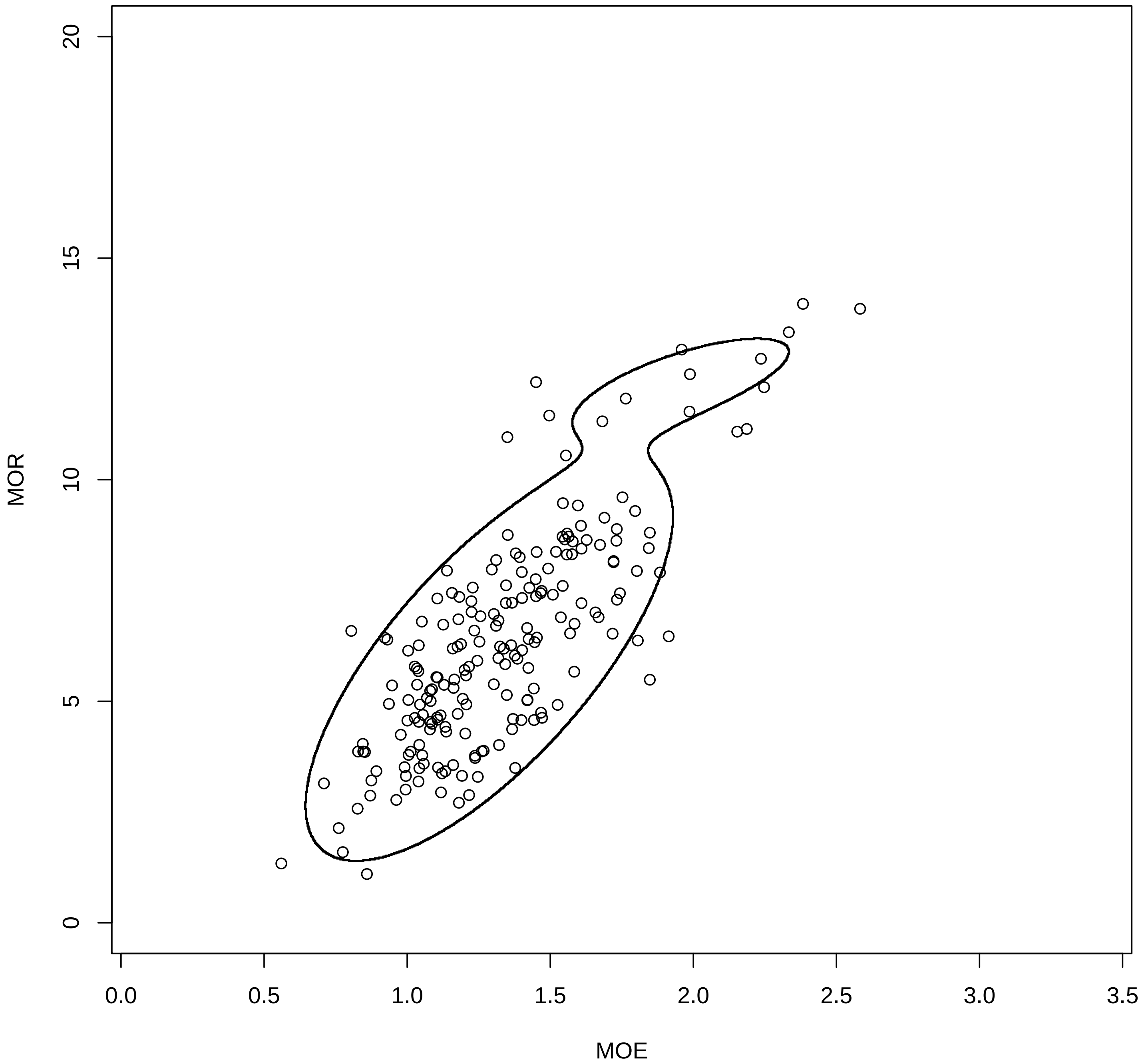


Figure 22: Winter, mill 2, 0.90 contour, 199 data points

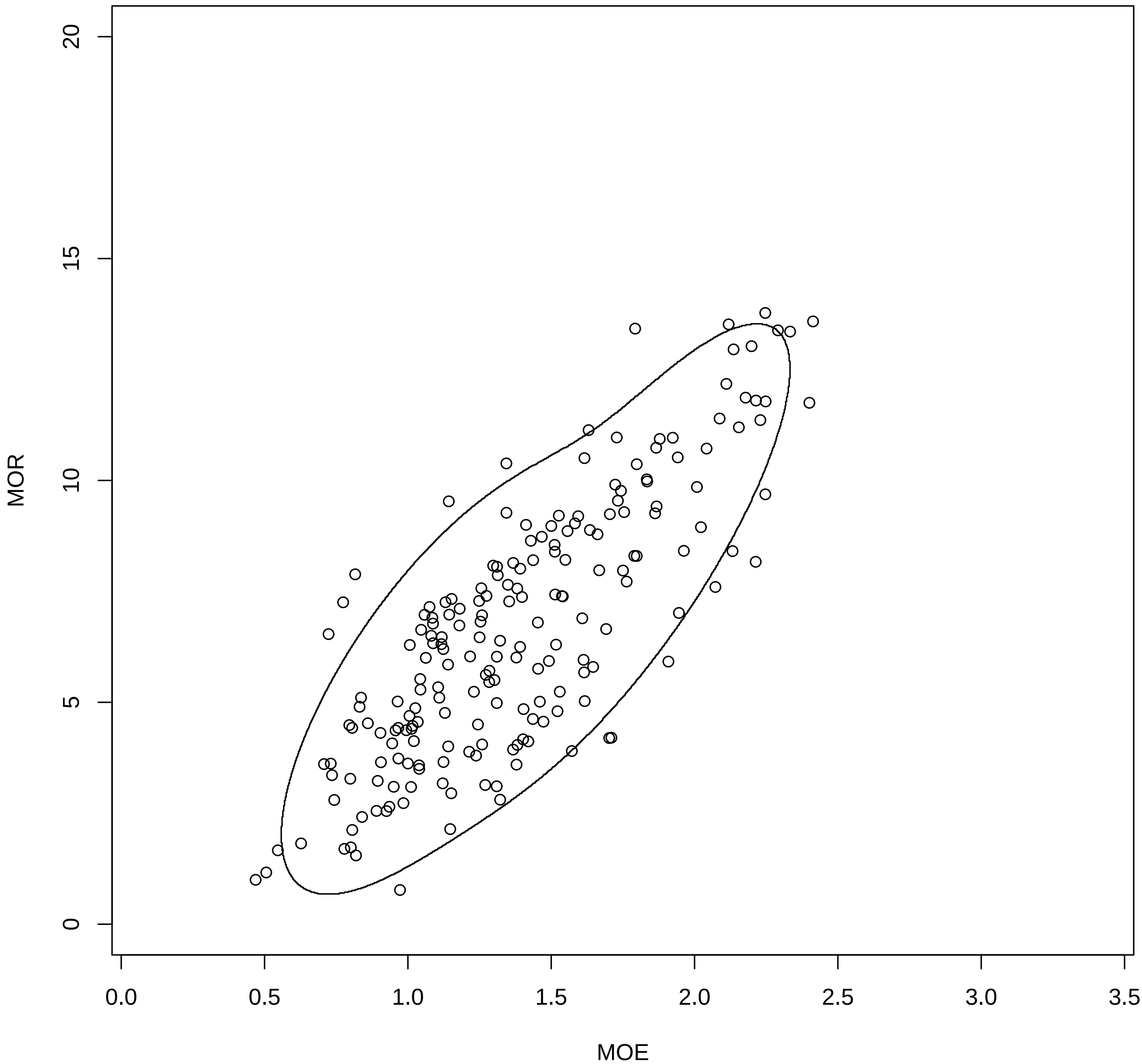


Figure 23: Winter, mill 3, 0.90 contour, all 200 data points

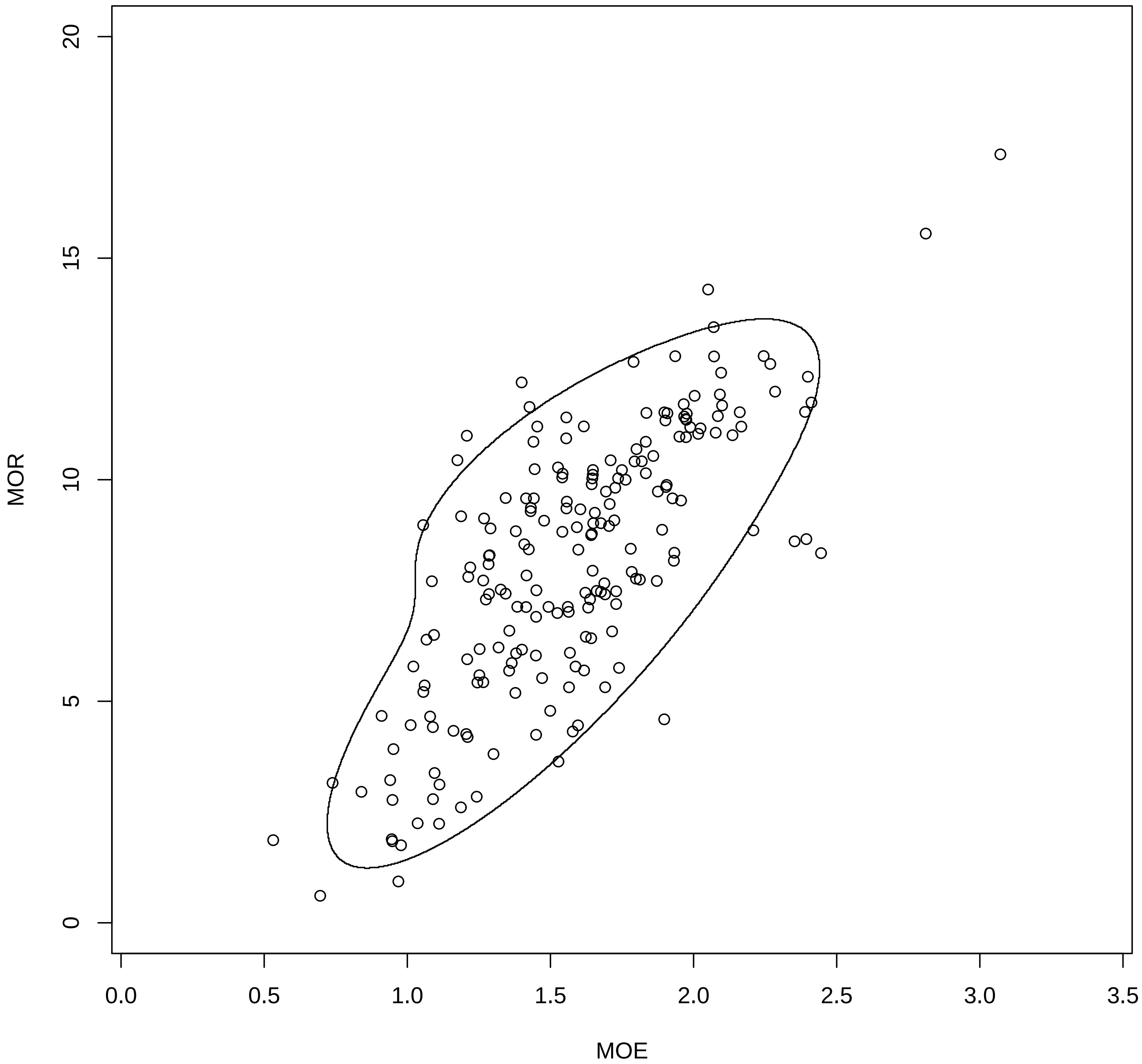


Figure 24: Winter, mill 4, 0.90 contour, all 199 data points