APPENDIX B

SUMMARY OF STATISTICAL ANALYSIS OF PARTICLEBOARD DRYER PM DATA

APPENDIX B

SUMMARY OF STATISTICAL ANALYSIS OF PARTICLEBOARD DRYER DATA

B.1 BACKGROUND

In terms of classic engineering experimentation problems, the data from the wood products industry, and in particular, those from particleboard that were examined fit the pattern of a multi-way factorial experiment with continuous covariates of interest. A multi-way factorial experiment is one in which there are multiple factors (such as wood type, dryer type, fuel, and air pollution control device in the particleboard example), with each of the factors have a small number (at least 2 and usually less than 10) of different levels. In the factorial experiment, at least one test is run at each combination of the different levels of the different factors. In such an experimental study, an analysis of variance (ANOVA) procedure can be used to examine the effects of each of the factors and on the various orders of interactions of those factors (two-way, three-way, etc.) on the outcome variable (in our example emissions). An interaction occurs when the effect of one factor varies across different levels of another factor. For example, if difference between the emissions from core and surface drying differ by type of fuel, then material and fuel would exhibit a two-way interaction. If the experiment also involves continuous variables such as temperature and moisture, which are not controlled but are allowed to vary naturally, then they are considered along with the factors in the analysis by using an analysis of covariance (ANACOVA) type model.

For the techniques described above to work reasonably well when applied directly, there must be at least one data point in most of the cells generated by cross-classifying on all of the levels of all of the factors of interest. For example, for two types of material [core and surface], two types of fuel [gas and wood], three types of dryer/firing [indirect, direct single pass, direct triple pass], and four levels of control [uncontrolled, EFB, multiclone, and WESP], there must be at least one "test" in each of the 48 cells created by the cross classifications of these variables. Because the particle board data were not created in an experimental procedure designed to look at all of the factors that we wanted to consider, it was not possible to proceed directly to the analyses using ANOVA or ANACOVA procedures. Consequently, many of the analyses that were performed were designed to determine where in the multi-factorial framework there were data and to explore the data in a preliminary fashion to see whether grouping of levels and possible of factors could reduce the "experimental design" to a manageable size.

The following numbered paragraphs pertain to the analyses performed, the commands and output for which are presented in the output log that follows. The analysis was performed using STATA statistical software. The paragraphs below are tabulated to relate directly to the numbers on the STATA output that were generated during the analysis. Each paragraph corresponds to a handwritten number that is in the right hand margin of the output log.

<u>Comments on the Morning Particleboard Log</u>

1. In all of the STATA statements , the "if Poll==1" component of the statement is used to inform STATA that the analyses are to be performed only for filterable particulate matter (PM), which is coded as 1.

2. The tabulation of dryer type by firing type indicates that essentially all of the PM data are for direct-fired dryers (DFIRE). Only three tests for a rotary triple-pass (RTP) dryer are available for dryers that are both direct- and indirect-fired. These results suggest that effect of firing type on emissions cannot be evaluated with the available data, so firing type can be ignored in the remaining analyses.

Although not performed in this analysis, one might want to look at the emission levels for those three tests for "BOTH" in comparison to the 30 for direct-fired for triple pass. If no difference, lump the data from "BOTH" into the remaining analyses. If different, exclude them.

3. At this point, because firing type is relatively unimportant due to lack of adequate distribution of observations, the data were examined as a function of dryer type using box plots. Although the box plots do not appear on the log, the on-screen review suggested no real difference in distribution other than for tube type dryers. These preliminary results suggest that the possibility of collapsing data across rotary dryer types might be considered for later analyses. However, some additional exploratory analyses were first performed.

4. These two cross-tabulations examine the number of tests in the cells defined by different levels of (1) dryer type and control device and (2) dryer type and material (core or surface). The first tabulation shows that only EFB's are available on rotary single-pass (RSP) dryers but that RTP dryers have a spectrum of controls. The second tabulation suggests that differentiating the effect of material type as defined by core or surface might be difficult as only 12 of the 43 tests used a specific material, and material type is unknown for 18 of the 43 tests.

5. In order to obtain a better picture of what can be interpreted from the combination of dryer type and material type, a new variable was created that combines core/surface material with the use of an RTP dryer. After the new variable was created, it was cross-tabulated with control device. The results of the cross tabulation indicated that for tests in which only core or surface material were known to be dried all were uncontrolled.

6. In the next step the emissions from RTP driers as a function of material type were compared using a box plot without regard to the type of control device . The plot, which is not in the log, seemed to indicate substantial overlap. In particular, the levels of 1 and 2 for the new variable, which depicted uncontrolled core and surface material showed little difference. To confirm the results for the boxplots, which suggested little difference in the groups, a one-way analysis of variance was used to compare the four groups. The p-value (0.28) for the overall F-tests indicates that these data do not provide evidence of a difference in groups. (Note that a p value less than 0.1 would indicate weak evidence of some difference. For illustrative purposes, individual differences in means were estimated using bonferroni correction, but in practice, the nonsignificance of the F-test indicates that individual comparisons need not be performed.

7. The description of the rationale and the results from this box plot seem sufficient. The objective was to look at EFB control as a function of dryer type. Material type was ignored because the analyses immediately above indicated that it was not an important variable.

8. A new variable is created that combines wood type (1=hard, 2=soft, 3=mixed, 4=unknown) with RTP dryers (i.e. this variable identifies tests by wood type on RTP dryers). The cross-tabulation shows the difficulty of examining the effect of wood type on control, as the types of control differ distinctly by wood type. In classic AP-42 terms, this means that controlled emission factors could not be developed as a function of wood type for specific control options.

9. The above analyses suggest that to avoid confounding the problem, control device performance must be examined by wood type. Consequently, separate box plots were developed for each of the two primary wood types; the notes in the log describe the results.

10. Cross-tabulation of dryer type with fuel type indicates a reasonable distribution of wood types across the RTP category of dryer. Consequently, a variable was generated that indicate fuel used in RTP dryers. Note that the value of this variable will be missing if another type of dryer is used.

11. Cross-tabulation of the newly created variable with control device shows reasonable distribution of controls for dryers that use wood as fuel (rtpfuel=1) but limit variation in the other fuel types.

12. Generate graph of uncontrolled (i.e. control=cyclone) emissions as a function of fuel type. After looking at these graphically, test for differences using the STATA procedure "oneway". Because of variability in the data, results do not differ statistically.

This concluded the morning session. The analyses were principally exploratory, but they did indicate that the data on material type (surface/core) and dryer type were inconclusive. These preliminary findings are carried forward in the afternoon session to develop a model for the data.

Comments for the Afternoon Particleboard Log

1. Because the preliminary analyses suggested that wood type might be important and because the effect of dryer type is likely to be important to NCASI, a cross-tabulation was run and a new variable was created to define a variable with a distinct level for each cell in the cross-tabulation for which there are data. The first page of the log is devoted to creating this table.

2. The variable created in (1) was cross-tabulated with control device and fuel type to develop an understanding of the distribution of tests available for analyses. The results indicate that emissions from multiclones and WESP's can be evaluated for only one dryer/wood combination each. Uncontrolled emission data are available for five combinations, but only two of those have more than one test, while limited data are available for several dryer/wood combination for EFB's. For the fuel analyses, both RTP with hardwood (Category 5) and RTP with softwood (Category 6), differences by fuel type could be examined.

3. Graphical comparison using box plots show the effect of fuel for drying hardwood in an RTP. The apparent differences illustrated in the box plots do not account for differences in control level.

4. On the next two pages, an ANOVA analysis is run. These models contain only the main effects of dryer/wood type, control device, and fuel type. One assumption imbedded in this model is that any effects are additive. The ANOVA table indicates that the model explains slightly more than 45 percent of the variance in emissions for these dryers (r^2=0.4576). However, the only factor that showed much relationship was control (based on the size of the p-values in the first table). Note that small p-values signify a result other than chance while large p-values are indicative of no effect. The regression results show that not much is happening, but that is not totally surprising as these data exhibit a substantial amount of variability, and a total of 18 parameters are being estimated.

5. Because the variable typewood showed no effects whatsoever, it was examined further graphically. The plot suggested that an effect might be determined from only the RTP dryer data. That possibility is examined in the next ANOVA model. This model again indicates absolutely no effect for the variable (p>0.95), so a decision was made to drop that variable from the model.

6. The revised model with only fuel type and control device type shows control device to have a strong effect (p=0.03) so the "regress" statement was used to examine the effects of individual levels. The

resulting table has an entry, "_cons" at the beginning of the table. This value represents the mean emission factor for the cell that represents level 6 for control (incinerator) and level 4 for fuel type (unknown). The other parameters in the table represent additive effects relative to that "reference cell". For example, to estimate uncontrolled PM emissions (level 4) from a system using wood (level 1), simply add 0.217, 2.628, and 0.920.

7. The next model considers the effects of control device, fuel, and their interaction (the Cont*Fuel term). The lack of significance for this term (p=0.93) suggests that leaving it out was reasonable.

8. Because the interaction was not significant, it was removed from the model, and the model was rerun. This new model showed fuel to be not significant (p=0.37) so it was removed, and a model that included only control was run. This model was run and showed a significant control effect, so the "regress" command was issued to obtain the individual effects. Again, "_cons" is the estimate of PM emissions for an incinerator controlled RTP dryer, and the other emission levels can be obtained by adding their parameters to those from the "reference cell."

9. Because the ANOVA model was reduced to a single factor, the "oneway" command can be used to compute individual differences and to test the significance of those. Ignoring the bonferroni correction (i.e., divide the p-values in the table by 10) the emissions from EFB's are different from those of multiclones and cyclones but not from the emissions from WESP's. Similarly, multiclones and cyclones do not appear to differ. Based on these results, the groups were combined as indicated in the log (9*).

10. To examine whether emissions from EFB's differ from emissions from WESP's, a new variable was created that included only those two control levels, and a simple two sample t-test was run with the command "test". The results confirm that the two control levels do not result in significantly different emissions, given the variability. However, note that the point estimates do vary somewhat (0.22 for EFB's and 0.48 for WESP's) using only data from RTP dryers.

11. A summary table shows the descriptive statistics for the different control types for RTP dryers. Again, ignoring the PBA category, the results suggest two emission factors are appropriate.

12. A series of boxplots was generated for various groups of data to show that other factors did not matter once we controlled for dryer type and control device.

13. A new variable was created that has a single level for each combination of dryer type (using only RSP and RTP dryers) and control device for which we had PM data. The variables moisture and temperature were examined graphically for each of these categories, and (13*) new variables were created that have the effect of comparing actual temperature to 200°F and actual moisture change to 50 percent. The purpose of these transformations is to generate models that have intercepts that can be interpreted meaningfully, as intercepts in the models now correspond to emissions at temperatures of 200°F and moisture changes of 50 percent.

14. An analysis of covariance model was run with the command ANOVA. Note that the option "cont" was used to signify that these variables are continuous and not factors with different levels. In the first model, temperature was very insignificant (p=0.8), so it was removed and a reduced model was run. In the second model, moisture was not at all significant (p=0.5). Taken together, these results suggest that temperature and moisture are not strongly related, so there is no need to consider them in developing emission factors.

This ends the analysis.