



- (51) **International Patent Classification:**
G05B 13/02 (2006.01) **G06Q 40/00** (2012.01)
G06F 17/40 (2006.01)
- (21) **International Application Number:**
 PCT/US2016/041182
- (22) **International Filing Date:**
 6 July 2016 (06.07.2016)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
 62/189,110 6 July 2015 (06.07.2015) US
- (71) **Applicant:** THE UNIVERSITY OF WYOMING RE-
 SEARCH CORPORATION D/B/A WESTERN RE-
 SEARCH INSTITUTE [US/US]; 3474 North 3rd Street,
 Laramie, WY 82072 (US).
- (72) **Inventors:** GLASER, Ronald, R.; 153 North Adams
 Street, Laramie, WY 82070 (US). TURNER, Thomas, F.;
 507 South Grant Street, Laramie, WY 82070 (US).

PLANCHE, Jean-Pascal; 4231 Bobolink Lane, Laramie,
 WY 82070 (US).

- (74) **Agent:** WIEDMANN, Alfred, K.; Santangelo Law Of-
 fices, P.C., 125 S. Howes Street, 3rd Floor, Fort Collins,
 CO 80521 (US).

- (81) **Designated States** (*unless otherwise indicated, for every
 kind of national protection available*): AE, AG, AL, AM,
 AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,
 BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM,
 DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
 HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR,
 KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG,
 MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM,
 PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC,
 SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN,
 TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

- (84) **Designated States** (*unless otherwise indicated, for every
 kind of regional protection available*): ARIPO (BW, GH,
 GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ,
 TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU,
 TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE,

[Continued on next page]

- (54) **Title:** METHOD FOR CORRELATING PHYSICAL AND CHEMICAL MEASUREMENT DATA SETS TO PREDICT
 PHYSICAL AND CHEMICAL PROPERTIES

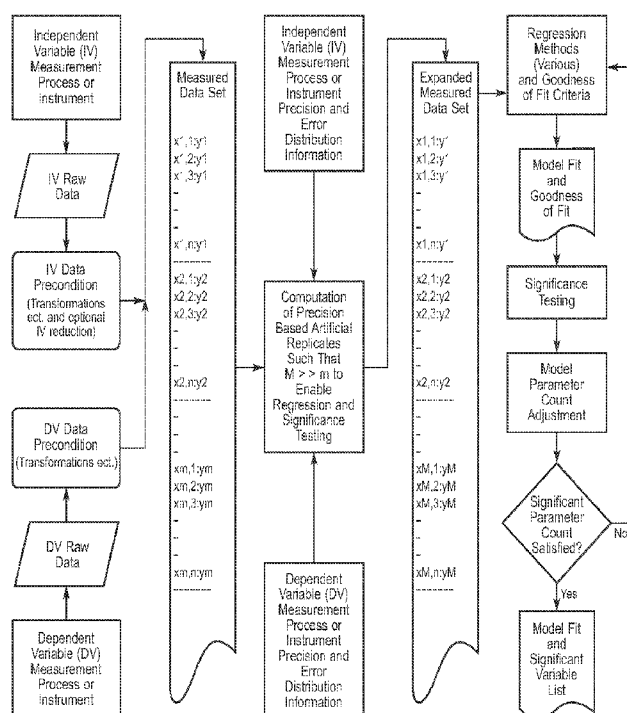


Fig. 1

- (57) **Abstract:** The present invention is generally related to the correlation of physical and/or chemical measurements with other physical and/or chemical measurements and the application of the correlation to transform a product or process (e.g., to formulate, mix, blend compounds or materials of various natures and origins) upon predicting/estimating certain property(ies) and/or performance index(ices) as indicated by a dependent variable estimate. Embodiments of the inventive technology applies specifically to the problem of producing a correlation when the independent variables of interest exceed the number of observations. This situation is common in many fields of science and technology, such as, but not limited to, spectroscopy, calorimetry, thermogravimetric, chromatography and others. A perhaps primary advantage of embodiments of the inventive method over prior art is the ability to generate correlations directly in terms of measured variables.



DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, **Published:**

LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE,

SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA,

GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

— *with international search report (Art. 21(3))*

Declarations under Rule 4.17:

- *as to applicant's entitlement to apply for and be granted
a patent (Rule 4.17(ii))*

METHOD FOR CORRELATING PHYSICAL AND CHEMICAL MEASUREMENT DATA SETS TO PREDICT PHYSICAL AND CHEMICAL PROPERTIES

5 CROSS-REFERENCE TO RELATED APPLICATIONS:

This international patent application claims priority to and the benefit of US Provisional Application 62/189,110, filed July 6, 2015, said provisional application incorporated herein in its entirety.

10

STATEMENT REGARDING FEDERAL RIGHTS

This invention was made with government support under contract DTFH61-07-D-00005 awarded by the U.S. Department of Transportation. The government has certain rights in the invention.

15

TECHNICAL FIELD

The inventive technology disclosed herein is especially useful where insufficient observations are available compared to the number of independent measurement variables available. This situation is common in many fields of science and technology, such as spectroscopy, calorimetry, thermogravimetric, chromatography, and others.

20

BACKGROUND ART

25

Modern analytical techniques often rapidly produce quite large data sets, the most common are those data sets generated using a spectrometer and are usually described as “spectra”. However, observations can be made using a wide variety of instruments/methods to generate different data; often one observation using a single instrument or can generate many pieces of data (e.g., a single IR spectrometer observation can generate 4000 absorbance data for 4000 different wave numbers). Any set of data that can be formulated as a response as a function of an index (time,

30

wave number, temperature, etc.) can be treated as a “spectrum”, although in some scientific parlance that term is reserved for spectrometer generated data.

Consequently, a thorough examination of the relationships between a given spectrum (or other arbitrary data matrix) type and an independently measured material property can generally expressed by the following general relationship:

$$y = f(x_0, x_1, x_2, \dots, x_n) \quad (1)$$

Where:

y is the dependent variable, e.g. complex modulus (as but one of many examples, including generally, but not limited to, either chemical or physical properties; and durability from properties measured at various aging stages, unaged and aged; see additional discussion below)

x_i is the independent variable(s), e.g. IR absorbance at wave number I (when the measuring instrument is, e.g., an IR spectrometer)

If $f(x)$ is assumed to be algebraically linear, and we have only 3 spectra representing 3 materials or conditions (a,b,c, such as a first asphalt, a second asphalt, and a third asphalt; a first crude oil, a second crude oil and a third crude oil, as but two of many different examples, or three different temperatures or other conditions), along with their dependent properties of interest, the equation set is:

$$y_a = k_0 x_{a0} + k_1 x_{a1} + k_2 x_{a2} + \dots + k_n x_{an} \quad (2)$$

$$y_b = k_0 x_{b0} + k_1 x_{b1} + k_2 x_{b2} + \dots + k_n x_{bn} \quad (3)$$

$$y_c = k_0 x_{c0} + k_1 x_{c1} + k_2 x_{c2} + \dots + k_n x_{cn} \quad (4)$$

where k is the proportionality constant for each wave number (or for each oil fraction, e.g.) 0 through n .

Since this is a curve fitting problem, the x and y pairs are known, and we seek k 's that satisfy the equation set. Such a deterministic solution is impossible if $n+1$ exceeds the number of observations. When the number of observations is exactly equal to $n+1$, then the fit is perfect, meaning no statistical evaluation of the fit quality is possible. This is analogous to the situation in two dimensions where you are fitting a line to 2 data points, obtaining a correlation coefficient of 1. To obtain a statistically meaningful test of a multidimensional fit, the observations should exceed the independent variable count by some factor, the larger the better, but generally conceded to be a multiplier of 7. Typical mid-infrared spectra will contain nearly 4000 wave numbers, so the examination of each and every wave number for significance when combined with the others would require 28000 measurements, clearly not practical. This situation is a recurring problem with spectral data and other extensive xy data sets as well, as the inclusion of all of the data results in an equation system with excessive adjustable parameters that is impossible to solve. A number of approaches exist for addressing this problem with a variety of strategies aimed at essentially reducing the number of effective k 's (independent variable fit parameters) to be discovered.

The WRI chemometric software is especially useful for applications where insufficient observations are available compared to the number of independent measurement variables available. This situation is common in many fields of science and technology, such as spectroscopy, calorimetry, thermogravimetric, chromatography and others.

A few of the methods used to address the problems outlined above are briefly described in the next three sections. These methods, particularly the process of correlating spectral data to other process or property variables, have been used successfully in a wide range of applications, but do not produce a closed form equation in terms of measured quantities, limiting their usefulness in fundamental scientific studies.

Multiple Linear Regression Multivariable regression is a time-honored technique going back to Pearson's 1901 use of it. Multivariable regression can establish that a set of independent variables explains a proportion of the variance in a dependent variable at a significant level (through a significance test of R^2), and can establish the relative predictive importance of the independent variables (by comparing beta weights). Variable transformations (most common is the logarithm) can be applied to independent or dependent variables to explore some curvilinear effects, and polynomials can be fit as well by expanding independent variables into a power series.

Multivariable linear regression can solve the matrix $Y=MX+B$, provided sufficient measurements of Y exist to obtain all of the coefficients in vector M . To be statistically meaningful, measurements of Y in excess of measurements of X must be available, meaning that a spectra of 3500 wave numbers would require at least 3500 measurements of, say, complex modulus. To be statistically reliable, 35000 would be better. It is generally impossible to apply multivariable linear regression directly to correlation studies involving data rich spectral data. However, the preconditioning of individual data points to related groups (spectral peaks, for example) is helpful to reduce the independent variable count. However, this is usually not sufficient unless a very extensive data set (many observations) is available. A variety of computation approaches have been developed in recent years that address this problem by projecting the data in one way or another into a smaller list of independent variables. These include Principle Component Analysis (PCA), Partial least Squares (PLS), and others.

Principle Component Analysis and Principle Component Regression Principle Component Analysis techniques are applied to the problem of too many x measurements relative to y measurements by searching for so-called latent variables. The covariance of XX' is examined and parameter space axis rotations are employed to arrive at new coordinates based on eigenvectors of the XX' matrix. In simple terms this means that independent variables that appear to change in a similar fashion are grouped. The translated x variables (often called indicator variables) are projected into a smaller parameter space of latent variables. It is implicitly assumed that these fictitious latent variables somehow describe a truer "latent structure" to the system. Recall that the underlying mathematical model for the entire data set is

linear, often patently untrue in chemical systems. This technique results in latent variable data sets with improved variance in the hopes of improving signal to noise ratios. Often, however, irrelevant data included in the translations pass spurious noise to the latent variables.

- 5 Principle components regression (PCR) is the application of ordinary linear regression methods to the latent variables developed from the principle components analysis. The difficulty with this method is that the complex axis rotations make understanding what the latent variables represent in terms of measurable quantities difficult. Interpretation of the results in terms of chemistry and physics is difficult and requires sensitivity testing by varying the input data. While useful for
10 calibration within the testing range of the data employed, using this method for understanding the underlying science is difficult.

Partial Least Squares PCR is based on the spectral decomposition of XX' to select latent variables for regression, while PLS is based on the singular value decomposition of $X'Y$. In
15 practice, PLS usually fairs better than PCR since the reduction of parameter space dimensions is accomplished through comparison of the independent variables with the dependent variables. PCR, on the other hand, focuses mainly on what can be thought as the signal strengths of the independent variables alone for parameter space reduction, and is therefore more prone to the introduction of irrelevant signals into the regression. As with PCR, PLS suffers from the
20 difficulty that the complex axis rotations make understanding what the latent variables represent in terms of chemistry and physics difficult and requires sensitivity testing by varying the input data. While useful for calibration within the testing range of the data employed, using this method for understanding the underlying science is difficult.

25 **Other Methods** Many other algorithms have been developed in recent years, including neural networks and artificial intelligence. While these “black box methods” can work extremely well over the calibration range used, we still are faced with the difficulty of understanding how the input variables relate directly to dependent variable without sensitivity testing.

30 Because of the difficulty of latent variable methods to demonstrate the correlation in terms of directly measured variables, we developed our own methods to address the issues associated with

impossible and/or unfavorable parameter to observation ratios. In the simplest of terms, two strategies can be employed to make the problem tractable; reduce the independent variable count, or increase the number of observations. Once a statistically meaningful correlation can be computed, a method for selecting the most important independent variables must be applied to remove irrelevant signals and find those responses that significantly affect the quality of the fit. When applying this technique to infrared spectroscopy-to-rheology correlations, the independent variables are spectral wave numbers and represent vibrational modes of functional groups. Consequently, important clues about how chemical changes cause rheological changes can be obtained.

SUMMARY OF THE INVENTION

This method provides a process to generate correlations between physical and chemical measurements, chemical and chemical measurements, and physical and physical measurements when sufficient observations are not available to perform the correlation while examining all of the measurements at once. Indeed, embodiments of the inventive “chemometric” software are especially useful for applications where insufficient observations are available compared to the number of independent measurement variables available. Unlike prior art, which reduces the independent variable count to “latent variables” through dimensional reduction accomplished by complex rotations and projects, as is the case for partial least squares and principle components analysis schemes, embodiments of the inventive method produce correlations that are expressed in closed form mathematical equations in terms of the measured values of significance.

Stepwise multivariable regression also produces correlations in measured value terms, but is unable to examine all combinations in the independent variable list at once; hence some combinations are not tested. Prior art focuses upon independent variable reduction schemes, while this method uses independent variable reduction scheme cast explicitly in terms of the measured values, and, uniquely can also expand the data set by producing additional artificial observations based upon the known (or determined or estimated) precision of the measurement methods. This expansion of the regression data set provides a key method for not only producing a “fit” of the data, but also assessing the significance of the parameters used using any of a

variety of well-established statistical methods for estimating parameter significance and parameter rejection criteria.

The invention comprises using an approach employing the new chemometrics software with data from a one or more chemical and spectroscopic analysis methods to generate relationships with selected physical properties. The results of the correlations will provide equations that could be interpreted in a manner enabling an understanding of how the analysis results reflect the physical behavior. This approach can be used to evaluate current properties and to predict changes in properties following aging or treatment.

The present invention is generally related to the correlation of physical and/or chemical measurements with other physical and/or chemical measurements. This method applies specifically to the problem of producing a correlation when the independent variables of interest exceed the number of observations. The advantage to this method over prior art is the ability to generate correlations directly in terms of measured variables. The WRI chemometric software is especially useful for applications where insufficient observations are available compared to the number of independent measurement variables available. This situation is common in many fields of science and technology, such as spectroscopy, calorimetry, thermogravimetric, chromatography and others.

BRIEF DESCRIPTION OF DRAWINGS

Figure 1 is a flow chart of the chemometric method to obtain relationships between independent variables measured and dependent variables measured.

Figure 2 is an example of grouping of IR spectra absorbances with absorbance at 2000 cm^{-1} .

Figure 3 is a graph showing one example of modified automated SAR-AD separation profile of an asphalt.

Figure 4 is a graph showing one example of size exclusion chromatography (RI detector) profiles for eight asphalts.

Figure 5 is a graph showing one example of penetration (PEN) correlation coefficients.

DESCRIPTION OF EMBODIMENTS OF THE INVENTIVE TECHNOLOGY

5 As mentioned earlier, the present invention includes a variety of aspects, which may be combined in different ways. The following descriptions are provided to list elements and describe some of the embodiments of the present invention. These elements are listed with initial embodiments, however it should be understood that they may be combined in any manner and in any number to create additional embodiments. The variously described examples and
10 preferred embodiments should not be construed to limit the present invention to only the explicitly described systems, techniques, and applications. Further, this description should be understood to support and encompass descriptions and claims of all the various embodiments, systems, techniques, methods, devices, and applications with any number of the disclosed elements, with each element alone, and also with any and all various permutations and
15 combinations of all elements in this or any subsequent application.

An assigned linear dependence of a dependent variable on a plurality of independent variables may be as follows:

$$20 \quad y_a = k_0 x_{a0} + k_1 x_{a1} + k_2 x_{a2} + \dots + k_n x_{an} \quad (2)$$

$$y_b = k_0 x_{b0} + k_1 x_{b1} + k_2 x_{b2} + \dots + k_n x_{bn} \quad (3)$$

$$y_c = k_0 x_{c0} + k_1 x_{c1} + k_2 x_{c2} + \dots + k_n x_{cn} \quad (4)$$

25

where k is the proportionality constant for each wave number (if the measuring instrument is an IR spectrometer, or for, e.g., each oil fraction, (if the measuring instrument is a SAR-AD analyzer), as but two examples, 0 through n. Note that a single observation can produce “n” measurements (e.g., where the observation instrument is an IR spectrometer, perhaps 4000
30 measurements (“n” = 4000) are made during that single observation (one for each wavelength); where the observation instrument is SAR-AD, perhaps 16 measurements are made (“n”=16).

Note also that while certain embodiments may include the step of performing “p” number of observations to obtain “p” number of measurements for each said dependent variable and said independent variables, wherein “p” is less than the sum of “n” + 1, other embodiments may include the step of performing “p” number of observations to obtain “p” number of
5 measurements for each said dependent variable and said independent variables, wherein “p” is less than “n.”

Where a relationship between a dependent variable (e.g., of a product, process, ingredient of a product, material that is acted on or used in any way in a step of a process, etc.) and a plurality “n” of independent variables is either known to be linear, suspected as linear, presumed linear
10 (whether to test a fit or other reasons), or in any way treated as linear (whether by computer, software, operator, etc.), it is said that linear dependence of the dependent variable on “n” number of independent variables is assigned. Note that a computer that in any way treats, mathematically, e.g., via coded instructions, said relationship as linear is said to assign such linear dependence. Even where it may eventually appear that some of the independent variables
15 do not have a linear impact on the dependent variable, that does not prevent the fact that at the initial stages of the inventive protocol disclosed herein, such linear dependence was assigned.

X Measurements (measurements of the independent variables) include but are not limited to the following, or measurements of the following phenomenon/properties, or measurements made using the following analysis/instruments:

- 20 - temperature, asphaltene %, IR wave number/length, UV Absorbance;
- Spectroscopy: IR, NIR, MIR wavelengths and band intensities, NMR displacement, peak intensity, UV, RAMAN, SAX, SANS and XRay diffraction...;
- Composition: elemental analysis, metal content;
- Microscopy and image analysis, Electronic, Optical, Atomic (AFM), tomography, MRI;
- 25 - Thermal properties: DSC glass transition temperature, crystallinity, TGA weight loss, HP DSC oxidation induction time;

- Separation: All SAR-AD, WAX-AD, SARA fractions and indices, Automated Flocculation Titrimeter (AFT) indices, GPC molecular weight or retention times and intensities, IEC (note that more details regarding the SAR and SARA separation may be found in Boysen, Ryan and Schabron, John, Automated HPLC SAR-AD Separation; Fundamental Properties of Asphalts and Modified Asphalts III Product: FP 01, March 2015, which is incorporated herein by reference in its entirety);

- Olefin index;

- Acidity-Basicity: TAN, TBN;

Standard procedures may be used to estimate the instrument or method measurement precision, if it is not known (e.g., as provided by the instrument manufacturer or process standardizing body such as ASTM), which may include the precision distribution, of the methods used for the collection of independent and dependent data. These measurements may be spectrographic, chromatographic, calorific, gravimetric, thermo-gravimetric, or even any measurement process that produces a numerical value. A flow chart of the chemometric method is provided in Figure 1. Indeed, these steps may describe identically an embodiment of the inventive technology. In addition to other measuring instruments/analytical tools indicated above and elsewhere herein, measurements (particularly of chemical properties) may be obtained using, e.g., using the following analytical tools/measuring instruments alone or in combination, as examples and non-exclusively :

- IR, NIR, MIR;
- SAR-AD, WAD and other SARA methods;
- NMR (1H and 13C);
- GPC / SEC;
- DSC;
- IEC; and
- AFT, as but a few examples.

More particularly with regard to physical properties, the following are merely examples of the many instruments/analytical tools that could be used to generate measurements of such properties: DSR, BBR, ABCD, DMA, mechanical test, and fouling apparatus, as but a few examples.

- 5 Materials that may be measured, materials for which a dependent variable may be estimated, materials that may be transformed, and materials as to which a process may be transformed using the inventive method (e.g., by estimating a dependent variable) include but are not limited to: petroleum, coal, and biomass products, fuel, medication, dietary supplements, cosmetics, food, lubricants, and any other materials indicated in this application, or in references incorporated
- 10 herein. Such material(s) may certainly be related to the product or process that is transformed (indeed, such material may be that product); that material(s) may be an ingredient in a process, anyway involved in a step of the process, may be a part of the product, a material that the product is a part of, as but a few examples.

- Y measurements (and indeed parameters that can be estimated upon determination of coefficients
- 15 of significant independent variables) include but are not limited to the following parameters/indexes/properties, and/or parameters/indexes/properties related to the following phenomenon, and/or measurements thereof:

- Any measurement related to crude oil / petroleum fouling, coking, emulsion ability, stability, instability;
- 20 - Gas cetane, octane numbers;
- Any measurements related to lubricants: anti-wear properties, viscosity index, oxidation resistance, fluidity, tribology;
- Asphalt penetration, ring and ball softening point, fraass brittle point, viscosity, modulus, phase angle, Superpave properties, DSR, BBR critical temperatures, oxidation resistance
- 25 short term and long term;
- Material fatigue resistance, brittleness, hardness, elasticity, plasticity, deformation, roughness, density;

- Any index related to organic, inorganic material Oxidation, weatherability, durability inflammability, explosiveness, carcinogenicity, mutagenicity;
- Metal corrosion;
- Liquid or paste fluidity, thixotropy, viscosity, density;
- 5 - perfume smell, spraying ability.
- medication efficiency/effectiveness,
- Viscosity, material hardness, reflectivity

Often, but not always, X measurements are measurements of a property, phenomenon, etc. of the same material that the Y measurements relate to.

10

Preconditioning The raw data may be preconditioned by any number of variable transformations, ratios, normalization as deemed useful for data interpretation. Grouping/Consolidation may be viewed, in certain embodiments, as a type of preconditioning.

- 15 **Grouping/Consolidation of Independent Variables** Additionally perhaps, and possibly but not necessarily as a preconditioning step, the independent variable list count may be reduced by correlating an independent variable with one or more of the remainder of the independent variables, and consolidating/grouping, where appropriate, them into a single variable if the quality of fit of a given variable pair (or generally, grouping, if there are more than two
- 20 independent variables that sufficiently correlate) exceeds a user defined value (e.g., where R^2 corresponding to the two or more independent variables is above a certain value). An example of this is provided in Figure 2. The consolidation of several independent variables may be based upon any of a number of criteria, such as best signal to noise ratio, averages, weighted averages, geometric means, or other formulations best suited to the type of measurement involved. Note
- 25 also that the term independent variables includes even those variables that, after analysis (e.g., grouping analysis), are found not to be entirely independent of one another (e.g., when one increases, another increases in linear relation).

Grouping combines independent variables together if they contain the same information. This reduces the number of independent variables without reducing information content. Grouping may find particular application to spectra where there are thousands of data points, many of which are most likely not relevant. F-Test (Add/Reject) may remove the least significant variables, i.e., those with no statistically relevant information. Grouping may be a type of preconditioning. In certain embodiments, if grouping does not reduce the number of independent variables enough, then generation of artificial replicates may be required.

Of course, the highest quality data is desired. Grouping (like artificial replication) cannot improve data quality; instead, its main benefit may be speeding the analyses. At present, large datasets require days to analyze using a computer (and indeed using the inventive software where independent variables have not been sufficiently reduced), and days for us to evaluate the results, and more days to write a report. Without grouping, at times, weeks could be required to obtain results; with grouping (and perhaps other steps such as independent variable reduction using, e.g., F-Test (Add/Reject)); with grouping and perhaps other data reduction, results can be obtained much faster.

Artificial Replication In order to make the correlation mathematically tractable, the number of observations must exceed the number of independent variables by one. Statistical validity requires a larger number, generally considered to be at least seven times the number required to solve the equation matrix. For simple systems, one can design the experiment or measurement program to collect sufficient observations, but often this is not practical. By knowing the precision of the measurement method (e.g., of the instrument used to measure, and/or the method or protocol (e.g., ASTM method) used), and the distribution of the precision error (collectively these may be referred to as measurement precision), a set of artificial measurements can be generated that looks very much like the set that would be created through physical measurement. These artificial replicates are created by averaging physically produced replicates to get the best mean value, if available, and then each independent variable measurement is used as a basis for creating an additional artificial replicate using “measurement precision”, which, in particular embodiments, suggests adding or subtracting a randomly generated value within the range of precision of the measurement at a frequency determined by the (known or estimated) distribution

that is characteristic of the particular measurement type used. Note that in some embodiments, one example of accomplishing the step of replication of artificial data at a measurement instrument's/measuring method's frequency may be conceptualized as selecting an unmeasured value from a series of concentric circles or annular rings centered on a measured value (or a mean/median, etc. of several measured values), each circle or ring corresponding to a range of values (e.g., uppermost and lowermost, diametrically opposed portions of one ring may correspond to a range from 17.05 to and including 17.23) and to a respective probability (e.g., a smallest radius ring may have a probability of 15%, a next largest (perhaps having an identical radial width) may have a probability of 9%, the next largest ring (perhaps having an identical radial width) may have a probability of 3%, etc.); whenever this conceptual model suggests datum generation within a certain ring, a random number generator may be used to generate a number within the range represented by that ring. Accordingly, in some embodiments, replicating artificial data using measurement precision may involve the steps of adding a number selected from certain appropriate ranges at frequencies corresponding to those ranges (for example, in the 17.05-17.23 range example given above, if a measurement within this range is expected to occur 5% of the time, then a random number between 0.01 and 0.18 may be selected and then added to 17.05 for 5% of the artificially generated measurements). However, this is merely one example of many different ways in which random numbers and a frequency profile could be used to artificially replicate data. The replication of artificial data process may continue until sufficient artificial replicates are generated to expand the experimental matrix to a size suitable for statistical study and mathematical tractability. Artificial replication, like grouping, cannot improve data quality, but can make an analysis possible. Typically, measurement precision of both the independent variables and the dependent variables is of concern and is considered.

Model Fitting Any form of regression, or curve fitting can be employed with the expanded data set. Any number of linear and non-linear algorithms may be employed.

Determining Statistically Significant Variables / Parameter Reduction A number of parameter reduction criteria exist in the prior art, including p test, F test, rate of decline of goodness of fit, and others. These often require multiple fits to compare the correlation quality

with and without the particular independent variable being tested. So, parameter reduction to determine the important factors in the phenomena under study usually involves repeated fitting and quality of fit comparisons, with less significant variables being rejected one by one. Independent variables that have a statistically insignificant impact on the dependent variable can
5 be ignored without having a statistically significant impact on results (or without impairing results to an unacceptable degree), but those having a statistically significant impact on the dependent variable are considered, and coefficients for them are later determined (coefficients for statistically insignificant independent variables may be set to zero). This differs from stepwise regression in that all variables are correlated initially, and removed one by one, rather
10 than being added to the model one at a time. Rejection criteria can vary and often require the judgment of the investigator. After it is determined which independent variables are statistically significant (or which are not, which could yield similar or identical information), the independent variable count may be adjusted.

15 The end result of this method is a series of models with decreasing parameter counts and quality of fit metrics, all in terms of measured quantities. The advantage to this over prior art is particularly acute in fundamental research where causality can be studied with further testing. Note that grouping of independent variables is not always required to obtain a soluble equation set, and this method does require grouping in all cases. In addition, prior art approaches to
20 independent variable reduction, such as Principle Components analysis, Partial least Squares, Neural Networks or Artificial Intelligence and others can also be used to reduce the independent variable count and discover significant measurements to apply to the multivariable regression step.

25 Upon determining coefficients for each of the statistically significant independent variables, a closed form mathematical relation may be developed (it may have fewer than n independent variables, each represented by “ x ”, coefficients for each of such variables, and a dependent variable. This relationship may be truncated (in other words, abbreviated or shortened) in that it has fewer than “ n ” independent variables (because it may only include statistically significant
30 variables (whether they be consolidated/grouped or not)). Accordingly, particularly as compared with large data sets (with “ n ” total measurements), results may be generated more quickly, even

where the relationship (and perhaps the entire inventive protocol) are computer implemented. The truncated relationship may be used to generate estimates of the dependent variable upon input of measurements (e.g., as numerical data) of statistically significant independent variables. That estimate can then be used to transform a process or product from what that process or product would be without consideration of that dependent variable estimate. Such transformation of process or product may be as described in more detail elsewhere in this disclosure. Because it may be known that a certain dependent variable value (e.g., within a certain range) suggests that a certain step be taken or acts be taken to achieve a certain benefit (e.g., such as using a particular additive, adding ingredients in a certain ratio, heating to a certain temperature, as but a few examples) to achieve a desired benefit (e.g., improved wearability, resistance to UV induced fading, coking risk mitigation, etc.), the dependent variable estimate (e.g., achieved using the truncated relationship) can be used to modify a process or product to achieve an improvement in that process or product.

EXAMPLE OF CORRELATIONS WITH PHYSICAL PROPERTIES:

A brief description of the use of the correlation method which is the subject of this invention is provided below for eight unaged asphalt binders. For this example, results of binder penetration (PEN) tests were correlated with results from several analysis techniques which are described below. Many other chemical and spectroscopic analyses and many other physical properties can be correlated in this manner. Some results from the analyses provided good correlations, and others did not.

Fourier Transform Infrared (FTIR) Spectroscopy FTIR spectra were obtained using an Agilent Cary 630 FTIR spectrometer to conduct analyses for this project. Solutions were 1.2 weight percent asphalt binder in tetrachloroethylene. Absolute peak absorbance values were used for the correlations.

Saturates, Aromatics, Resins-Asphaltene Determinator (SAR-ADTM) Separation The automated SAR separation coupled with automated AD separation (SAR-AD) is described by Boysen and Schabron (2013). The combined system, SAR-AD, generates *saturates*, *aromatics*, and *resins* (SAR) chromatographic fractions and elutes cyclohexane soluble, toluene soluble, and

methylene chloride-methanol soluble asphaltene subfractions. The separation couples a high performance liquid chromatography (HPLC) based SAR separation with a previously-developed asphaltenes analysis method called the Asphaltene Determinator[®] (Schabron et al. 2010) which characterizes asphaltenes by solubility. One observation using the SAR-AD may yield several (e.g., 16) measurements, each corresponding perhaps to a single different fraction. The separation was further modified to separate the resins fraction into two fractions. Solutions of asphalt were prepared as 10% (wt/vol) in chlorobenzene. The solutions were filtered through 0.45 micron syringe filters into autosampler vials. Portions of 20 µL were injected for the SAR-AD separation. All separation profiles were electronically blank subtracted prior to peak integration. A representative SAR-AD separation profile is given in Figure 3.

Peak Descriptions (from left to right on an ELSD separation profile (see Fig. 3))

Peak 1. Saturates: Elutes through all four columns with heptane, fully saturated alkyl molecules (model compound cholestane is in this fraction),

Peak 2. Naphthene Saturates: Elutes through all four columns with heptane, but the elution time is retarded by the activated silica. This material absorbs some light at 230 nm and 260 nm and very little at 290 and 310 nm indicating this material may contain some hydrocarbons with one or two aromatic ring structures with significant amounts of alkyl side chains.

Peak 3. Cyclohexane Soluble Asphaltenes: Highly alkyl substituted, polar, pericondensed aromatics

Peak 4. Toluene Soluble Asphaltenes: Polar, more pericondensed aromatics

Peak 5. CH₂Cl₂:MeOH Soluble Asphaltenes: Pre-coke, polar, most pericondensed aromatics

Peaks 6 and 7 Combined. Aromatics: Total aromatics. The cut between these peaks is very sensitive to the activity of the aminopropyl bonded silica, which can change with temperature

humidity, and solvent purity. These peaks are combined to increase precision in the total aromatics fraction.

Peak 8. Resins: Polar heptane soluble material that elutes with CH₂Cl₂:MeOH (98:2 v:v) from the amino-propyl bonded silica and glass bead columns; some of this material absorbs visible light at 500 nm

Calculated Parameters

Coking Index: Ratio of peak areas of cyclohexane soluble asphaltenes to CH₂Cl₂:MeOH soluble asphaltenes which is a measure of pyrolysis severity history. Values below 1.0 for 500 nm peak areas indicate the presence of coke.

Asphalt Aging Index: Ratio of the toluene soluble asphaltenes 500 nm peak area to the sum the resins and aromatics fractions 500 nm peak areas. Absorbance at 500 nm is due to the presence of extended pi systems that impart brown color to oil, which increase with oxidation.

Total Pericondensed Aromatics (TPA): The approximate weight percent of material in the sample that absorbs 500 nm (visible) light.

Elemental and Metals Analyses Table 1 contains the elemental and metals results for eight asphalt samples. CHNOS analyses were performed on the neat asphalts by Huffman Laboratories, Golden, Colorado. Metals analyses at Huffman Laboratories were performed on the 10% nitric acid solutions from the wet ash / dry ash procedure performed at WRI. A quality control sample was submitted with the metals solutions. The results indicated that the sample prep and analysis were in control.

Table 1. Elemental and Metals Analyses Results.

Analysis	Asphalt Binder							
	B1	B2	B3	B4	B5	B6	B7	B8

weight %								
Carbon	84.27	84.74	83.10	83.21	83.75	83.99	85.24	84.28
Hydrogen	9.94	9.98	9.86	9.60	9.93	10.21	10.24	10.08
Nitrogen	0.68	0.45	0.44	0.45	0.58	0.51	0.50	0.43
Oxygen	0.64	0.56	0.44	0.43	0.53	0.60	0.50	0.50
Sulfur	4.45	4.59	6.30	6.30	5.02	4.62	3.64	4.28
<i>Total</i>	<i>99.98</i>	<i>100.32</i>	<i>100.14</i>	<i>99.99</i>	<i>99.81</i>	<i>99.93</i>	<i>100.12</i>	<i>99.57</i>
H/C ratio	1.41	1.40	1.41	1.37	1.41	1.45	1.43	1.43
μg/g (ppm)								
Copper	< 0.35	< 0.35	< 0.35	< 0.35	< 0.35	< 0.35	< 0.35	< 0.35
Iron	27.8	31.4	16.9	17.6	28.4	23.6	35.3	35.9
Nickel	107	35.8	36.1	38.9	50.0	84.6	45.6	36.2
Vanadium	413	115	129	144	174	251	141	115
V/Ni ratio	3.86	3.21	3.57	3.70	3.48	2.97	3.09	3.18

Automated Flocculation Titrimetry (AFT) The titration method is described in ASTM D6703, which was developed at WRI. Titrations were conducted using toluene solutions of asphalt titrated with heptane using an automated system. Calculated parameters include the state of peptization, P, which has a theoretical lower limit of 1.0 (highly instable), however values of P commonly vary between 2.5 to 10 for unmodified or neat asphalts. Low P values indicate internally incompatible material. Values in P are calculated as a function of two parameters that relate to the peptizability of the asphaltenes p_a and the solvent power of the maltenes, p . The AFT data are listed in Table 2.

Table 2. Automated Flocculation Titration Calculated Parameters for Eight Asphalts.

Binder	AFT Parameter	Trial		Binder	AFT Parameter	Trial	
		1 st	2 nd			1 st	2 nd
B1	p_a	0.629	0.631	B5	p_a	0.665	0.660
	p_o	1.03	1.01		p_o	1.22	1.30
	P	2.76	2.73		P	3.65	3.83
B2				B6			

	p _a	0.691	0.694		p _a	0.655	0.662
	p _o	1.05	0.98		p _o	1.14	1.04
	P	3.41	3.22		P	3.29	3.07
B3	p _a	0.714	0.710	B7	p _a	0.702	0.695
	p _o	1.02	1.21		p _o	0.88	1.00
	P	3.56	4.18		P	2.96	3.27
B4	p _a	0.715	0.716	B8	p _a	0.693	-
	p _o	1.12	1.05		p _o	1.05	-
	P	3.93	3.70		P	3.43	-

Size Exclusion Chromatography (SEC) Sample solutions were prepared by dissolving 0.30 ± 0.0005 g in tetrahydrofuran (THF) and bringing to volume in 10 mL volumetric flasks to generate 3 wt/vol% solutions. Solutions were filtered through 0.45 μm syringe filters and 30 μL aliquots were injected into a high performance liquid chromatography (HPLC) system equipped with a 7.8 X 300 mm, 5μm, 50Å Phenogel column thermostatted to 35 °C and THF eluent at 0.5 mL/min. The THF was HPLC grade stabilized with butylated hydroxytoluene (BHT). A differential refractive index (RI) detector was used to record the separation profiles. A second order curve fit of polystyrene standards of peak molecular weights (MW) 3000, 1300, 890 and 370 Da (g/mol) respectively was used to calibrate the system. Chromatograms were split into slices for analysis consisting of material >2966 Da, material <2966 and >1000 Da, material <1000 and >700 Da, material <700 and >370 Da, and material <370 Da. It is very likely that material <400 Da does not exist within asphalt and detection of material this size is likely due to reversible adsorption effects by polar type compounds on the column. Figure 4 shows the RI profiles for the eight binders, and Table 3 summarizes the data.

Table 3. Size Exclusion Chromatography Data Summary.

Binder	Parameter	<u>Slice 1</u> >3050	<u>Slice 2</u> 3050-1000	<u>Slice 3</u> 1000-700	<u>Slice 4</u> 700-370	<u>Slice 5</u> <370
B1	M _n (Da)	3513	1554	836	513	280

	M _w (Da) % in Peak	3577 9.06	1715 41.12	845 15.52	531 21.67	287 12.63
B2	M _n (Da)	3541	1575	837	514	280
	M _w (Da)	3608	1741	846	532	287
	% in Peak	11.55	42.74	14.66	19.66	11.39
B3	M _n (Da)	3495	1580	837	515	281
	M _w (Da)	3555	1747	847	534	288
	% in Peak	10.75	44.61	14.93	19.32	10.4
B4	M _n (Da)	3523	1599	837	514	281
	M _w (Da)	3585	1770	846	533	288
	% in Peak	12.52	44.33	14.17	18.69	10.28
B5	M _n (Da)	3539	1588	839	518	282
	M _w (Da)	3603	1755	848	537	289
	% in Peak	12.89	46.51	14.48	17.43	8.69
B6	M _n (Da)	3606	1601	838	516	281
	M _w (Da)	3682	1774	847	534	288
	% in Peak	16.64	43.42	13.6	17.13	9.2
B7	M _n (Da)	3457	1547	839	516	281
	M _w (Da)	3510	1697	848	535	288
	% in Peak	7.15	47.22	15.68	19.39	10.55
B8	M _n (Da)	3519	1564	837	514	280
	M _w (Da)	3582	1727	846	533	288
	% in Peak	10.4	42.99	15.13	20.11	11.37

Multiple Linear Regressions The correlations were evaluated by WRI's advanced multiple linear regression software. This program was designed to investigate relationships between independent and dependent variables using standard multivariable linear regression algorithms, but with some added features. In addition to solving classical multivariable data fitting problems, additional methods are available for fitting data sets with unfavorable dependent-to-independent variable ratios. This method provides a process to generate correlations between physical and chemical measurements, chemical and chemical measurements, and physical and physical measurements when sufficient observations are not available to perform the correlations while examining all of the measurements at once. An example of SAR-AD correlation parameters that may be used is provided in Table 4.

Table 4. Example of SAR-AD[®] Correlation Parameters.

SAR-AD Fraction	ELSD	500 nm Absorption
Total Saturates	x	
Total Aromatics	x	X
Resins	x	X
Cyclohexane Asphaltenes	x	X
Toluene Asphaltenes	x	X
CH ₂ Cl ₂ /MeOH Asphaltenes	x	X
Calculated Metrics		
ELSD Toluene Asphaltenes to Aromatics Ratio		
500 nm Toluene Asphaltenes to Aromatics Ratio		
ELSD Resins to Aromatics Ratio		
ELSD Resins to Total Asphaltenes Ratio		
ELSD Total Asphaltenes		
ELSD Resins plus Total Asphaltenes		
ELSD Total Saturates plus Total Aromatics		
Total Pericondensed Aromatics (TPA) The weight percent of material in the sample that absorbs 500 nm		
500 nm Coking Index Ratio of 500 nm Cyclohexane Asphaltenes to 500 nm CH ₂ Cl ₂ :MeOH Asphaltenes. Values below 1.0 for 500 nm peak areas indicate presence of coke.		
ELSD Coking Index Ratio of ELSD Cyclohexane Asphaltenes to ELSD CH ₂ Cl ₂ :MeOH Asphaltenes.		

Absorbance Aging Index
Ratio of 500 nm Toluene Asphaltenes to 500 nm Resins.

Unlike most multiple linear regression programs, which reduce the independent variable count to “latent variables” through dimensional reduction by complex rotations and projections (as is the case for partial least squares and principle components analysis schemes), this method produces correlations that are expressed in closed form mathematical equations in terms of the significant measured values.

The correlation software algorithm uses two test methods to reject or accept parameters in a model. The F-Test Reject method fits data using all independent variables and then removes the least significant variable based on the F-test. This cycle of fitting and rejecting parameters is repeated until only one parameter is remaining. The R^2 values for all the fits are plotted versus the number of parameters to produce an “Over fit” plot. This plot provides a visual clue for how many parameters are relevant for a given fit. Having too many parameters results in over fitting, where the model is meaningless.

The F-Test Reject method, because it rejects variables one at a time, sometimes rejects independent variables that, in combination, would produce acceptable models. The F-Test Add method was developed to approach the fitting process from the other end and to reduce the chances of missing important correlations. In this method, the fitting starts with a fixed set of independent variables (usually selected from the best single variable fits) and increases the independent variable count one at a time based on the significance of the added variable to the model using the F-test.

Parameter Transformations Several parameter transformations were used to aid in the search for relevant correlations. Table 2 summarizes the whole correlation effort and shows the algebraic forms of the most common transformations used. All logarithmic transformations used natural logarithms.

Example of Correlation Results Results for the eight unaged asphalt binders were examined for this example. These represent a small sample set and interpretations and applications of the correlations should be used with care. A three or four metric correlation may be meaningful or not.

5

The WRI multivariate regression software (MLS) was used to look for correlations with penetrometer test data for these binders. In MLS, all independent variables are used at the start. The variables are removed one by one based on how much they influence the overall fit. Caution must be used in properly grouping the independent variables to obtain the most meaningful relationships.

10

Correlations for the penetration data are shown in figure 5. The best single variable correlations, colored black in figure 5, are the following:

15

IR absorption at 3754 cm^{-1} with an R^2 of 0.82, and
SAR-AD (Nap ELS) with an R^2 of 0.86.

The best multivariable correlations are also shown in figure 5 with the gray bars.

20

APPLICATIONS OF THE INVENTION

Embodiments of the present invention generally relate to the use of the specialized WRI chemometric software for the determination of mathematical relationships between physical and chemical measurements. The invention may be used to estimate numerical values for coefficients (e.g., linear) for each statistically significant independent variable to generate a closed form mathematical equation which can be used to predict/estimate a dependent variable based on knowledge (e.g., upon measurement) of only such statistically significant independent variables. Ease, simplicity, and in many cases, speed of results, without impairing results to an unacceptable degree, may be particularly valuable benefits of the inventive technology, which often may be embodied in a computer program or software, and applied to a particular problem

30

upon user input of measured data. Another may be, at times, elimination of the need to measure the dependent variable.

This method can be used to predict various physical properties of asphalt, for example. The software can be used to correlate chemical measurement data such as (but not limited to) near- or mid- infrared (IR) spectroscopy, gel permeation chromatography, asphaltene flocculation titration, ion exchange chromatography, asphaltene solubility subfraction profile analysis, chromatography, or the fully automated saturates, aromatics resins- Asphaltene DeterminatorTM (SAR-AD) analysis.

The use of this method can be combined with a wide assortment of analytical techniques to provide predictions of many physical properties of oil (petroleum or non-petroleum derived), asphalts, polymers, biological materials or any material for which chemical or spectroscopic measurements can be made. A detailed description of the development of, and certain aspects of, the specialized WRI chemometric software, a key component to this invention, is provided in Glaser et al. report to FHWA (2015) (see Appendix 1), and is incorporated herein by reference. Note that certain aspects of embodiments of the inventive technology, particularly regarding the application of the technology, may be discussed in only rudimentary terms in this Glaser report. Note also that any step of any of the claims can be combined with any step of any of the other claims to describe a particular embodiment of the inventive technology.

Applications for asphalt can include but are not limited to any rheological or empirical mechanical properties, for any type of asphalt binder included modified, roofing, paving, or sealing. More generally, the inventive approach disclosed herein can be used for many other types of materials, and related processes, also. And this approach generates information (e.g., estimate of a dependent variable) that can be used to transform a process relating to any of such materials, or a product that includes any of such materials. A non-exhaustive list of some of these processes (or materials that such processes relate to), and materials, is follows:

- Processes relating to durability, as indicated by properties measured at various aging stages, unaged and aged
- Product formulation in general

- blending proportions based on those properties
- designing an additive based on those properties
- determining the type and/or amount of additive to be added
- Compatibility and phase separation in asphalt binder and consequences in terms of stability, either for asphalt made of blends from refining bases (residues from straight run distillation, solvent deasphalting airblowing, visbreaking, hydrotreating, cracking or coking units), or for any of those blends further modified with any semi-compatible additives, including but not restricted to polymers, acids, waxes, rubbers, amines, and derivatives.
- Asphalt and petroleum emulsion ability, storability, breaking, coalescence and curing, and any physical properties of these emulsions and their residues after recovery process.
- Asphalt binder and flux aging, short term and long term, w/ w/o UV and moisture (to address both paving and roofing coatings)
- Long term durability and performance of highway and roofing materials
- Blending properties of asphalts with aged asphalts from recycled paving materials or recycled roofing materials
- Asphalt specification parameters
- Asphalt binder physical properties, rheological properties in particular, such as complex modulus, phase angle or any combinations or derivatives.
- Properties and performance of asphalt binder, asphalt aggregate mixture or chip seals, asphalt shingles or other industrial applications.
- Reactivity characteristics of petroleum or petroleum derived fractions or materials for various processes including production, heating, distillation, hydrotreating, coking and others.
- Refining an asphalt (or other materi) blend/mix; selecting a bitumen therefor; modifying a blend recipe; determining an ingredient amount

- Fouling characteristics of crude oils in upstream and downstream applications and oil derived materials including fuels and asphalts.
- Investigating and predicting properties of polymers, biological materials, biofuels, asphalt binder sealants, asphalt binder rejuvenators
- 5 • Investigating and predicting properties or effects (whether intended or not) of cosmetics, surfactants, medications and food materials
- Hydrocarbon, asphalt, any type of oil, petroleum, coal, and biomass products, fuel, medication, dietary supplements, cosmetics, food, lubricants, and any other materials indicated in this application, or in references incorporated herein.

10

Additional exemplary applications of embodiments of the inventive software/method may be found, in detail, in: Delfosse, F., et al, Impact of the Bitumen Quality on the Asphalt Mixes Performances, E&E Congress 2016, 6th Eurasphalt & Eurobitume Congress, EE.2016.049 (see Appendix 2); and Glaser, R., et al., Relationships Between Solubility and Chromatographically
15 Defined Bitumen Fractions and Physical Properties, E&E Congress 2016, 6th Eurasphalt & Eurobitume Congress, EE.2016.337 (see Appendix 3), both said papers incorporated herein, by reference, in their entirety.

Again, knowledge obtained from this invention (e.g., estimates of a dependent value) can be used to transform any of such above referenced processes or processes involving any of the above-
20 mentioned materials, in addition to any other process disclosed or indicated herein, or related to any material disclosed or indicated herein. For example, this technology can be used to formulate, blend and mix more cost efficiently hydrocarbons such as long term performing asphalt materials, lubricants, greases, crude oils or any petroleum products, or more generally chemical products, including additives and polymers, making them easier to produce and handle
25 avoiding trial and error based empirical methods. Estimates can be used to transform any of a variety of processes (e.g., modifying a formulation of a hydrocarbon mixture, emulsion, or blend; or a blend or mix of hydrocarbons; or modify a process involving an additives(s) (e.g., designing and selecting better additives), as but a few examples). Where a process is carried out in any manner that, due to information (particularly regarding the independent variable) generated upon

use of the inventive software/method, is different from that manner in which the process would be carried out in the absence of such information, said process is said to have been transformed. Similarly, estimates of a dependent variable can be used to transform a product (e.g., one defined or qualified by a dependent variable estimate) from what it would be without such dependent variable information (e.g., as where an estimate of a dependent variable of a certain material/product is used to determine how much or what type of an additive to add to that product to achieve a certain result (e.g., prevent coking)). Transformation of the product or process, in preferred embodiments, results in an improvement, typically to that product or to a product that the process relates to (e.g., a product that the process generates). For example, transformation of an asphalt may lead to an increase in the constituent amounts of one of its ingredients, resulting in an asphalt with improved durability and/or better aging; transformation to an asphalt blending process may lead to a decrease in one ingredient and an increase in another ingredient resulting in an asphalt with better UV resistance. Transformation of a hydrocarbon processing method may involve the use of an additive that would otherwise not be used, or use of an additive in amounts that otherwise would not be observed, to better avoid coking, or allow for higher processing temperatures with confidence that no coking will occur, resulting in more efficient processing.

Note that the inventive technology is not limited to inventive methods, as indeed a system for transforming a process or product may describe, generally, an aspect of the inventive technology. Such system may comprise the following: a linear dependence assignment element that assigns a linear dependence of a dependent variable on “n” number of independent variables; an observation element that yields “p” number of observations to obtain “p” number of measurements for each said dependent variable and said independent variables, wherein “p” is less than the sum of “n” + 1; an artificial data generation element that generates artificial data using measurement precision, for at least some of said variables; statistically significant independent variable determiner that determines statistically significant independent variables, wherein said statistically significant independent variables have a statistically significant impact on said dependent variable, and are fewer in number than “n”; a coefficients generator that generates coefficients for each of said statistically significant independent variables; a truncated, closed form mathematical relationship generator that generates a relationship according to which

said dependent variable linearly depends from only said statistically significant independent variables, wherein said truncated, closed form mathematical relationship yields results that are sufficiently precise; a dependent variable estimator that uses said relationship, and at least one measurement of each of said at least said statistically significant independent variables to obtain
5 a dependent variable estimate; and a transformation of a process or a product from what said process or said product would be without consideration of said dependent variable estimate. Note that each of said elements may be a subroutine, e.g., a series of encoded instructions, as indicated in the Additional Information section herein. Apparatus/system versions of all method claims filed herewith are disclosed either explicitly herein, or upon consideration of the fact that
10 the disclosure of the steps of generating, determining, producing, developing, truncating, estimating, etc., is deemed disclosure of corollary apparatus steps of a determinator, producer, developer, truncator, estimator, etc., respectively; any and all disclosure particulars that relate to each specific step is also deemed to describe each corollary apparatus componentry.

15 The WRI chemometric software is especially useful for applications where insufficient observations are available compared to the number of independent measurement variables available. This situation is common in many fields of science and technology, such as spectroscopy, calorimetry, thermogravimetric, chromatography and others.

20 Additional Information: As can be easily understood from the foregoing, the basic concepts of the present invention may be embodied in a variety of ways. It involves both correlation techniques as well as devices to accomplish the appropriate correlation. In this application, the correlation techniques are disclosed as part of the results shown to be achieved by the various devices described and as steps which are inherent to utilization. They are simply the natural
25 result of utilizing the devices as intended and described. In addition, while some devices are disclosed, it should be understood that these not only accomplish certain methods but also can be varied in a number of ways. Importantly, as to all of the foregoing, all of these facets should be understood to be encompassed by this disclosure.

The discussion included in this application is intended to serve as a basic description. The reader
30 should be aware that the specific discussion may not explicitly describe all embodiments possible; many alternatives are implicit. It also may not fully explain the generic nature of the

invention and may not explicitly show how each feature or element can actually be representative of a broader function or of a great variety of alternative or equivalent elements. Again, these are implicitly included in this disclosure. Where the invention is described in device-oriented terminology, each element of the device implicitly performs a function.

5 Apparatus claims may not only be included for the device described, but also method or process claims may be included to address the functions the invention and each element performs. Neither the description nor the terminology is intended to limit the scope of the claims that will be included in any subsequent patent application.

10 It should also be understood that a variety of changes may be made without departing from the essence of the invention. Such changes are also implicitly included in the description. They still fall within the scope of this invention. A broad disclosure encompassing the explicit embodiment(s) shown, the great variety of implicit alternative embodiments, and the broad methods or processes and the like are encompassed by this disclosure and may be relied upon when drafting the claims for any subsequent patent application. It should be understood that
15 such language changes and broader or more detailed claiming may be accomplished at a later date (such as by any required deadline) or in the event the applicant subsequently seeks a patent filing based on this filing. With this understanding, the reader should be aware that this disclosure is to be understood to support any subsequently filed patent application that may seek examination of as broad a base of claims as deemed within the applicant's right and may be
20 designed to yield a patent covering numerous aspects of the invention both independently and as an overall system.

Further, each of the various elements of the invention and claims may also be achieved in a variety of manners. Additionally, when used or implied, an element is to be understood as encompassing individual as well as plural structures that may or may not be physically
25 connected. This disclosure should be understood to encompass each such variation, be it a variation of an embodiment of any apparatus embodiment, a method or process embodiment, or even merely a variation of any element of these. Particularly, it should be understood that as the disclosure relates to elements of the invention, the words for each element may be expressed by equivalent apparatus terms or method terms -- even if only the function or result is the same.
30 Such equivalent, broader, or even more generic terms should be considered to be encompassed in

the description of each element or action. Such terms can be substituted where desired to make explicit the implicitly broad coverage to which this invention is entitled. As but one example, it should be understood that all actions may be expressed as a means for taking that action or as an element which causes that action. Similarly, each physical element disclosed should be understood to encompass a disclosure of the action which that physical element facilitates. Regarding this last aspect, as but one example, the disclosure of a "correlation" should be understood to encompass disclosure of the act of "correlating" -- whether explicitly discussed or not -- and, conversely, were there effectively disclosure of the act of "correlating", such a disclosure should be understood to encompass disclosure of a "correlation" and even a "means for correlating." Such changes and alternative terms are to be understood to be explicitly included in the description. Further, each such means (whether explicitly so described or not) should be understood as encompassing all elements that can perform the given function, and all descriptions of elements that perform a described function should be understood as a non-limiting example of means for performing that function.

Any acts of law, statutes, regulations, or rules mentioned in this application for patent; or patents, publications, or other references mentioned in this application for patent are hereby incorporated by reference. Any priority case(s) claimed by this application is hereby appended and hereby incorporated by reference. All claims filed herewith are incorporated into this application. Any appendices filed with this application are hereby incorporated into this application. In addition, as to each term used it should be understood that unless its utilization in this application is inconsistent with a broadly supporting interpretation, common dictionary definitions should be understood as incorporated for each term and all definitions, alternative terms, and synonyms such as contained in the Random House Webster's Unabridged Dictionary, second edition are hereby incorporated by reference. Finally, all references listed in the list of References To Be Incorporated By Reference In Accordance With The International Patent Application or other information list or statement filed with the application are hereby appended and hereby incorporated by reference, however, as to each of the above, to the extent that such information or statements incorporated by reference might be considered inconsistent with the patenting of this/these invention(s) such statements are expressly not to be considered as made by the applicant(s).

Thus, the applicant(s) should be understood to have support to claim and make a statement of invention to at least: i) each of the correlation devices as herein disclosed and described, ii) the related methods disclosed and described, iii) similar, equivalent, and even implicit variations of each of these devices and methods, iv) those alternative designs which accomplish each of the functions shown as are disclosed and described, v) those alternative designs and methods which accomplish each of the functions shown as are implicit to accomplish that which is disclosed and described, vi) each feature, component, and step shown as separate and independent inventions, vii) the applications enhanced by the various systems or components disclosed, viii) the resulting products produced by such systems or components, ix) each system, method, and element shown or described as now applied to any specific field or devices mentioned, x) methods and apparatuses substantially as described hereinbefore and with reference to any of the accompanying examples, xi) an apparatus for performing the methods described herein comprising means for performing the steps, xii) the various combinations and permutations of each of the elements disclosed, xiii) each potentially dependent claim or concept as a dependency on each and every one of the independent claims or concepts presented, and xiv) all inventions described herein.

In addition and as to computer aspects and each aspect amenable to programming or other electronic automation, the applicant(s) should be understood to have support to claim and make a statement of invention to at least: xv) processes performed with the aid of or on a computer as described throughout the above discussion, xvi) a programmable apparatus as described throughout the above discussion, xvii) a computer readable memory encoded with data to direct a computer comprising means or elements which function as described throughout the above discussion, xviii) a computer configured as herein disclosed and described, xix) individual or combined subroutines and programs as herein disclosed and described, xx) a carrier medium carrying computer readable code for control of a computer to carry out separately each and every individual and combined method described herein or in any claim, xxi) a computer program to perform separately each and every individual and combined method disclosed, xxii) a computer program containing all and each combination of means for performing each and every individual and combined step disclosed, xxiii) a storage medium storing each computer program disclosed, xxiv) a signal carrying a computer program disclosed, xxv) the related methods disclosed and described, xxvi) similar, equivalent, and even implicit variations of each of these systems and

methods, xxvii) those alternative designs which accomplish each of the functions shown as are disclosed and described, xxviii) those alternative designs and methods which accomplish each of the functions shown as are implicit to accomplish that which is disclosed and described, xxix) each feature, component, and step shown as separate and independent inventions, and xxx) the various combinations and permutations of each of the above.

With regard to claims whether now or later presented for examination, it should be understood that for practical reasons and so as to avoid great expansion of the examination burden, the applicant may at any time present only initial claims or perhaps only initial claims with only initial dependencies. The office and any third persons interested in potential scope of this or subsequent applications should understand that broader claims may be presented at a later date in this case, in a case claiming the benefit of this case, or in any continuation in spite of any preliminary amendments, other amendments, claim language, or arguments presented, thus throughout the pendency of any case there is no intention to disclaim or surrender any potential subject matter. It should be understood that if or when broader claims are presented, such may require that any relevant prior art that may have been considered at any prior time may need to be re-visited since it is possible that to the extent any amendments, claim language, or arguments presented in this or any subsequent application are considered as made to avoid such prior art, such reasons may be eliminated by later presented claims or the like. Both the examiner and any person otherwise interested in existing or later potential coverage, or considering if there has at any time been any possibility of an indication of disclaimer or surrender of potential coverage, should be aware that no such surrender or disclaimer is ever intended or ever exists in this or any subsequent application. Limitations such as arose in *Hakim v. Cannon Avent Group, PLC*, 479 F.3d 1313 (Fed. Cir 2007), or the like are expressly not intended in this or any subsequent related matter. In addition, support should be understood to exist to the degree required under new matter laws -- including but not limited to European Patent Convention Article 123(2) and United States Patent Law 35 USC 132 or other such laws-- to permit the addition of any of the various dependencies or other elements presented under one independent claim or concept as dependencies or elements under any other independent claim or concept. In drafting any claims at any time whether in this application or in any subsequent application, it should also be understood that the applicant has intended to capture as full and broad a scope of coverage as legally available. To the extent that insubstantial substitutes are made, to the extent that the

applicant did not in fact draft any claim so as to literally encompass any particular embodiment, and to the extent otherwise applicable, the applicant should not be understood to have in any way intended to or actually relinquished such coverage as the applicant simply may not have been able to anticipate all eventualities; one skilled in the art, should not be reasonably expected to have drafted a claim that would have literally encompassed such alternative embodiments.

Further, if or when used, the use of the transitional phrase “comprising” is used to maintain the “open-end” claims herein, according to traditional claim interpretation. Thus, unless the context requires otherwise, it should be understood that the term “comprise” or variations such as “comprises” or “comprising”, are intended to imply the inclusion of a stated element or step or group of elements or steps but not the exclusion of any other element or step or group of elements or steps. Such terms should be interpreted in their most expansive form so as to afford the applicant the broadest coverage legally permissible. The use of the phrase, “or any other claim” is used to provide support for any claim to be dependent on any other claim, such as another dependent claim, another independent claim, a previously listed claim, a subsequently listed claim, and the like. As one clarifying example, if a claim were dependent “on claim 20 or any other claim” or the like, it could be re-drafted as dependent on claim 1, claim 15, or even claim 25 (if such were to exist) if desired and still fall with the disclosure. It should be understood that this phrase also provides support for any combination of elements in the claims and even incorporates any desired proper antecedent basis for certain claim combinations such as with combinations of method, apparatus, process, and the like claims.

Finally, any claims set forth at any time are hereby incorporated by reference as part of this description of the invention, and the applicant expressly reserves the right to use all of or a portion of such incorporated content of such claims as additional description to support any of or all of the claims or any element or component thereof, and the applicant further expressly reserves the right to move any portion of or all of the incorporated content of such claims or any element or component thereof from the description into the claims or vice-versa as necessary to define the matter for which protection is sought by this application or by any subsequent continuation, division, or continuation-in-part application thereof, or to obtain any benefit of, reduction in fees pursuant to, or to comply with the patent laws, rules, or regulations of any country or treaty, and such content incorporated by reference shall survive during the entire

pendency of this application including any subsequent continuation, division, or continuation-in-part application thereof or any reissue or extension thereon.

LIST OF REFERENCES TO BE INCORPORATED BY REFERENCE INTO THIS INTERNATIONAL PATENT APPLICATION

5 I. US PATENTS

	Patent No.	Kind Code	Country Code	Date of Issuance	Patentee
1	5590218		US	1996-12-31	Ornstein
2	5752007		US	1998-05-12	Morrison
3	8370082	B2	US	2013-02-05	De Peinder et al.
4	4800279	A	US	1989-01-24	Indian University Foundation
5	4847793		US	1989-07-11	Cygnarowicz et al.

II. US PUBLICATIONS

	Publication No.	Kind Code	Country Code	Date of Issuance	Patentee
1	20130067991	A1	US	2013-03-21	Schabron et al.

III. FOREIGN REFERENCES

	Patent No.	Kind Code	Country Code	Date of Issuance	Patentee
1	2071329	A2	EP	2009-06-17	ENI S.P.A.
2	1997014953	A1	WO	1997-04-24	Shell Int Research
3	2645695	A1	CA	2009-06-13	Eni S. P. A.
4	1997001096	A1	WO	1997-01-09	Exxon Research and Engineering Company

10

IV. NON PATENT LITERATURE

NON PATENT LITERATURE DOCUMENTS	
1	Schabron, J.F., Turner, T.F., Boysen, R.B., Glaser, R.R., Loveridge, J.L., Miknis, F.P. Qin, Q., Rovani, J.F., Salmans, S.L., 2013, Laboratory and Field Asphalt Binder Aging: Chemical Changes and Influence on Asphalt Binder Embrittlement; Products: FP 19 Chemical Changes with Oxidative Aging; FP 20 Linear Amplitude Sweep Test; FP 21 Application of the Asphaltene Determinator to Asphalt, WRI Report to the Federal Highway Administration, Contract No. DTFH61-07-D-00005. 193 pages.
2	Lima, F.S.G., Leite, L.F.M., 2004, Determination of Asphalt Cement Properties by Near Infrared Spectroscopy and Chemometrics, Petroleum Science and Technology, 22:5-6 589-600. 14 pages.

3	De Peinder, D.D. Petrauskas, F. Singlenberg, F. Salvatori, T. Visser, F. Soulimani, B.M. Weckhuysen, 2008, Prediction of Long and Short Residue Properties form Crude Oils form their Infrared and Near-Infrared Spectroscopy, <i>Applied Spectroscopy</i> , 62(4), 414-422. 9 pages.
4	Glaser, R.R., Beemer, A., Turner, T.F., 2015, Chemo-Mechanical Software, Fundamental Properties of Asphalts and Modified Asphalts III Product: FP 06, Western Research Institute Report to the Federal Highway Administration, Contract No. DTFH61-07-D-00005. rhttp://www.westernresearch.org/uploadedFiles/Transportation_Technology/FHWA_Research/Fundamental_Properties/Technical%20White%20Paper%20FP%2006--Chemo-Mechanical%20Software.pdf 54 pages.
5	Schabron, J.F., J.F. Rovani, and M.M. Sanderson, 2010, The Asphaltene Determinator Method for Automated On-column Precipitation and Re-dissolution of Pericondensed Asphaltene Components, <i>Energy and Fuels</i> , 24, 5984-5996.
6	Boysen, R. B., Schabron, J. F., 2013, The Automated Asphaltene Determinator Coupled with Saturates, Aromatics, Resins Separation for Petroleum Residua Characterization, <i>Energy Fuels</i> , 27: 4654–4661.
8	Delfosse, et al. Impact of the bitumen quality on the asphalt mixes performances. E&E Congress 106. Prague, Czech Republic. June 1-3, 2016. 13 pages.
9	Glaser et al. Relationships between solubility and chromatographically defined bitumen fractions and physical properties. E&E Congress 106. Prague, Czech Republic. June 1-3, 2016. 11 pages.
10	Wikipedia, "Synthetic Data". https://en.wikipedia.org/wiki/Synthetic_data . Page last modified April 1, 2016.
11	Albuquerque, Georgia et al. "Synthetic Generation of High-Dimensional Datasets". http://graphics.tu-bs.de/media/publications/Albuquerque2011SGH.pdf . Posted online 23 October 2011. 8 pages.
12	Whiting, Mark, et al. Creating Realistic, Scenario-Based Synthetic Data for Test and Evaluation of Information Analytics Software. https://www.purdue.edu/discoverypark/vaccine/assets/pdfs/publications/pdf/Creating%20Realistic,%20Scenario-Based.pdf . ©2008. 9 pages.
13	"Dataset generation". http://www.causality.inf.ethz.ch/data/dataset_generation.html . Date unknown.
14	"Artificial Data". https://www.cs.cmu.edu/afs/cs/project/jair/pub/volume18/thompson03a.html/node19.html . Cindi Thompson, 2003-01-02.
15	Brownlee, Jason. 8 Tactics to Combat Imbalanced Classes in Your Machine Learning Dataset. http://machinelearningmastery.com/tactics-to-combat-imbalanced-classes-in-your-machine-learning-dataset/ . August 19, 2015. 15 pages.
16	Boysen, Ryan and Schabron, John, Automated HPLC SAR-AD Separation; Fundamental Properties of Asphalts and Modified Asphalts III Product: FP 01, March 2015

Filed via EFS

IN THE UNITED STATES PATENT AND
TRADEMARK OFFICE

Application Number:
Filed:
Applicants: Ronald R. Glaser et al.
Title: Method for Correlating Physical and Chemical Measurement Data
Sets to Predict Physical and Chemical Properties
Assignee: The University of Wyoming Research Corporation d/b/a Western
Research Institute
Attorney Docket: WRI-CorrSoft-PCT
Customer No.: 33549
Confirmation No.:

APPENDIX 1

Technical White Paper

Chemo-Mechanical Software

Fundamental Properties of Asphalts and Modified Asphalts III Product: FP 06

March 2015

Prepared for
Federal Highway Administration
Contract No. DTFH61-07-D-00005

By
Ronald R Glaser
Appy Beemer
Thomas F. Turner
Western Research Institute
3474 North 3rd Street
Laramie, WY 82072
www.westernresearch.org

WesternResearch
INSTITUTE

TABLE OF CONTENTS

INTRODUCTION	1
BACKGROUND	1
Multiple Linear Regression.....	2
Principle Component Analysis and Principle Component Regression.....	3
Partial Least Squares.....	3
Other Methods	4
SPECTRELATE SOFTWARE DEVELOPMENT	4
PROGRAM IMPLEMENTATION	19
Data Input.....	19
Dependent Variable Selection.....	23
Regression Type.....	23
Results and Output.....	43
RESULTS AND DISCUSSION	47
CONCLUSIONS	47
RECOMMENDATIONS	48
ACKNOWLEDGMENTS	48
DISCLAIMER	48
REFERENCES	48

LIST OF TABLES

Table 1. Independent variable counts at different grouping threshold values	14
---	----

LIST OF FIGURES

Figure 1. Graph. The correlation between wave numbers 3200 and 1035 for oxidized AAD-1 asphalt binder	5
Figure 2. Graph. The correlation between wave numbers 3600 and 1035 for oxidized AAD-1 asphalt binder	6
Figure 3. Screen capture. The regression spectra for wave number 1035 for oxidized AAD-1 asphalt binder	7
Figure 4. Screen capture. The regression spectrum for wave number 1035 for oxidized AAB-1 asphalt binder	8
Figure 5. Screen capture. The regression spectrum for wave number 1035 for oxidized asphalt binder data set	9
Figure 6. Screen capture. Mid infrared spectra with non-linear response areas removed	10
Figure 7. Screen capture. Mid infrared change spectra with non-linear response areas removed	11
Figure 8. Screen capture. The regression spectrum for wave number 1035 for oxidized asphalt binder data set change spectra	12
Figure 9. Screen capture. The regression spectrum for wave number 1035 for oxidized asphalt binder data set with a grouping range shown	13
Figure 10. Graph. Groups produced at different threshold values	15
Figure 11. Graph. Simple linear plot with insufficient measures to test confidence	16
Figure 12. Graph. Simple linear plot with 7 replicates per measurement, signal to noise ratio 0.5	17
Figure 13. Graph. Simple linear plot with 7 replicates per measurement, signal to noise ratio 5	18

LIST OF FIGURES (continued)

Figure 14. Graph. Simple linear plot with 7 replicates per measurement, signal to noise ratio 5	18
Figure 15. Graph. The regression coefficient dependence on the number of replications and the measurement precision	19
Figure 16. Screen capture. Load data button	20
Figure 17. Screen capture. Data file selection	21
Figure 18. Screen capture. Data display	22
Figure 19. Screen capture. Dependent data button	22
Figure 20. Screen capture. Dependent variable selection	23
Figure 21. Screen capture. Regression type selection	24
Figure 22. Screen capture. Regression spectra example	25
Figure 23. Screen capture. Regression spectra example (magnified)	26
Figure 24. Screen capture. Regression plot selection	26
Figure 25. Screen capture. Regression plot example	27
Figure 26. Screen capture. Regression plot example 2	28
Figure 27. Screen capture. Multivariate regression selection	29
Figure 28. Screen capture. Independent variable reduction method selection	29
Figure 29. Screen capture. Independent variable selection example 1	30
Figure 30. Screen capture. Independent variable selection example 2	31
Figure 31. Screen capture. Regression results plot of predicted and measured dependent variable values	32
Figure 32. Screen capture. Software notification of favorable observation to parameter ratio	33
Figure 33. Screen capture. Regression results plot, example 2	34

LIST OF FIGURES (continued)

Figure 34. Screen capture. Impossible observation to parameters ratio notification.....	35
Figure 35. Screen capture. Possible observation to parameters ratio notification, but statistically suspect.....	35
Figure 36. Screen capture. Computation failure error message.....	36
Figure 37. Screen capture. Successful computation, with suspect results warning.....	36
Figure 38. Screen capture. Regression results plot example 3	37
Figure 39. Screen capture. Regression results summary tab example 1	38
Figure 40. Screen capture. Regression results plot example 4	39
Figure 41. Screen capture. Regression results summary tab example 2	39
Figure 42. Screen capture. The Auto best independent variable list run option.....	40
Figure 43. Screen capture. Regression results summary tab example 3	40
Figure 44. Screen capture. Independent variable grouping options.....	41
Figure 45. Screen capture. Independent variable grouping threshold setting.....	41
Figure 46. Screen capture. Regression results summary tab example 4	42
Figure 47. Screen capture. Regression results summary tab example 5	42
Figure 48. Screen capture. Regression results summary tab example 6	43
Figure 49. Screen capture. Regression results summary tab example 7	43
Figure 50. Screen capture. Regression results summary tab copy and paste example	44
Figure 51. Screen capture. Complete run summary results copy and paste example.....	45
Figure 52. Screen capture. Right click context menu for plots.....	46
Figure 53. Screen capture. Plot attribute editing area.....	47

CHEMO-MECHANICAL SOFTWARE

INTRODUCTION

This technical report describes a software product designed to discover additive combinations of a wide range of independent variables that correlate with a limited set of dependent variables. A specific example would be the search for combinations of infrared spectra changes that correlate with rheological changes in an asphalt binder as it oxidizes. The original incentive for the development of this tool arose from the need to correlate a wide range of chemical measurements of asphaltic materials to the mechanical properties exhibited by those materials. The tool developed is not limited to our current application, and can be extended to many problems where a large number of independent variables are involved in a data set with limited observations of the dependent variable. Examples of possible applications would include, but not be limited to, correlation of spectral data, chromatographic data, or any data set that can be described as a list of x,y or t,y pairs against some other measured property. The crux of these problems is that many measurements are taken that are not related to the property of interest, but finding the relevant combinations is difficult. The problem we have focused on is discovering which changes in the mid-infrared spectra are most closely related to changes in an asphalt binder's rheological response as the material ages. Perhaps combinations of four or five spectral measurements are related to the property changes while the other areas of the spectra are irrelevant. This report describes the computational method and software application to discover the relevant measurements. Application of the method to asphalt problems is described in the respective technical white papers. This is not an experimental report. It is a product description that is essentially mathematical in nature.

BACKGROUND

Problem Definition: Modern analytical techniques often rapidly produce quite large data sets, the most common are those data sets usually described as "spectra". Any set of data that can be formulated as a response as a function of index can be treated as a "spectra".

Consequently, a thorough examination of the relationships between a given collection of spectra (or other arbitrary data matrix) type and an independently measured material property can generally be expressed by the following general relationship:

$$y = f(x_0, x_1, x_2, \dots, x_n) \quad (1)$$

where y is the dependent variable, e.g., Complex Modulus; and x_i is the independent variable, e.g., IR absorbance at wave number i .

If $f(x)$ is assumed to be algebraically linear, and we have only three spectra representing three asphalts (a , b , and c), along with their asphaltene contents, the equation set is:

$$y_a = k_0 x_{a0} + k_1 x_{a1} + k_2 x_{a2} + \dots + k_n x_{an} \quad (2)$$

$$y_b = k_0 x_{b0} + k_1 x_{b1} + k_2 x_{b2} + \dots + k_n x_{bn} \quad (3)$$

$$y_c = k_0 x_{c0} + k_1 x_{c1} + k_2 x_{c2} + \dots + k_n x_{cn} \quad (4)$$

where k is the proportionality constant at each wave number 0 through n . Since this is a curve fitting problem, the x and y pairs are known, and we seek k 's that satisfy the equation set. Such a deterministic solution is impossible if $n+1$ exceeds the number of measurements. When the number of measurements is exactly equal to $n+1$, then the fit is perfect, meaning no statistical evaluation of the fit quality is possible. This is analogous to the situation in two dimensions where you are fitting a line to two data points, obtaining a correlation coefficient of 1. To obtain a statistically meaningful test of a multidimensional fit, the observations should exceed the independent variable count by some factor, the larger the better. The actual size of the multiplier depends on the desired confidence in the answer, and the precision of the measurements. Typical mid-infrared spectra will contain nearly 4,000 wave numbers, so the examination of each and every wave number for significance when combined with the others would require 28,000 observations, clearly not practical. This situation is a reoccurring problem with spectral data and other extensive xy data sets as well, as the inclusion of all of the data results in an equation system with excessive adjustable parameters, impossible to solve. A number of approaches exist for addressing this problem with a variety of strategies aimed at essentially reducing the number of effective k 's (independent variable fit parameters) to be discovered.

Available methods: A few of the methods used to address the problems outlined above are briefly described in the next three sections (Sharaf et al. 1986; Wold 1991; Wold et al. 2001; Barros and Rutledge 2004; Garson 2007; Hasegawa 2006). These methods, particularly the process of correlating spectral data to other process or property variables, have been used successfully in a wide range of applications (Basu et al. 1998; Satya 2005; Satya et al. 2007; Chalmers and Everall 1996; Hasegawa 1999; Karstang et al. 1991; Lachenmeier 2007; Mark and Workman 2007; Sastry et al. 1998; Zagonel et al. 2004), but do not produce a closed form equation in terms of measured quantities, limiting their usefulness in fundamental scientific studies.

Multiple Linear Regression

Multivariable Linear Regression (MLR) is a time-honored technique going back to Pearson (1901). Multivariable regression can establish that a set of independent variables explains a proportion of the variance in a dependent variable at a significant level (through a significance test of R^2), and can establish the relative predictive importance of the independent variables (by comparing beta weights). Variable transformations (most common is the logarithm) can be applied to independent or dependent variables to explore some curvilinear effects. Polynomials can be fit as well by expanding independent variables into a power series.

Multivariable linear regression can solve the matrix $Y=MX+B$, provided sufficient measurements of Y exist to obtain all of the coefficients in vector M . To be statistically

meaningful, measurements of Y in excess of measurements of X must be available, meaning that a spectra of 3,500 wave numbers would require at least 3,500 measurements of, say, complex modulus. To be statistically reliable, 35,000 would be better. It is generally impossible to apply multivariable linear regression directly to correlation studies involving data rich spectral data. However, the preconditioning of individual data points to related groups (spectral peaks, for example) is helpful to reduce the independent variable count. However, this is usually not sufficient unless a very extensive data set (many observations) is available. A variety of computation approaches have been developed in recent years that address this problem by projecting the data in one way or another into a smaller list of independent variables. These include Principle Component Analysis, Partial least Squares, and others.

Principal Component Analysis and Principal Component Regression

Principal Component Analysis (PCA) techniques are applied to the problem of too many x measurements relative to y measurements by searching for so-called latent variables. The covariance of XX' is examined and parameter space axis rotations are employed to arrive at new coordinates based on eigenvectors of the XX' matrix. In simple terms this means that independent variables that appear to change in a similar fashion are grouped together. The translated x variables (often called indicator variables) are projected into a smaller parameter space of latent variables. It is implicitly assumed that these fictitious latent variables somehow describe a truer "latent structure" to the system. Recall that the underlying mathematical model for the entire data set is linear, often patently untrue in chemical systems. This technique results in latent variable data sets with improved variance in the hopes of improving signal to noise ratios. Often, however, irrelevant data included in the translations pass spurious noise to the latent variables.

Principal Components Regression (PCR) is the application of ordinary linear regression methods to the latent variables developed from the principle components analysis. The difficulty with this method is that the complex axis rotations make understanding what the latent variables represent in terms of measurable quantities difficult. Interpretation of the results in terms of chemistry and physics is difficult and requires sensitivity testing by varying the input data. While useful for calibration within the testing range of the data employed, using this method for understanding the underlying science is difficult.

Partial Least Squares

PCR is based on the spectral decomposition of XX' to select latent variables for regression, while the Partial Least Squares method (PLS) is based on the singular value decomposition of $X'Y$, meaning that the independent variables are compared to the dependant variables. In practice, PLS usually fairs better than PCR since the reduction of parameter space dimensions is accomplished through comparison of the independent variables with the dependent variables. PCR, on the other hand, focuses mainly on what can be thought as the signal strengths of the independent variables alone for parameter space reduction, and is therefore more prone to the introduction of irrelevant signals into the regression. As with PCR, PLS suffers from the difficulty that the complex axis rotations make understanding what the latent variables represent in terms of chemistry and physics is difficult and requires sensitivity testing by varying the input

data. While useful for calibration within the testing range of the data employed, using this method for understanding the underlying science is difficult.

Other Methods

Many other algorithms have been developed in recent years, including neural networks and artificial intelligences. While these “black box methods” can work extremely well over the calibration range used, we still are faced with the difficulty of understanding how the input variables relate directly to dependent variable without sensitivity testing. Because of the difficulty of latent variable methods to demonstrate the correlation in terms of directly measured variables, we developed our own methods to address the issues associated with impossible and/or unfavorable parameter to observation ratios. In the simplest of terms, two strategies can be employed to make the problem tractable; reduce the independent variable count, or increase the number of observations. Once a statistically meaningful correlation can be computed, a method for selecting the most important independent variables must be applied to remove irrelevant signals and find those responses that significantly affect the quality of the fit. When applying this technique to infrared to rheology correlations, the independent variables are spectral wave numbers and represent vibrational modes of functional groups. Consequently, important clues about how chemical changes cause rheological changes can be obtained.

SPECTRELATE SOFTWARE DEVELOPMENT

After reviewing the available methods for multivariate studies of spectral data, we found that understanding the important wave numbers was very difficult with methods utilizing the latent variable concepts. The time honored linear multivariable method does produce closed form correlations, but is only useful if intelligent selection of possible significant wave numbers is known a priori. The number of independent variables to use in the correlations is also limited based upon the number of actual observations used. Monte Carlo approaches would likely work, but the adaptation of the ideas present in PCR to reduce variable counts in a logical manner, along with a defensible method for expanding the observation matrix, is the strategy we chose to employ. The core strategy we employ involves reducing the number of independent variables into groups and increasing observations artificially if needed. In short, we use multivariable linear regression to examine the entire spectrum at one time. This is possible by preconditioning the data set by combining independent variable single measurements (that is, a single wave-number) that represent the same information to reduce the independent variable list and by increasing the observation matrix size by generating synthetic replicates from knowledge of measurement precision. The method precision provides an envelope for the application of random variations in the existing measurements to generate a probable collection of additional synthetic observations.

Consider a set of asphalt spectra aged at different conditions. We also have rheological measurements of these materials corresponding to the each spectrum measurement. For the time being we are not interested in the dependent variable (the rheology) but only the spectra. There will be variations in the spectra depending on the aging conditions. If we take a data set that contains infrared spectra of an asphalt that has been oxidized over a range of times, all other

factors held constant (pressure, temperature, concentrations) and then plot a single wave number against all the others, we find that for some wave numbers the plot is nearly a perfect line, and for others, the plot is quite scattered. Let's illustrate this idea with some real data from WRI's ambient atmosphere thin film aging study.

If we plot the 1035 wave number against 3200, for AAB-1, we get the plot shown in figure 1.

Each point in the plot represents those absorbances for that amount of oxidation. So, an increase in the 3200 absorbance is always accompanied by an increase in the 1035 absorbance by a factor of roughly five. For this binder, it appears that the 1035 response and the 3200 response are very closely related (correlation coefficient of 0.99) and describe either the same functional group in different vibrational modes, or possibly describe reaction products produced by the same mechanism. Other wave numbers have no correlation at all (correlation coefficient of 0.10), as this comparison of 1035 with 3600 demonstrates in figure 2.

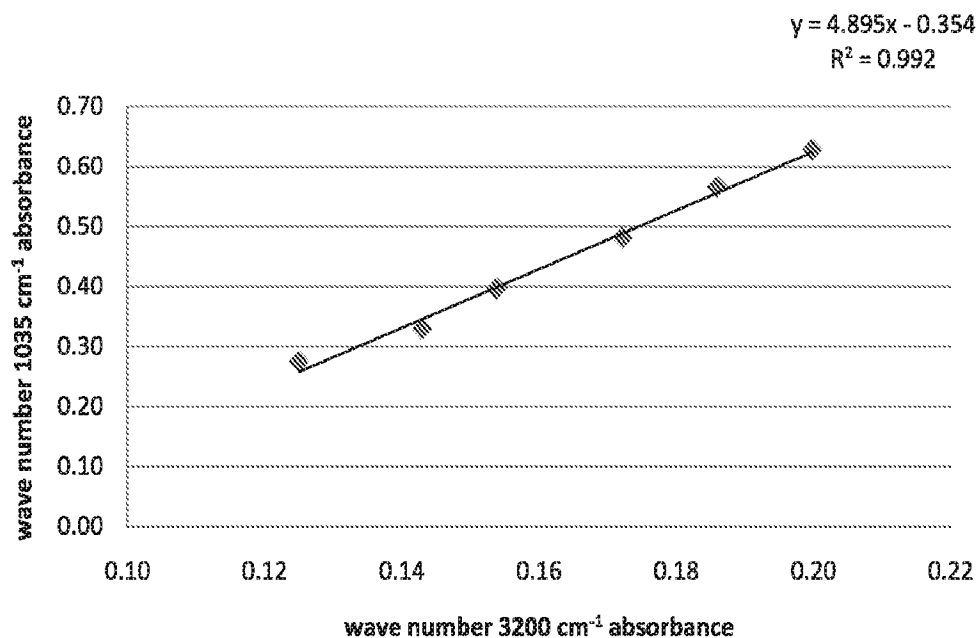


Figure 1. Graph. The correlation between wave numbers 3200 and 1035 for oxidized AAD-1 asphalt binder.

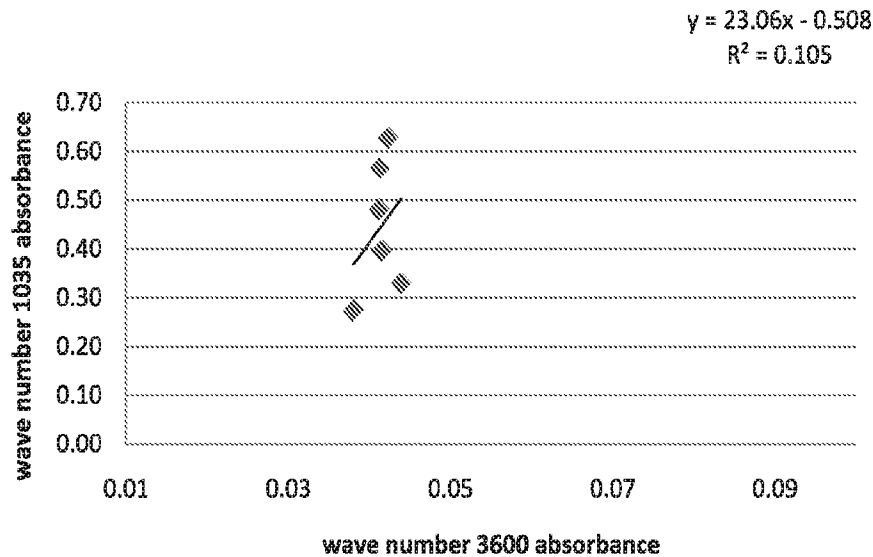


Figure 2. Graph. The correlation between wave numbers 3600 and 1035 for oxidized AAD-1 asphalt binder.

If we perform this exercise for each and every wave number, and record the correlation coefficient, we can construct “regression spectra” to investigate which wave numbers seem to contain the same information as the asphalt oxidized (these independent variable change in proportion to each other, and are not co-linear, but co-related). A co-linear relationship is on the same line. The slopes and intercepts are identical. For the binder AAD-1, we get the plot shown in figure 3 when the regression spectra are plotted.

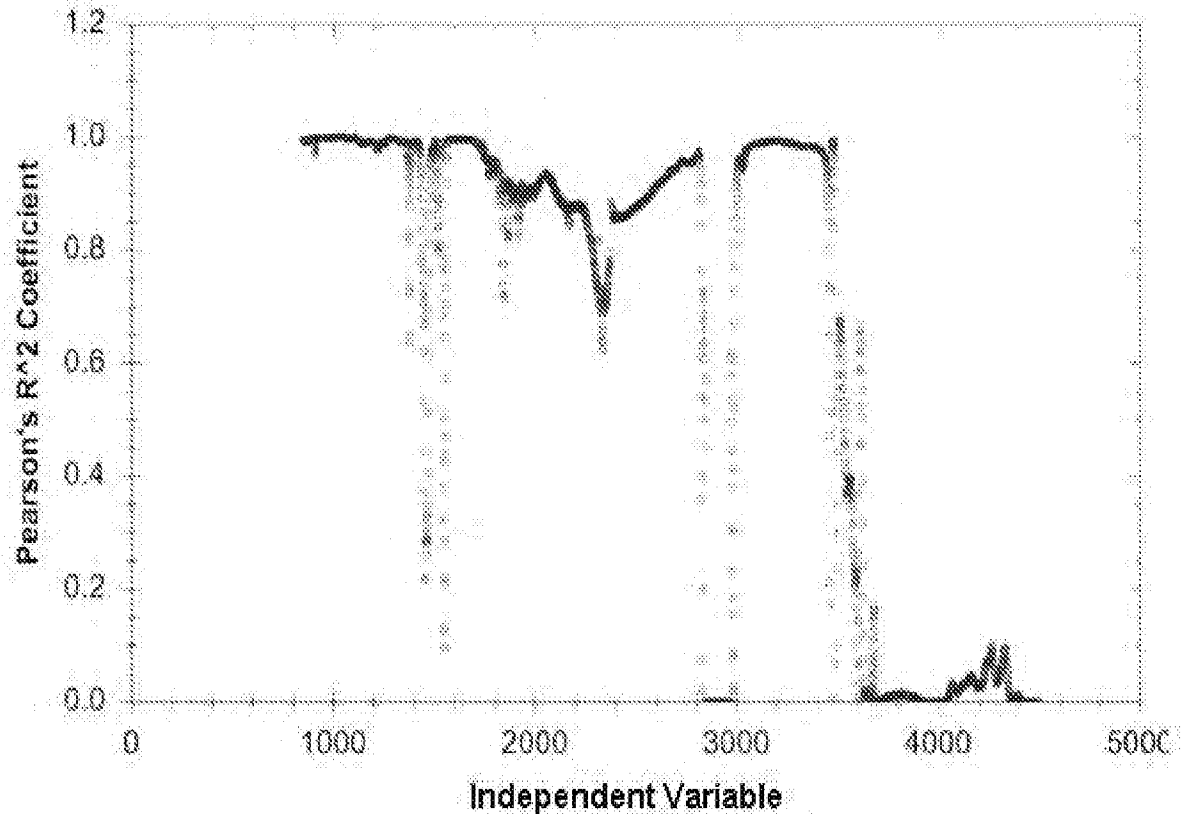


Figure 3. Screen capture. The regression spectrum for wave number 1035 for oxidized AAD-1 asphalt binder.

For this particular binder, if we were to choose a cross correlation threshold of 0.95, then most of the spectrum would fall into one group, greatly simplifying any regression efforts. However, when studying a collection of spectra for a variety of binders, we find that groupings vary from binder to binder. Compare a regression spectrum of AAB-1 (figure 4) with AAD-1 above (figure 3).

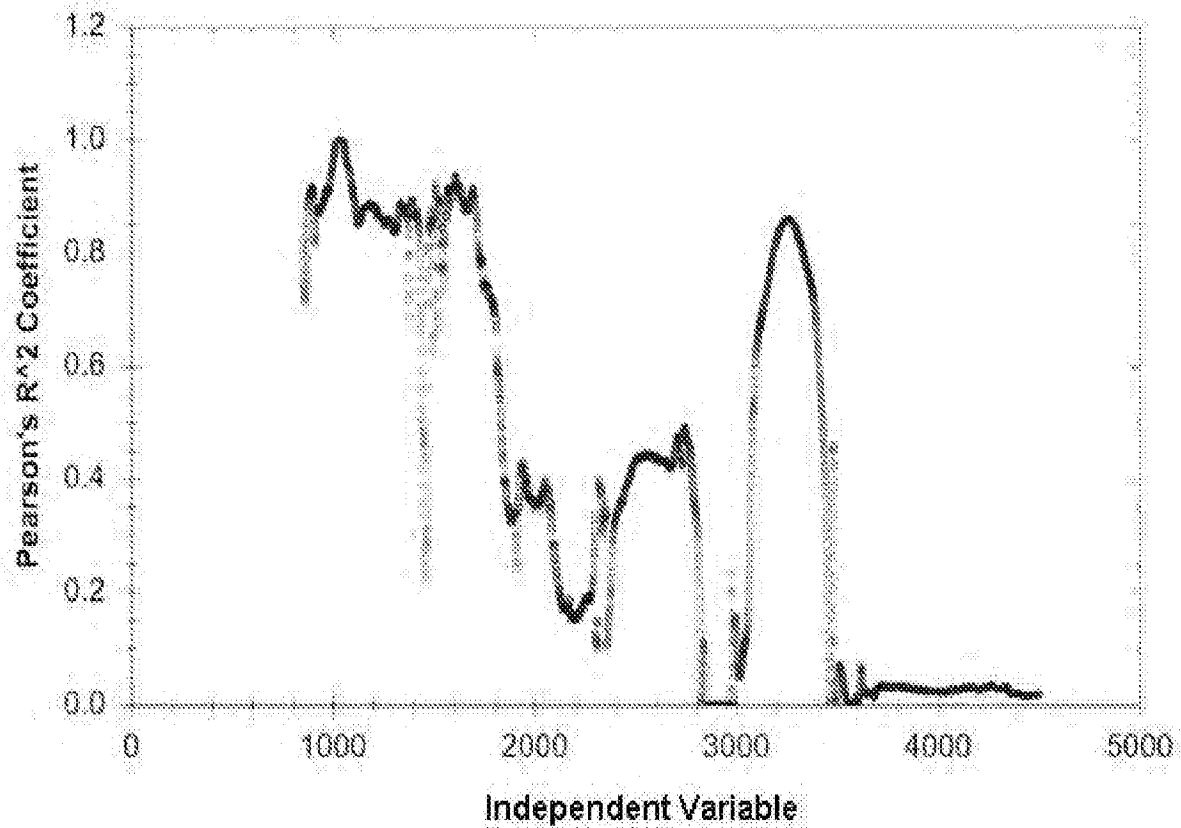


Figure 4. Screen capture. The regression spectrum for wave number 1035 for oxidized AAB-1 asphalt binder.

If we further extend the exercise and produce an “overall” grouping schema for many binders by creating a regression for this wave number (1035) we get the regression spectrum shown in figure 5.

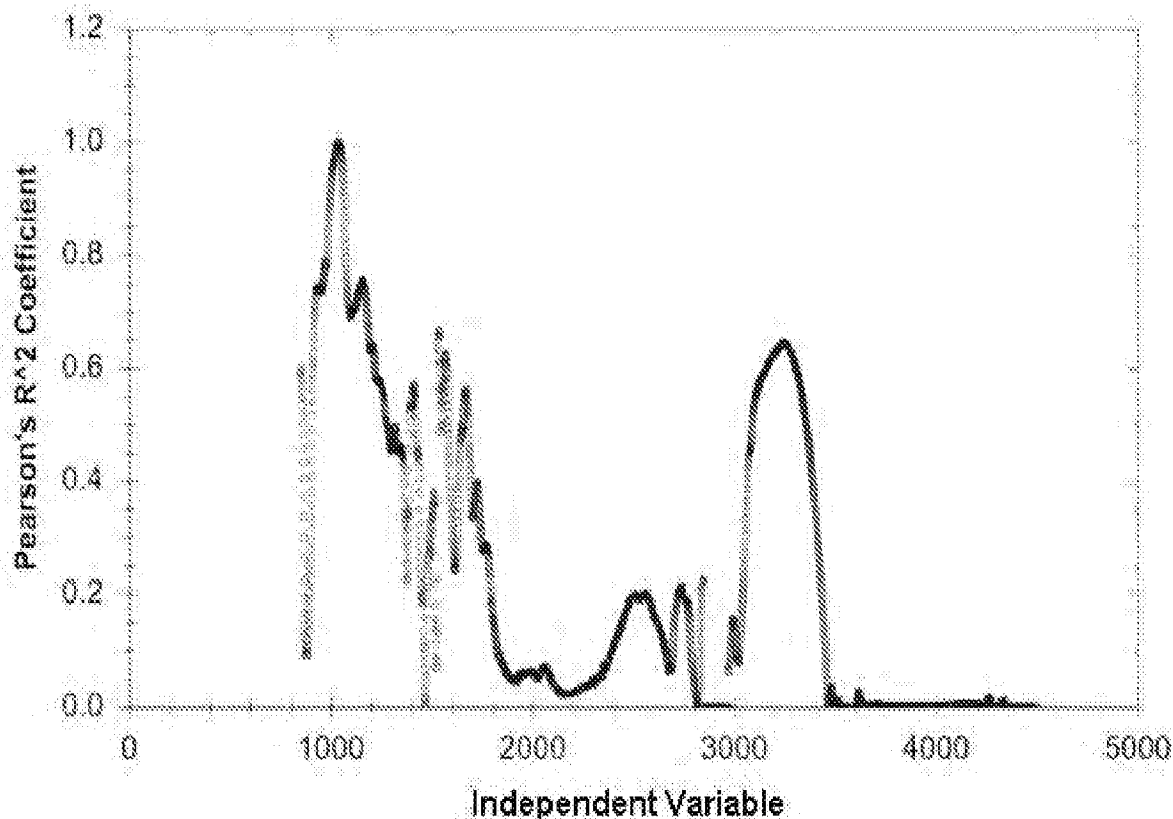


Figure 5. Screen capture. The regression spectrum for wave number 1035 for oxidized asphalt binder data set.

This plot suggests that nothing really cross correlates at any wave number with 1035 except adjacent wave numbers making up the primary peak. However, if we take our pre conditioning a step farther, and consider only the changes in the spectra by subtracting the RTFO spectra (time 0) from the remaining spectra aged at times up to 12 weeks, a very different pattern emerges. We are now only considering the changes that occur with oxidation.

The raw spectral data is shown in figure 6 with nonlinear and solvent affected areas set to zero:

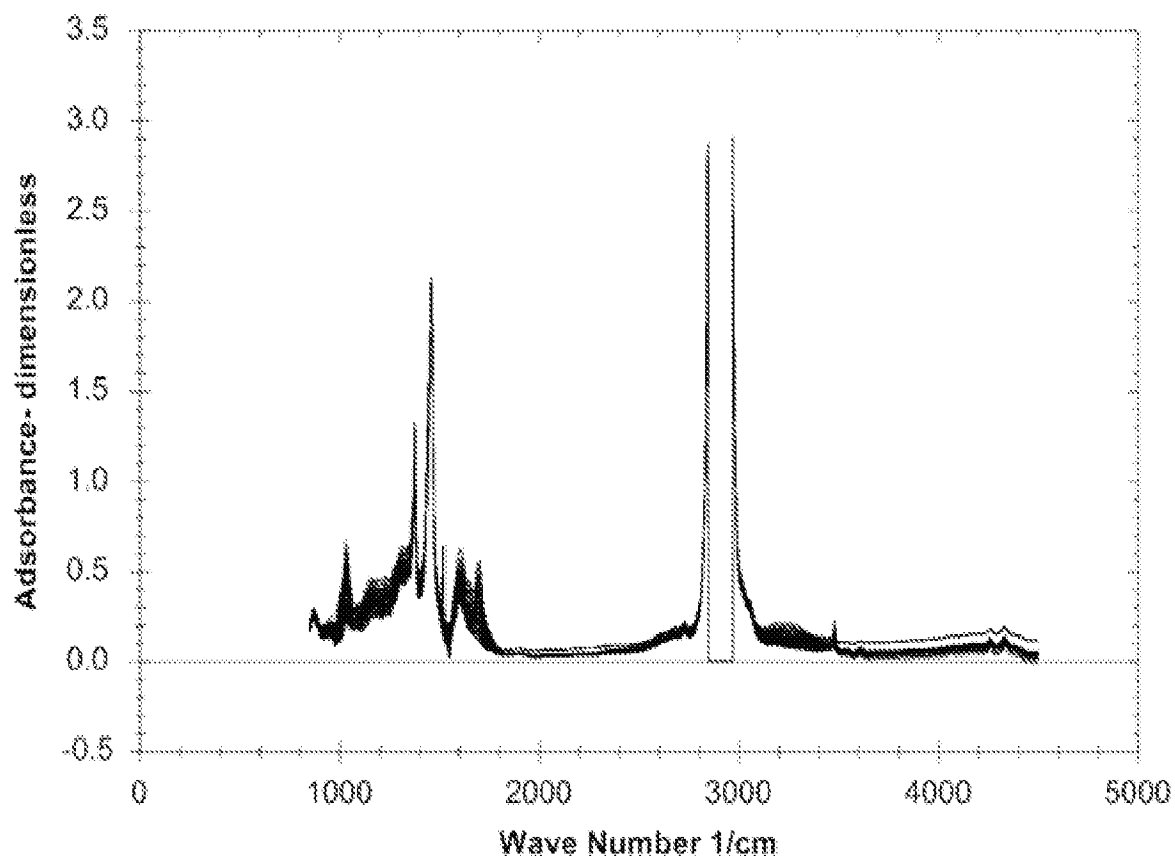


Figure 6. Screen capture. Mid infrared spectra with non-linear response areas removed.

The RTFO (zero oxidation time) subtracted spectra is shown in figure 7 (effectively eliminating the y axis intercept in the regressions and focusing on the changes).

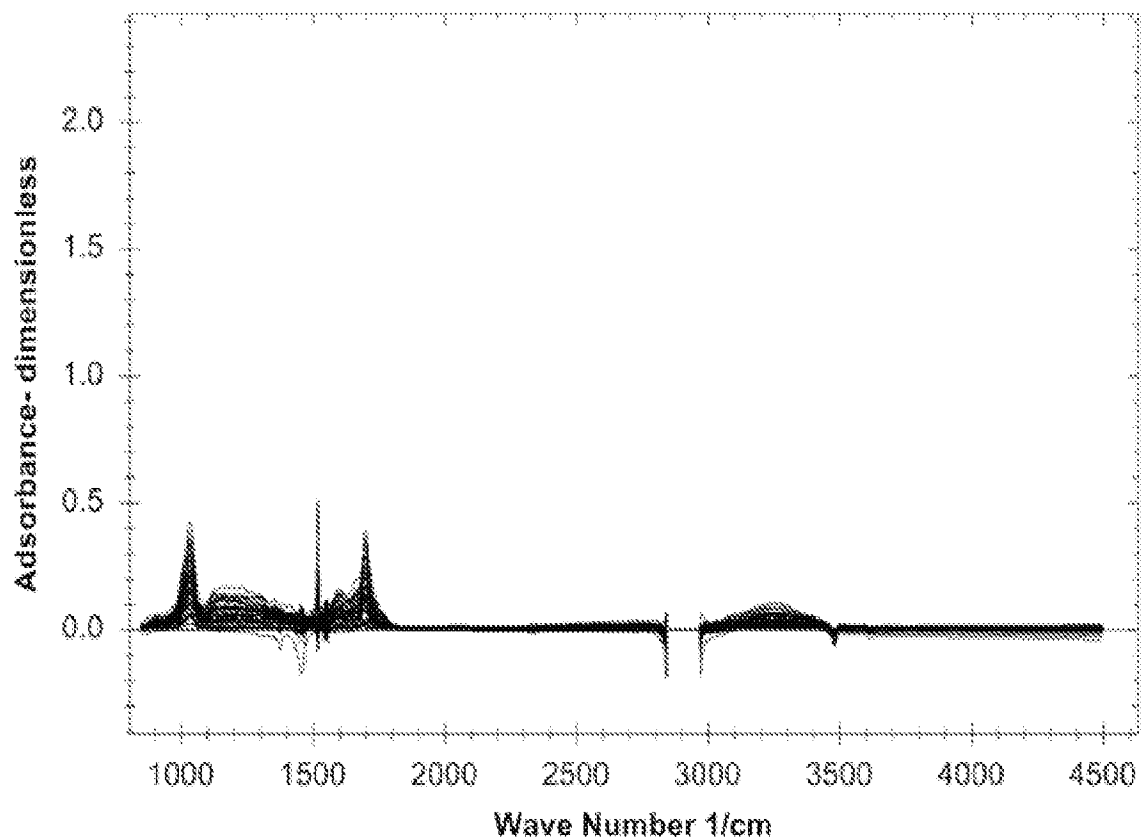


Figure 7. Screen capture. Mid infrared change spectra with non-linear response areas removed.

When 12 different binders are correlated, we get a the regression spectrum shown in figure 8.

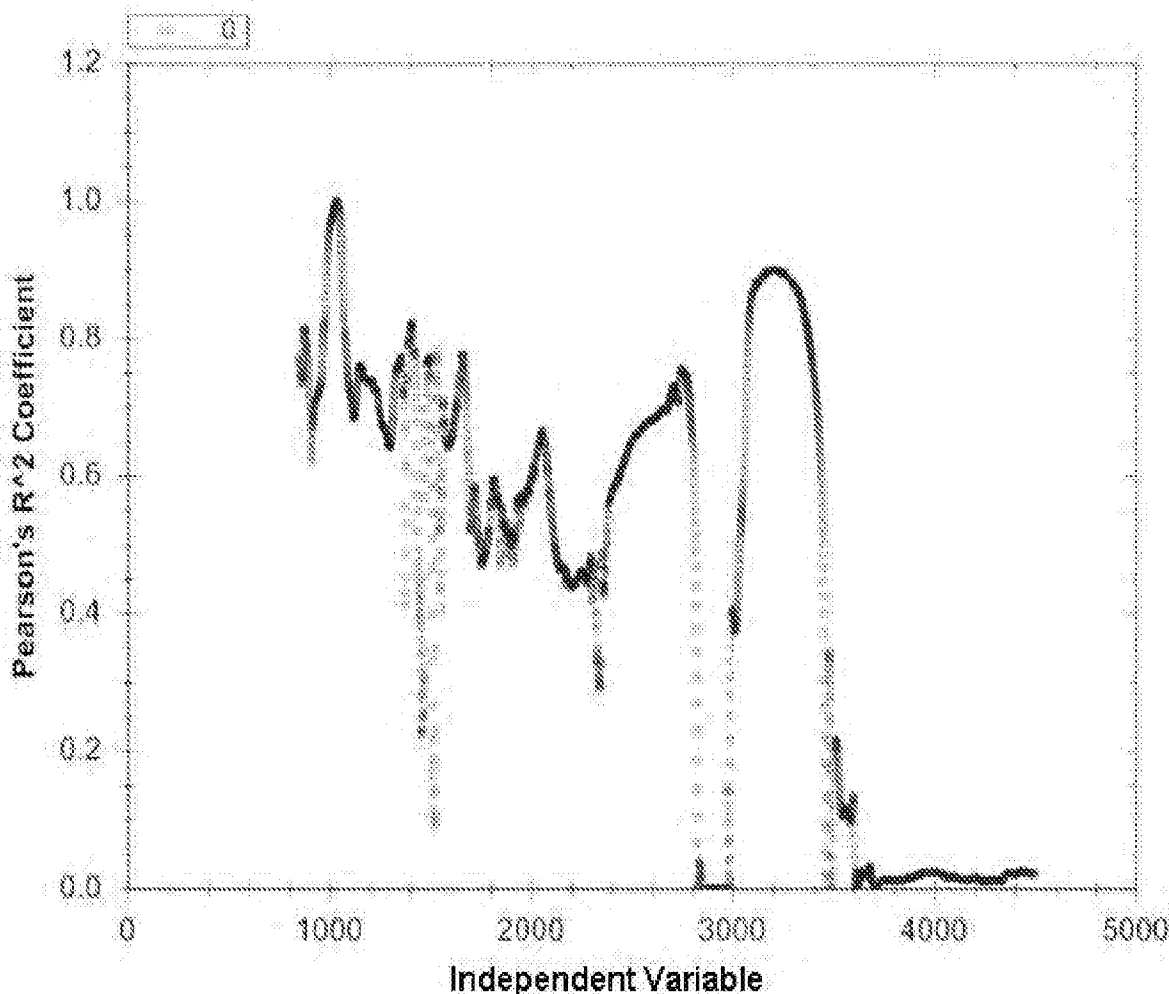


Figure 8. Screen capture. The regression spectrum for wave number 1035 for oxidized asphalt binder data set change spectra.

In figure 8 above, as compared to figure 5, the increase in sulfoxide (wave number 1035) and the increase in hydrogen bonding (wave number 3200) appears to be related. They both grow proportionally as oxidation progresses, providing clues to the underlying chemical mechanism.

When the program groups the related wave numbers, the user must specify a correlation threshold as criteria for group selection. If a regression threshold of $r^2=0.8$ is selected, then this group would contain all the wave numbers inside the grayed out area (figure 9), reducing the number of independent variables significantly. A higher threshold produces more groups, with fewer members in each group. A threshold above $r^2=0.95$ or so would not include any of the 3200 peak in the group. For example, if try to correlate each wave number in the collection of spectra with every other wave number with an excessively strict regression threshold of 1, we would have 3200 groups(spectral range of 800 to 4000) with a single member in each group. As we reduce the criteria for assuming the wave number represents the same information as others in the spectra, then those wave numbers near the wave number of interest begin to be included, describing the region of the spectral band that “moves” with the wave number of interest in the data set. If the threshold for acceptance into the group becomes low enough, we begin to see

other regions in the spectrum some distance from the immediate peak or band that either describes other vibrational modes of the same functional group, or in the case of a reaction series like oxidation, these distance related responses might include other functional groups participating in similar reaction paths. The decision on how many groups is reasonable is essentially a matter of judgment, perhaps best tempered with experimenting with different threshold values in the regression analyses.

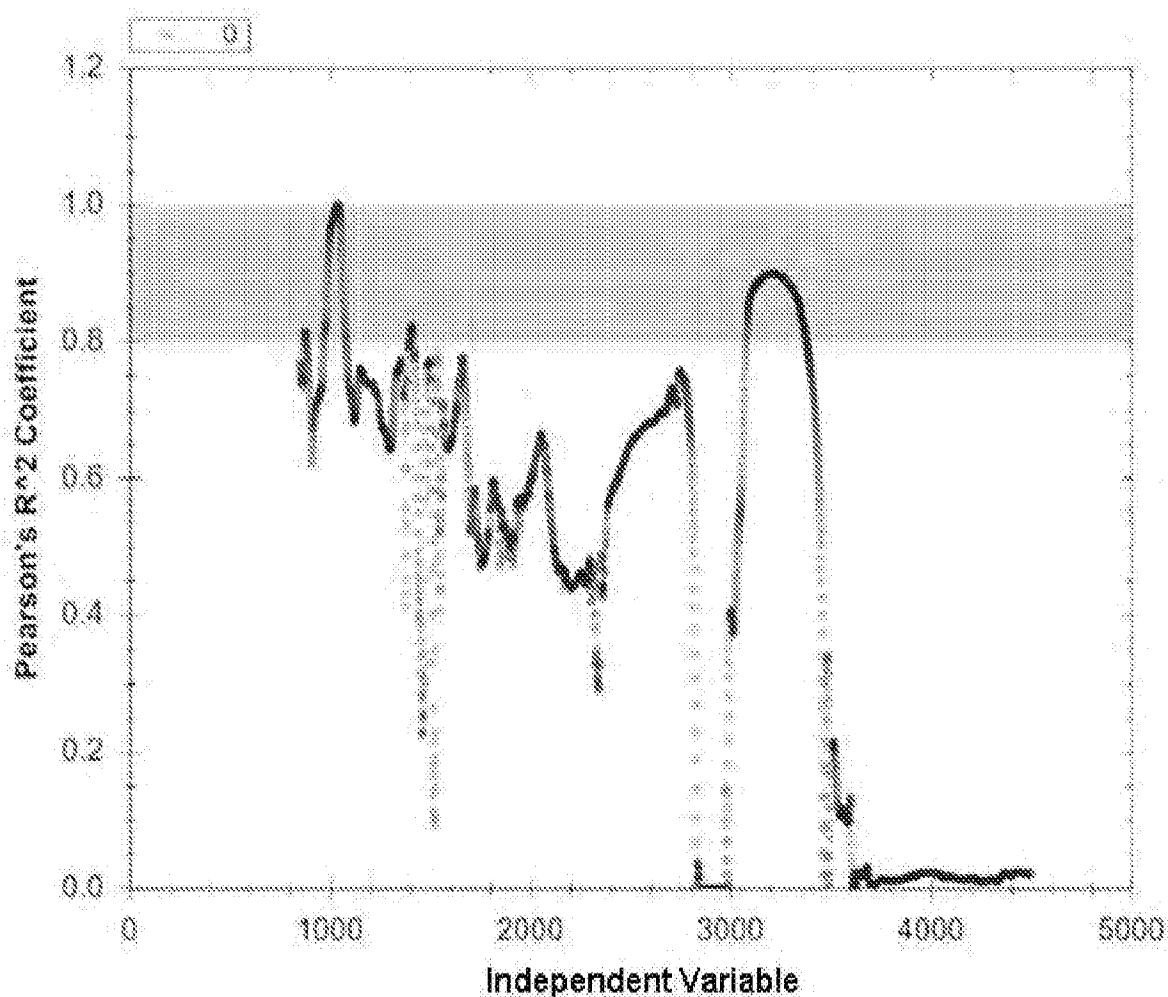


Figure 9. Screen capture. The regression spectrum for wave number 1035 for oxidized asphalt binder data set with a grouping range shown.

Using this same data set, which is 70 oxidation spectra obtained from 12 binders, the number of groups identified with this technique is shown in table 1.

Table 1. Independent variable counts at different grouping threshold values.

Net Change Spectra (12 Binders-70 Spectra)	
R^2 Threshold	Number of Groups
0.99	284
0.95	101
0.9	51
0.85	35
0.8	28
0.75	22
0.7	20
0.65	16
0.6	12
0.55	11
0.5	9
0.45	8
0.4	7
0.35	6
0.3	5

The question then arises as to how groups are selected to properly characterize the data without including differentiation caused by method noise. A defensible theory is not currently known to the authors, but a semi-log plot of the table above does provide some guidance with regard to a point of diminishing return appearing to exist near the threshold of $r^2=0.9$ (figure 10).

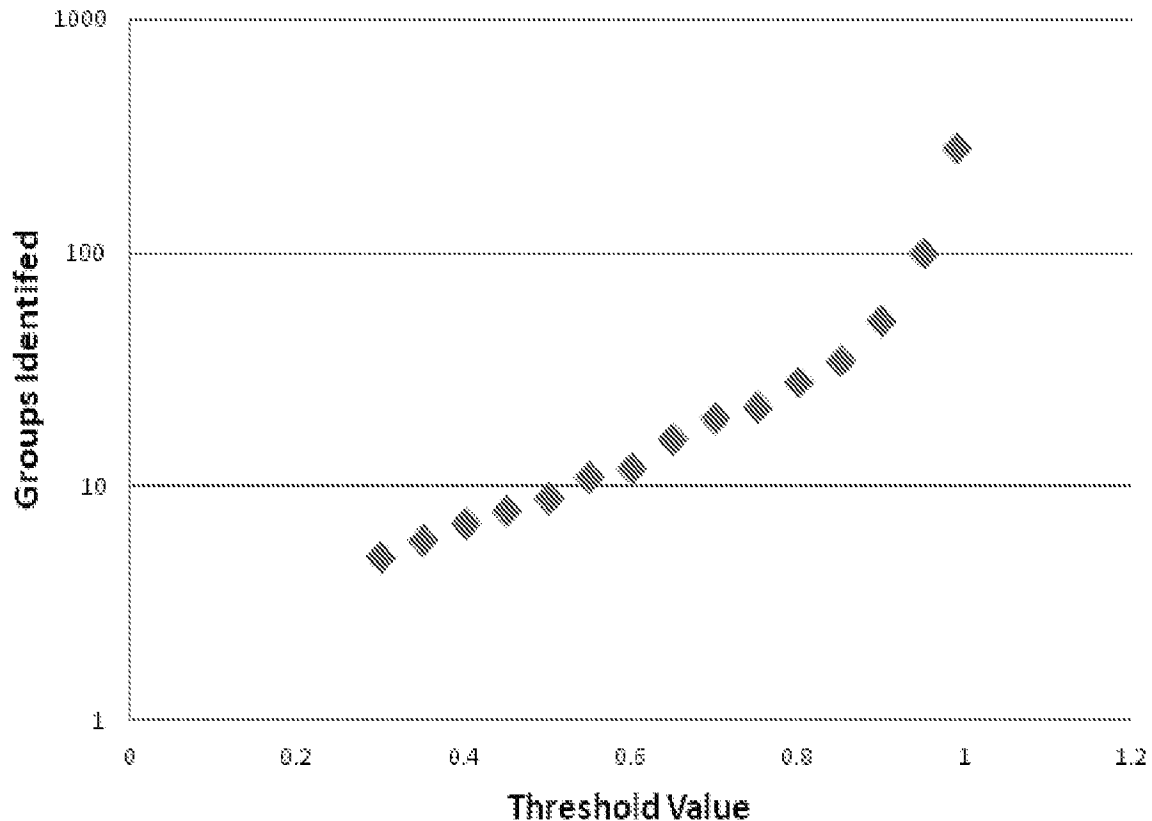


Figure 10. Graph. Groups produced at different threshold values.

Providing the method above produces a reasonable guide to identify meaningful signals from the data, we have 51 independent variables to examine in combination for significance with only 70 observations available when using a grouping threshold of 0.9. The observation ratio is only slightly greater than 1, so any fitting procedure would be a 51 dimensional analogue to drawing a line through two points in 2 dimensions. If we had 357 observations, (O/P, observations/parameters = 7) then we could begin to believe the statistical validity of our correlation attempts. We accomplish this by generating synthetic independent data and dependent data (in this example, spectra and rheological measurements) by randomly varying the measurements within the method precision envelope. The idea here is that if it were practical to create such a huge data base, the data would look very much like the synthetically generated set. Once the regression is performed using a commercially available multivariable linear regression package, the significance of the individual variables can be ranked based upon the F test for that variable (calculated by the statistics package). The F test is a numerical representation of the quality of a fit with and without the independent variable of interest. Those variables that test with a higher value are more significant than those that test with a lower value. The formula used to calculate F test is

$$F = \frac{\left(\frac{RSS_1 - RSS_2}{p_2 - p_1} \right)}{\left(\frac{RSS_2}{n - p_2} \right)} \quad (5)$$

RSS_1 and RSS_2 are residual sum of squares of model 1 and model 2 (residual sum of squares is the squared difference between measured and model values, and in our case the two models are those with and without the parameter in question). The numbers of parameters used for each model are p_1 and p_2 , and n is the number of observations. So, in simple terms this is a ratio of goodness of fit with and without the parameter in question. The F-test value provides a means of ranking the significance of the parameters in the models proposed, which can be used as a rejection criteria.

Our software performs the correlations by first grouping the independent variables, then creating, if needed, sufficient observations by synthetic replication to make the matrix not only tractable, but statistically significant. The idea of using synthetic replication is rather simple, but difficult to describe concisely in multidimensional problems. Fortunately, the principles at work can be shown in simple two dimensional linear fitting problems with only two adjustable parameters, the line slope and the intercept. Clearly, at least two points are required to obtain some idea of the relationship existing between the measurements, if any. Figure 11 is a plot of two points generated from the relationship $y=0.25x+1$.

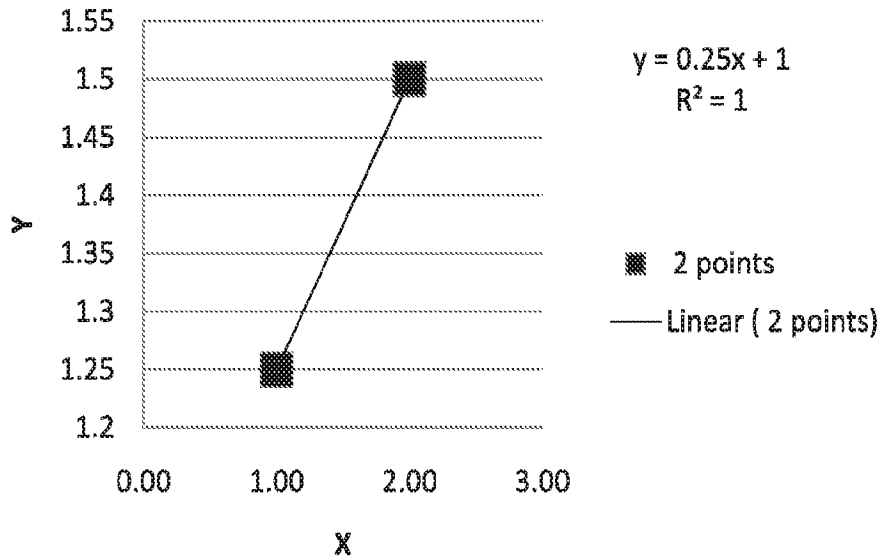


Figure 11. Graph. Simple linear plot with insufficient measures to test confidence.

If the measurements are perfect (as they are in figure 11) the relationship between the variables is clear. Notice that Pearson's regression coefficient, r^2 , is perfect in this case and that this metric gives us no indication of confidence in the data. Suppose we have a very low signal to ratio for these measurements of only 0.5, which is a proportional precision of +/- 200%. If we repeat our measurements 7 times for each point, we get the following plot:

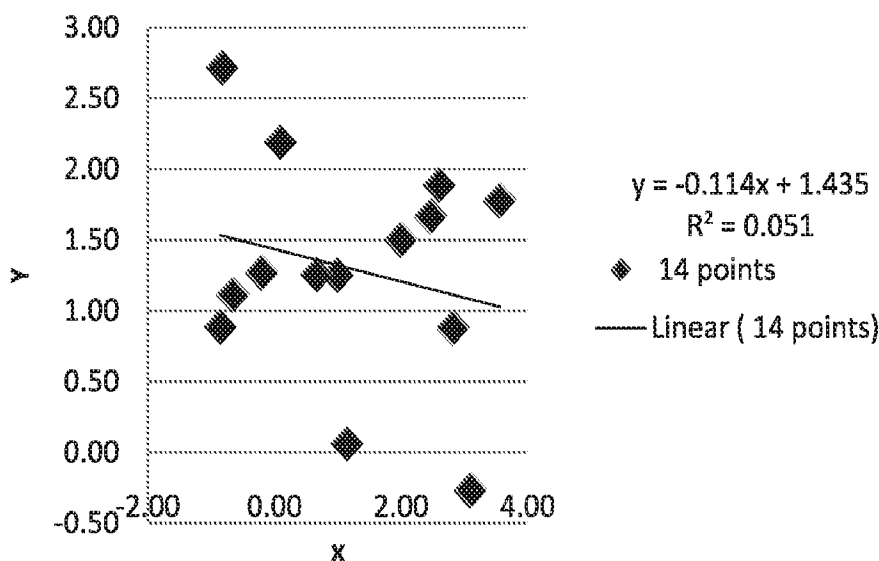


Figure 12. Graph. Simple linear plot with 7 replicates per measurement, signal to noise ratio 0.5.

In this case, the correlation coefficient is quite small, indicating very little confidence in the data. If this data were part of a multidimensional fit, the F test, which is also based upon the sum of squared residuals, would low. The important point to realize is even though there is a correlation in this set, we cannot get an accurate fit without running many, many replicates. The other point to keep in mind, is that synthetic replicates assumes the actual measured value is near the mean, but we have no way of knowing without real replicates. The actual fit is poor in this case as well. This rarely matters in our software implementation, since we are checking several thousand dimensions (independent variables) in the spectra to see if they are statistically believable. If not, they are thrown out. The situation improves significantly with better precision and/or areas of the spectra with a strong response. Keep in mind that the proportional precision in spectra varies with measurement strength. Low absorbance signals have huge error bounds, while strong signals have small proportional error bounds. The absolute error is approximately constant and the software provides that option for specifying precision. Consider a signal noise ratio of 5 (20% precision) as shown in figure 13, which would apply to a stronger absorbance measurement than seen in figure 12:

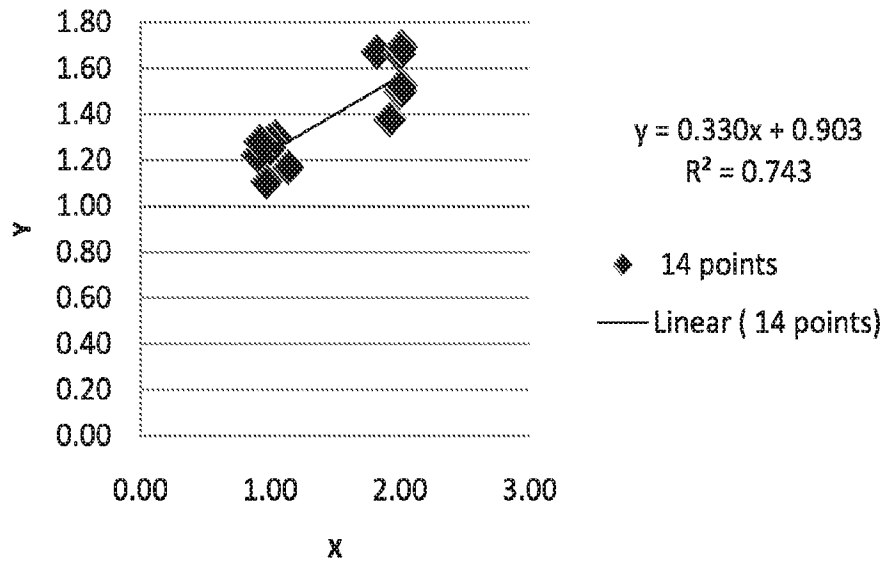


Figure 13. Graph. Simple linear plot with 7 replicates per measurement, signal to noise ratio 5.

Further improvement in the signal to noise ratio to the value of 20 (5% precision) leads to a high confidence correlation:

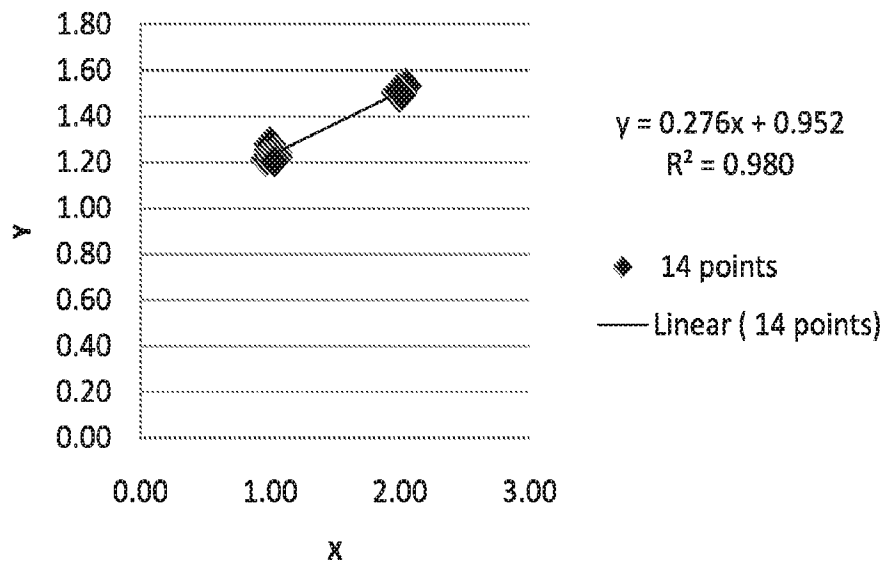


Figure 14. Graph. Simple linear plot with 7 replicates per measurement, signal to noise ratio 20.

For the hypothetical, simple data set presented here, we have plotted in figure 15 the relationship between the number of replicates, and regression coefficient for measurement precisions ranging from 5% to 200%. We typically begin to have a reasonable estimate of the correlation coefficient at when the number of replicates exceeds 7. This behavior carries over to multivariable correlations as well.

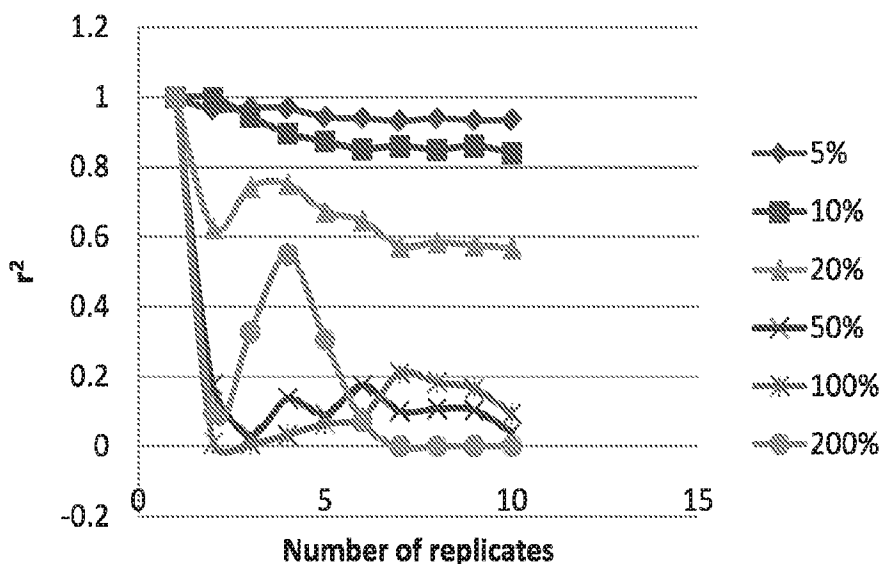


Figure 15. Graph. The regression coefficient dependence on the number of replications and the measurement precision.

To recap, the Spectrelate program reduces the number of independent variables by grouping measurements that correlate with each other into a single value. In the event the ratio of observations to parameters (or observations to independent variables +1) is unfavorable for matrix solution or more commonly meaningful regression coefficient or F test calculations, synthetic replication can be used to increase the O/P ratio in order to judge which variables have insufficient precision and signal strength to justify their use. These are eliminated from consideration. The initial multivariable correlation is performed assuming all of the independent variables are needed in the model. The F-test ranking is then used to discard the variables one by one, performing repeated correlations until there is only one independent variable remaining. The Pearson's correlation coefficient is plotted against the number of model parameters, and a judgment made, guided by the shape of this curve, of how many independent variables are needed to explain the variance within the precision of the data. If properly used, the resulting selected independent variables will produce the proper correlation without grouping or synthetic replication provided the O/P ratio for the final model is large enough.

PROGRAM IMPLEMENTATION

Data Input

As with any computational software, experimentally obtained measurements must be entered to perform the calculations. The data is entered in the form of two files, one containing the independent variable measurements, and a second file containing the dependent file measurements.

In the example files provided, the independent variables are the measured absorbances at each frequency, expressed as wave number. The data is read as a comma delimited file, which most

commercially available spectrometers can produce. Any data that can be placed into a Microsoft Excel spreadsheet can be saved as a comma delimited file. These files can also be opened in Excel for editing or perform alternative calculations such as variable transformations. The first row is arbitrary and can contain any information of the user desires. The second row serves as a label for the data in each column, and for the example datasets this will be the name of the infrared file. This text entry will appear in the observation list in the software. The first column is an exception as it contains the wave number list. Each spectrum must be the same length and have the same spacing between wave number readings.

The dependent variable file is of similar form to the independent variable file. The first row can have arbitrary entries (not used by the software). The second row identifies the samples from which the measurements (or measurements) were taken that are found in the column below. Any number of measurements can exist in the column below, but there must be at least one. The second rows should indicate the same sample was used for both the infrared measurements in the columns below it and also the same physical sample was used to obtain the dependent variable measurement (in our case, rheological measurements). The first column is the labels of each dependent variable measurement. If user wishes to investigate variable transformations, such as the logarithm of a measurement, it is easy to do here by entering the measured values for each sample in row 1, and the transformed value in row 2. The label for these entries (each row) will be displayed in the dependent variable selection list. This selection must be made before any kind of regression can be attempted. The structure of the independent and dependent variable files are easily understood by examination. They can be opened in excel, or a text editor.

The process of reading the data and placing it into numerical arrays in the program is initiated by first selecting the "Load Independent Data" button (figure 16).

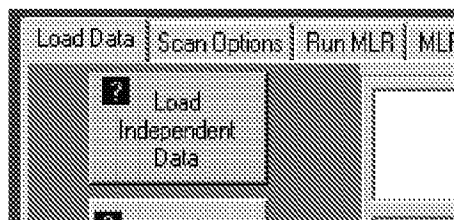


Figure 16. Screen capture. Load data button.

An open file dialog box appears (figure 17), and the user can navigate the computer file structure to find the desired file.

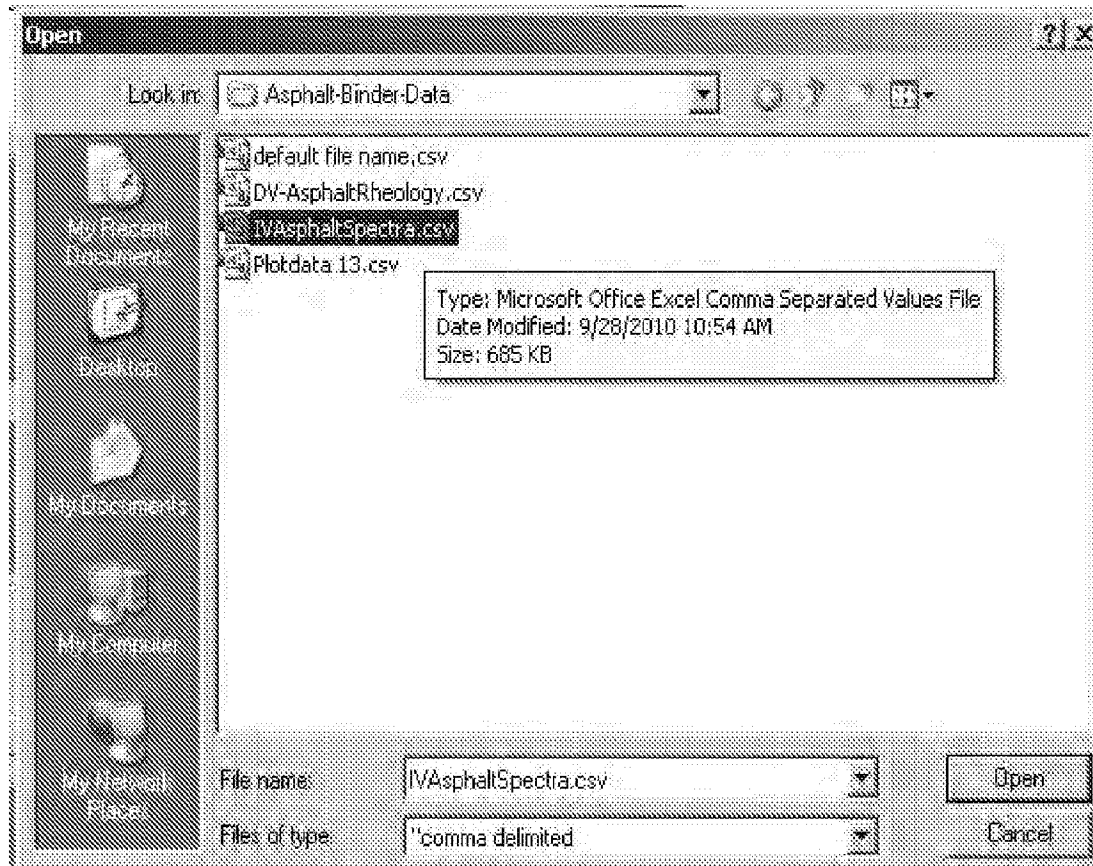


Figure 17. Screen capture. Data file selection.

Once the data has been successfully read, the independent variable sets will be plotted in two of the plots to the right (raw data and reduced data, see figure 18). At this point, these are identical, but will change if cross correlation grouping is used.

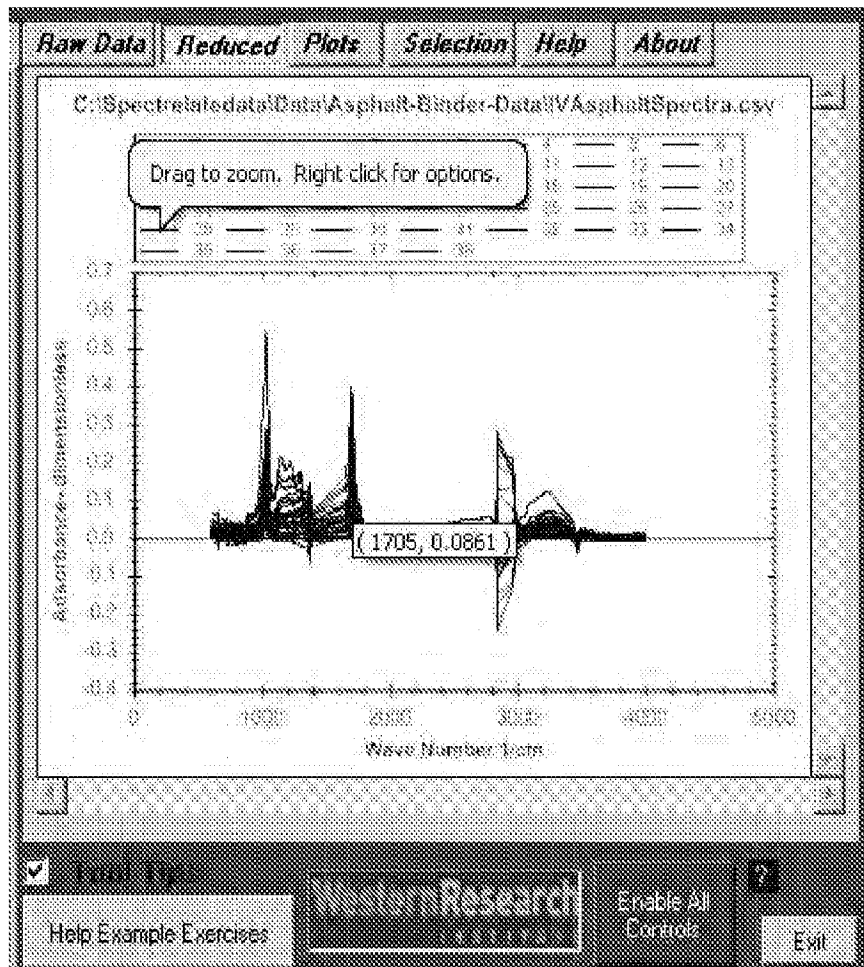


Figure 18. Screen capture. Data display.

A similar procedure for reading the dependent data file is employed. It is started with a click on the “Load Dependent File” button (figure 19).

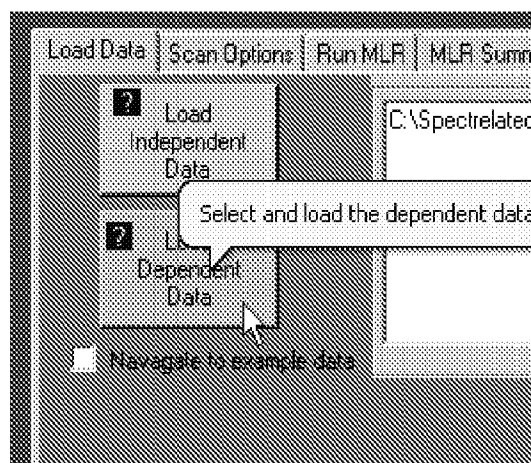


Figure 19. Screen capture. Dependent data button.

The dependent data are not plotted. However, the dependent variable selection list box will be enabled after the program verifies that sample counts match for both independent and dependent files. If you have several candidate dependent variables in your file, then there will be a list here. At least one item must be selected to proceed with any regression attempts (figure 20).

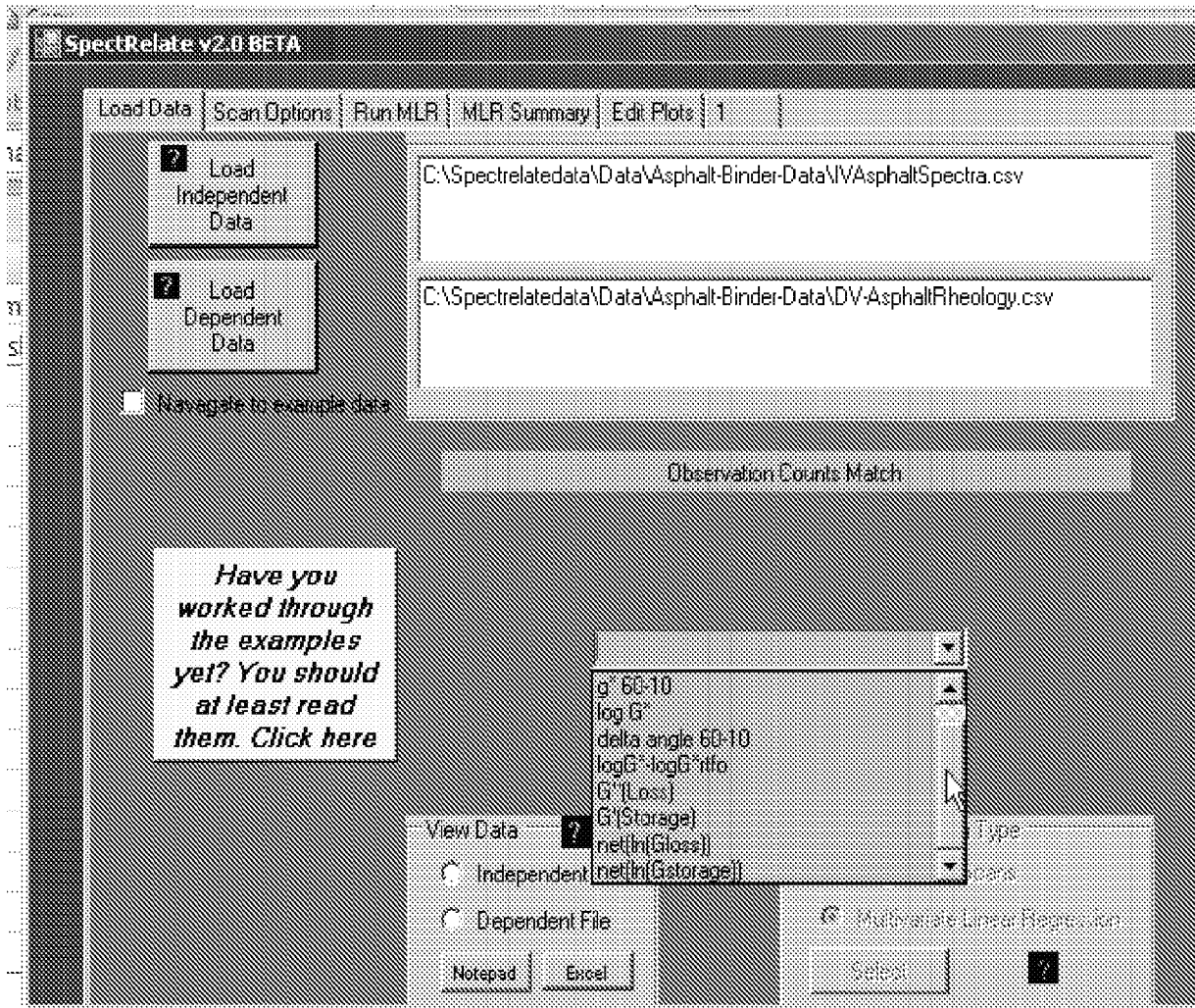


Figure 20. Screen capture. Dependent variable selection.

Dependent Variable Selection

A dependent variable must be selected to enable additional controls. Here we select the change in the logarithm of the complex modulus.

Regression Type

Two options exist regarding regression type. The regression scan goes through the entire list of independent variables and does a single variable regression on each one of them (figure 21).

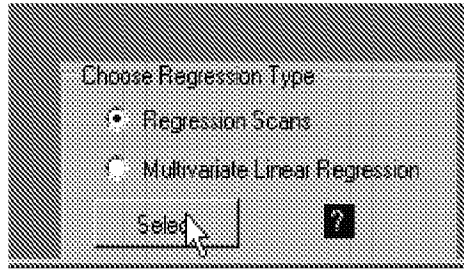


Figure 21. Screen capture. Regression type selection.

Upon selection, the tab control shifts to the regression scan page, and a scan is performed on the current data set using one variable at a time. The regression coefficient, the intercept, or the slope of the line for each one of these linear correlations is plotted on the y axis under the plots tab. The example below in figure 22 shows the regression coefficient.

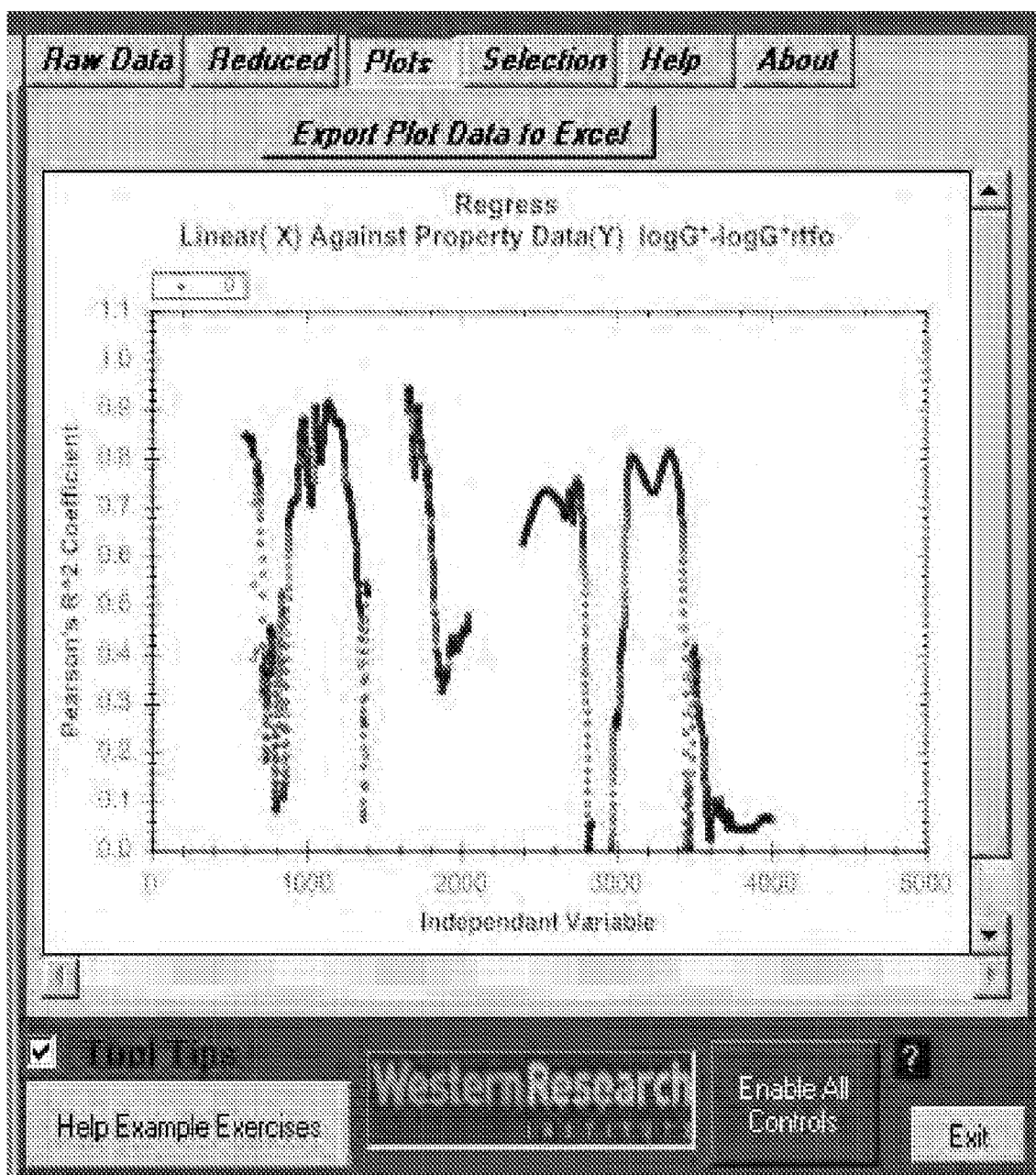


Figure 22. Screen capture. Regression spectra example.

The regression coefficients for individual independent variable can be queried from the plot with a mouse hover over the point. This is quite helpful if you want to examine the actual regression using the individual plots feature (figure 23).

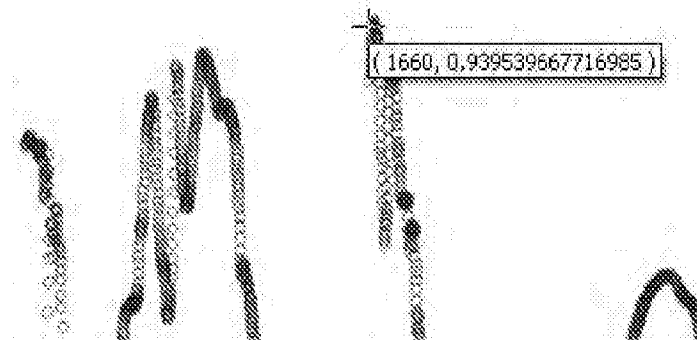


Figure 23. Screen capture. Regression spectra example (magnified).

In this example, we select the wave number 1660 and its value and the associated regression correlation coefficient is shown. To obtain the actual regression plot, the independent variable is selected from the drop down list for individual plots (figure 24).

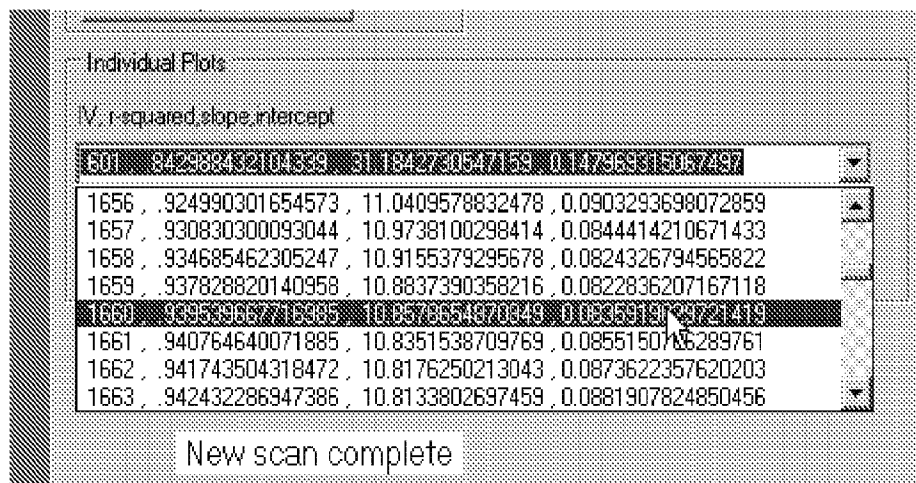


Figure 24. Screen capture. Regression plot selection.

A click of the single iv regression plot button produces the plot for examination (figure 25). The legend lists the measured data points and the regression line.

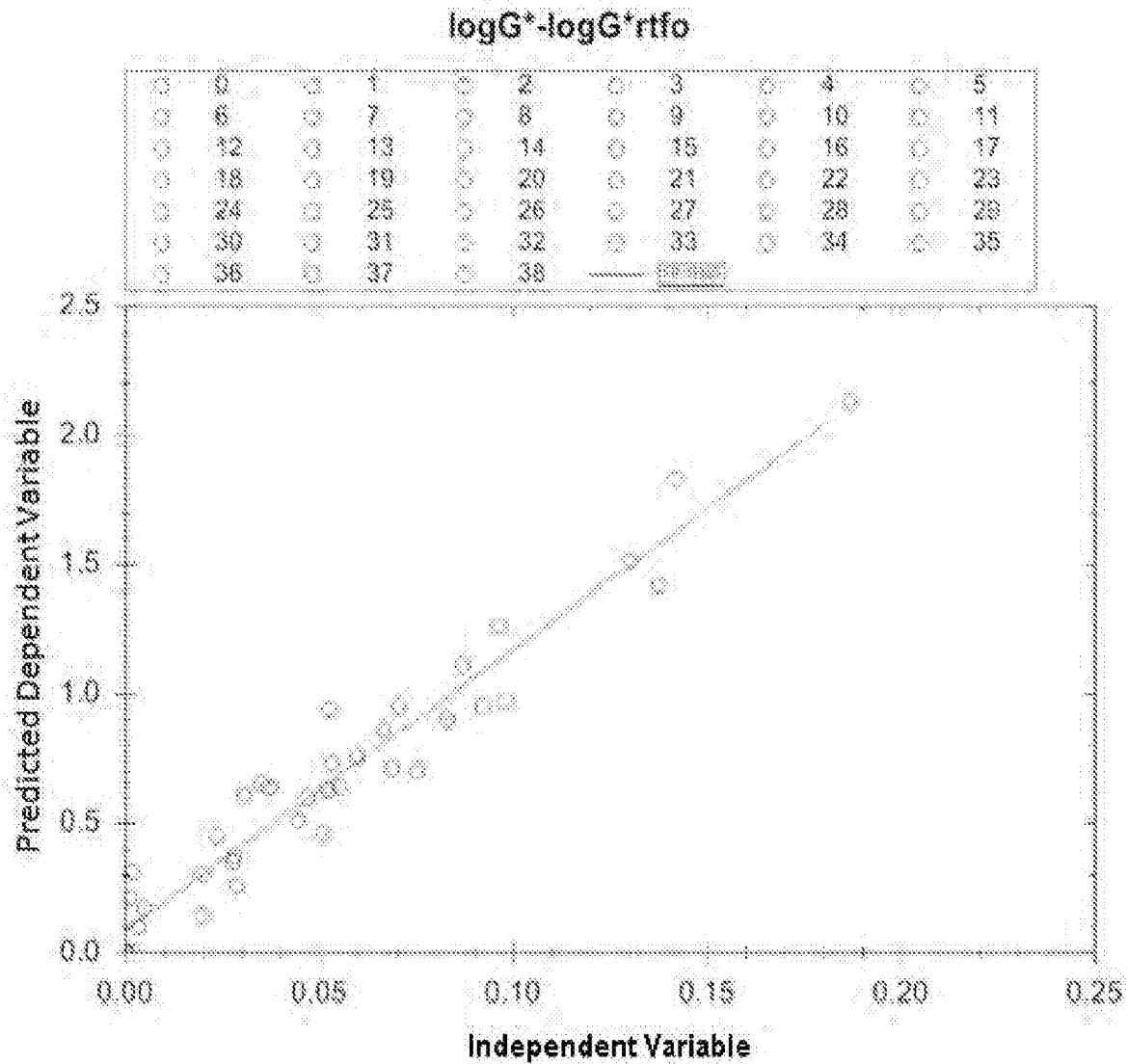


Figure 25. Screen capture. Regression plot example.

The poorly correlating wave number 1878 produces the following plot (figure 26). The examination of the individual plots is helpful to investigate the possibility of a relationship that is not linear and therefore appears to correlate poorly when a variable transformation to linearize the relationship may provide a better indication of the relationship. Obviously, such a situation does not exist in figure 26.

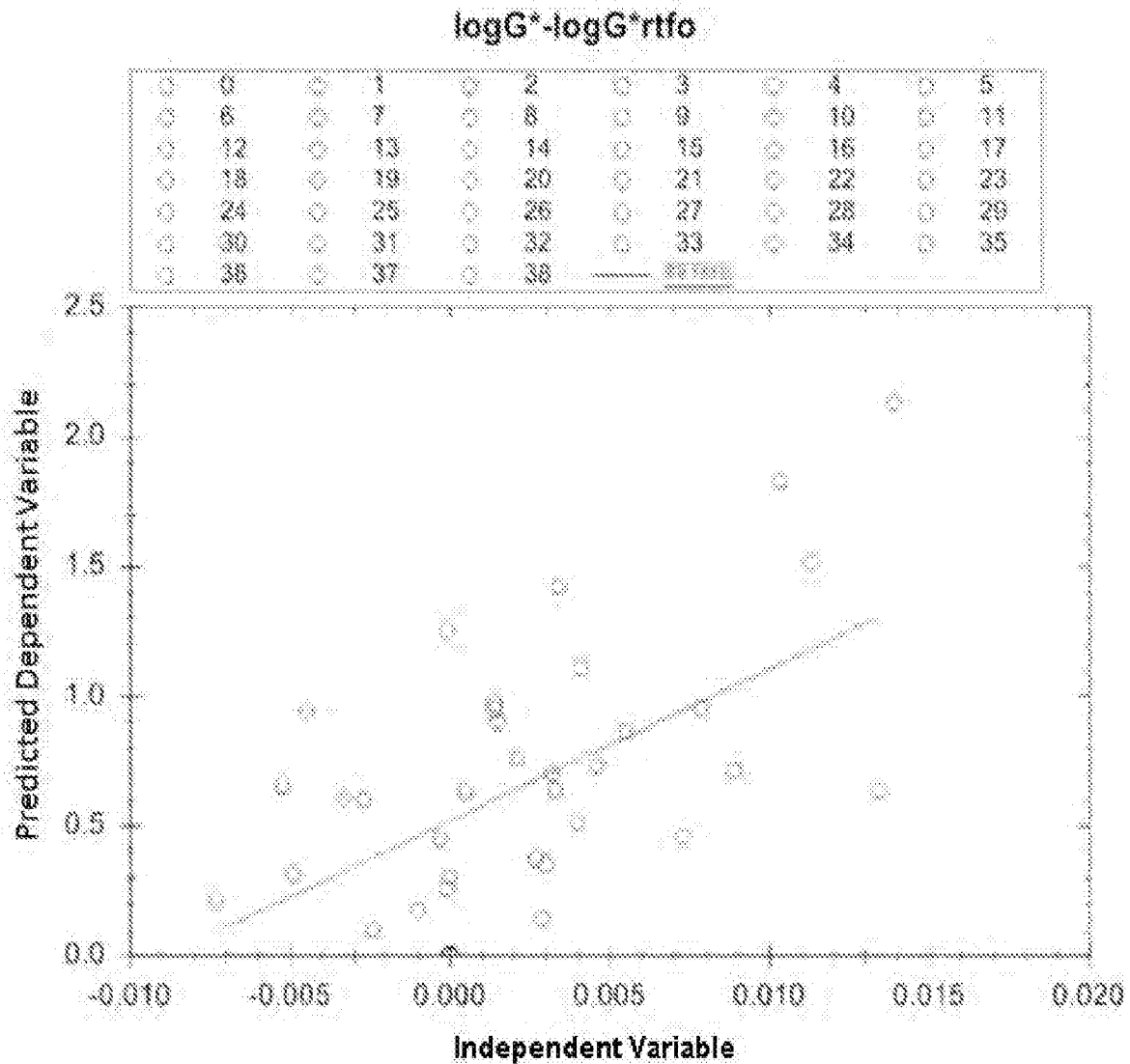


Figure 26. Screen capture. Regression plot example 2.

The discussion to this point described the capabilities of using the regression scan option to investigate single independent variable correlations of the form for each and every independent variable in the set:

$$y = ax + b \quad (6)$$

The other option, found on the Load Data tab, is multivariate linear regression, which is used to investigate additive correlations of more than one independent variable.

$$y = b + a_1x_1 + a_2x_2 + a_3x_3 + \dots + a_nx_n \quad (7)$$

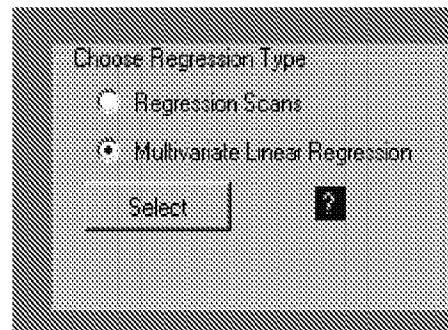


Figure 27. Screen capture. Multivariate regression selection.

When this option is selected (figure 27), the next step is to reduce the number of independent variables through some kind of selection process. The automated grouping process has been discussed in some detail earlier, but in some cases, other criteria may be more useful in choosing the list of independent variables likely to explain the system variance. A conceptual model may be available or highly correlating wave numbers from the scan study may produce better correlations when combined. In the example we have been showing, we notice several wave numbers correlate fairly well, and it might be worth trying some of those. So, in the iv reduction selection list, we would select manual (figure 28).

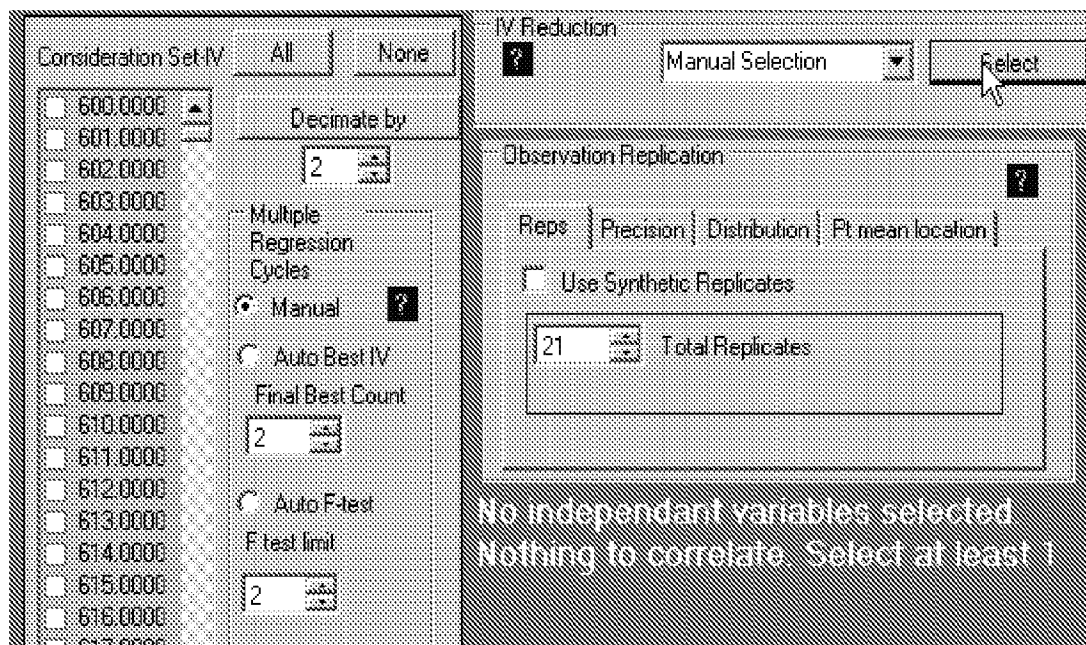


Figure 28. Screen capture. Independent variable reduction method selection.

The consideration set on the left shows every independent variable in the data set, in this case IR wave numbers spaced apart by 1 unit. Notice that the Observation replication function is disabled, and the status window informs that nothing is selected to correlate. If we select all of the wave numbers, then we discover the correlation cannot possibly be done (figure 29).

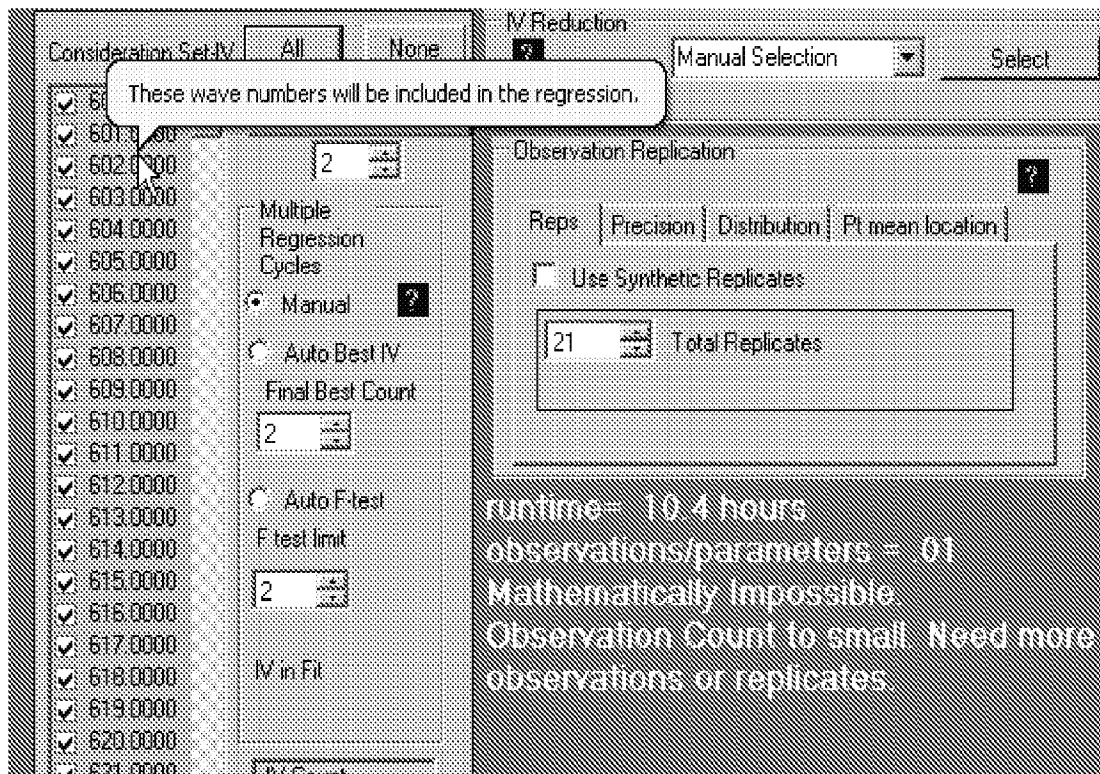


Figure 29. Screen capture. Independent variable selection example 1.

If we deselect all of the wave numbers, and then select just 1663, we get this screen (figure 30).

Load Date | Scan Options | Run MLR | MLR Summary | Edit Plots | 1

Consideration Set-IV: All | None

Decimate by: 2

Multiple Regression Cycles: ☒ Manual ☐ Auto Best IV

Final Best Count: 2

Auto F-test: ☐ F test limit: 2

IV in Fit: ☐ IV Count: 1

Auto Deselect Insignificant IV

IV Reduction: ? Manual Selection Select

Observation Replication: ?

Reps | Precision | Distribution | Pt mean location

☐ Use Synthetic Replicates

21 Total Replicates

runtime= 0.5 seconds
 observations/parameters = 19.5
 Observation count meets the accepted ratio for a relevant regression. Please proceed. If using synthetic replicates, reproducibility of fits will be improved with more replicates

RUN REGRESSION

Done-Restore List

Export to Excel and View

Figure 30. Screen capture. Independent variable selection example 2.

It is now possible to calculate a regression, the result being a simple linear one variable regression shown in figure 31.

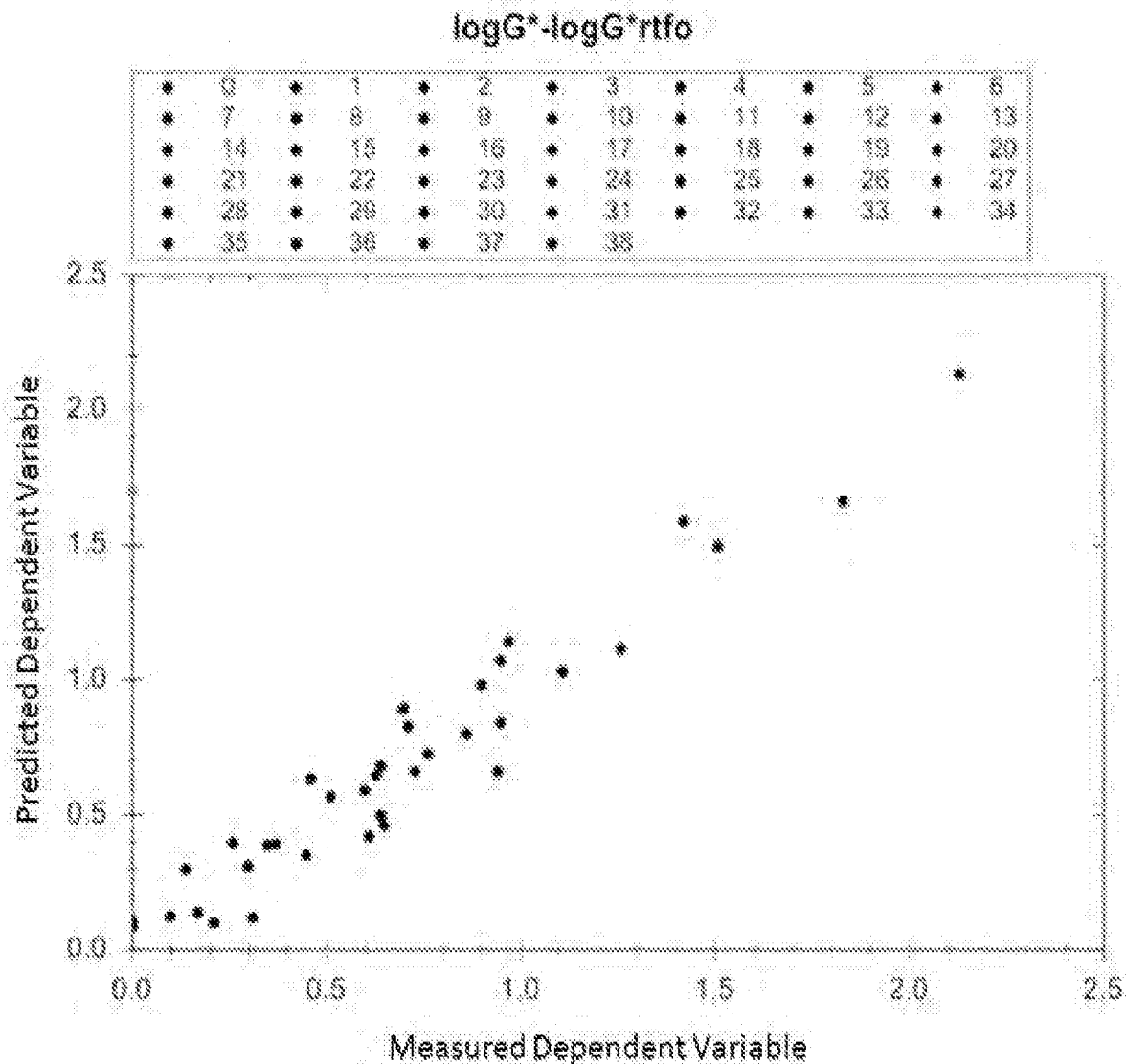


Figure 31. Screen capture. Regression results plot of predicted and measured dependent variable values.

If we select 6 independent variables, the O/P ratio now falls to 5.57 (above 7 would statistically be better), and the regression can proceed (figure 32).

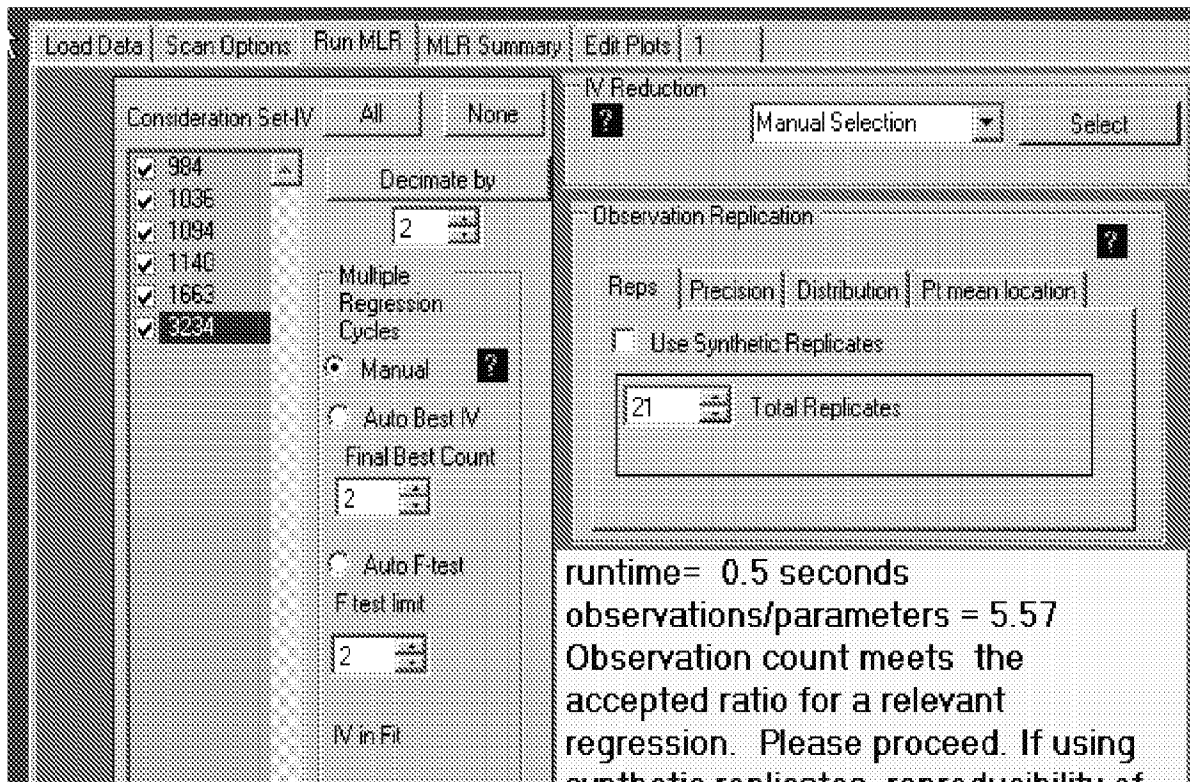


Figure 32. Screen capture. Software notification of favorable observation to parameter ratio.

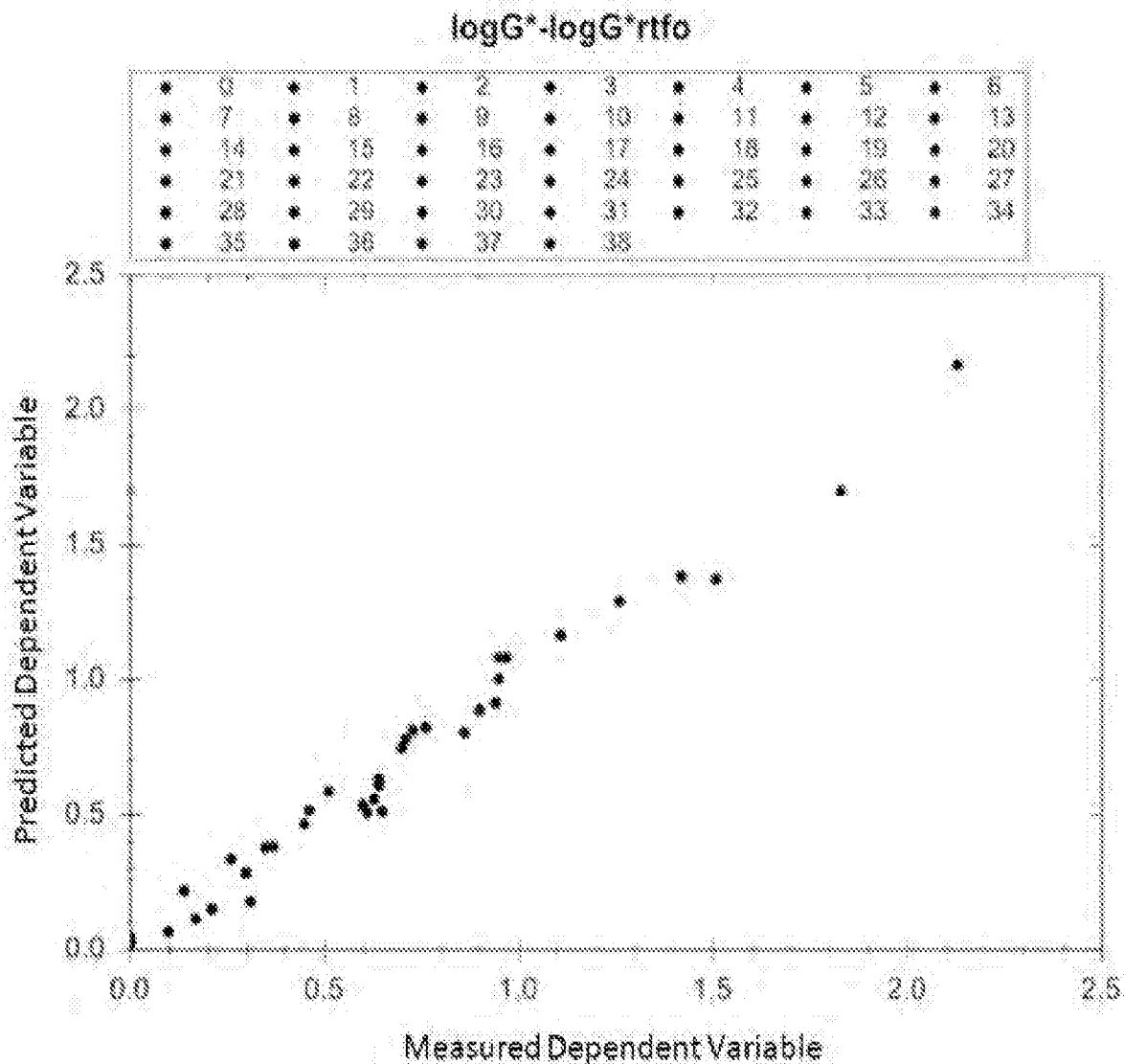


Figure 33. Screen capture. Regression results plot, example 2.

That worked quite well (figure 33), so perhaps more wave numbers would be better. When the count exceeds 38, the screen turns red, indicating the computation is impossible (figure 34).

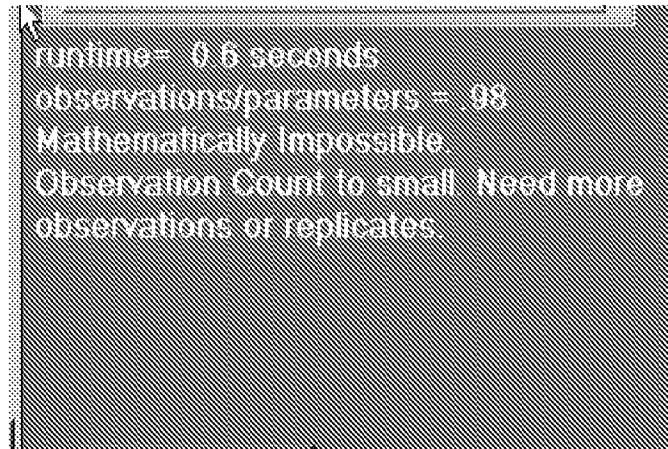


Figure 34. Screen capture. Impossible observation to parameters ratio notification.

If we deselect a couple of wave numbers, we get a yellow screen (figure 35).

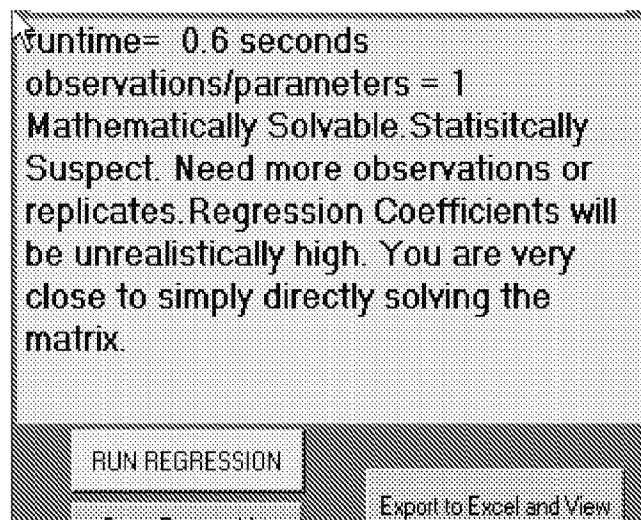


Figure 35. Screen capture. Possible observation to parameters ratio notification, but statistically suspect.

An attempt to run the correlation failed in the dynamic linked library as the matrix still could not be solved (in theory, this would have been soluble if there were no collinear variables). This situation raises an error message that is displayed to the user in figure 36. When you get this message, you must either choose fewer independent variables, group the independent variables to reduce the count, run more experiments to get more observations, or synthetically produce more observations.

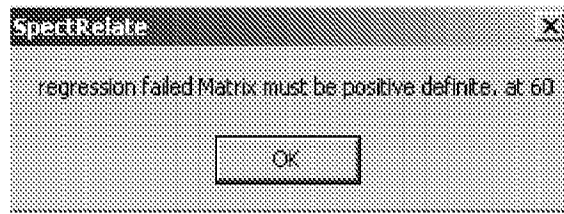


Figure 36. Screen capture. Computation failure error message.

We deselect a few more wave numbers. With 33 independent variables, the O/P ratio is still yellow at 1.15, but maybe we can solve the matrix anyway. The program runs, and the user is warned (figure 37).

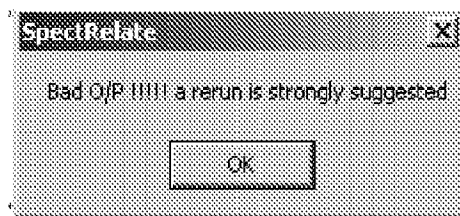


Figure 37. Screen capture. Successful computation, with suspect results warning.

The results are simply just stunning! (figure 38).

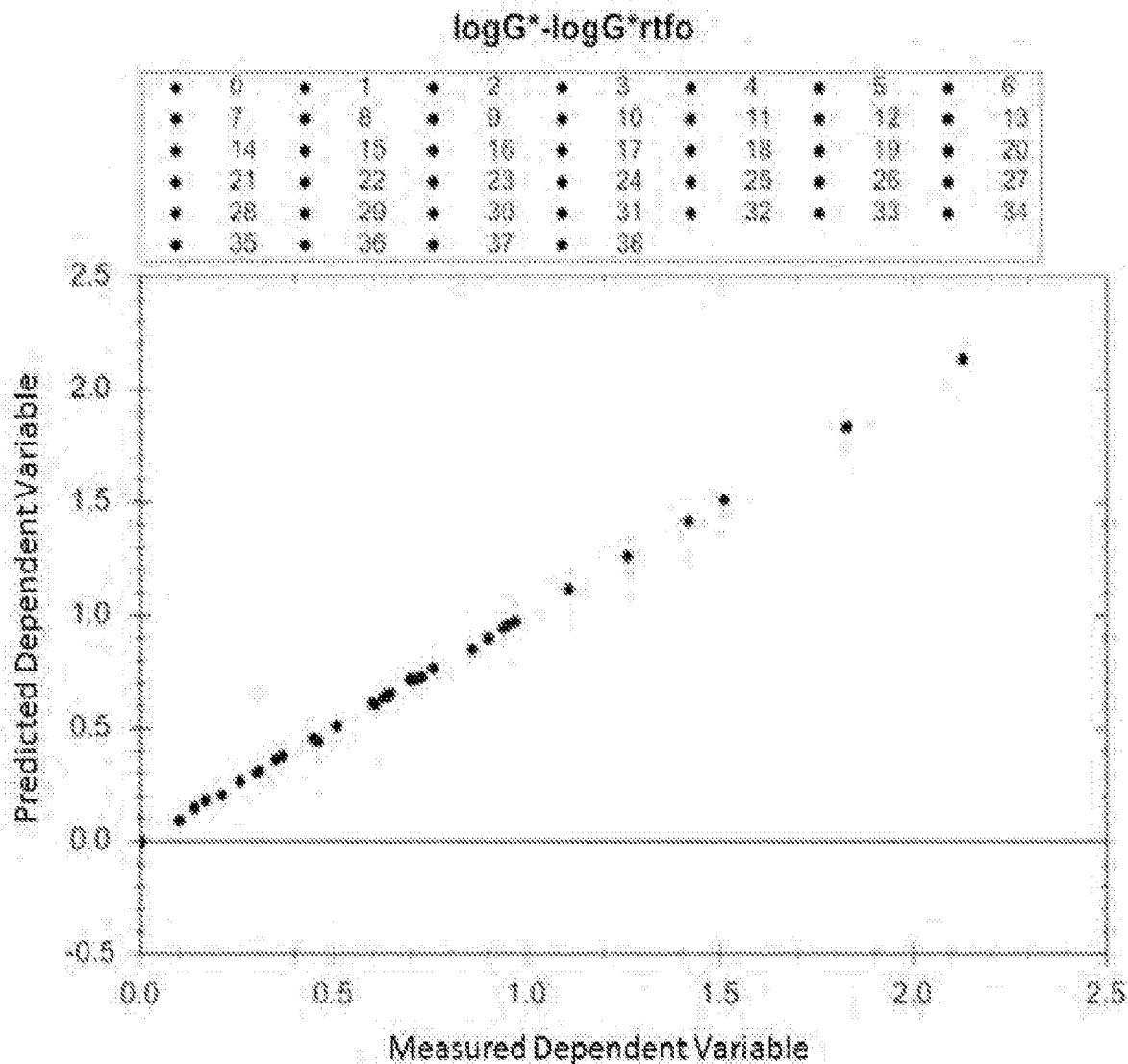


Figure 38. Screen capture. Regression results plot example 3.

The regression coefficient is .9998. With such a great correlation, why is the user warned? (This phenomenon, is often the reason PLS, CPA, and other chemometric methods produce such high correlation coefficients.) The easiest way to explain the problem here is to consider a simple linear regression with just two points in the data set. Any fit of a straight line to two points will yield a perfect correlation coefficient. In multidimensional space, the O/P ratio can be thought of as the number of points *in each dimension*. In this example, we have essentially drawn a straight line through two points in multiple dimensions. This results in a nearly perfect correlation! This also happens with data sets consisting of randomly generated data. We have two options at this point. We can examine the F test values for each variable and discard those that are insignificant. This is most easily accomplished using the auto deselect button. We now have 20 variables and 21 parameters in our model. The O/P ratio is now 1.86, a bit better than the 1.15 of the previous, but still far from being statistically believable. We still get a near perfect regression. The F test limit can be adjusted from the default value of 2. Let's try an F test limit of 3. The variable count

drops to 18, O/P still only 2. An F test limit of 10 gives us an O/P of only 2.44. At an F test limit of 20, we have an O/P of 6.5, with only the 5 best fitting variables. The results are summarized below in figure 39.

Observations List MLR results Best Search plot				
Rsqr Measured, 0.9668 Rsqr Replicated, 0.9668 Rep Rsqr/ M Rsqr, 1.0000 AIC(rep)small best -263.66641270326				
regression data				
IV x	Coefficient	group no.	F	GOOD
x(0)	1.496E-001			
x5	2.350E001	1140.0000	00026.553	YES
x6	-9.167E000	1212.0000	00005.193	NO
x17	-1.625E001	2655.0000	00004.141	NO
x18	-7.910E-001	3014.0000	00000.063	NO
x24	-8.067E000	3441.0000	00007.968	NO
Coefficient of determination (rsqr ACTUAL MATRIX) =			0.96679	
Overall regression is SIGNIFICANT				

Figure 39. Screen capture. Regression results summary tab example 1.

The other option would be to use synthetic replication. Up to this point we have only used classic multivariable correlation methods and they remain at the heart of our method. Out of over 3000 available measurements in the spectra, classical methods limit us to examining just a few at a time. In this example, we can only examine additive relationships using about 1% of the data in order to get a calculation to be possible. The added constraint of statistical validity lowers that number even further to a fraction of a percent, so, the chance of missing an important combination is quite high.

Returning to the original variable list, We enable the synthetic replication and adjust the total replicates to 7, resulting in an O/P of 7.18. The correlation coefficient is now 0.9749, slightly higher than the 5 variable results we obtained before. But now we have 37 variables, most of them insignificant (figures 40 and 41).

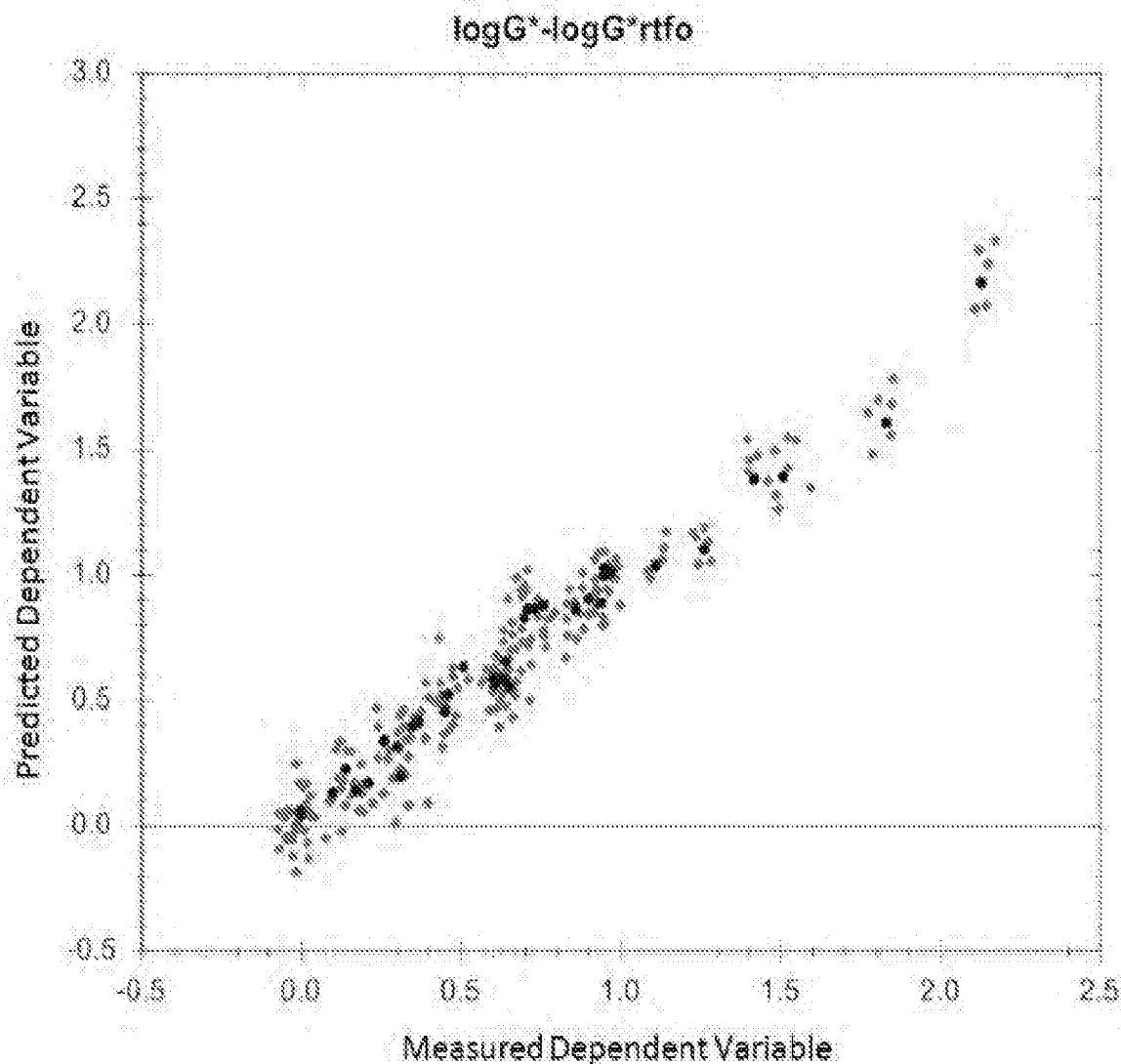


Figure 40. Screen capture. Regression results plot example 4.

Observations List	MLR results	Best Search	plot	
Rsqr Measured, 0.9749 Rsqr Replicated, 0.9505 Rep Rsq/ M Rsq, 0.9750 AIC(rep)small best -2275.78771472376				
regression data				
IV x	Coefficient	group no.	F	GOOD
x(0)	4.465E-002			
x1	-7.265E-002	0984.0000	00000.078	NO
x2	1.696E000	1036.0000	00130.507	YES
x3	-1.178E-001	1094.0000	00000.248	NO
x4	5.895E-001	1135.0000	00006.364	NO
x5	9.084E-001	1140.0000	00014.804	NO
x6	4.913E-001	1212.0000	00004.229	NO
x7	1.084E-001	1298.0000	00000.213	NO
x8	-2.222E-001	1353.0000	00000.889	NO
x9	8.022E-001	1663.0000	00011.338	NO
x10	1.829E000	1701.0000	00118.286	YES
x11	0.400E-001	1825.0000	00010.220	NO

Figure 41. Screen capture. Regression results summary tab example 2.

We can now use the Auto Best IV feature to reduce the model to any desired number of variables, and for comparison with the previous results, we will go down to 5 (figure 42).

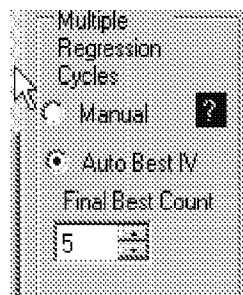


Figure 42. Screen capture. The Auto best independent variable list run option.

The results are shown figure 43.

Observations List MLP results Best Search plot				
Rsqr Measured, 0.9659 Rsqr Replicated, 0.9263 Rep Rsqr/ M Rsq, 0.9590 AIC(rep)small best -2231.3393019866				
regression data				
IV x	Coefficient	group no.	F	GOOD
x(0)	3.415E-002			
x2	2.008E000	1036.0000	00358.633	YES
x4	9.264E-001	1135.0000	00012.669	YES
x6	1.007E000	1212.0000	00015.293	YES
x9	7.707E-001	1663.0000	00009.634	YES
x10	1.763E000	1701.0000	00145.671	YES
Coefficient of determination (rsqr ACTUAL MATRIX) =			0.92631	
Overall regression is SIGNIFICANT				

Figure 43. Screen capture. Regression results summary tab example 3.

The regression correlation coefficient we obtained is nearly identical using classic multivariable regression, and there is some agreement concerning the wave numbers of importance. Some of the wave numbers appearing in the two results are likely correlated, and describe the same measurement. We have not used grouping before the correlation, and that minimizes, limiting complexity. In theory, we could apply synthetic replication to this data set without grouping first. With some data sets, that contain few independent variables, this makes sense, but with this set, the size of the matrix that must be solved would be excessive, and very long execution times would result. In addition to reducing the size of the matrix for solution, grouping also eliminates the problem of several areas of the spectra representing the same phenomena.

To reduce the independent variable list, we chose one of the grouping methods in the IV reduction selection list (figure 44).

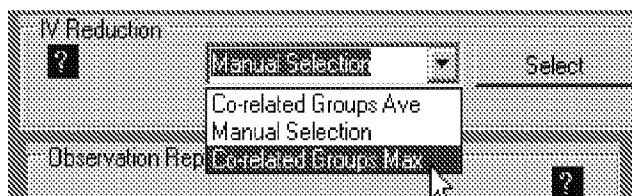


Figure 44. Screen capture. Independent variable grouping options.

We have examined the use of the Manual selection option for classic multivariable regression solutions, and the possibility to use synthetic replication to advantage without grouping first. The two grouping options work the same, as described earlier with the only exception of how the group measurements are reduced to a single number. The two methods either average the values of the wave numbers in that group, or find the largest average measurement (largest wave number over the entire range of data measured in that grouping. This can vary in individual spectrum. The idea here is to obtain the best signal to noise ratio) and use the value for that single wave number. For our studies using IR independent variable, we have found slightly better correlations using the highest peak method, but one could imagine situation where an average may better suit the physics and/or chemistry behind the relationships sought.

When a method is selected, the user is prompted for a regression threshold. For reasons discussed previously (see figure 10), this data set should probably use a value no larger than 0.9 (figure 45). The user may use higher values for more groups and selectivity.

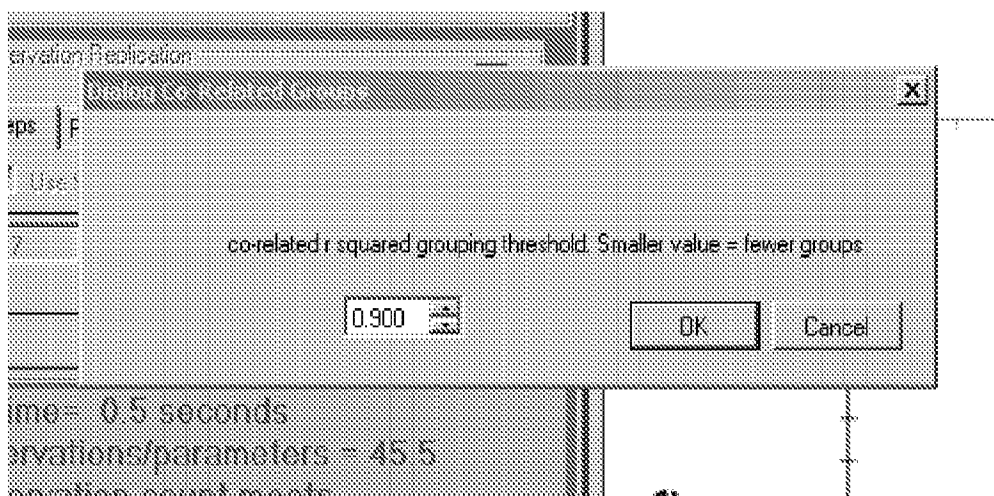


Figure 45. Screen capture. Independent variable grouping threshold setting.

When the prompt is dismissed by clicking "OK" the variables are grouped. The list of groups appears in the consideration set, with 36 groups identified. The ratio is only 1.05, and this might actually run, but would be statistically suspect, so enabling synthetic replicate at a value of 7 sets us up for a meaningful regression. Using all of the groups, we get a correlation coefficient of 0.9675. Deselecting the F test values less than 2 we reduce the number to 15 independent

variables. Running the regression on this smaller model actually produces a slightly better fit with a correlation coefficient of .9691. Eliminating F-tests less than 2, we have 8 variables to fit, resulting in a correlation coefficient of 0.9688, essentially as good as fit as 15. With F test elimination still set at 2, we reduce the number of variables in the model to 6 with essentially as good a fit. Continuing the process of discarding the variables one at a time based upon the lowest F-test, we get these results for 5 variables in figure 46. The “Rsqr Measured” and reported by the software, is the value obtained without replication. The “Rsqr Replicated” includes the replicated data. Next is the ratio of the two. AIC is a statistical measure of model size we found to be either useless or miscalculated. “rsqr ACTUAL MATRIX” is identical to “Rsqr Measured” with the exception of being computed by the regression programming object while the “Rsqr Measured” is calculated by code written by the software authors.

Observations List MLR results Best Search plot				
Rsqr Measured, 0.9662 Rsqr Replicated, 0.9250 Rep Rsq/ M Rsq, 0.9574 AIC(rep)small best -2226.50759633759				
regression data				
IV x	Coefficient	group no.	F	GOOD
x(0)	4.44E-002			
x8	8.795E-001	1108.0000	00010.177	YES
x14	9.003E-001	1025.0000	00017.649	YES
x15	1.526E000	1035.0000	00065.633	YES
x16	2.371E000	1701.0000	00384.772	YES
x25	-3.993E-001	2971.0000	00008.307	YES
Coefficient of determination (rsqr ACTUAL MATRIX) =			0.92499	
Overall regression is SIGNIFICANT				

Figure 46. Screen capture. Regression results summary tab example 4.

And for 4 variables there is little reduction in the quality of the fit (figure 47).

Observations List MLR results Best Search plot				
Rsqr Measured, 0.9611 Rsqr Replicated, 0.9253 Rep Rsq/ M Rsq, 0.9627 AIC(rep)small best -2229.59412706994				
regression data				
IV x	Coefficient	group no.	F	GOOD
x(0)	2.181E-002			
x8	7.124E-001	1108.0000	00008.041	YES
x14	1.103E000	1025.0000	00023.640	YES
x15	1.461E000	1035.0000	00063.961	YES
x16	2.369E000	1701.0000	00496.823	YES
Coefficient of determination (rsqr ACTUAL MATRIX) =			0.92529	
Overall regression is SIGNIFICANT				

Figure 47. Screen capture. Regression results summary tab example 5.

The 3 most significant wave numbers are (figure 48):

Observations List	MLR results	Best Search	plot
Rsqr Measured, 0.9610 Rsqr Replicated, 0.9028 Rep Rsq/ M Rsq, 0.9395 AIC(rep)small best -2159.85852818992			
regression data			
IV x	Coefficient	group no.	F
x(0)	2.224E-002		
x8	5.326E-001	1108.0000	00003.142
x15	2.352E000	1035.0000	00480.833
x16	2.370E000	1701.0000	00371.273
Coefficient of determination (rsqr ACTUAL MATRIX) =			0.90284
Overall regression is SIGNIFICANT			

Figure 48. Screen capture. Regression results summary tab example 6.

The best 2 variables are (figure 49):

Observations List	MLR results	Best Search	plot
Rsqr Measured, 0.9610 Rsqr Replicated, 0.9103 Rep Rsq/ M Rsq, 0.9472 AIC(rep)small best -2183.57292138146			
regression data			
IV x	Coefficient	group no.	F
x(0)	3.284E-002		
x15	2.298E000	1035.0000	00544.379
x16	2.548E000	1701.0000	00799.743
Coefficient of determination (rsqr ACTUAL MATRIX) =			0.91027
Overall regression is SIGNIFICANT			

Figure 49. Screen capture. Regression results summary tab example 7.

These two wave numbers are generally accepted to represent carbonyl and sulfoxide responses in asphalt binder oxidation. The fit coefficients are nearly the same for these two wave numbers. A simple sum of these absorbencies should work well as oxidation extent of reaction indicator.

Results and Output

When the user is satisfied with the regression results, the software provides a number of options to export the data for use in spreadsheets and publications.

A right mouse click in many areas of the program interface produces a context menu for cut and paste operations of highlighted text (in blue). The following screen capture illustrates a copy operation from the "MLR results" tab (figure 50).

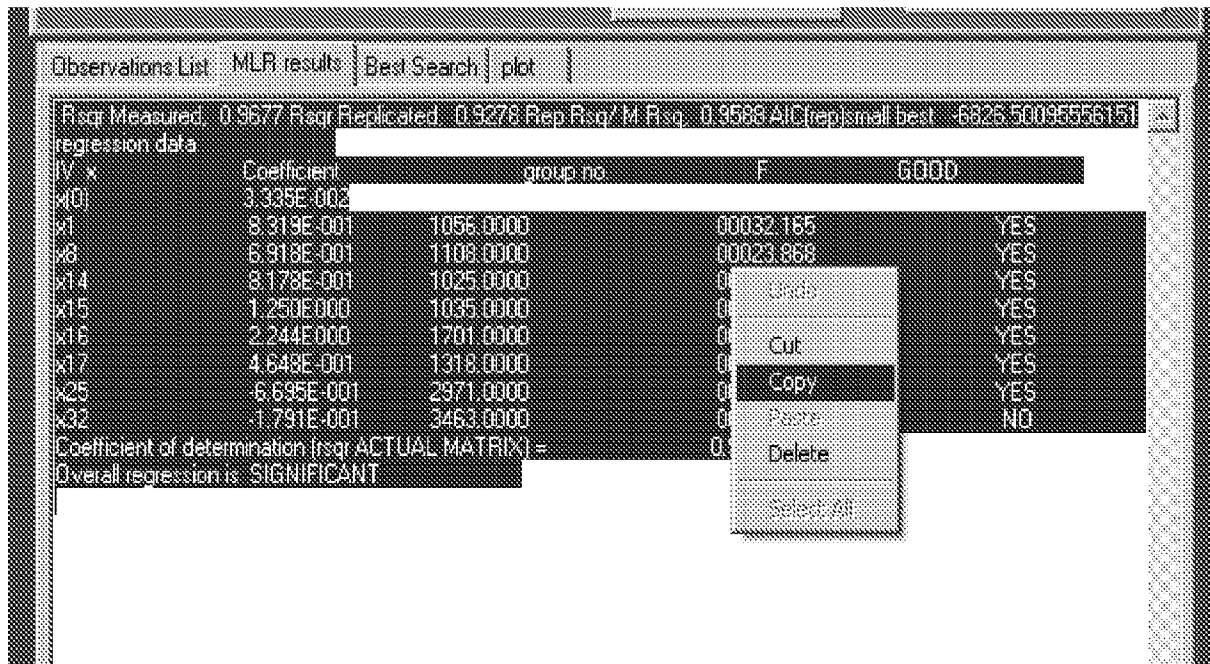


Figure 50. Screen capture. Regression results summary tab copy and paste example.

For a more detailed document for archiving purposes, the MLR Summary tab in the upper left corner fully documents the run with data, files used, and an ANOVA table. This information (must be highlighted) can be copied through the right click context menu (figure 51).

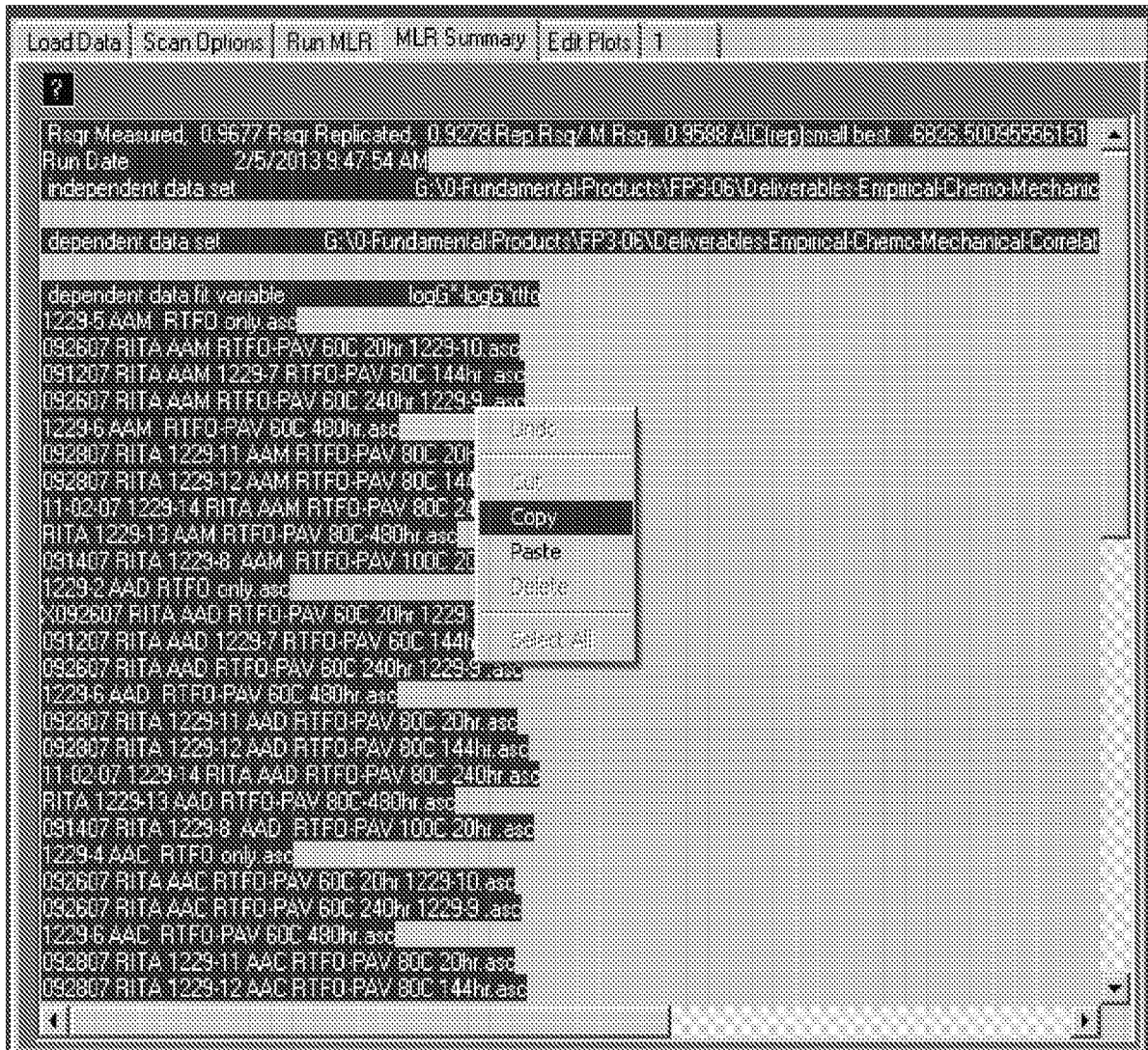


Figure 51. Screen capture. Complete run summary results copy and paste example.

The comparison plot of the regression predicted dependent variable and the actual measured values can be exported as an image, or the data can be exported to a spreadsheet for re-plotting with other software (figure 52).

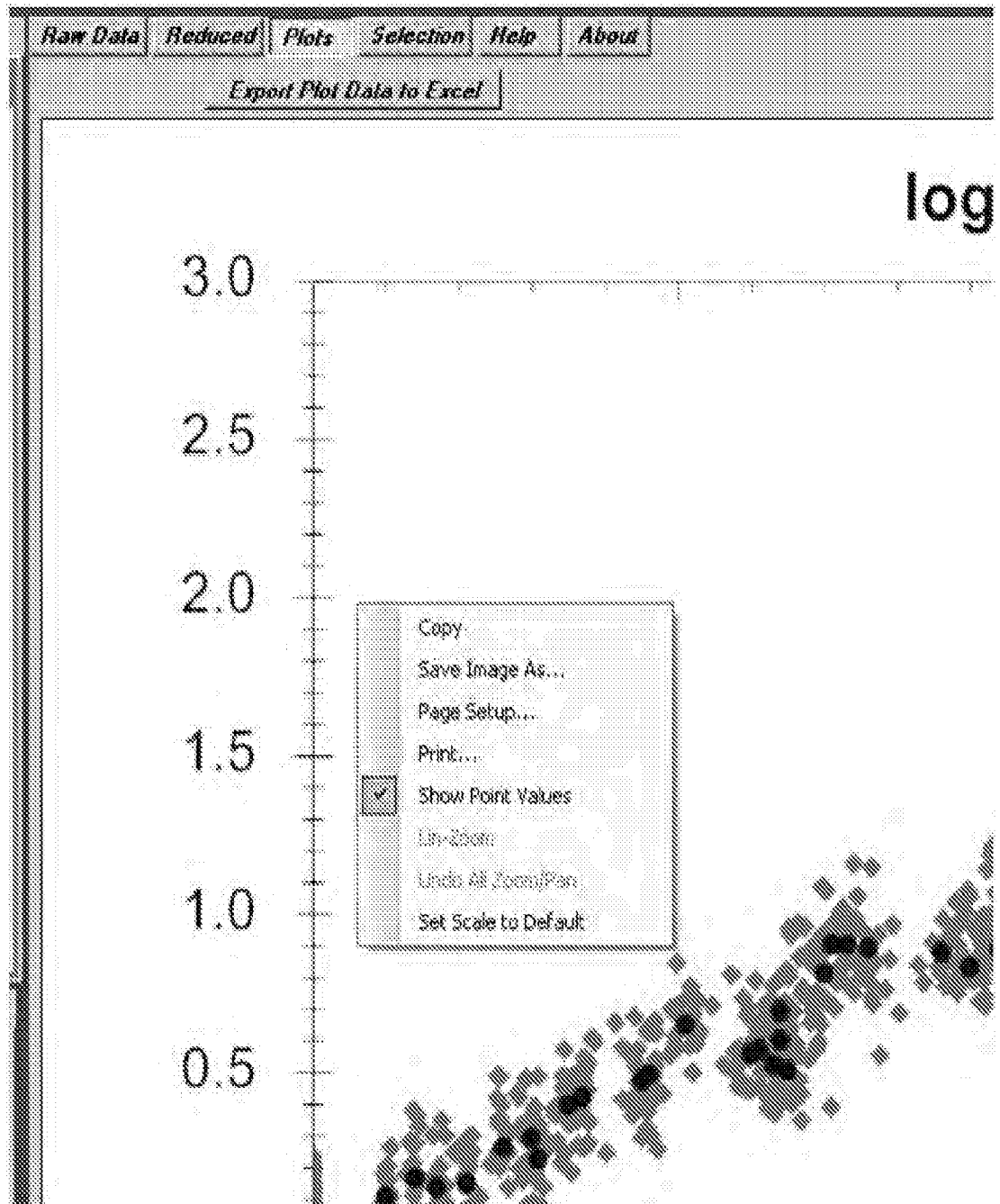


Figure 52. Screen capture. Right click context menu for plots.

The Edit Plots tab is where the plot appearance can be edited to suit the user's needs (figure 53).

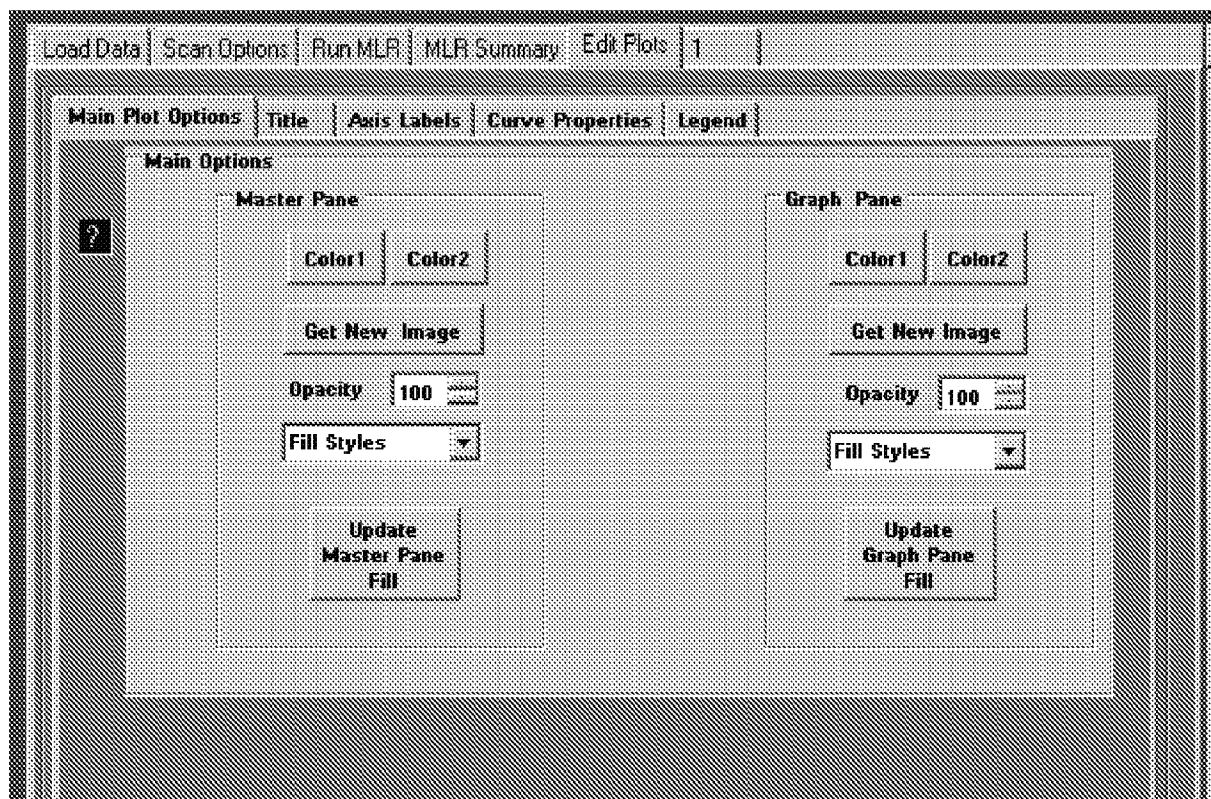


Figure 53. Screen capture. Plot attribute editing area.

RESULTS AND DISCUSSION

The use of time honored multivariable linear regression with provisions to deal with excessive independent variables based upon measurement method precision has been developed. This has proven to be extremely useful in our inquiries into asphalt binder oxidation chemistry. A software package, called Spectrelate has been developed that guides the user with regard to avoiding common errors often encountered by users of multivariable methods without the usual training normally required to understand statistical validity.

CONCLUSIONS

Multivariable empirical approaches to data analysis are a valuable tool for discovering significant variables involved in chemo-physical processes. This software package has proven quite useful in the study of rheological changes due to asphalt binder oxidation, and the details of these significant advances can be found in other fundamental properties product technical white papers. The product has many other potential applications in other areas of scientific investigation and provides a number of advantages to other often employed methods. These advantages include the production of an easily understood relationship in terms of measured variables using a time honored method with a rich history of statistical validity study.

RECOMMENDATIONS

The methods employed here could be expanded to non-linear regression problems.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the Federal Highway Administration, U.S. Department of Transportation, for financial support of this project under contract no. DTFH61-07D-00005.

DISCLAIMER

This document is disseminated under the sponsorship of the Department of Transportation in the interest of information exchange. The United States Government assumes no liability for its contents or use thereof.

The contents of this report reflect the views of Western Research Institute which is responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views of the policy of the Department of Transportation.

REFERENCES

- Barros, A. S., and D. N. Rutledge, 2004, Principle Components transform-partial least squares: a novel method to accelerate cross-validation in PLS regression. *Chemometrics and Intelligent Laboratory Systems*, 73 (2): 245-255.
- Basu, B., D. Saxena, V. Kavl, M. I. S. Sastry, and R. T. Mookken, 1998, Prediction of Oxidation Stability of Inhibited Base Oils from Chemical Composition using an Artificial Neural Network (ANN). *Lubrication Science*, 10: 121-134.
- Chalmers, J. M., and N. J. Everall, 1996, FTIR, FT-Raman and chemometrics: applications to the analysis and characterization of polymers. *Trends in analytical chemistry*, 15 (1): 18-24.
- Garson, G. David (n.d.). "Popular algorithms," from Statnotes: Topics in Multivariate Analysis. Retrieved 11/21/2007 from <http://www2.chass.ncsu.edu/garson/PA765/statnote.htm> .
- Hasegawa, T., 2006, Spectral Simulation Study on the Influence of the Principle Component Analysis Step on Principle Component Regression. *Applied Spectroscopy*, 60 (1): 95-98.
- Hasegawa, T., 1999, Detection of Minute Chemical Species by Principal-Component Analysis. *Anal. Chem.*, 71 (15): 3085-3091.
- Karstang, T. V., A. A. Christy, B. Dahl, and O. M. Kvaheim, 1991, Diffuse reflectance Fourier-transformed infrared spectroscopy in petroleum exploration: a multivariate approach to maturity determination. *Journal of Geochemical Exploration*, 41 (1-2): 213-226.

- Lachenmeier, D. W., 2007, Rapid Quality control of spirit drinks and beer using multivariate data analysis of Fourier transform infrared spectra. *Food Chemistry*, 100 (2): 825-832.
- Mark, H., and J. Workman Jr., 2007, Chemometrics in Spectroscopy What Can NIR Predict? *Spectroscopy*, 22(6): 20-26, www.spectroscopyonline.com, Nov. 21, 2007.
- Pearson, K., 1901, On lines and planes of closest fit to systems of points in space. *Phil. Mag. Ser. 6*, 2: 559-72.
- Sastry, M. I. S., A. Chopra, A. S. Sarpal, S. K. Jain, S. P. Srivastava, and A. K. Bhatnagar, 1998, Determination of Physiochemical Properties and Carbon-Type Analysis of Base Oils Using Mid-IR Spectroscopy and Partial Least-Squares Regression Analysis. *Energy & Fuels*, 12 (2): 304-311.
- Satya, S., 2005, *Chemometrics: A Tool to Predict Crude Oil Properties*, PhD. Dissertation, Dept. of Chemical Engineering, University of Utah.
- Satya, S., R. M. Roehner, M. D. Deo, and F. V. Hanson, 2007, Estimation of Properties of Crude Oil Residual Fractions Using Chemometrics. *Energy & Fuels*, 21 (2): 998-1005.
- Sharaf, M. A., D. L. Illman, and B. R. Kowalski, 1986, *Chemometrics*, Volume 82 in Chemical Analysis, Elving, P. J., and J. D. Winefordner, eds., John Wiley & Sons, New York.
- Wold, S., 1991, Chemometrics, why, what and where to next. *Journal of Pharmaceutical & Biomedical Analysis*, 9 (8): 589-596.
- Wold, S., H. Trygg, A. Berglund, and H. Antti, 2001, Some Recent Developments in PLS Modeling. *Chemometrics and Intelligent Laboratory Systems*, 58 (2): 131-150.
- Zagonel, G. F., P. Peralta-Zamora, and L. P. Ramos, 2004, Multivariate monitoring of soybean oil ethanolysis by FTIR. *Talanta*, 63 (4): 1021-1025.

Filed via EFS

IN THE UNITED STATES PATENT AND
TRADEMARK OFFICE

Application Number:

Filed:

Applicants:

Title:

Assignee:

Attorney Docket:

Customer No.:

Confirmation No.:

Ronald R. Glaser et al.

Method for Correlating Physical and Chemical Measurement Data
Sets to Predict Physical and Chemical PropertiesThe University of Wyoming Research Corporation d/b/a Western
Research Institute

WRI-CorrSoft-PCT

33549

APPENDIX B

Impact of the bitumen quality on the asphalt mixes performances

Frédéric Delfosse^{1, a}, Ivan Drouadaine^{1, b}, Stéphane Faucon-Dumont¹, Sabine Largeaud^{1, c},
Bernard Eckmann², Jean Pascal Planche^{3, d}, Fred Turner³, Ron Glaser³

¹ Eurovia, Mérignac, France

² Eurovia, Rueil Malmaison, France

³ Western Research Institute, Wyoming, United States

^a frederic.delfosse@eurovia.com

^b ivan.drouadaine@eurovia.com

^c sabine.largeaud@eurovia.com

^d jplanche@uwyo.edu

Digital Object Identifier (DOI): dx.doi.org/10.14311/EE.2016.049

ABSTRACT

European refining, French in particular, is currently going through a phase of rationalization and search for maximum flexibility in crude supplies. For users of bitumen, this creates concerns about the quality and consistency of products delivered, especially as the European standard EN 12591 appears to them as insufficient to ensure satisfactory performance of the finished products, particularly in the case of specialty products such as high modulus asphalt, polymer modified bitumen, and bitumen emulsions.

In this context, the search for correlations between bitumen properties and the performance of the finished product is more relevant than ever. The study presented here is focused on asphalt made with pure bitumen. It was based on a standard design, but with two different types of aggregates. After a preliminary selection, 8 bitumen (20/30, 35/50 and 50/70 pen. grades) were selected. The characterization of asphalt mixes covers all the usual characteristics (stiffness modulus, resistance to rutting and fatigue, resistance to thermal cracking, water sensitivity). The characterization of binders, besides conventional testing, includes the rheological properties (DSR, MSCR, and BBR tests) and the compositional analysis, particularly infra-red spectroscopy and SARA analysis. These tests were performed on the original binders, after RTFO, after RFFO + PAV as well as on the binders recovered from asphalt.

This project was conducted as a collaboration between Eurovia and the Western Research Institute (WRI) which performed the compositional analysis of binders, including the SAR-AD™ (WRI improved SARA separation technique) and the chemometrics analysis using their software ExpliFit™.

Keywords: Ageing, Asphalt, Chemical properties, Low-Temperature, Rheology

1. INTRODUCTION

In recent years, a significant evolution on the European bitumen market has been observed. French and European refining is currently in a phase of rationalization and search for maximum flexibility in crude supplies.

Road contractors such as Eurovia are worried about bitumen quality and have observed new problems on field.

The current European standard EN 12591 appears insufficient to ensure satisfactory performance of the finished products, particularly in case of specialty products such as high modulus asphalt (modulus, fatigue), polymer modified bitumen, and bitumen emulsions (settling tendency, viscosity).

The paving industry is therefore more and more confronted with the same problem: how to evaluate the quality of a given bitumen in relation to its intended use.

The search and validation of performance-related bituminous binder properties continues to be a key issue for the paving industry in Europe, as well as in the rest of the world. With the Superpave system implementation in the US, important progress has been achieved and is still on-going. In Europe, the development of 2nd generation product standards appears to be more necessary than ever.

In this context, Eurovia and the Western Research Institute (WRI) in Laramie, Wyoming (USA) launched a research program in 2013 to search for correlations between bitumen properties and the performance of the finished product.

2. RESEARCH PROGRAM

The research program was based on a carefully selected experimental matrix.

Eight bitumen samples (all unmodified) were selected: B1 to B8. With these binders, 12 asphalts were manufactured (8 with diorite and 4 with limestone aggregates). Each asphalt had a 4.9% bitumen content.

Table 1 presents the main characteristics of these bitumen samples and the different asphalt designs.

Table 1: Bitumen characteristics and asphalt designs

	B1	B2	B3	B4	B5	B6	B7	B8
Penetration (1/10 mm) : NF EN 1426	40	37	40	22	26	28	55	57
Ring and Ball Temperature (°C) : NF EN 1427	53.4	53	52	59	57.2	61	49	49.2
Superpave performance grading (PG)	70-22	70-16	70-16	76-16	76-16	76-10	64-22	64-16
HMA with diorite	X	X	X	X	X	X	X	X
HMA with limestone	X		X	X			X	

The analysis program for the bitumens (neat, after RTFO, recovered, after RTFO + PAV) included:

- 1- Chemical analysis: infrared, SAR-AD: Saturates, Aromatics, Resins and Asphaltene Determinator [1], SEC: Size exclusion chromatography, DSC: Differential Scanning Calorimetry (glass transition, wax content)
- 2- Superpave rheological tests: Bending Beam Rheometer (BBR), DSR: master curves, crossover, R parameter [2]...
- 3- Advanced rheological tests (LAS tests [3]...)
- 4- Asphalt Binder Cracking Device (ABCD) test [4],
- 5- Conventional European tests: penetration, ring and ball temperature, Fraass breaking point [5],...

Table 2 presents the different tests performed to analyze asphalts.

Table 2: Asphalt tests performed

Test	Standard
Gyratory shear compactor	NF EN 12697-31
Water sensitivity	NF EN 12697-12 A
	AASHTO T283
Wheel tracking test	NF EN 12697-22
Stiffness modulus	NF EN 12697-26 E
	NF EN 12697-26 C
	LC 26/700
Fatigue	NF EN12697-24
Thermal Stress Restrained Specimen (TSRST)	AASHTO TP 10
Void content/sample	Gamma bench

From the overall research program launched in 2013, this article will present only some chemometric results [6], and the correlation between bitumen properties at low temperature (BBR, ABCD test, glass transition, Fraass breaking point) and asphalt properties at low temperature (Thermal Stress Restrained Specimen Test) with diorite aggregates. Other articles will be published in the future to present more results in detail.

3. BITUMEN SELECTION

The key starting point in a chemometric correlation is based on the quality of bitumen selection. The first step of the program before launching the analyses was to verify that the chemical composition and rheological properties of the selected bitumen samples were significantly different.

3.1 Chemical composition

The SAR-AD [1] analysis is a novel approach, developed by the Western Research Institute. The main principle of this approach is an on-column precipitation followed by a sequence of re-dissolutions, using selected solvents and columns at the various stages of the separation. In practice, it combines the Automated Asphaltene Determinator (AD) separation with an automated SAR (saturates, aromatics and resins) separation to provide a fully integrated rapid automated SARA (saturated, aromatics, resins and asphaltenes) separation using milligram sample quantities. The combined SAR-AD separation utilizes high performance liquid chromatography (HPLC) equipment with multiple columns and solvents switching valves to conduct the highly complex automated separation. The solvents are selected on polarity and include n-heptane, cyclohexane, toluene, dichloromethane / methanol blend.

Figure 1 presents the chemical compositions of the 8 bitumens.

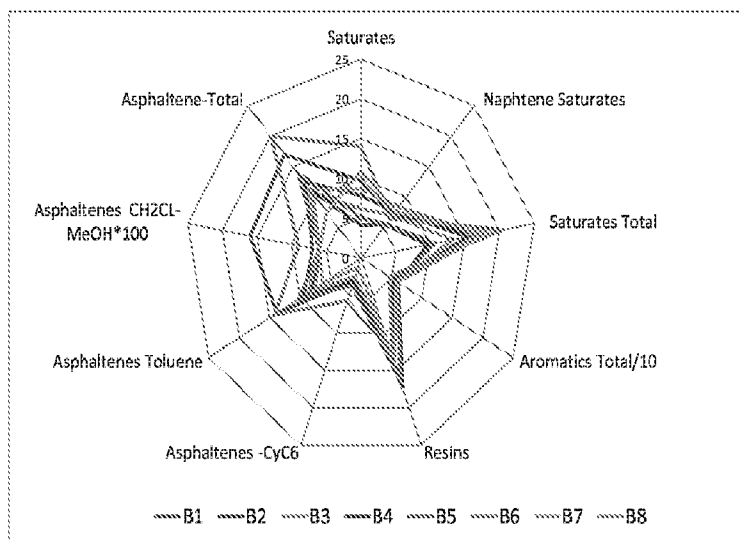


Figure 1: Bitumen composition for B1 to B8

3.2 Rheological properties

Figures 2a and 2b present the isotherms at 15°C from 1 to 30Hz.

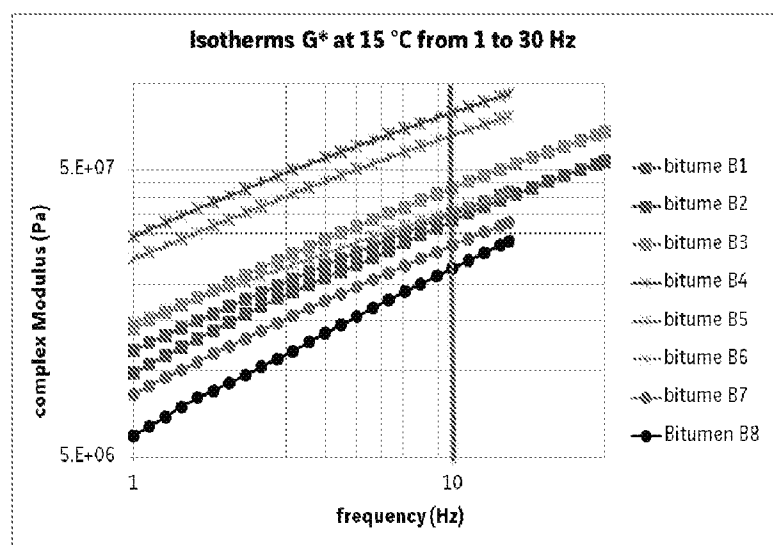


Figure 2a : Isotherms G^* at 15°C from 1 to 30 Hz

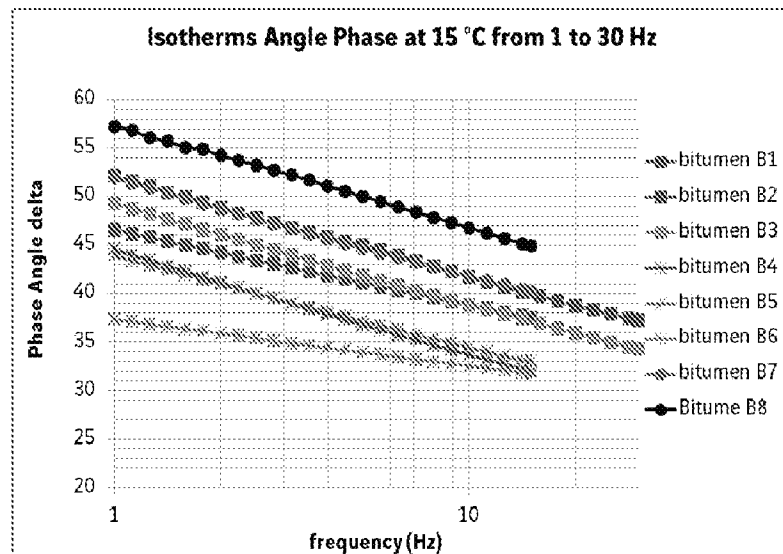


Figure 2b : Phase angle at 15°C from 1 to 30 Hz

3.3 Comments

These analyses enabled the validation of the bitumen choice, showing significant differences both in chemical composition and rheological properties. For example, the asphaltene content varies from 7 to 20 % and the saturate content from 9 to 21 %, whereas the bitumen rheological properties feature differences both in terms of stiffness level and thermal susceptibility. Thus, it is worth mentioning the atypical rheological behavior of sample B6, and of B1 to a lesser degree.

4. CHEMOMETRIC CORRELATIONS

4.1 Chemometric Software

A new software [5] has been used to investigate relationships between independent and depend variables using standard multivariable linear regression algorithms. This software was developed at Western Research Institute.

The dependent variables data are all analyses performed on bitumen or asphalt.

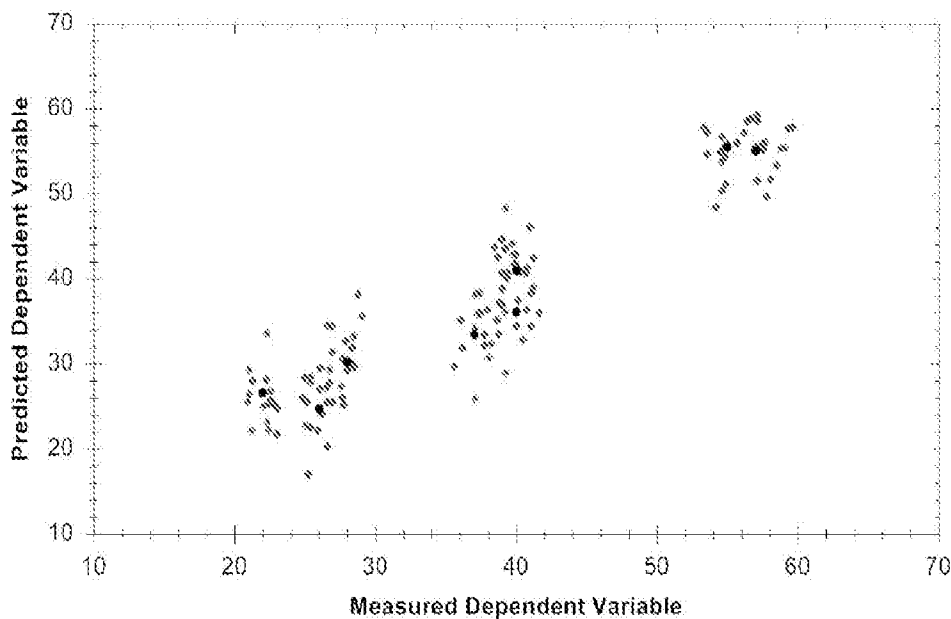
The independent variables data are typically measured to predict the dependent variables data. In this research program, the independent data include infrared (IR) spectra measurements, SAR-AD compositions, and distribution of the particle sizes by SEC.

Example: If we try to correlate bitumen IR measurements with bitumen penetration, in a first step, the software will find out which of the wavenumbers are significant when combined additively with other significant wavenumbers. This step will enable a reduction in the number of relevant wavenumbers. In a second step, the software will propose a correlation equation such as:

$$\text{Bitumen penetration (1/10 mm)} = c_0 + c_1 [\text{Abs}_1] + c_2 [\text{Abs}_2] + \dots + c_n [\text{Abs}_n]$$

c_1 = Fit coefficient 1, Abs_1 = Absorbance to the wavenumber 1 ...

In addition to the actual measurements, a precision file for each independent and dependent variables data set must also be prepared. The program requires these files to create the data clouds needed for computation. From these values, the software represents automatically the correlation between predicted values and measured values (figure 3): black dots represent the actual measurements, red dots are created data points.



**Figure 3 : Chemometric correlation obtained by Explifit® software.
Calculation of bitumen grade from IR measurements**

In this case the software gives the following equation:

$$\text{Bitumen penetration (1/10 mm)} = 27.4 + 1211 [\text{Abs}_{1035}] - 1363 [\text{Abs}_{1570}] + 13.71 [\text{Abs}_{2929}] - 1053 [\text{Abs}_{3854}]$$

The linear regression coefficient (R^2) is 0.90.

4.2 Results

Different chemometric correlations are presented here:

4.2.1 Correlation between bitumen properties from Infrared measurements

All the IR measurements are performed using a PerkinElmer Spectrum 400 infrared spectrometer.

The bitumen concentration is 3 w% in perchloroethylene. For all the chemometric correlations, we determined the linear regression coefficient (R^2) and an equation of correlation with 2 to 4 (maximum) independent variables. Figure 4 features the R^2 correlation coefficient according to the different bitumen tests. From only 8 bitumen samples, interesting fair correlations are obtained with R^2 ranging from 0.55 to 0.88, most above 0.75.

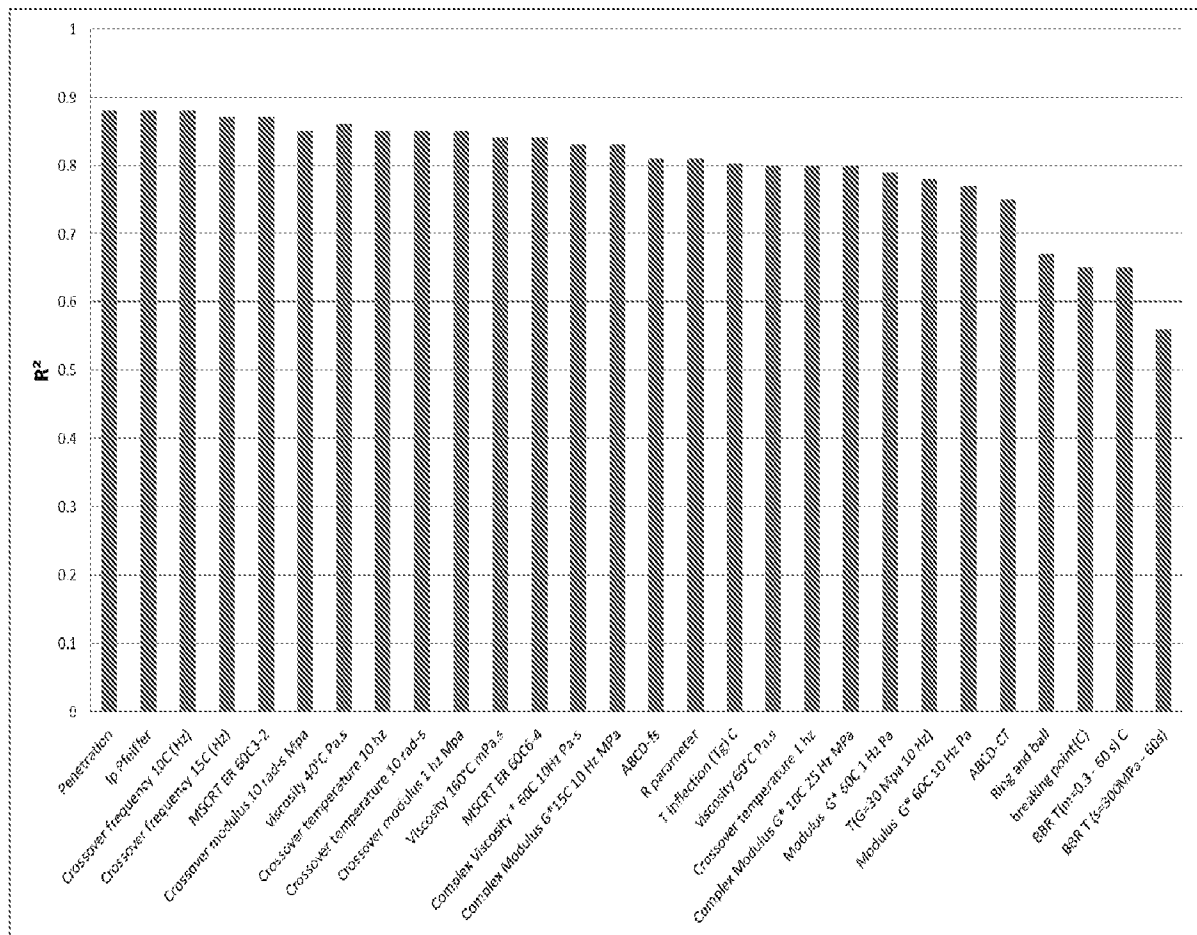


Figure 4 : Correlation between IR measurements and bitumen tests

4.2.2 Correlation between bitumen properties and SAR-AD compositions

Figure 5 presents correlations obtained between SAR-AD bitumen composition and bitumen tests.

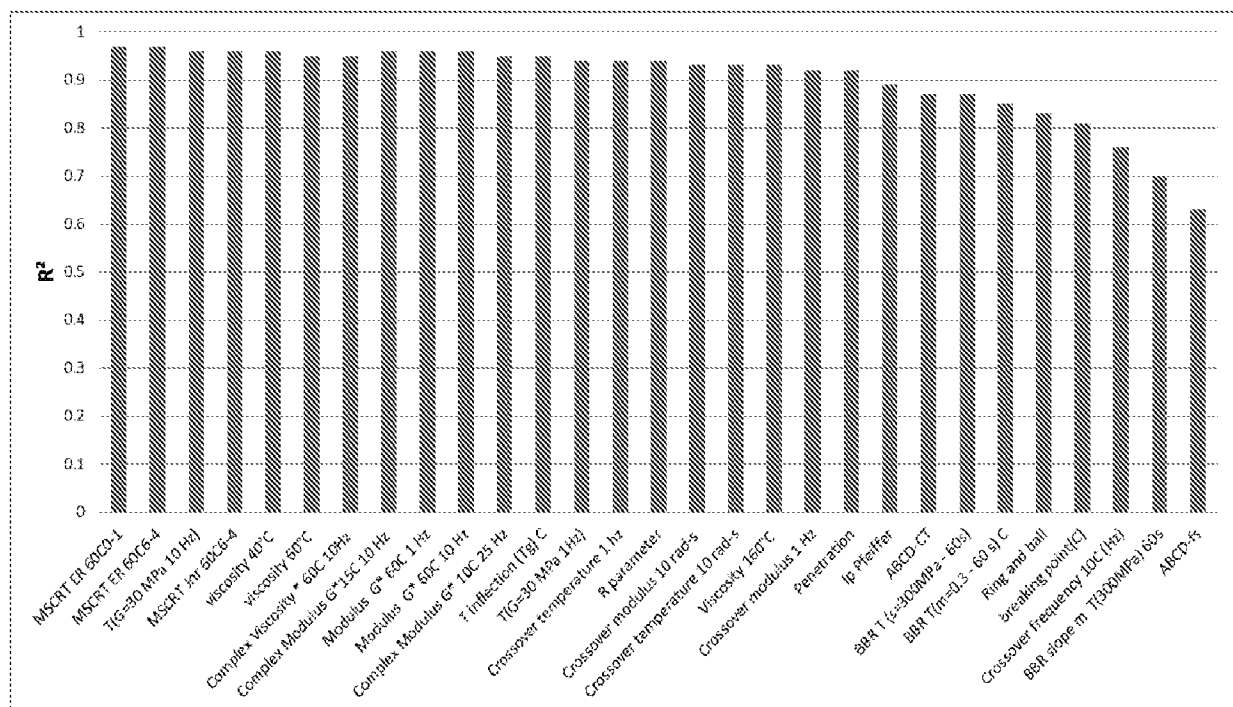


Figure 5 : Correlation between SAR-AD measurements and bitumen tests

From SAR-AD compositions, more interesting and stronger correlations are obtained with R^2 ranging from 0.62 to 0.98, most above 0.80, including very good correlations ($R^2 > 0.85$), using 3 or 4 independent variables, with a lot of tests such as MSCR, dynamic viscosity (40 to 160 °C), complex viscosity (60°C), complex modulus (10 to 60°C for different frequencies), crossover modulus, crossover temperature, R-parameter, penetration test, IP Pfeiffer...

For some tests, the correlation is less relevant (R^2 between 0.6 and 0.85): Ring and Ball temperature, Fraass breaking point, BBR results, ABCD results...

Among the various fractions determined by SAR-AD measurements, 2 fractions often appear in correlations: naphthene saturates and asphaltene soluble in cyclohexane.

5. CORRELATION BETWEEN ASPHALT PROPERTIES (AC 10) AND SAR-AD COMPOSITIONS OF THE CORRESPONDING BITUMEN

A lot of asphalt tests are impacted by the mix design and the nature of the aggregates (shape, petrography...): gyratory shear compactor, rutting, sensitivity to water or stiffness modulus according to test procedure.

To be able to predict the impact of the chemical composition of bitumen, the mix design was fixed in terms of void content, bitumen content, and aggregate nature (diorite). This research was based on the following tests: rutting, various stiffness modulus tests all using cylindrical specimens (NF EN 12697-26 C, applying indirect tension at 10 °C for 124ms application time, NF EN 12697-26 E applying direct tension at 15°C, for 0.02 s, LC 26/700 applying direct tension compression at 15 °C, 10 Hz), fatigue and TSRST.

Figure 6 presents the chemometric correlations obtained from SAR-AD composition of neat bitumen and bitumen after RTFO.

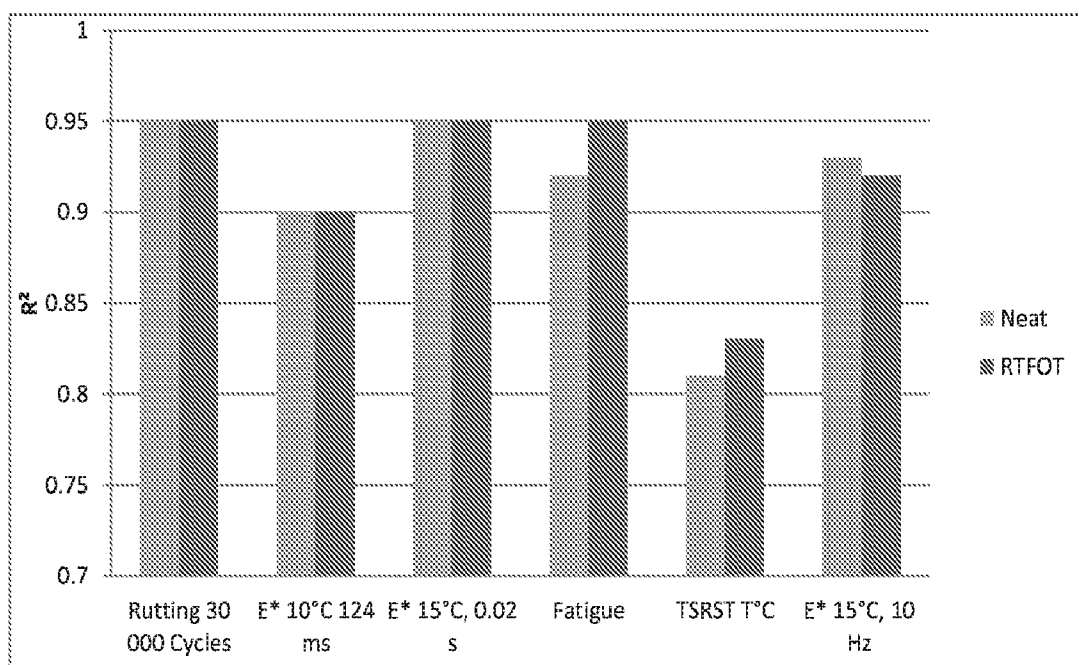


Figure 6 : Correlation between chemical composition of bitumens (SAR-AD) with asphalt performances

Very good correlations are obtained from SAR-AD composition except for the TSRST tests. Concerning low temperature performance as measured by TSRST, the lower quality of the correlation may be connected to the lower quality of the correlations obtained for low temperature tests (Fraass, ABCD, BBR) on bitumen (Figure 5).

6. RELATIONSHIPS BETWEEN LOW TEMPERATURE BEHAVIOR OF BITUMEN AND ASPHALT PERFORMANCE (TSRST)

The Fraass test is the common test in Europe to qualify bitumen at low temperature. This test is known for its poor reproducibility and its use for PmB characterization is often disputed [7]. The bending beam Rheometer (BBR) [8],

developed during the SHRP program, enables the determination of two criteria: the stiffness modulus and the ability of stress relaxation of the bitumen ($m\text{-value} = -d(\log S)/d(\log t)$). The relaxation rate is a function of the loading time dependency of the stiffness and directly related to the creep rate. The higher the creep rate, the faster the relaxation of stresses. These two parameters are numerically called:

- $T_{S=300\text{ MPa}}$ = temperature at which the stiffness modulus (S) equals 300 MPa for the loading time of 60 s
- $T_{m=0.3}$ = temperature at which $m = 0.3$ for the loading time of 60 s

In the literature, the correlation between BBR test results on bitumen and polymer modified bitumen with the TSRS test on asphalt are contrasted [9,10]. Other authors found interesting correlations between BBR results and field behavior after several years [11] or validated that linear viscoelastic properties play an important role in the energy dissipation during crack opening but highlighted that this test is insufficient to predict the temperature of brittleness of bitumen [12].

In order to have an overall view of the different tests used to characterize the low temperature behavior we added two other tests in this research program: the ABCD test [4,13] and DSC to determine the glass transition [14].

The ABCD test is a fairly new test method, using a simple testing device that can provide the overall low temperature cracking potential of a bitumen. A circular bitumen specimen is prepared on the outside of an Invar ring. Invar is a steel alloy with near zero coefficient of thermal expansion/contraction. As the temperature is lowered, the thermal stress within the bitumen increases until fracture. For the tests, the cooling rate was fixed at 10°C/h .

Figure 7 presents the results obtained on the 8 neat bitumen samples.

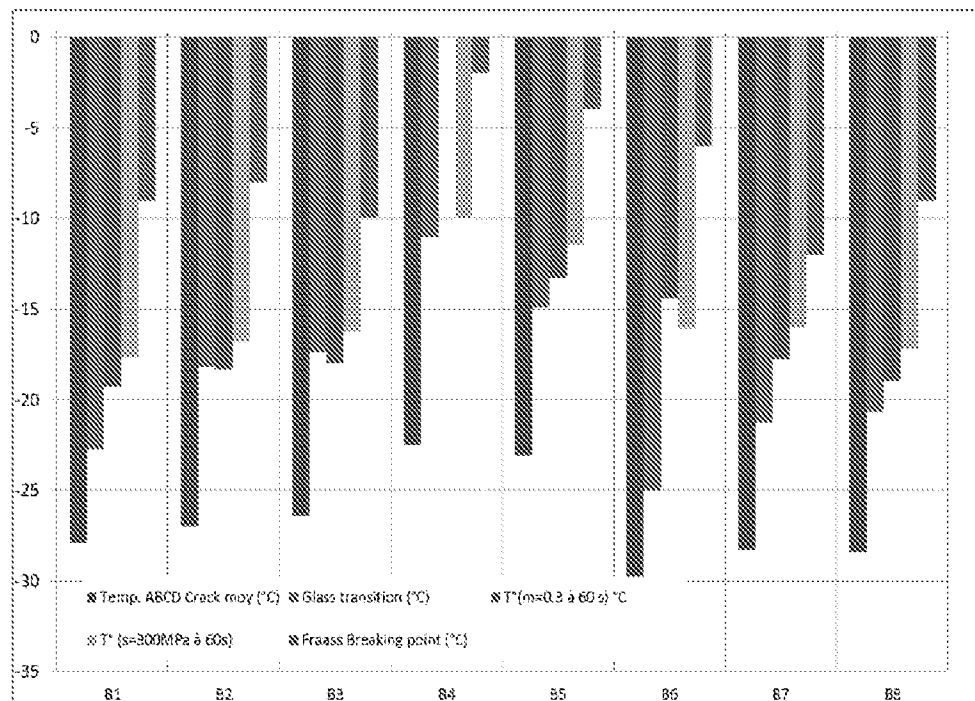


Figure 7 : Comparison of different tests to qualify the low temperature behavior of bitumens

Taking into account the reproducibility of the ABCD test ($R=3$), the bitumen classification in ascending order of temperature gives the following result : B6 (-30°C) < B7, B8, B1 (-28°C) < B2, B3 (-27°C) < B4, B5 (-22.5°C). The glass transition measurement, gives a temperature between 5 and 11°C higher. The bitumen ranking remains the same except for sample B4 which becomes the most brittle. The BBR test results do not show significant difference according to either $T_{S=300\text{ MPa}}$ or $T_{m=0.3}$. The main difference in comparison to the glass transition or ABCD test concerns sample B6. With the BBR test, this sample seems to be more brittle than samples B7, B8, B1, B2 and B3. The level of temperature with this test is close to the glass transition temperature. The Fraass breaking point ranks bitumen according to their grade at a significantly higher level of temperature: 20°C vs. the ABCD test, and 5 to 10°C vs. BBR or the glass transition. Figure 8 compares these results with TSRST critical temperature values.

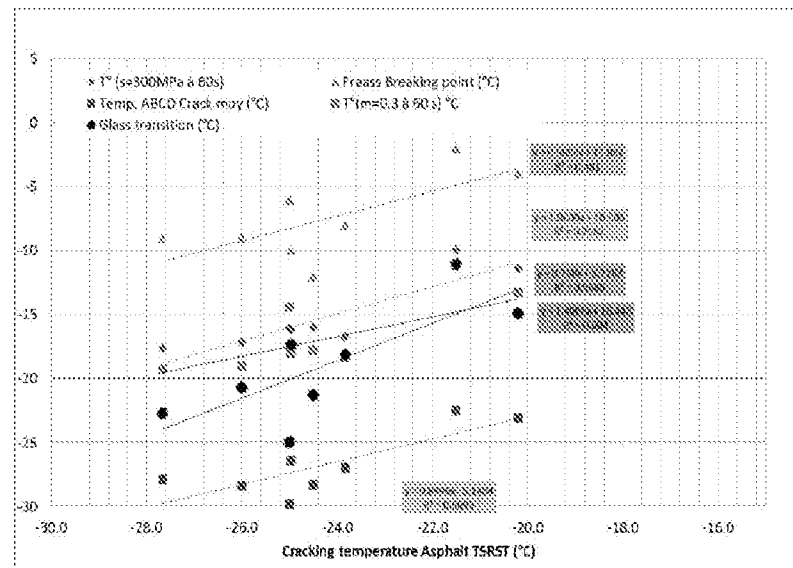


Figure 8: Correlation between TSRST test and different bitumen tests before aging

The best correlation is obtained with the BBR test on the S criterion ($T_{S=300\text{MPa}}$) but the ABCD test gives cracking temperatures closest to the TSRS test critical ones. Subtracting 10 °C to the BBR temperatures as applied in the Superpave specifications, the gap with the TSRST values becomes very low.

There is no correlation between the TSRS test and the Fraass values (between 10 to 20°C of difference).

Figure 9 presents the correlation between the bitumen tests after aging (RTFOT + PAV) with the TSRS test.

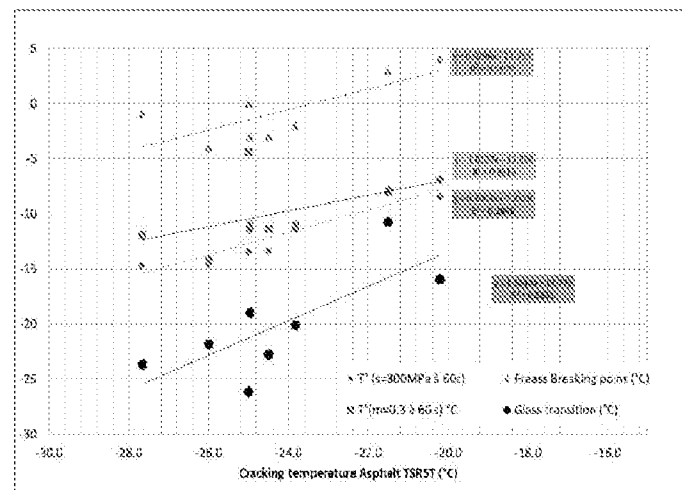
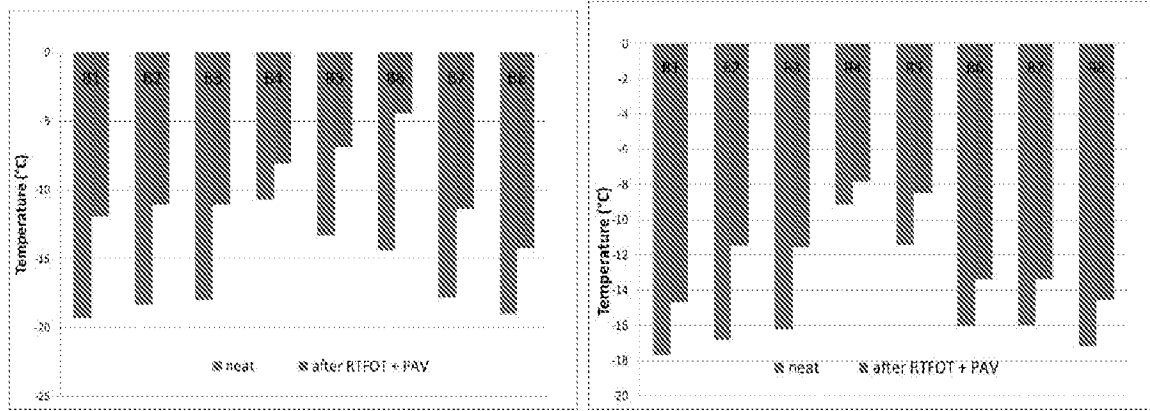


Figure 9: Correlation between TSRST test and different bitumen tests after RTFOT+PAV aging

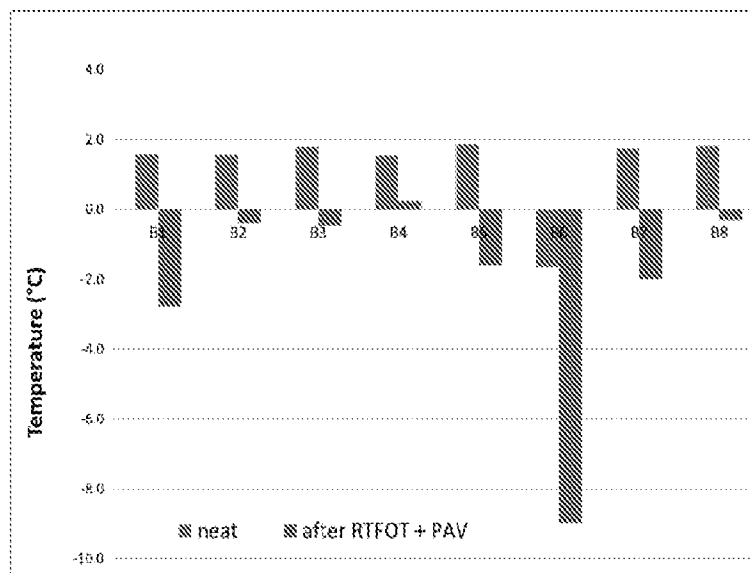
PAV aging does not have the same impact on the different BBR criteria. $T_{S=300\text{ MPa}}$ correlates well with the TSRS test; the evolution of temperature (after PAV versus unaged) is between 15 to 30% for all bitumen samples. $T_{m=0.3}$ on neat bitumen and even more on bitumen after RTFO + PAV doesn't correlate with the TSRS critical temperature.

Figures 10 a and 10 b present the difference for all bitumen samples after aging on the BBR tests.

Figure 10a : $T_{m=0.3}$ before and after RTFOT + PAVFigure 10b : $T_{S=300 \text{ MPa}}$ before and after RTFOT + PAV

These results highlight the importance of analyzing the low temperature properties of bitumen after aging. Some bitumen are sensitive to oxidation phenomena and their low temperature characteristics are strongly impacted. According to bitumen origin, aging reduces greatly the relaxation potential of bitumen while the effect on the low temperature stiffness is only moderate. This is particularly the case of bitumen B6 [15].

The difference of temperature (ΔT) between $T_{S=300 \text{ MPa}}$ and $T_{m=0.3}$ after PAV could be a good indicator to illustrate this bitumen oxidation sensitivity (figure 11). The higher the $|\Delta T|$ value, the higher the sensitivity to oxidation and lesser the durability of the road could possibly be.

Figure 11: Evaluation of the bitumen oxidation ΔT ($T_{S=300 \text{ MPa}} - T_{m=0.3}$)

The PG (Superpave Performance Grade) of the bitumen B6 is conditioned by its property after PAV; B1 also shows some of the same behavior but to a lesser extent. This is not reflected by the mechanical properties measured on laboratory made asphalt which lead to very good results (fatigue, TSRST...). However, these asphalt properties are measured with no long term aging conditioning. This is likely to explain this discrepancy.

Additional tests are ongoing at the Eurovia research centre. More asphalts will be laboratory manufactured and aged according to the procedure defined by the RILEM ATB/SIB technical committee [16] using some of the same bitumen samples, keeping the same VMA and VFA. The TSRS test will be then performed to correlate with the BBR results after PAV.

7. CONCLUSION

The collaboration Eurovia/Western Research Institute enabled to evaluate the potential of applying chemometric correlations to predict bitumen or asphalt properties from the infrared spectrum or /and the SAR-AD generic composition.

Correlations to date are more relevant for SAR-AD composition than Infrared data. An Infrared spectrum contains more than 3400 wavelengths, the selection of 3 or 4 wavelengths from only 8 bitumens is too complex. Additional investigations on a wider bitumen population may enable to improve some of these correlations.

From SAR-AD compositions, good correlations were found with several bitumen viscosity or rheological tests ($R^2 > 0.85$), including the viscoelastic behavior characterized by the R-parameter, crossover modulus, and penetration. For BBR or ABCD low temperature properties, the correlations are less relevant ($R^2 < 0.8$).

Chemometric correlations with asphalt properties are obviously more complicated. For many tests, the aggregate impact is significant (sensitivity to water, gyratory shear compactor...). From the SAR-AD test limited to 2 or 3 fractions, very interesting and significant correlations are obtained to predict fatigue, stiffness modulus...

Additional analyses are ongoing to improve or validate these correlations but also to search for new correlations to understand the impact of bitumen origins on their mechanical performances after long term aging.... About 10 new bitumen samples are going to be added to this study.

The second part of this collaboration is devoted to the comparison between bitumen tests and asphalt tests to propose new and more relevant performance indicators for product standards in Europe. In this filed, the present article has been focused on low temperature properties.

The results highlight the importance of analyzing the low temperature behavior of bitumen after aging to take into account the impact of oxidation. According to bitumen origin, aging can greatly reduce the relaxation potential of bitumen while the effect on the low temperature stiffness is only moderate. It is to be reminded here that some publications correlate indeed stress relaxation to the cracking phenomenon on field.

The fact that asphalt tests are always done at the early stage could explain the bad correlation with the BBR results after PAV. Additional tests will be performed on asphalt after aging to evaluate the ability to correlate with BBR results after PAV.

These trials may enable to answer recurring questions related to low temperature performance, such as the possibility to correlate a stiffness test on bitumen with a thermal stress restrained test on asphalt, and the selection of more relevant test methods to predict cracking phenomena in the field after aging.

REFERENCES

- [1] *The Automated Asphaltene Determinator Coupled with Saturates, Aromatics, and Resins Separation for Petroleum Residua Characterization*, Ryan B. Boysen and John F. Schabron, Energy Fuels 2013, 27, 4654–4661
- [2] *Interpretation of Dynamic Mechanical Test Data for Paving Grade Asphalt*, Christensen, D.W. and Anderson, D.A. Journal of the Association of Asphalt Paving Technologists, 61, 67-116, 1992
- [3] AASHTO TP 101-12-UL, *Estimating Damage Tolerance of Asphalt Binders Using the Linear Amplitude Sweep*, available on the MARC website: <http://uwmarc.wisc.edu/linear-amplitude-sweep>
- [4] *Asphalt Binder Cracking Device to Reduce Low-Temperature Asphalt Pavement Cracking*, S. Kim, Final Report, FHWA-HIF-11-029, 2010
- [5] NF EN 12593: *Determination of the Fraass breaking point*
- [6] *Chemo- mechanical software, Fundamental properties of asphalts and modified asphalts product*, R. Glaser, A. Beemer, T.F. Turner, march 2015

- [7] *Checking low temperature properties of polymer modified bitumen – is there a future for Fraass Breaking point ?*, B. Eckmann, M. Mazé, Y. Le Hir, O. Harders, G. Gautier, B. Brulé, 3rd Eurasphalt & Eurobitume congress, Vienne, 2004
- [8] NF EN 14771: *Determination of the flexural creep stiffness: Bending beam rheometer (BBR)*
- [9] *Caractérisation du comportement à basse température des liants bitumineux*, S. Largeaud, B. Eckmann, S. Faucon Dumont, Y. Hung, L. Lapalu, G. Gauthier, *Revue générale des routes et de l'aménagement*, n° 928, juin 2015
- [10] *Combined traffic and climate effects on durability of pavement mixture with polymer modified binder*, M. Ould-Henia, A.G Dumont, M. Pittet, J.P Planche, S. Dreessen, *Proceedings ISAP 2010 congress*, Nagoya, JP, p 1594.
- [11] *Durability study: Field aging of conventional and polymer modified binders*, S; Dreessen, M. Ponsardin, JP. Planche M. Pittet, A.G Dumont, TRB, 2010
- [12] *Relationships between low temperature properties of asphalt binders*, Babadopoulos, Le Guern, Chailleux, Dreessen, ISAP 2012
- [13] *Determination of low temperature thermal cracking of asphalt binder by ABCD*, S.S Kim, Z.D. Wysong, J. Kovach, TRB, 2006
- [14] *Characterization of paving asphalt by differential scanning calorimetry*, P. Claudy, J.M Létoffé, G.N King, J.P Planche, B. Brulé, *Fuel science and technology international* 9, 1991.
- [15] *Performance indicators for low temperature cracking*, H. Soenen, A. Vanelstraete, Rilem, 2003
- [16] *Advances in interlaboratory testing and evaluation of bituminous materials*, M.N. Partl et al., Rilem state of art report 9, 2013

Filed via EFS

IN THE UNITED STATES PATENT AND
TRADEMARK OFFICE

Application Number:

Filed:

Applicants:

Title:

Assignee:

Attorney Docket:

Customer No.:

Confirmation No.:

Ronald R. Glaser et al.

Method for Correlating Physical and Chemical Measurement Data
Sets to Predict Physical and Chemical PropertiesThe University of Wyoming Research Corporation d/b/a Western
Research Institute

WRI-CorrSoft-PCT

33549

APPENDIX 3

Relationships between solubility and chromatographically defined bitumen fractions and physical properties

R. Glaser^{1, a}, J.-P. Planche^{1, b}, F. Turner^{1, c}, R. Boysen^{1, d}, J. F. Schabron^{1, e}, F. Delfosse^{2, f}, I. Drouadaine^{2, g}, S. Faucon-Dumont^{2, h}, S. Largeaud^{2, i}, B. Eckmann^{3, j}

¹ Western Research Institute, Laramie, United States

² Eurovia Research Centre, Merignac, France

³ Eurovia Technical Department, Paris, France

^a rglaser@uwyo.edu

^b jplanche@uwyo.edu

^c fturner@uwyo.edu

^d rboysen@uwyo.edu

^e jfschabra@uwyo.edu

^f frederic.delfosse@eurovia.com

^g Ivan.drouadaine@eurovia.com

^h Stephane.Faucon-Dumont@eurovia.com

ⁱ Sabine.Largeaud@eurovia.com

^j Bernard.Eckmann@eurovia.com

Digital Object Identifier (DOI): dx.doi.org/10.14311/EE.2016.337

ABSTRACT

An understanding of how bitumen chemical composition influences mechanical behavior is critical to addressing a number of practical issues concerning bitumen utilization. Using simple chemical tests to assess bitumen quality is of practical value to the purchaser, but other applications exist as well. Blending to achieve material design objectives is obviously of huge industrial and commercial value, as is designing and selecting better additives. This is also key to understanding physical changes related to bitumen oxidation and predicting performance.

Western Research Institute (WRI), in partnership with Eurovia, has examined several paving grade bitumens using an automated asphaltene solubility fractionation method developed by WRI under contract with the United States Federal Highway Administration (FHWA) that is an expansion of the traditional SARA method (SAR-AD™). The results of these chemical characterization studies were then correlated to a wide range of bitumen properties (Penetration, Ring and Ball softening point, Dynamic Shear Rheometry, and others) using WRI's multivariable significance search algorithms (ExpliFit™). In general, most properties can be explained with a high coefficient of correlation by considering a balance between mobile and relatively immobile constituents, as well as interactions induced by polarity and polyaromaticity. This paper focuses on the micro-structural explanation of the significant parameters affecting the storage shear modulus over the range of temperatures investigated. The correlation results suggest that as temperature changes, the amount of the mobile fractions in the bitumen dominates low temperature behavior, while at high temperatures, multiple fractions must be considered.

Keywords: Ageing, Asphalt, Chemical properties, Compatibility, Mechanical Properties

1. INTRODUCTION

In recent years, a significant evolution on the European market has been observed. European refining, French in particular, is currently in a phase of rationalization and search for maximum flexibility in crude supplies.

The European standard EN 12591 appears as insufficient to ensure satisfactory performance of the finished products, particularly in case of specialty products such as high modulus asphalt regarding stiffness modulus and fatigue resistance, polymer modified bitumen, and bitumen emulsions with respect to settling tendency and viscosity.

The search and validation of performance-related bituminous binder properties continues to be a key issue for the paving industry in Europe, as well as in the US and the rest of the world. With the Superpave system implementation in the US, important progress has been achieved and is still on-going. In Europe, the development of 2nd generation product standards appears to be more necessary than ever.

In this context, Eurovia and the Western Research Institute (WRI) launched a research program, in 2013, to search for correlations between bitumen properties and the performance of the finished asphalt product.

In the short term, the identification of robust correlations between bitumen composition and mechanical properties has obvious practical value in material evaluation and blending to meet given specifications. The information obtained from these correlations can also be applied to testing and improving the understanding of the fundamental concepts of how bitumen composition gives rise to the observed physical properties. Bitumen has long been considered to behave similarly to colloidal systems, and the idea of a colloid-like microstructure has existed at least as long as the turn of the century when asphaltenes were first identified [1]. Since that time, a wide range of conceptual models of the bitumen microstructure have been proposed, at various levels of detail. Although not comprehensive, several references are provided to illustrate some of the work done in this area historically [2-37]. Nearly all of these propose that bitumen is not homogenous at some scale above molecular dimensions. It is also generally conceded there exists some relationship between solubility defined fractions and the resulting micro-structure. This study is an effort to quantify these relationships. The micro-structure, in turn, is primarily responsible for the mechanical properties of interest to the design of a number of bitumen containing products. This work correlating solubility defined fractions to rheological properties suggests, as expected, that the important fractions defining the mechanical behavior change with temperature. At low temperatures, much of the material exists as a relative immobile glass or associated gel-like material, with the content of saturates, the last fraction to solidify into a glass, being the most significant one in defining the mechanical behavior. As the material warms, the portioning of mobile and immobile phases, along with a change from gel-like to sol-like behavior changes according to temperature dependant solubility characteristics. Consequently, empirical correlations of solubility defined fractions with mechanical properties will not show a consistent set of fractions primarily defining the mechanical properties. At low temperatures the most mobile fractions are the most significant where gel-behavior is observed. At high temperatures where sol behavior is observed, multiple fractions are required to define the system, with the suspension defining fraction, the asphaltenes, being the most significant. In rheological terms, low phase angle can be described with a few parameters, while higher phase angle properties depend more strongly on a range of solubility fractions.

2. RESEARCH PROGRAM

This research program was launched by Eurovia in collaboration with the Western Research Institute (Wyoming/ USA). 8 bitumens (all unmodified) were selected: B1 to B8. With these bitumens, 12 asphalts were manufactured (8 with a diorite and 4 with limestone aggregates). For each asphalt, the bitumen content was 4.9 %. Table 1 presents the main characteristics of these bitumens and the different asphalt designs.

Table 1: Bitumen characteristics and asphalt designs

	B1	B2	B3	B4	B5	B6	B7	B8
Penetration (1/10mm): NR EN 1426	40	37	40	22	26	28	55	57
Ring and Ball Temperature (°C) : NF EN 1427	53.4	53	52	59	57.2	61	49	49.2
Superpave Performance Grading (PG)	70-22	70-16	70-16	76-16	76-16	76-10	64-22	64-16
HMA with diorite	X	X	X	X	X	X	X	X
HMA with limestone	X		X	X			X	

The analysis program for the bitumens (neat, after RTFO, recovered, after RTFO + PAV) :

- 1- Chemical analysis : infrared, SAR-AD : Saturates, Aromatics, Resins and Asphaltene Determinator [38], SEC : Size exclusion chromatography, DSC : Differential Scanning Calorimetry (to assess glass transition, wax content)
- 2- SHRP tests: Bending Beam Rheometer (BBR), DSR: to determine master curves, crossover, R-parameter... [43],[44]
- 3- Advanced rheological tests (LAS tests...) [42]
- 4- Asphalt Binder Cracking Device (ABCD) test [39],
- 5- Conventional European tests: penetration, ring and ball temperature, Fraass breaking point,...

This paragraph presents the overall research program launched in 2013, but this article will present only some chemometric results [40], the correlation between bitumen rheological properties, specifically the storage modulus, over a range of temperature (10° to 60°C). Other articles will be published in the future to present more results in detail.

3. BITUMEN SELECTION

The key point in a chemometric correlation is based on determination of the quality of bitumen selection. The first step of the program before launching the analyses was to verify that the chemical composition and rheological performances of these bitumens were significantly different.

3.1 Chemical composition

The SAR-AD [38] test is a novel approach, developed by the Western Research Institute, which combines the Automated Asphaltene Determinator (AD) separation with an automated SAR (saturates, aromatics and resins) separation to provide a fully integrated rapid automated SARA (saturated, aromatics, resins and asphaltenes) separation using milligram sample quantities. The combined SAR/AD separation utilizes high performance liquid chromatography (HPLC) equipment with multiple columns and solvent switching valves to conduct the highly complex automated separation. Figure 1 presents the chemical compositions of the 8 bitumens. The sample set represents considerable variation in the solubility defined fractions.

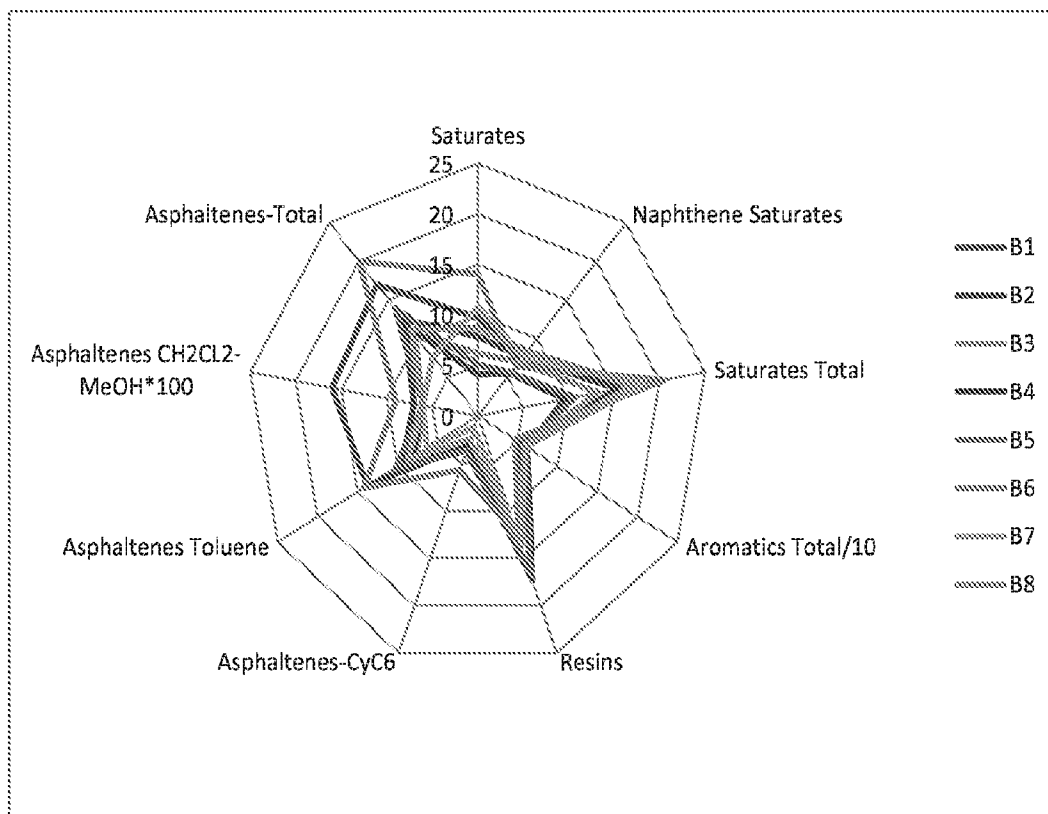


Figure 1: Bitumen composition B1 to B8

3.2 Rheological properties

The bitumen samples selected show a wide range of rheological behavior. Figure 2 illustrates the variation in the complex modulus isotherms at 15 °C from 1 to 30 Hz.

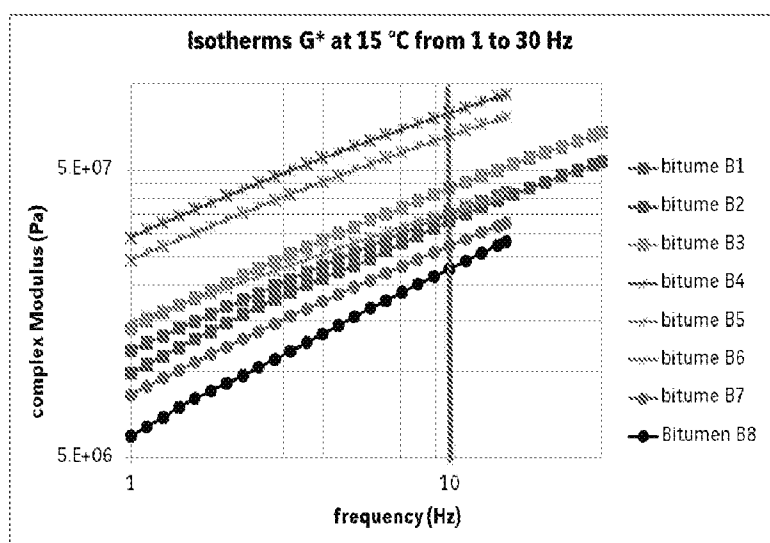


Figure 2: Isotherms G^* at 15°C from 1 to 30 Hz

4. CHEMOMETRIC CORRELATIONS

4.1 Explifit® Software

Explifit ® [40] is a software program designed to investigate relationships between independent and dependent variables using standard multivariable linear regression algorithms adapted for under determined problems. An under determined problem is a situation where the count of possibly significant independent variables exceeds the number of observations. For example, measuring 24 chemical properties to correlate with 8 bitumen's is under determined and not tractable with traditional methods. This software was developed at Western Research Institute.

The dependent variables data are all analyses performed on bitumen or asphalt.

The independent variables data are typically measured to predict the dependent variables data. In this research program, the independent data are: infrared (IR) spectra measurements, SAR-AD compositions and distribution of the particle sizes by SEC.

Example: If we try to correlate bitumen IR measurements with the bitumen penetration, in a first step, the software will find out which of the wavenumbers are significant when combined additively with other significant wavenumbers. This step will enable a reduction in the number of relevant wavenumbers. In a second step, the software will propose an equation such as:

$$\text{Bitumen penetration (1/10 mm)} = c_0 + c_1 [\text{Abs}_1] + c_2 [\text{Abs}_2] + \dots + c_n [\text{Abs}_n]$$

c_i = Fit coefficient i , Abs_i = Absorbance to the wavenumber i ...

In addition to the actual measurements, a precision file for each independent and dependent variables data set must also be prepared. The program requires these files to create the data clouds needed for computation.

4.2 Results

The research program carried out at Eurovia consists of an enormous collection of measurements for the bitumen samples selected. Only a subset of that data will be studied in detail here, addressing the temperature dependant role of solubility fractions measured by SAR-AD. Seven solubility fractions are measured using 2 detectors, Evaporative Light Scattering Detector (ELSD) to measure mass fractions and a 500 nm UV detector to study polycondensed aromatics. The following ELSD measurements are used in this discussion (Table 2):

Table 2: SAR-AD Measurements correlated with rheology.

Saturates
Naphthene saturates (ring structures)
Aromatics
Resins
Cyclohexane soluble asphaltenes (least polar)
Toluene soluble asphaltenes (moderately polar)
Methylene chloride-methanol soluble asphaltenes (highly polar)
Total asphaltenes

To find the significant bitumen fractions responsible for the observed rheological properties, and the effect of temperature on their relative importance, multivariable correlations of the ELSD measurements were fit against the complex, storage and loss shear moduli measured at 10°C, 15°C, 20°C, 30°C, 40°C, 50°C and 60°C using a dynamic

shear rheometer. Due to space limitations, only the Shear storage modulus (G') at 1Hz frequency will be described in detail here. Similar relationships exist for the complex (G^*) and loss (G'') moduli and for measurements at other frequencies.

Initially, 4-parameter fits at each temperature range were completed using the eight ELSD measurements. The qualities of the fits are shown in Table 3:

Table 3: Initial ELS detector fits for G'

Temperature °C	Pearson's r^2
10	0.870
15	0.864
20	0.879
30	0.898
40	0.947
50	0.976
60	0.986

The significance of the parameters compared using the F test, which is a ratio of the fit residuals without the parameter of interest divided by the residuals with the parameter of interest. Large F test values are more significant than small ones. An examination of the significance of the parameters appearing in 4-parameter fits suggest that 3 parameters may be sufficient for a robust correlation across the temperature range studied here. (note: F-tests for negative fit coefficients are shown negative)

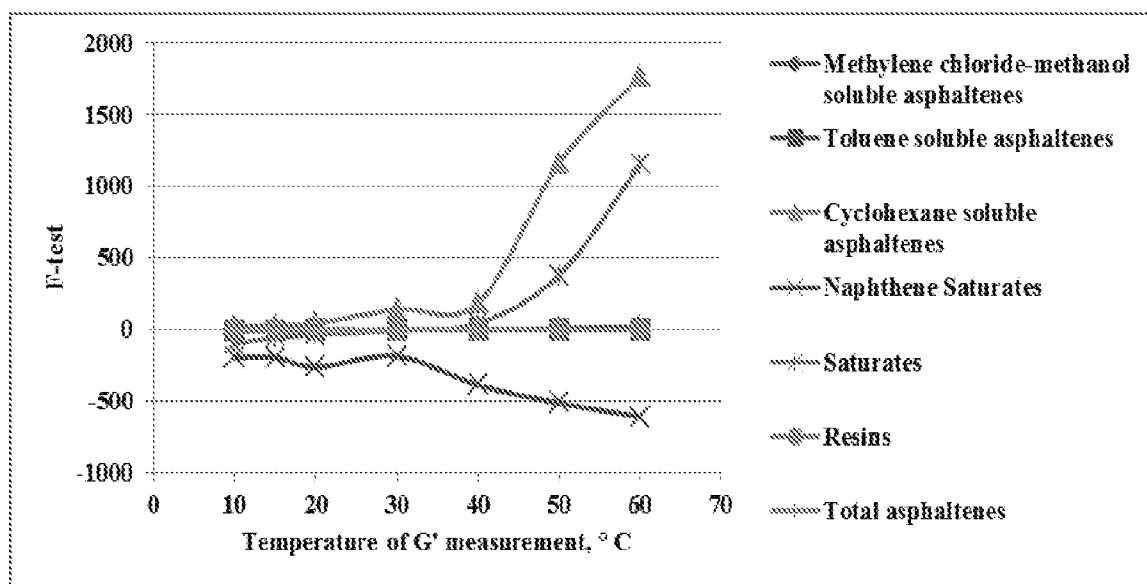


Figure 4 : F-test significance from ELSD fits

Based upon the results depicted in figure 4, the regressions were repeated using only cyclohexane-soluble asphaltenes, saturates, and naphthene saturates values. The fits shown in table 4 are similar in quality to those obtained using the entire ELSD measurement list to generate a 4-parameter model.

Table 4: Correlation fit quality of 4 parameter and 3 consistent parameter models.

Temperature °C	4 parameter Pearson's r^2	3 consistent parameter Pearson's r^2
10	0.870	0.871
15	0.864	0.849
20	0.879	0.843
30	0.898	0.894
40	0.947	0.951
50	0.976	0.973
60	0.986	0.984

Now that a simple consistent set of independent SAR-AD variables is established for all seven correlations at each temperature, it is possible to determine the temperature dependency of the coefficients and produce one equation that describes G' in SAR-AD compositional terms over the entire temperature range. The same approach can be used with the complex and loss moduli. Newtonian materials follow an Arrhenius form,

$$\eta = Ae^{\frac{E_a}{RT}}$$

η is viscosity

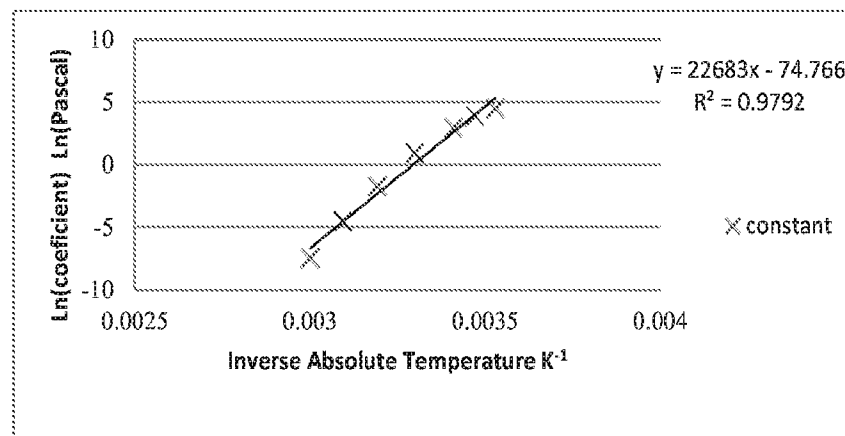
E_a is the activation energy

R is the universal gas constant

T is the absolute temperature

A plot of $\ln(a_n)$ vs $1/T$ should yield a straight line with a slope of E_a / R and an intercept of $\ln(A)$.

The constants are nearly Arrhenius, although a slight curvature can be seen in the plots, not unlike the similar master curve shift function shapes observed for bitumens, suggesting that a more sophisticated temperature function such as WLF may produce a higher precision calibration.

Figure 5: Intercept ($\ln(A)$) Arrhenius plot

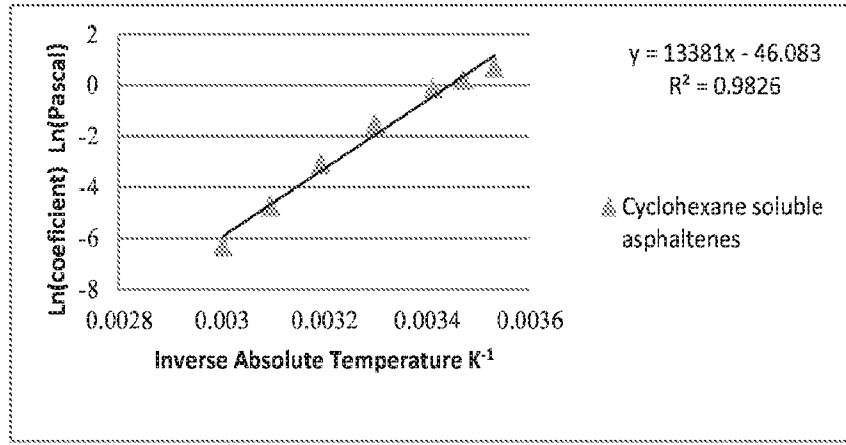


Figure 6 : Cyclohexane soluble asphaltenes fit constant Arrhenius plot

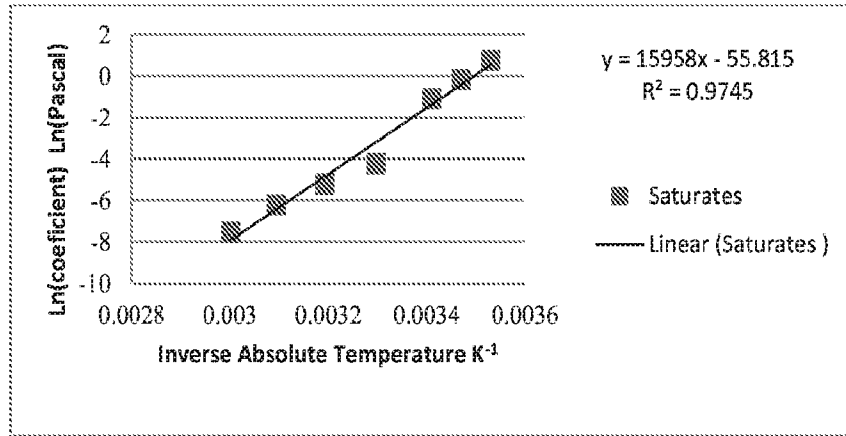


Figure 7 : Saturates fit constant Arrhenius plot

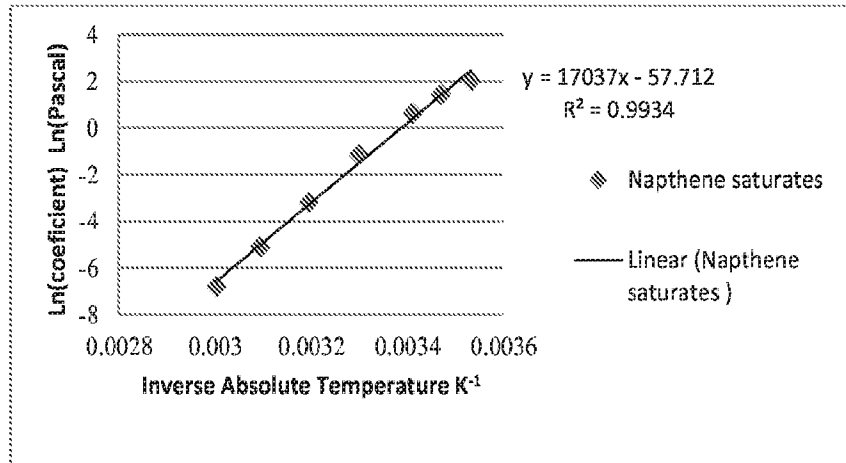


Figure 8: Naphthene Saturates fit constant Arrhenius plot

The resulting correlation over the entire temperature range is:

$$G'(T)_{Hz} = A_0 e^{\left(\frac{E_a}{R}\right)_0 \frac{1}{T}} + A_1 e^{\left(\frac{E_a}{R}\right)_1 \frac{1}{T}} (\% \text{ Cyclohexane Asphaltenes}) \\ + A_2 e^{\left(\frac{E_a}{R}\right)_2 \frac{1}{T}} (\% \text{ Saturates}) - A_3 e^{\left(\frac{E_a}{R}\right)_3 \frac{1}{T}} (\% \text{ Naphthene Saturates})$$

where the fit constants are listed in table 5.

Table 5: Temperature SAR-AD function for G' fit constants

A ₀	3.41E-33	(E _a /R) ₀	22683
A ₁	9.72E-21	(E _a /R) ₁	13381
A ₂	5.78E-25	(E _a /R) ₂	15958
A ₃	8.65E-26	(E _a /R) ₃	17037

4.3 Discussion

The results of the investigation of the SAR-AD fractions needed to estimate the storage modulus (G') over the temperature range examined suggest that at higher temperatures where a sol-like structure is presumed to exist, several fractions exist in the mobile phase and are significant to the estimate. However, at lower temperatures where many fractions exist in an associated gel-like phase, very good estimates of the storage modulus can be obtained using the mobile naphthene saturate fraction alone (see Figure 4 where the naphthene saturates is the only fraction with a significant F-Test value at low temperature). The data and resulting correlations also suggest, that for the eight bitumens studied, the associated materials have similar mechanical characteristics, and since this variance is small, the variance in the storage modulus can be explained at all temperatures with a knowledge of composition of the mobile fraction. This is consistent with differential scanning calorimetry results that show the mobile fraction anchors the low end of the glass transition region regardless of asphaltene content [41].

5. CONCLUSION

The collaboration work by Eurovia/Western Research Institute enabled the evaluation of applying chemometric correlations to predict bitumen or asphalt properties from the infrared spectrum or /and the SAR-AD composition. Using multivariable techniques, quite good correlations are often obtained with coefficient-squared values often 0.9 or above.

The correlation of chemical composition with rheological properties generally requires a knowledge of several components in the bitumen, as these components interact to define the structure and ultimately, the mechanical properties. Understanding the relationships between the solubility defined fractions can lead better quality control methods and bitumen composition adjustments to produced desired mechanical properties.

REFERENCES

- [1] *Mémoires de J.-B. Boussingault*, Boussingault, Jean-Baptiste (1892–1903), Paris: impr. de Chamerot et Renouard, 1892
- [2] Direct observation of the asphaltene structure in paving-grade bitumen using confocal laser-scanning microscopy, Bearsley, S., Forbes, A. and Haverkamp, R. G., *Journal of Microscopy*, Vol. 215, Issue 2, pp. 149-155, 2004
- [3] The glass transition temperature and the mechanical properties of asphalt, Breen, J. J. and Stephens, J. E., *Proceedings of The Annual Conference of the Canadian Technical Asphalt Association*, Report No. JHR-67-16, 1967
- [4] A molecular interpretation of the toughness of glassy polymers, Brown, H. R., *Macromolecules*, Vol. 24, pp. 2752-2756, 1991
- [5] A new interpretation of time dependent physical hardening in asphalt based on DSC and optical thermoanalysis, Claudy, P., Létoffé, J. M., Rondelez, F., Germanaud, L., King, G. N. and Planche, J. P., *ACS Symposium on Chemistry and Characterisation of Asphalts*, Washington, D.C. pp. 1408-1426, 1992
- [6] Electron microscopic investigations on the nature of petroleum asphaltics, Dickie, J. P., Haller, M. N. and Yen, T. F., *Journal of Colloid and Interface Science*, Vol. 29, pp. 475-484, (1969)
- [7] A kinetic investigation of association in asphalt, Ensley, E. K., *Journal of Colloid and Interface Science*, Vol. 53, No. 3, pp. 452-460, 1975
- [8] *Flory-Huggins Theory*, Franzen, S., North Carolina State University: Department of Chemistry <http://chsfpc5.chem.ncsu.edu/~franzen/CH7951/lectures/fht/sld001.htm>, Date accessed 7 September 2009, 2004

- [9] Asphaltenic aggregates are polydisperse oblate cylinders, Gawrys, K. L. and Kilpatrick, P. K., *Journal of Colloid and Interface Science*, Vol. 288, pp. 325-334, 2005
- [10] Influence of composition of paving asphalts on viscosity, viscosity-temperature susceptibility, and durability, Griffen, R. L., Simpson, W. C. and Miles, T. K., *Journal of Chemistry and Engineering Data*, Vol. 4, pp. 349-354, 1959
- [11] *Hansen solubility parameters: a user's handbook*, Hansen, C. M. Boca Raton: CRC Press, pp. 1-7, (2007)
- [12] The asphalt model: results of the SHRP asphalt research program, Jones, D. R. and Kennedy, T. W., *Proceedings of the Conference, Strategic Highways Research Program and Traffic Safety on Two Continents, Part Four*, Gothenberg, 1991
- [13] Models of an asphaltene aggregate and a micelle of the petroleum colloid system, Jovanović, A., *Chemistry and Industry*, Vol. 54, No. 6, pp. 270-275, 2000
- [14] Nature of Asphaltic Substances, Katz, D. L. and Beu, K. E., *Industrial and Engineering Chemistry*, Vol. 37, No. 2, pp.195-200, 1945
- [15] Identification of chemical species and molecular organisation of bitumens, Le Guern, M., Chailleux, E., Farcas, F., Dreessen, S., Planche, J.-P. and Débarre, D., *Proceedings of the RILEM Technical Committee TC 231 workshop on micro- and nano-characterisation and modelling of bituminous materials*, Duebendorf, 2011
- [16] A structure-related model to describe asphalt linear viscoelasticity, Lesueur, D., Gerard, J.-F., Claudy, P., Letoffe, J.-M., Planche, J.-P. and Martin, D., *The Journal of Rheology*, Vol. 40, pp. 813-836, 1996
- [17] Colloid chemistry of asphalts, Mack, C. J., *Journal of the Institute of Petroleum Technologists*, Vol. 36, pp. 2901-2914, 1932
- [18] Modelling of asphaltene and other heavy organic deposits, Mansoori, G. A., *Journal of Petroleum Science and Engineering*, Vol. 17, pp. 101-111, 1997
- [19] Bitumen microstructure by modulated differential scanning calorimetry, Masson, J.-F. and Polomark, G.M., *Thermochimica Acta*, Vol. 374, No. 2, pp. 105-114, 2001
- [20] The colloidal structure of crude oil and the structure of oil reservoirs, Mullins, O. C., Betancourt, S. S., Cribbs, M. E., Dubost, F. X., Creek, J. L., Andrews, A. B. and Venkataramanan, L., *Energy and Fuels*, Vol. 21, pp. 2785-2794, 2007
- [21] The constitution of asphalt, Nellensteyn, F. J., *Journal of the Institute of Petroleum Technology*, Vol. 10, pp. 311-325, 1924
- [22] The effect of asphaltenes, naphthenic acids and polymeric inhibitors on the pour point of paraffin solids, Oliveira, G. E., Mansur, C. R. E., Lucas, E. F., González, G. and de Souza, W. F., *Journal of Dispersion Science and Technology*, Vol. 28, pp. 349-356, 2007
- [23] SANS study of asphaltene aggregation, Overfield, R. E., Sheu, E. Y., Sinha, S. K. and Liang, K. S., *Fuel Science and Technology International*, Vol. 7, Issue 5-6, pp. 611-624, 1989
- [24] Asphalt solidification theory, Pauli, T., Beemer, A. and Miller, J., *Proceedings of the 43rd Peterson Asphalt Research Conference*, Laramie, 2005
- [24] Chemo-mechanics of bituminous materials, Pauli, T., Grimes, W., Boysen, R. and Kringos, N., *Proceedings of the RILEM Technical Committee TC 231 workshop on micro- and nano-characterisation and modelling of bituminous materials*, Duebendorf, 2011
- [25] Chemical composition of asphalt as related to asphalt durability, Peterson, J. C. In T. F. Yen and G. V. Chilingar (Eds.), *Asphaltenes and asphalts*. London: Elsevier Science B.V., pp. 363-399, 2000
- [26] Asphaltic bitumen as colloid system, Pfeiffer, J.P.H. and Saal, R.N.J., *The Journal of Physical Chemistry*, Vol. 44, No. 2, pp. 139-149, 1939
- [27] Asphaltene macrostructure by small angle neutron scattering, Ravey, J. C., Ducouret, G. and Espinat, D., *Fuel*, Volume 67, pp. 1560-1567, 1988
- [28] The structure of asphaltenes in bitumen, Redelius, P., *International Journal of Road Materials and Pavement Design*, Special Issue: EATA 2006, pp. 143-162, 2006
- [29] Hansen solubility parameters of asphalt, bitumen and crude oils, Redelius, P., *Hansen solubility parameters: a user's handbook*. Boca Raton: CRC Press, pp. 151-158, 2007
- [30] *Relation between bitumen chemistry and performance*, Redelius, P., *Proceedings of the RILEM Technical Committee TC 231 workshop on micro- and nano-characterisation and modelling of bituminous materials*, Duebendorf, 2011
- [31] Correlation between bitumen polarity and rheology, Redelius, P. and Soenen, H., *International Journal of Road Materials and Pavement Design*, Vol. 6, No. 3, pp. 385-405, 2005
- [32] *Chemical properties of asphalts and their relationship to pavement performance*, Robertson, R. E., Washington, D.C.: Western Research Institute, Strategic Highways Research Program, National Research Council, 1991
- [33] Development of a performance related chemical model of petroleum asphalt for SHRP, Robertson, R. E., Branthaver, J. F. and Peterson, J. C., *American Chemical Society Division of Fuel Chemistry Preprints*, Vol. 37, Issue 3, pp. 1272-1278, 1992
- [34] Relationship of asphalt properties to chemical constitution, Simpson, W. C., Griffen, R. L. and Miles, T. K., *Journal of Chemical and Engineering Data*, Vol. 6, No. 3, pp. 426-429, 1961
- [35] Microstructure-based identification of bitumen performance, Stangl, K., Jäger, A. and Lackner, R., *International Journal of Road Materials and Pavement Design*, Vol. 5, pp. 111-142, 2004

- [36] Observation of bitumen microstructure changes using scanning electron microscopy, Stulirova, J. and Pospisil, K., *International Journal of Road Materials and Pavement Design*, Vol. 9, No. 4, pp 745-754, 2008
- [37] The colloidal nature of asphalt as shown by its flow properties, Traxler, R. N. and Coombs, C.E., *The Journal of Physical Chemistry*, Vol. 44, pp. 149-165, 1935
- [38] Automated HPLC SAR-AD Separation, Fundamental properties of asphalts and modified asphalts product, J. Schabron, R. Boysen, March 2015
- [39] Asphalt Binder cracking device to reduce low temperature Asphalt pavement cracking: Final report, Federal Highway administration, July 2010
- [40] Chemo- mechanical software, Fundamental properties of asphalts and modified asphalts product, R. Glaser, A. Beemer, T.F. Turner, March 2015
- [41] DSC Studies of Asphalts and Asphalt Components, Turner T.F. and Branthaver, J.F., *Asphalt Science and Technology*, Usmani A.M. ed., New York: Marcel Dekker Inc., pp. 59-101, 1997.
- [42] Estimating Damage Tolerance of Asphalt Binders Using the Linear Amplitude Sweep, AASHTO TP 101-12-UL, Estimating Damage Tolerance of Asphalt Binders Using the Linear Amplitude Sweep), available on the MARC website: <http://uwmarc.wisc.edu/linear-amplitude-sweep>, 2011
- [43] Interpretation of Dynamic Mechanical Test Data for Paving Grade Asphalt, Christensen, D.W. and Anderson, D.A., *Journal of the Association of Asphalt Paving Technologists*, 61, 67-116, 1992
- [44] NF EN 14771: Determination of the flexural creep stiffness: Bending beam rheometer (BBR)

CLAIMS:

What is claimed is:

1. A method for transforming a process or product, comprising the steps of:
 - 5 - assigning linear dependence of a dependent variable on “n” number of independent variables;
 - performing “p” number of observations to obtain “p” number of measurements for each said dependent variable and said independent variables, wherein “p” is less than the sum of “n” + 1;
 - 10 - generating artificial data, using measurement precision, for at least some of said variables;
 - determining statistically significant independent variables, wherein said statistically significant independent variables have a statistically significant impact on said dependent variable, and are fewer in number than “n”;
 - 15 - generating coefficients for each of said statistically significant independent variables;
 - developing a truncated, closed form mathematical relationship according to which said dependent variable linearly depends from only said statistically significant independent variables, wherein said truncated, closed form mathematical relationship yields results that are sufficiently precise;
 - 20 - performing at least one observation to obtain at least one measurement of each of at least said statistically significant independent variables;
 - using said truncated, closed form mathematical relationship, and said at least one measurement of each of said at least said statistically significant independent variables to obtain a dependent variable estimate; and

- using said dependent variable estimate to transform a process or a product from what said process or said product would be without consideration of said dependent variable estimate.

2. A method for transforming a process or product as described in claim 1 wherein said
5 observations are made using an IR instrument or a SAR-AD instrument.

3. A method for transforming a process or product as described in claim 1 wherein said step
of generating artificial data for at least some of said variables comprises the step of generating
artificial data for said dependent variable.

4. A method for transforming a process or product as described in claim 1 wherein said step
10 of generating artificial data for at least some of said variables comprises the step of generating
artificial data for a plurality of said independent variables.

5. A method for transforming a process or product as described in claim 4 wherein said step
of generating artificial data for at least some of said variables comprises the step of generating
artificial data for all "n" of said independent variables.

15 6. A method for transforming a process or product as described in claim 1 wherein said step
of generating artificial data, using measurement precision, for at least some of said variables,
comprises the step of generating artificial data using measurement error distribution information.

7. A method for transforming a process or product as described in claim 1 wherein said step
of generating artificial data comprises the step of artificially generating enough data so that said
20 linear dependence is mathematically tractable.

8. A method for transforming a process or product as described in claim 7 wherein said step
of generating artificial data comprises the step of artificially generating observations such that
the total number of observations, actual and artificial, is equal to or greater than said sum of "n"
+ 1.

25 9. A method for transforming a process or product as described in claim 1 wherein said step
of generating artificial data, using measurement precision, for at least some of said variables
comprises the steps of delineating, for each said independent variables and said dependent

variable, a plurality of ranges centered around a measured variable value; assigning a frequency to each of said ranges according to a known or estimated frequency for each of said ranges; randomly determining a first value within each of said ranges; and generating a plurality of said data for each of said independent variables and said dependent variable according to said
5 frequencies of said ranges for each of said variables.

10. A method for transforming a process or product as described in claim 1 further comprising the step of determining whether any combinations of two or more independent variables have a statistically significant impact on said dependent variable

11. A method for transforming a process or product as described in claim 1 further
10 comprising the step of assessing whether all possible combinations of two independent variables have a statistically significant impact on said dependent variable.

12. A method for transforming a process or product as described in claim 11 further comprising the step of assessing whether all possible combinations of two or more independent variables have a statistically significant impact on said dependent variable.

15 13. A method for transforming a process or product as described in claim 1 wherein said steps are performed in the order shown.

14. A method for transforming a process or product as described in claim 1 wherein said steps are not performed in the order shown.

15. A method for transforming a process or product as described in claim 1 where at least
20 two of said steps are performed simultaneously.

16. A method for transforming a process or product as described in claim 1 wherein said method is at least partially computer implemented.

17. A method for transforming a process or product as described in claim 16 wherein said steps of generating artificial data, determining statistically significant independent variables,
25 generating coefficients, developing a truncated, closed form mathematical relationship, and using said truncated, closed form mathematical relationship are performed through use of a computer.

18. A method for transforming a process or product as described in claim 1 further comprising the step of preconditioning said measurements for at least some of said independent variables.

19. A method for transforming a process or product as described in claim 18 wherein said
5 step of preconditioning comprises the step of consolidating at least some of said independent variables.

20. A method for transforming a process or product as described in claim 1 further comprising the step of preconditioning said measurements for said dependent variable.

21. A method for transforming a process or product as described in claim 1 further
10 comprising the step of consolidating at least some of said independent variables.

22. A method for transforming a process or product as described in claim 1 further comprising the step of determining whether an acceptably low number of said statistically significant independent variables provide sufficiently precise results when measurements thereof are applied in said truncated, closed form mathematical relationship.

23. A method for transforming a process or product as described in claim 1 further
15 comprising the step of determining whether a minimum precision of results corresponds with an acceptably low number of statistically significant independent variables.

24. A method for transforming a process or product as described in claim 1 wherein said independent variables related to a property selected from the group consisting of: temperature,
20 asphaltene percent, asphaltene fraction percentage, IR wave number/length, UV absorbance, spectroscopy, IR spectroscopy, NIR spectroscopy, MIR band intensities, MIR wavelengths, NMR displacement, spectroscopic peak intensity, UV spectroscopy, RAMAN analysis, SAX analysis, SANS analysis, XRay diffraction, composition, elemental analysis, metal content, microscopy and image analysis property, electronic microscopy, image analysis property, optical
25 microscopy and image analysis property, atomic (AFM) microscopy and image analysis property, tomography, MRI, thermal properties, DSC glass transition temperature, crystallinity, TGA weight loss, HP DSC oxidation induction time, oil component fractions, SAR-AD measured properties, WAX-AD measured properties, SARA fractions, SARA indices, AFT

indices, GPC molecular weight, GPC molecular retention times, GPC molecular retention intensities, IEC related property, olefin index, acidity-basicity property, TAN, TBN, elemental analysis property, microscopy and image analysis property, electronic microscopy and image analysis, optical microscopy and image analysis, atomic (AFM) microscopy and image analysis, tomographic property, and MRI.

25. A method for transforming a process or product as described in claim 1 wherein said step of performing “p” number of observations is accomplished at least in part through the use of a method or instrument selected from the group consisting of: IR spectrometer, NIR spectrometer, MIR spectrometer, SAR-AD analyzer, WAD analyzer, any SARA method, DSR, BBR, ABCD, DMA, mechanical test, fouling analyzer, NMR (1H and 13C), GPC / SEC, DSC, IEC and AFT.

26. A method for transforming a process or product as described in claim 1 wherein said step of performing “p” number of observations comprises the step of performing observations of a material selected from the group consisting of: petroleum product, coal product, hydrocarbonaceous material, biomass product, asphalt, bitumen, fuel, medication, dietary supplements, cosmetics, food, and lubricant.

27. A method for transforming a process or product as described in claim 1 wherein said dependent variables relate to a property, phenomenon or parameter selected from the group consisting of crude oil property, petroleum fouling parameter, coking, emulsion ability, stability, emulsion instability, gas/fuel cetane number, gas/fuel octane numbers, lubricant property, anti-wear property, viscosity index, oxidation resistance, fluidity, tribology, asphalt penetration, ring and ball softening point, fraass brittle point, viscosity, modulus, phase angle, superpave properties, DSR, BBR critical temperatures, oxidation resistance short term and long term, material fatigue resistance, brittleness, product formulation, hardness, elasticity, plasticity, deformation, roughness, density, organic or inorganic material oxidation, material weatherability, material durability, material inflammability, explosiveness, carcinogenicity, mutagenicity, metal corrosion, liquid or paste fluidity, thixotropy, viscosity, material density, perfume smell, spraying ability, medication efficiency/effectiveness, product formula, product formulation, fluid viscosity, material hardness and material reflectivity.

28. A method for transforming a process or product as described in claim 1 wherein said step of using said dependent variable estimate to transform a process or product comprises the step of using said dependent variable estimate to transform a process or product selected from the group consisting of: processes relating to durability measured at various aged and unaged aging stages, blending process, blending proportions, blending proportions based on durability, product formulation, additive design, additive amount for addition to hydrocarbon or other product, additive type for addition to hydrocarbon or other product, compatibility and phase separation in asphalt binder and consequences in terms of stability, either for asphalt made of blends from refining bases (residues from straight run distillation, solvent deasphalting airblowing, visbreaking, hydrotreating, cracking or coking units), or for any of those blends further modified with any semi-compatible additives, including but not restricted to polymers, acids, waxes, rubbers, amines, and derivatives, asphalt and petroleum emulsion ability, storability, breaking, coalescence and curing, and any physical properties of these emulsions and their residues after recovery process, asphalt binder and flux aging, short term and long term, with and without UV and moisture (to address both paving and roofing coatings), long term durability and performance of highway and roofing materials, blending properties of asphalts with aged asphalts from recycled paving materials or recycled roofing materials, product formulation, asphalt specification parameters, asphalt binder physical properties, rheological properties in particular, such as complex modulus, phase angle or any combinations or derivatives, properties and performance of asphalt binder, asphalt aggregate mixture or chip seals, asphalt shingles or other industrial applications, reactivity characteristics of petroleum or petroleum derived fractions or materials for various processes including production, heating, distillation, hydrotreating, coking and others, refining an asphalt (or other material) blend/mix; selecting a bitumen therefor; modifying a blend recipe; determining an ingredient amount, fouling characteristics of crude oils in upstream and downstream applications and oil derived materials including fuels and asphalts, investigating and predicting properties of polymers, biological materials, biofuels, asphalt binder sealants, asphalt binder rejuvenators, investigating and predicting properties or effects (whether intended or not) of cosmetics, surfactants, medications and food materials, hydrocarbon, asphalt, any type of oil, petroleum, coal, and biomass products, fuel, medication, dietary supplements, cosmetics, food, lubricants.

29. A method of transforming a product or process comprising the steps of:

- performing at least one observation to obtain measured data that includes at least one measurement of each of independent and dependent variables;

- generating artificial data, using measurement precision, for at least some of said variables, said step of generating artificial data comprising the steps of

5 - delineating, for each of at least some of said independent variables, and for said dependent variable, a plurality of ranges centered around a measured variable value;

- assigning a frequency to each of said ranges according to a known or estimated frequency for each of said ranges;

- randomly determining a first value within each of said ranges; and

10 - generating a plurality of said artificial data for each of said at least some of said independent variables and said dependent variable according to said frequencies of said ranges for each of said at least some of said variables;

said method further comprising the steps of:

15 - using said artificial data and said measured data to determine coefficients of a linear relationship between said at least some of said independent variables and said dependent variable, thereby determining a closed form mathematical relationship between said at least some of said independent variables and said dependent variable;

- performing at least one observation to obtain at least one measurement of said each of said at least some of said independent variables;

20 - using said closed form mathematical relationship, and said at least one measurement of each of at least some of said independent variables, to obtain a dependent variable estimate; and

25 - using said dependent variable estimate to transform a process or a product from what said process or said product would be without consideration of said dependent variable estimate.

30. A method of transforming a product or process as described in claim 29 further comprising the step of determining statistically significant independent variables, wherein said statistically significant independent variables have a statistically significant impact on said dependent variable.

5 31. A method of transforming a product or process as described in claim 30 wherein said at least some of said independent variables comprises said statistically significant independent variables.

32. A method of transforming a product or process as described in claim 29 further comprising the step of consolidating at least some of said independent variables.

10 33. A method of transforming a product or process as described in claim 29 wherein said step of generating artificial data for at least some of said variables comprises the step of generating artificial data for all “n” of said independent variables.

34. A method of transforming a product or process as described in claim 29 wherein said step of generating artificial data comprises the step of artificially generating enough data so that said
15 linear dependence is mathematically tractable.

35. A method of transforming a product or process as described in claim 29 wherein said steps are performed in the order shown.

36. A method of transforming a product or process as described in claim 29 wherein said steps are not performed in the order shown.

20 37. A method of transforming a product or process as described in claim 29 wherein at least two of said steps are performed simultaneously.

38. A method of transforming a product or process as described in claim 29 wherein said method is at least partially computer implemented.

39. A method of transforming a product or process as described in claim 29 further
25 comprising the step of preconditioning said measurement data for said at least some of said independent variables.

40. A method of transforming a product or process as described in claim 39 wherein said step of preconditioning comprises the step of consolidating at least some of said independent variables.

41. A method of transforming a product or process as described in claim 29 further comprising the step of preconditioning said measurements for said dependent variable.

42. A method of transforming a product or process as described in claim 29 wherein said independent variables related to a property selected from the group consisting of: temperature, asphaltene percent, asphaltene fraction percentage, IR wave number/length, UV absorbance, spectroscopy, IR spectroscopy, NIR spectroscopy, MIR band intensities, MIR wavelength, NMR displacement, spectroscopic peak intensity, UV spectroscopy, RAMAN analysis, SAX analysis, SANS analysis, XRay diffraction, elemental analysis, metal content, microscopy and image analysis property, electronic microscopy, image analysis property, optical microscopy and image analysis property, atomic (AFM) microscopy and image analysis property, tomography, MRI, thermal properties, DSC glass transition temperature, crystallinity, TGA weight loss, HP DSC oxidation induction time, oil component fractions, SAR-AD measured properties, WAX-AD measured properties, SARA fractions, SARA indices, AFT indices, GPC molecular weight, GPC molecular retention times, GPC molecular retention intensities, IEC related property, olefin index, acidity-basicity property, TAN, TBN, elemental analysis property, microscopy and image analysis property, electronic microscopy and image analysis, optical microscopy and image analysis, atomic (AFM) microscopy and image analysis, tomographic property, and MRI.

43. A method of transforming a product or process as described in claim 29 wherein said step of performing at least one observation is accomplished at least in part through the use of a method or instrument selected from the group consisting of: IR spectrometer, NIR spectrometer, MIR spectrometer, SAR-AD analyzer, WAD analyzer, any SARA method, DSR, BBR, ABCD, DMA, mechanical test, fouling analyzer, NMR (1H and 13C), GPC / SEC, DSC, IEC and AFT.

44. A method of transforming a product or process as described in claim 29 wherein said step of performing at least one observation comprises the step of performing at least one observation of a material selected from the group consisting of: petroleum product, coal product,

hydrocarbonaceous material, biomass product, asphalt, bitumen, fuel, medication, dietary supplements, cosmetics, food, and lubricant.

45. A method of transforming a product or process as described in claim 29 wherein said dependent variables relate to a property, phenomenon or parameter selected from the group consisting of crude oil property, petroleum fouling parameter, coking, emulsion ability, stability, 5 emulsion instability, gas/fuel cetane number, gas/fuel octane numbers, lubricant property, anti-wear property, viscosity index, oxidation resistance, fluidity, tribology, product formula, product formulation, asphalt penetration, ring and ball softening point, fraass brittle point, viscosity, modulus, phase angle, superpave properties, DSR, BBR critical temperatures, oxidation 10 resistance short term and long term, material fatigue resistance, brittleness, hardness, elasticity, plasticity, deformation, roughness, density, organic or inorganic material oxidation, material weatherability, material durability, material inflammability, explosiveness, carcinogenicity, mutagenicity, metal corrosion, liquid or paste fluidity, thixotropy, viscosity, material density, perfume smell, spraying ability, medication efficiency/effectiveness, fluid viscosity, material 15 hardness and material reflectivity.

46. A method of transforming a product or process as described in claim 29 wherein said step of using said dependent variable estimate to transform a process or product comprises the step of using said dependent variable estimate to transform a process or product selected from the group consisting of: processes relating to durability measured at various aged and unaged aging stages, 20 blending process, blending proportions, blending proportions based on durability, additive design, additive amount for addition to hydrocarbon or other product, additive type for addition to hydrocarbon or other product, compatibility and phase separation in asphalt binder and consequences in terms of stability, either for asphalt made of blends from refining bases (residues from straight run distillation, solvent deasphalting airblowing, visbreaking, 25 hydrotreating, cracking or coking units), or for any of those blends further modified with any semi-compatible additives, including but not restricted to polymers, acids, waxes, rubbers, amines, and derivatives, asphalt and petroleum emulsion ability, storability, breaking, coalescence and curing, and any physical properties of these emulsions and their residues after recovery process, asphalt binder and flux aging, short term and long term, with and without UV 30 and moisture (to address both paving and roofing coatings), long term durability and

performance of highway and roofing materials, blending properties of asphalts with aged asphalts from recycled paving materials or recycled roofing materials, product formulation, asphalt specification parameters, asphalt binder physical properties, rheological properties in particular, such as complex modulus, phase angle or any combinations or derivatives, properties
5 and performance of asphalt binder, asphalt aggregate mixture or chip seals, asphalt shingles or other industrial applications, reactivity characteristics of petroleum or petroleum derived fractions or materials for various processes including production, heating, distillation, hydrotreating, coking and others, refining an asphalt (or other material) blend/mix; selecting a bitumen therefor; modifying a blend recipe; determining an ingredient amount, fouling
10 characteristics of crude oils in upstream and downstream applications and oil derived materials including fuels and asphalts, investigating and predicting properties of polymers, biological materials, biofuels, asphalt binder sealants, asphalt binder rejuvenators, investigating and predicting properties or effects (whether intended or not) of cosmetics, surfactants, medications and food materials, hydrocarbon, asphalt, any type of oil, petroleum, coal, and biomass products,
15 fuel, medication, dietary supplements, cosmetics, food, lubricants.

47. A system for transforming a process or product, comprising the steps of:

- a linear dependence assignment element that assigns a linear dependence of a dependent variable on “n” number of independent variables;
- an observation element that yields “p” number of observations to obtain “p” number of
20 measurements for each said dependent variable and said independent variables, wherein “p” is less than the sum of “n” + 1;
- an artificial data generation element that generates artificial data using measurement precision, for at least some of said variables;
- statistically significant independent variable determiner that determines statistically
25 significant independent variables, wherein said statistically significant independent variables have a statistically significant impact on said dependent variable, and are fewer in number than “n”;

- a coefficients generator that generates coefficients for each of said statistically significant independent variables;
- a truncated, closed form mathematical relationship generator that generates a relationship according to which said dependent variable linearly depends from only said statistically significant independent variables, wherein said truncated, closed form mathematical relationship yields results that are sufficiently precise;
- a dependent variable estimator that uses said relationship, and at least one measurement of each of said at least said statistically significant independent variables to obtain a dependent variable estimate; and
- a transformation of a process or a product from what said process or said product would be without consideration of said dependent variable estimate.

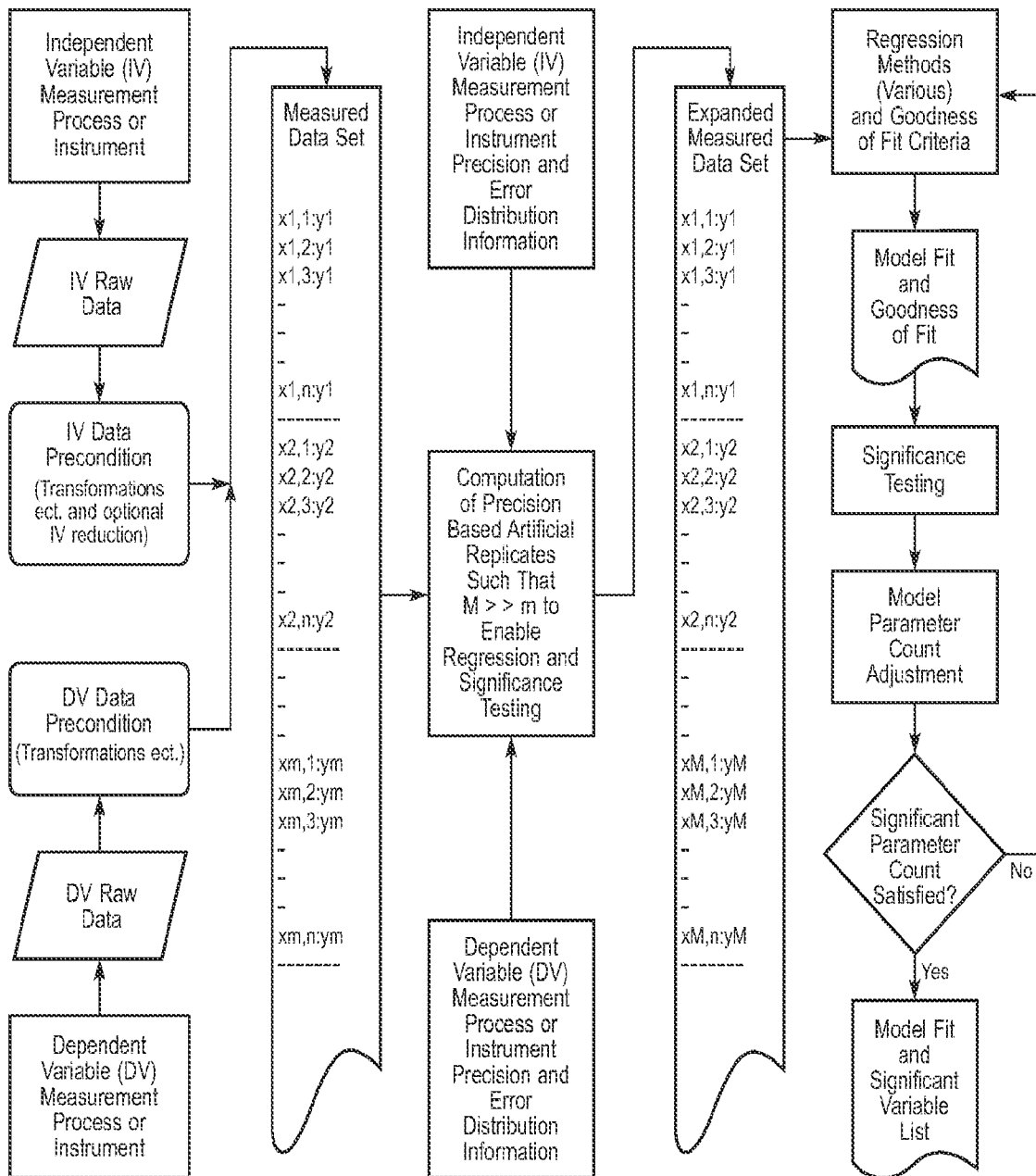


Fig. 1

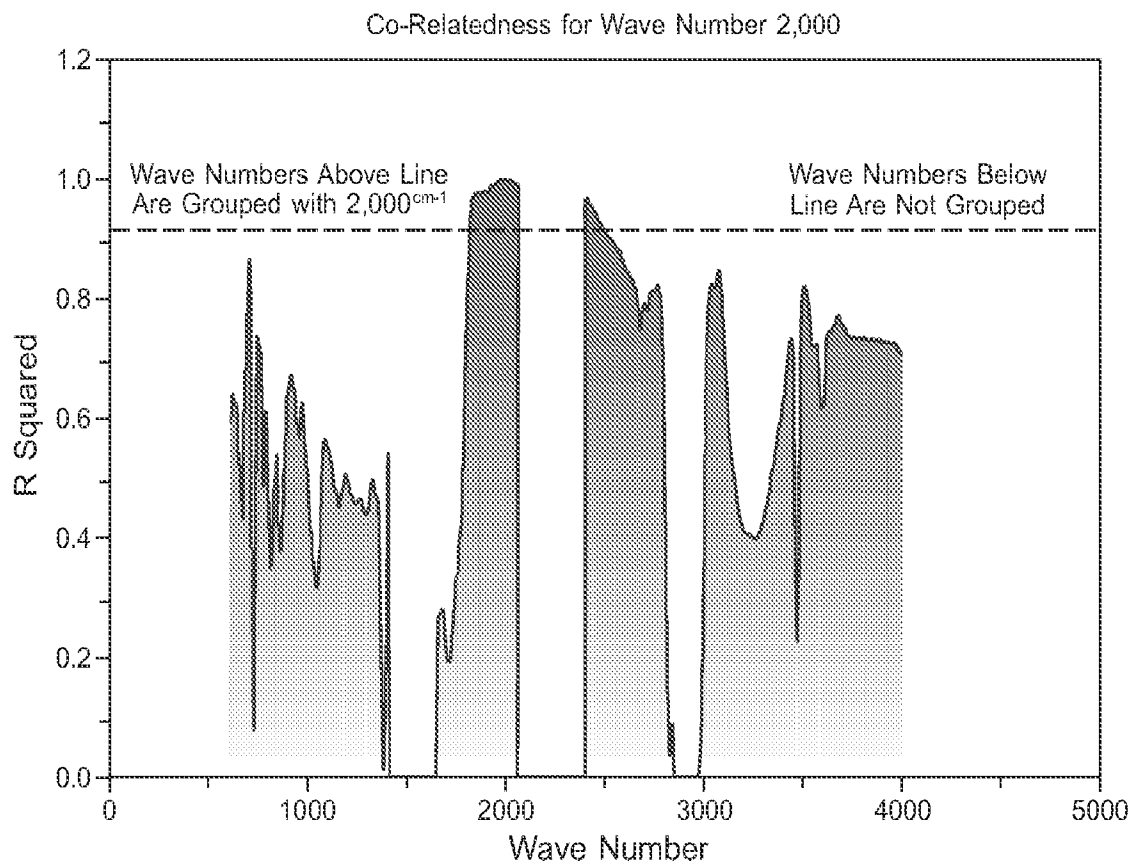


Fig. 2

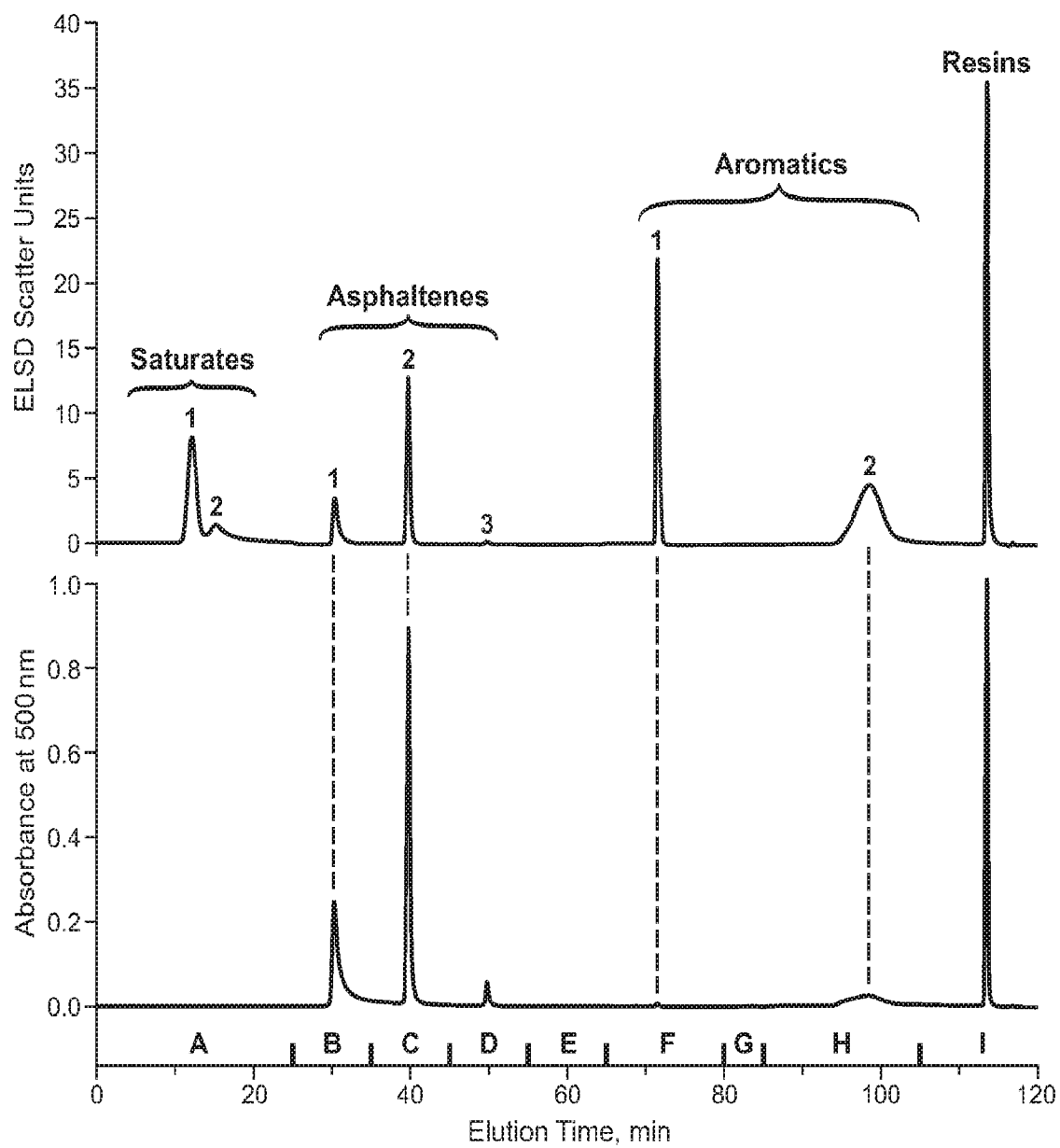


Fig. 3

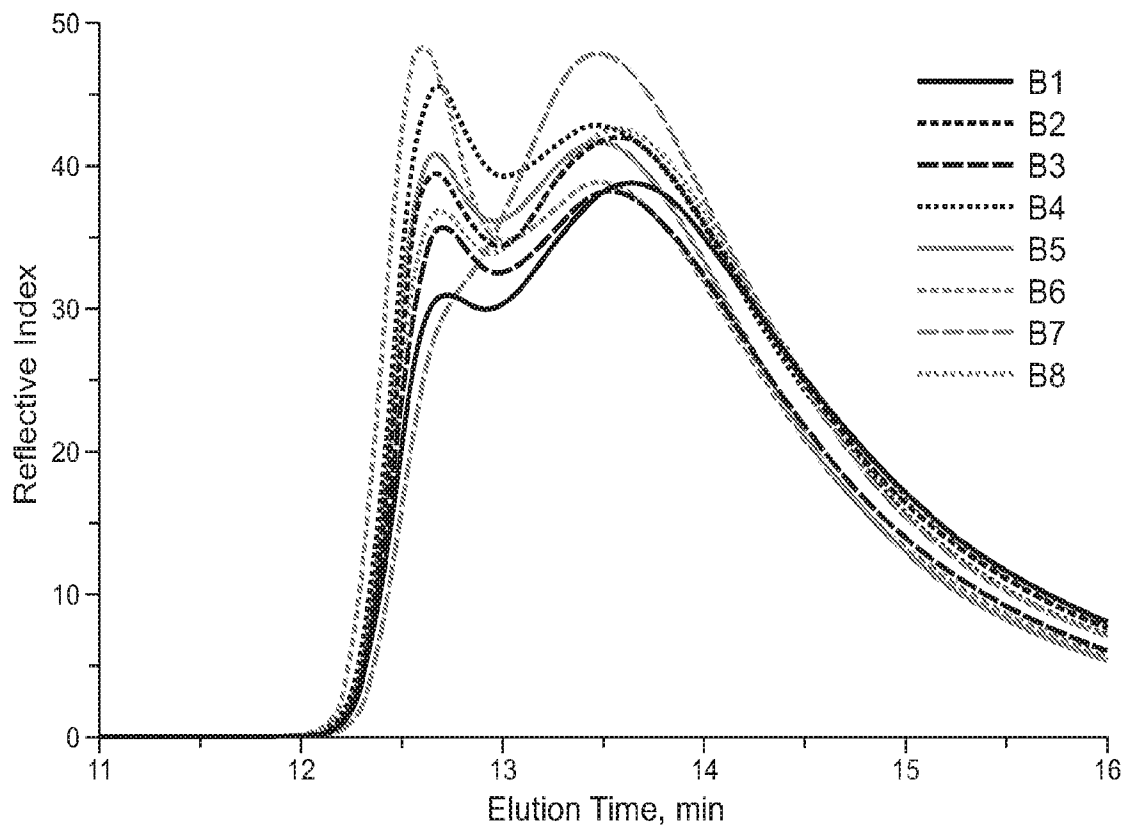


Fig. 4

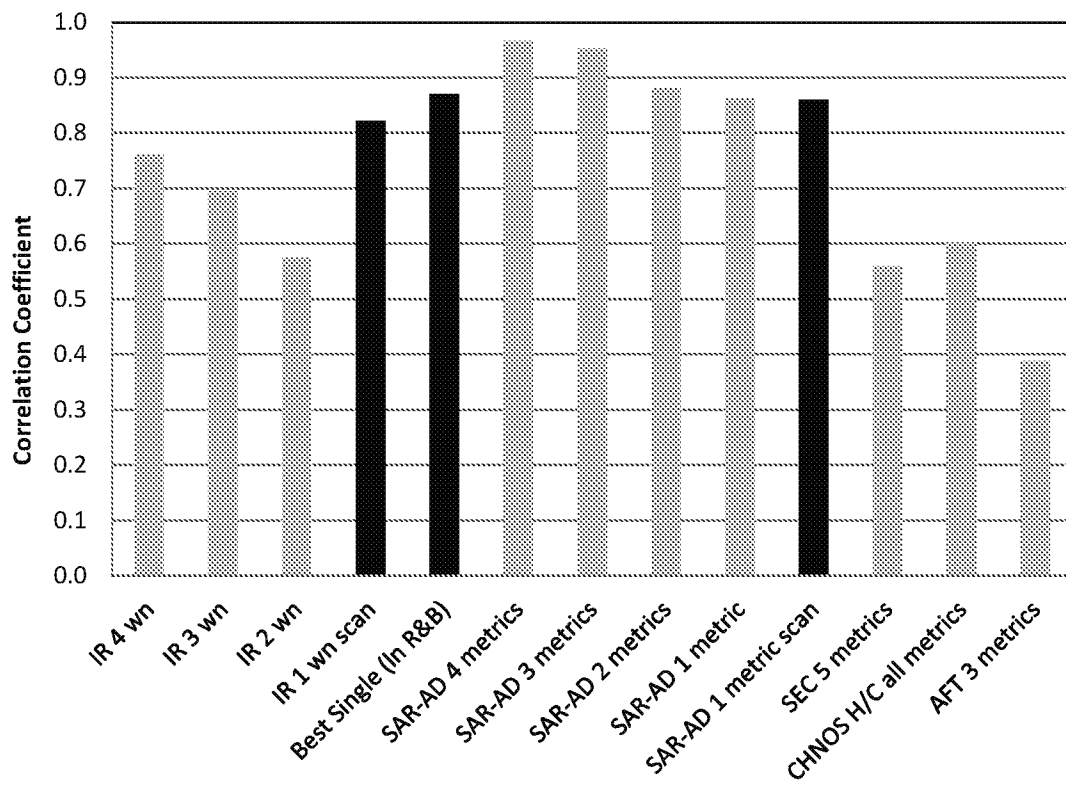


FIG. 5

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2016/041182

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - G05B 13/02, G06F 17/40, G06Q 40/00 (2016.01)

CPC - G05B 13/0205, G06F 19/707, G06Q 40/06 (2016.08)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC - G05B 13/02, G06F 17/40, G06Q 40/00

CPC - G05B 13/0205, G06F 19/707, G06Q 40/06

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

USPC - 700/36, 702/19, 702/22, 702/32, 702/86, 707/748, 707/E17.008, 707/E17.033 (keyword delimited)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Orbit, Google Patents, Google Scholar, Google

Search terms used: transform process, product, assign linear dependence, dependent variable, independent variables

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WRIGHT STATE UNIVERSITY. "Regression Models." 10 May 2012 (10.05.2012) Retrieved from <http://www.wright.edu/~thaddeus.tarpey/es714reg.pdf>, entire document	1-47
Y	US 2005/0173299 A1 (MCADAMS et al) 11 August 2005 (11.08.2005), entire document	1-47
Y	AKRITAS et al. "Linear regression for astronomical data with measurement errors and intrinsic scatter." In: arXiv preprint astro-ph/9605002. 01 May 1996 (01.05.1996) Retrieved from <http://arxiv.org/pdf/astro-ph/9605002.pdf>, entire document	1-47
Y	WO 1994/008225 A1 (ASHLAND OIL, INC.) 14 April 1994 (14.04.1994), entire document	2, 25, 43
Y	US 6,089,078 A (CHELVEDER et al) 18 July 2000 (18.07.2000), entire document	9, 29-46
Y	US 5,170,367 A (MACKAY et al) 08 December 1992 (08.12.1992), entire document	18-21, 32, 39-41
A	US 2012/0221580 A1 (BARNEY) 30 August 2012 (30.08.2012), entire document	1-47
A	US 5,155,677 A (KURTZBERG et al) 13 October 1992 (13.10.1992), entire document	1-47
A	US 6,188,969 B1 (MINOR) 13 February 2001 (13.02.2001), entire document	1-47
A	WO 2015/094545 A1 (MUN) 25 June 2015 (25.06.2015), entire document	1-47

☐ Further documents are listed in the continuation of Box C.☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

02 September 2016

Date of mailing of the international search report

26 SEP 2016

Name and mailing address of the ISA/

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, VA 22313-1450

Facsimile No. 571-273-8300

Authorized officer

Blaine R. Copenheaver

PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774