Research & Development

PREDICTION OF HOFFMANN ANALYTES USING TWO DIFFERENT MODELLING METHODS

Jon Sheppard, Nigel Warren and Paul Case

British American Tobacco

British American Tobacco

Aim

- Produce Multiple Regression Equations which can predict Hoffmann Analytes in mainstream smoke accurately

Factors / Regimes

- Tip Ventilation (TV)
- Paper Permeability (PP)
- Filter Pressure Drop (FPD)

- Squared terms (TV^2, PP^2, FPD^2)
- Interaction terms (TV*PP, TV*FPD,PP*FPD, TV*PP*FPD)

- Model 1 – ISO (35/2/60) & Intense (55/2/30) regimes
- Model 2 – ISO (35/2/60) regime only

Data & Blend Model 1

- Virginia tobacco blend only
- Used 16 data points from the central composite design
- 8 corner points, 6 star points and 1 centre point which was duplicated

Starting Multiple Regression Equation Model 1

- Predicted Value = e.g. NFDPM

$$a + b*TV + c*PP + d*FPD + e*TV^2 + f*PP^2 + g*FPD^2 + h*TV*PP + j*TV*FPD + k*PP*FPD + m*TV*PP*FPD + \xi$$

- TV = %Tip Ventilation
- PP = Paper Permeability
- FPD = Filter PD Vents Closed
- ξ = error
- a = constant
- b, c, … m are coefficients.

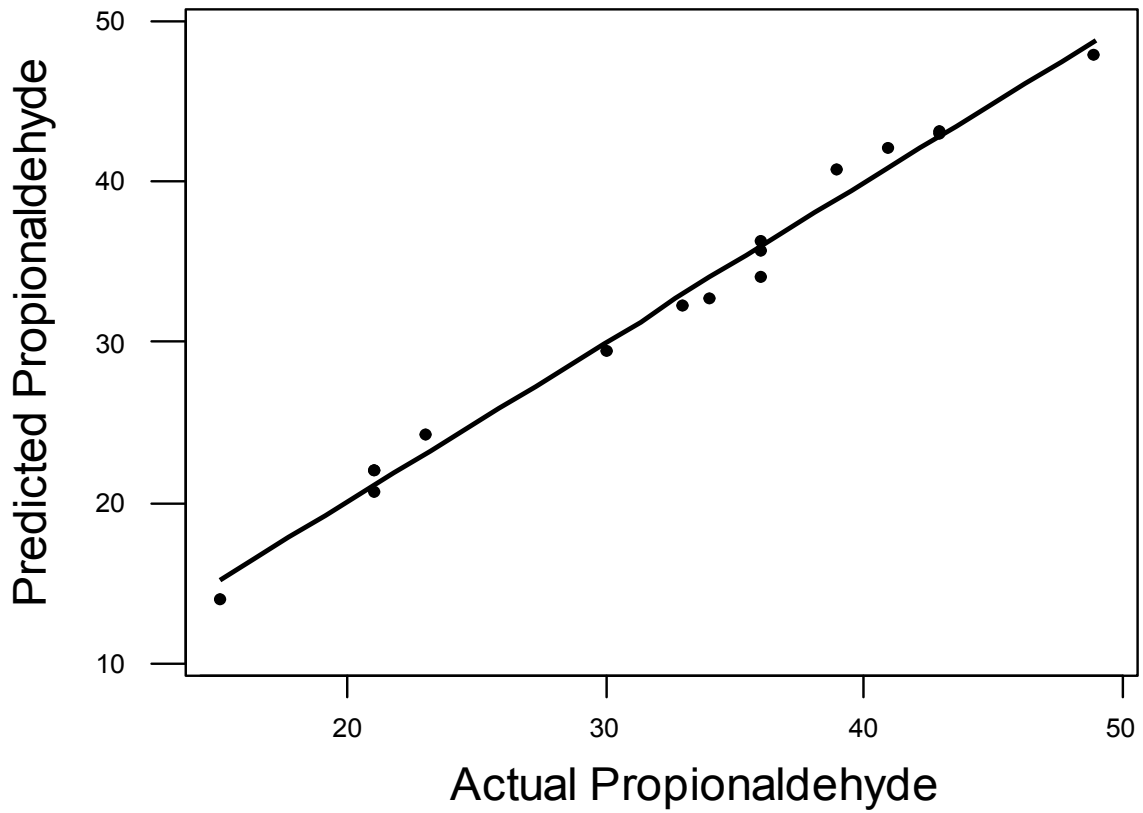
Multiple Regression : Model 1

- Using a stepwise backward elimination regression method
- Run starting multiple regression equation. Examine p-values of 10 factors. Remove factor with the largest p-value ≥ 0.05
- Re-run regression equation with 9 factors. Examine p-values, remove largest p ≥ 0.05
- Continue the process of removing factors, until all remaining factors have a p-value < 0.05
- Obtain best predictive multiple regression equation for the Hoffmann analytes

Results Model 1
Virginia Blend Only

R-Sq ≥ 90.0% Analyte	ISO R-Sq %	Intense R-Sq %	80% ≤ R-Sq ≤ 89.9% Analyte	ISO R-Sq %	Intense R-Sq %	R-Sq ≤ 79.9% Analyte	ISO R-Sq %	Intense R-Sq %	No Equation Analyte	ISO R-Sq %	Intense R-Sq %
Acetaldehyde	99.9	95.9	Arsenic	89.3	97.3	NNK	79.8	78.9	1-aminonaphthalene	10.6	59.5
Acrolein	99.7	95.5	Crotonaldehyde	87.2	74.3	Quinoline	78.1	82.8	2-aminonaphthalene	9.2	6.8
NFDPM	99.1	97.9	NAT	85.2	91.9	Chromium	77.6	39.4	4-aminobiphenyl	8.3	0.7
Propionaldehyde	98.9	98.9	Ammonia	80.7	87.0	Mercury	76.3	3.9	3-aminobiphenyl	5.2	2.1
Methyl Ethyl Ketone	97.4	79.0				Lead	76.0	10.7	Toluene	-	13.0
Styrene	97.0	86.9				Resorcinol	74.3	49.5	NAB	-	82.4
Nicotine	96.9	96.7				Acrylonitrile	74.3	41.0			
CO	96.4	96.2				Benzene	73.5	38.0			
HCN	96.0	84.1				Isoprene	67.1	51.3			
Phenol	95.9	90.9				Selenium	65.7	28.3			
Formaldehyde	95.8	92.4				1, 3-Butadiene	56.3	45.7			
o-Cresol	95.2	86.9				Nickel	40.0	4.8			
Catechol	95.1	81.3									
p-Cresol	94.6	85.6									
Butyraldehyde	94.5	82.2									
Hydroquinone	94.5	84.0									
Cadmium	94.3	81.4									
Benzo*(a)* pyrene	93.0	93.3									
NNN	91.9	85.9									
Puff Number	91.1	88.5									
m-Cresol	90.8	77.4									
Acetone	90.7	94.5									
Pyridine	90.4	81.8									

Predicted vs Actual Plot



Check Point Comparison Model 1

	ISO Check Point		Intense Check Point	
Hoffmann Analyte	**Actual**	**Predicted**	**Actual**	**Predicted**
Acrolein	46.0	44.4	103.0	99.2
Acetone	177.0	176.5	367.0	374.9
Phenol	25.2	25.2	50.6	46.6
Benzo(a)Pyrene	12.6	10.9	25.4	22.1
Styrene	5.3	5.3	18.6	18.1
NAT	26.2	25.3	50.0	49.8
Crotonaldehyde	14.0	17.6	50.0	35.2
Chromium	9.8	16.3	45.8	26.5
Isoprene	262.8	240.0	468.4	561.5
Toluene	44.3	-	74.3	-
2-aminonaphthalene	7.0	-	16.7	-
4-aminobiphenyl	1.0	-	3.2	-

Data & Blend Model 2

- Virginia, Burley, Oriental and Mix (50% Virginia, 50% Burley)
- Used 64 data points for NFDPM, Nicotine, CO and Puff Number
- Used 44 data points for the other Hoffmann analytes

Proportionality Test

- Test the hypothesis that the ratios are independent of the four tobacco blends using analysis of variance (ANOVA)

- Calculate ratios of the designed experimental points to the centre points for each tobacco blend

- Normalise the ratios using the following equation

$$\frac{H_i}{H_c} \times \frac{TV_i}{TV_c} \times \frac{PP_i}{PP_c} \times \frac{FPD_i}{FPD_c}$$

ANOVA Example

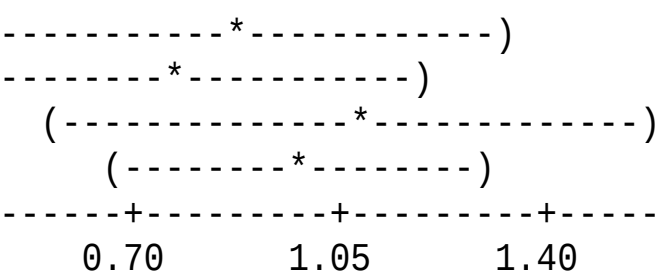
Analysis of Variance for Benzo(a)pyrene

Source	DF	SS	MS	F	P
Tob Blen	3	0.376	0.125	0.35	**0.789**
Error	32	11.424	0.357		
Total	35	11.800			

Level	N	Mean	StDev
Burley	8	0.8855	0.6089
Mix	8	0.7664	0.5800
Oriental	6	1.0685	0.6548
Virginia	14	0.9776	0.5772

Pooled StDev = 0.5975

Individual 95% CIs For Mean
Based on Pooled StDev



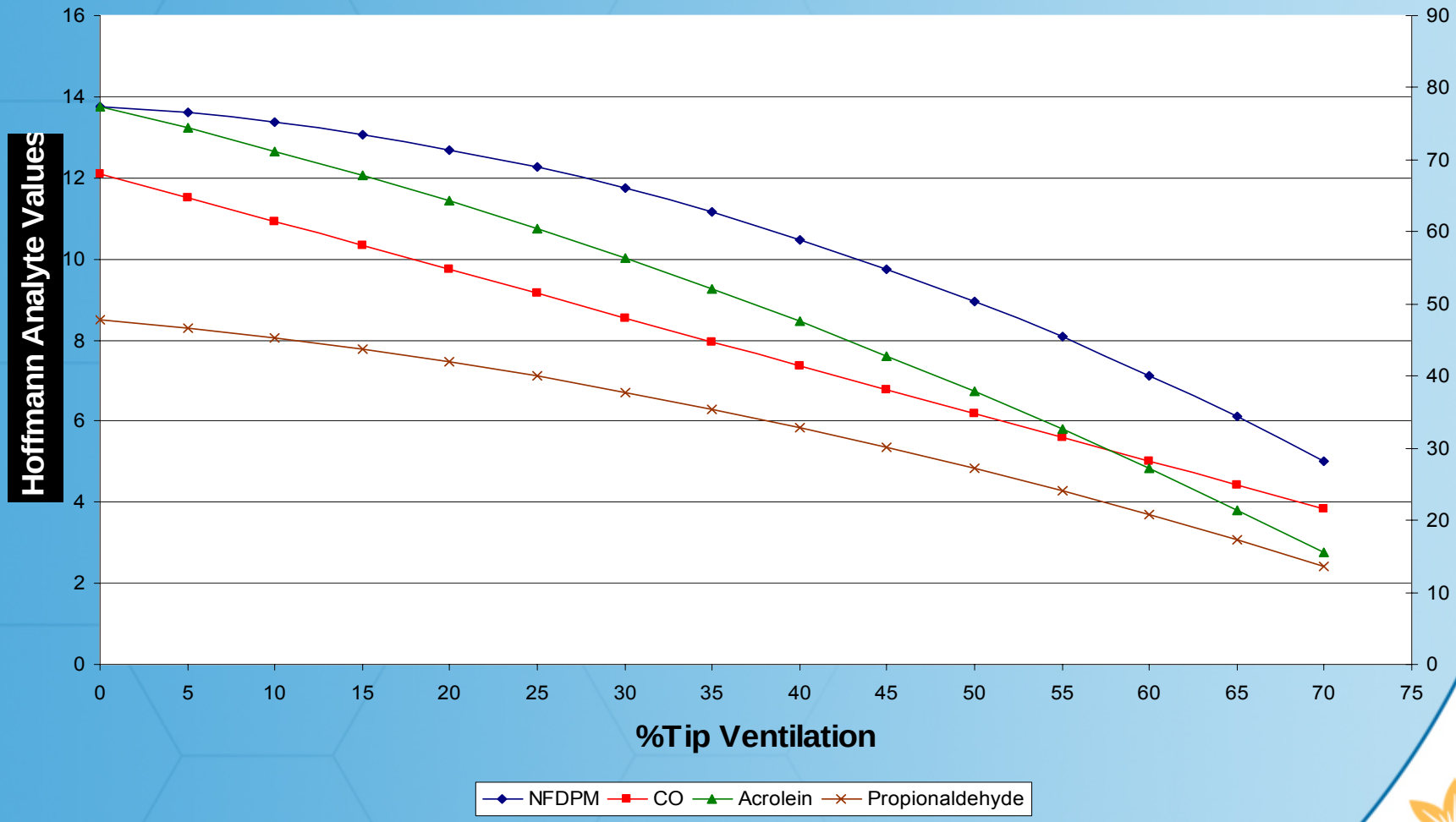
Proportionality Conclusion

- The ratios are independent of the blend type
- The four tobacco blends can be treated as one data set for the multiple regression analysis

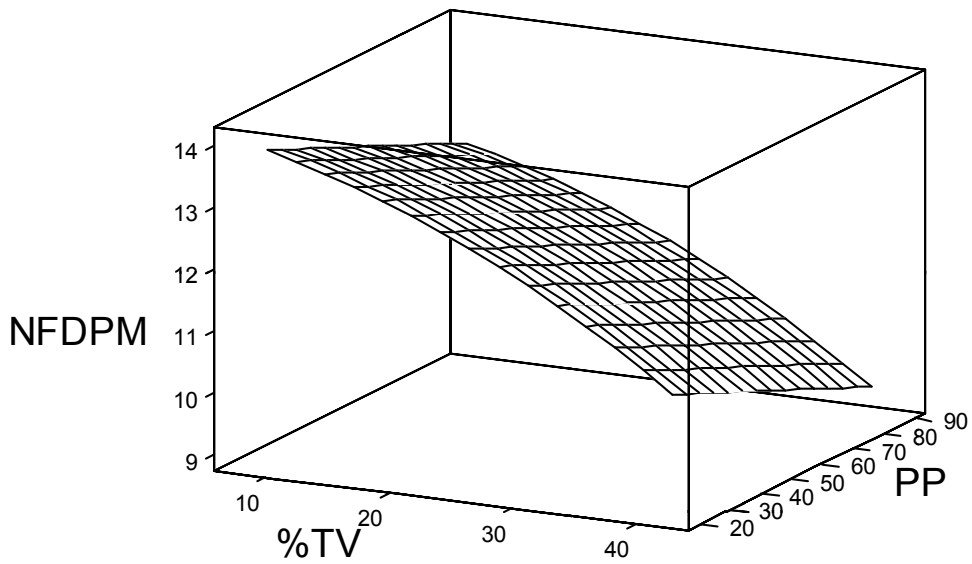
Results Model 2
All Blend Types

R-Sq ≥ 90.0% Analyte	ISO R-Sq (%)	80% ≤ R-Sq ≤ 89.9% Analyte	ISO R-Sq (%)	R-Sq ≤ 79.9% Analyte	ISO R-Sq (%)
Acrolein	97.1	Propionaldehyde	89.7	Quinoline	79.9
HCN	96.7	Acetone	89.3	MEK	79.0
Acetaldehyde	96.7	NOx	88.9	Toluene	78.3
Nicotine	95.2	NNK	88.8	NAT	78.2
CO	95.0	Hydroquinone	88.5	Crotonaldehyde	75.8
Styrene	94.7	Catechol	88.4	Butyraldehyde	75.1
NFDPM	94.2	NO	87.8	Resorcinol	73.6
Pyridine	92.6	NNN	87.7	NAB	71.4
Formaldehyde	90.7	Phenol	85.4	Isoprene	70.9
		m-Cresol	85.4	1, 3 Butadiene	70.5
		o-Cresol	84.6	Ammonia	52.3
		p-Cresol	84.5	1-aminonapthalene	44.3
		Benzene	84.2	2-aminonapthalene	41.1
		B(a)P	81.8	3-aminobiphenyl	33.4
		Acrylonitrile	80.8	4-aminobiphenyl	30.9

Comparison of four Hoffmann Analyte values when %Tip Ventilation is increased (Virginia Blend PP = 40mg FPD = 80mg)



Wireframe Plot of NFDPM



Conclusion \ Limitations

- Given various cigarette design parameters it is possible to predict the resultant Hoffmann analyte yields for cigarettes
- Most multiple regression equations produced for Models 1 and 2 give good prediction equations. Predict most Hoffmann analytes very accurately using both models
- A percentage change in Hoffmann analytes should be consistent across all blend types
- The results can be illustrated in a number of different meaningful ways
- The multiple regression equations produced are only as good as the analytical data provided