

Oxidation KInetics Data Analysis

- User Manual -

```
.d88888b.  888    d8P   d8b  8888888b.          88888
d88P" "Y88b 888    d8P   Y8P  888  "Y88b        d88888
888    888 888  d8P           888    888        d88P888
888    888 888d88K         888    888    888        d88P 888
888    888 888 888b       888    888    888        d88P 888
888    888 888  Y88b      888    888    888        d88P 888
Y88b. .d88P 888  Y88b     888    888    d88  d8888888888
"Y88888P"  888  Y88b  888  88888888P"  d88P      88888
```

Written by: Dan Hallatt¹

Faculty of Energy Systems and Nuclear Science
Ontario Tech University

Oshawa, Ontario, Canada

¹Currently a doctoral student at Université de Lille, UMET – Matériaux Terrestres et Planétaires, Villeneuve d'Ascq, France: daniel.hallatt@univ-lille.fr

Table 1. The history of some notable revisions to the code of OKiDA.

Revision No.	Description	Author	Date
1	first version for D. Hallatt NuMat 2018 FeCrAl poster	D. Hallatt	06/10/2018
2	qualifying/debugging with Lipkina et al. FeCrAl data	D. Hallatt	06/06/2019
3	converting data treatment options into functions	D. Hallatt	28/10/2019
4	generalizing for flexible sample sizes/exp. designs	D. Hallatt	03/12/2019
5	adding propagated error data treatment option (<i>Ind Prop</i>)	D. Hallatt	15/01/2020
6	debugging <i>Ind Pooled</i>	D. Hallatt	26/07/2020
7	debugging <i>Ind Prop</i>	D. Hallatt	29/07/2020
8	generalizing for rate law coefficient	D. Hallatt	13/08/2020

Table 2. The history of some notable revisions to this manual.

Revision No.	Description	Author	Date
1	first version with regression modelling options explained	D. Hallatt	13/12/2019
2	adding user-experience flowchart	D. Hallatt	10/06/2020
3	editing descriptions and spell-check	D. Hallatt	26/07/2020

Contents

Abbreviations	v
1 An Introduction to OKiDA	1
2 The Oxidation Kinetics Analyzed by OKiDA	3
3 The Structure of OKiDA	11
4 Handling Repeated Data–Sets in Regression	21
4.1 Independent Regression (<i>Ind</i>)	21
4.2 Independent Regression with Pooled Variance (<i>Ind Pooled</i>)	23
4.3 Independent Regression with Propagated Variance (<i>Ind Prop</i>)	24
4.4 Regression on Data Averaged Between–Group (<i>BGAvg</i>)	25
4.5 Weighted Regression on Data Averaged Between–Group (<i>WtBGAvg</i>)	27
4.6 Regression on All Data (<i>All</i>)	28
4.7 Hierarchical Linear Modelling (<i>HLM</i>)	30
5 How to Use OKiDA	33

Abbreviations

HLM Hierarchical Linear Modelling.

MATLAB MATrix LABoratory.

OKiDA Oxidation KInetics Data Analysis.

OLS Ordinary Least Squares.

TGA ThermoGravimetric Analysis.

WLS Weighted Least Squares.

CHAPTER 1

An Introduction to OKiDA

Oxidation KInetics Data Analysis (OKiDA) automates the study of a material's steady-state isothermal oxidation kinetics from available mass-time data. OKiDA can calculate the rate law coefficient (n), reaction rate coefficient (K_p), and reaction rate intercept (M_o) from the steady-state portion of a single isothermal data-set, while also attempting to deduce and illustrate the transient behaviour from the entire data-set. If a collection of isothermal data-sets are provided at different temperatures, the user may also ask OKiDA to calculate a material's oxidation activation energy (E_a) and Arrhenius pre-exponential factor (A) within the temperature range spanned by the data. As the basis of such analyses is linear regression, OKiDA presents the user with the opportunity to identify which linear regression model (Ordinary Least Squares [OLS], Weighted Least Squares [WLS], Hierarchical Linear Modelling [HLM] , etc.) may be suitably employed depending on the particular experimental design. Accordingly, OKiDA is structured in a way that makes it possible to handle multiple (*repeated*) data-sets at the same temperature. Considering these capabilities, OKiDA is intended to act as a foundation for a stand-alone oxidation data analysis code, quantifying the isothermal oxidation kinetics of a collection of data-sets spanning different and/or the

same temperatures. In this way, OKiDA has also been written to encourage and foster the consideration and expectation of sound statistical analysis as applied to oxidation kinetics measurements. By providing the software with repeat data-sets at a particular temperature, the confidence in an analysis is significantly improved.

OKiDA is a publicly available code (<https://github.com/DanHallatt/OKiDA>), written in MATrix LABoratory (MATLAB), which requires mass-time data in the form of comma-separated values (.csv) files. The mass-time data required by OKiDA is typically generated by ThermoGravimetric Analysis (TGA), where such data is commonly available in .csv format from the software of commercially available instruments. There is also an example data-set made publicly available, which the public version of OKiDA has user-defined inputs pre-defined for. This manual will explain those user-defined inputs whose values will likely be changed by the user for their own data-sets beyond the example data-set supplied at the link above (requiring different temperatures to be specified for example). In this way, the proper file format of the example data-set and the associated user-inputs in the publicly available version of OKiDA give this document a reference which can be learned from, and followed hand-in-hand with this manual, immediately after downloading. Details on OKiDA's required file formats for input data, the software's algorithm structure, the user experience, the approach taken to quantify oxidation kinetics, and the analysis techniques (regression based modelling options) of OKiDA compose the remainder of this document. Descriptions of the fundamental, oxidation kinetics analysis theory that OKiDA is based are not within the scope of this work. Instead, the reader will be referred to the appropriate references that inspired each aspect of OKiDA's analysis.

CHAPTER 2

The Oxidation Kinetics Analyzed by OKiDA

OKiDA is founded on the idea of exploring a material's temporal and thermal mass-change behaviour with a natural progression of basic oxidation kinetics relationships. With the aim of quantifiably characterizing the behaviour of material while subject to oxidizing conditions, OKiDA is mainly concerned with estimating the value of a series of standardized parameters to mathematically describe the behaviour with the known relationships. This progression consists of three main aspects, starting with the determination of the *rate law* (quantified by a parameter n). The rate law can help give insight into the nature of the physical mechanism of oxide film formation and its evolution through time, something vital to know if the specific rate at which a material's mass change is to be quantified at a particular temperature. The rate law of a process describes the general trend in mass-change, whether it be linear, parabolic, etc., that of which can be defined for the material in particular by the *reaction rate* (quantified by two parameters k_p and m_0). The rate of such a particular material's oxidation process is however commonly known to be thermally dependent, and likewise may be described mathematically as an Arrhenius relationship (quantified by the two parameters E_a and A). With the knowledge of the oxidation reaction's Arrhenius parameters, the change

in a material's mass, within that particular chemical environment, can be captured. The way that OKiDA calculates such oxidation kinetics parameters from a set of mass–time data is through the convenient mathematical rearranging of the fundamental oxidation kinetics relationships such that the mathematical relationships can be statistically fit to the available data and the value of the rearranged relationship's fitted parameters related to the standard oxidation parameters used to objectively define the behaviour of the material.

This user–manual does not intend to serve as a source of thorough derivation of any of the mathematical relationships or techniques used by OKiDA. Instead, this document hopes to shed light on the capabilities, intentions, and limitations of the OKiDA code, where high–level discussions on mathematical descriptions or oxidation theory are included as needed. The first insight into the change of a material's mass corresponding to an oxidation process is the proportionality between that change in mass and time as in

$$\Delta m^n \propto t$$

where m is mass, t is time, and the parameter n is known as the *rate law coefficient*. Here, it should be noted, that the change in mass (Δm) is easily defined from a given mass–time data–set by subtracting from each mass the initial mass of the material at $t = 0$ (hereunto stated as t_0). By rearranging such a relationship, a linear approximation of the proportionality may be constructed as in

$$\ln(\Delta m_{s.s.}) \propto n^{-1} \ln(t) \tag{2.1}$$

With the temporal mass change relationship quantified in such a way, the evident role of the rate law coefficient is clear. The rate law coefficient acts as a proportionality constant between a logarithmic change in mass and a logarithmic change in time. In the context of oxidation kinetic measurements, the reaction rate coefficient can be used as a parameter describing the nature of the material's mass change, whether it be linear with time (n being a value of

unity in this case), parabolic ($n = 2$), or something more complex like para-linear, cubic, etc.

If the nature of the proportionality is informed by the rate law coefficient, a differential equation can be inferred to describe the isothermal kinetics of the mass change. For the case of linear and parabolic behaviour ($n = 1$ or 2 respectively), the exact form of the differential equations governing the relation of a steady-state temporal change in a materials mass are respectively

$$\frac{d\Delta m_{s.s.}}{dt} = k_p$$

$$\frac{d\Delta m_{s.s.}}{dt} = \frac{k_p}{\Delta m}$$

where k_p is called the *reaction rate coefficient*. As an ordinary, first-order differential equation, these mathematical descriptions can be integrated to produce the general form of the mass change relationship with time, more clearly either linear or parabolic in nature,

$$\Delta m_{s.s.} = k_p t + C \tag{2.2}$$

$$\Delta m_{s.s.}^2 = k_p t + C$$

With some simple rearrangement, the parabolic relationship can be defined for the mass change itself (not the mass change squared as above), where the integration constant can now also adopt an equivalent physical interpretation to the above linear relationship as in,

$$\Delta m_{s.s.} = k_p^{1/2} t^{1/2} + m_0 \tag{2.3}$$

where m_0 is called the *reaction rate constant*. In this form, the constant of integration can be interpreted as the initial value of the mass change according to the steady-state behaviour (at $t = 0$), hence why it is given the symbol m_0 (as the intersection of the linear relation-

ship with the axis of the independent variable). With such an interpretation, the value of m_0 may be thought to hold a value of zero at first consideration. However, the steady-state assumption of the underlying differential relation means that such an interpretation is commonly incomplete, as it is dependant on the time period considered when relating the reaction rate constant to an observed behaviour by Equation 2.3. In particular, as a result of the formation of an oxide film and its initial growth, early changes to a sample coupon's mass may be influenced by a transient component in addition to the steady-state behaviour that is described by Equation 2.3. From an initial condition of a completely exposed and vulnerable surface, this time period corresponds to the experimental condition of an abrupt change in the conditions during a transition from a protective reducing atmosphere to the oxidizing one of interest. Because the isothermal oxidation behaviour of a material is commonly desired to be studied at elevated temperatures, it is inevitable that a time period of large driving forces for oxidation are experienced when the material is eventually subject to oxidizing conditions at the temperature of interest. A common experimental prescription for an isothermal oxidation measurement is illustrated in Figure 2.1. Here, transition to an environment where isothermal mass-time measurements are to be made in an oxidizing environment (from stage ii to iii) instigates the introduction of transient behaviour where the mass of the material drastically and non-uniformly changes with the transient growth of potentially competing oxide films before a single, thermodynamically dominant oxide covers the surface and likewise controls the subsequent oxidation kinetics (by mechanisms such as self-shielding in the case of parabolic steady-state behaviour). With t_0 defined at this transition point (between stage ii and iii in Figure 2.1), the reaction rate constant m_0 is clearly effected as it is defined as the intercept at the very earliest of times, well within this temporal domain where a material may experience significant transient behaviour during oxide film formation. Relating to Figure 2.1, the oxidation of the material is intended to occur within stage iii, and the associated analysis of mass-time kinetics measurements are required to only consider the steady-state fraction of this temporal domain, specifically starting from a time ($t_{s.s.}$) offset from its beginning (t_0). The presence of a transient oxide film growth in the initial mass-change data is attributable to the discrepancy between the interpretation

of the reaction rate constant (m_o) and the actual mass-change at t_0 which is obviously zero when defining t_0 as the onset of oxidizing conditions. As a result of an initial transient stage, the value of m_0 potentially calculated by an analysis like OKiDA may be larger or smaller than zero depending on the rate of mass-change within the steady-state domain. For samples which somehow already have a formed and developed oxide film, or for an analysis that instead redefines time as starting at a time after steady-state kinetics have been achieved, this discrepancy is not significant.

The following is a theoretical (idealistic) approach which OKiDA uses to estimate the transient-contributions to the behaviour in a data-set, shown for the case of a parabolic oxidation behaviour

$$\begin{aligned}\Delta m_{t.s.} &= \Delta m - \Delta m_{s.s.} \\ &= \Delta m - k_p^{1/2} t^{1/2} + m_0\end{aligned}\tag{2.4}$$

where $\Delta m_{t.s.}$ is the transient mass change and Δm is the raw mass-time data starting at $[0, t_0]$. Because m_0 should be a positive number, the trend of $\Delta m_{t.s.}$ should therefore begin at exactly $[-m_0, t_0]$. In this way the interpretation of the exact value of $m_{t.s.s.}$ is complicated by negative values, as $m_{s.s.}$ is defined by the steady-state behaviour that is potentially offset by the initial transient's $m_0 \neq 0$. Here the relative temporal domain between the steady- and transient-state contributions are possibly able to be graphically illustrated, where $\lim_{t \rightarrow 0} \Delta m_{t.s.} = 0$ can verify the assumption of steady-state behaviour required by all of the aforementioned oxidation kinetics calculations. To make the interpretation of a transient mass change more interpretable, the steady-state behaviour is normalized to an origin of $[m_0 = 0, t_0]$, as if the transient-behaviour does not exist, and is subtracted from the overall mass-change data as in

$$m_{t.s.} \approx \Delta m - k_p^{1/2} t^{1/2}\tag{2.5}$$

Here the estimated transient mass is not interpretable as a change in mass, but hypotheti-

cially as the actual mass itself or the transient-behaviour, where the steady-state assumption is valid when $m_{t.s.}$ is constant.

Regarding the other reaction rate parameter, the coefficient k_p , it is clear that for behaviours which can be described by the single termed proportionality of Equation 2.1, the rate law coefficient in this way clearly describes how the mass of a material reacts after an elapsed time during steady-state behaviour. It is clear that after the temporal decay of the transient-state behaviour (in other words, well within the steady-state regime, at $t \geq t_{s.s.}$), the reaction rate coefficient represents the linear slope of the mass-time data, as described by the relationship in Equation 2.1. If a transient-behaviour is first experienced by the material (and likewise included in the data-set), prediction at $t \geq t_{s.s.}$, based on this reaction rate coefficient, is effectively offset by the non-zero value of m_0 as an initial mass change during transients of film formation.

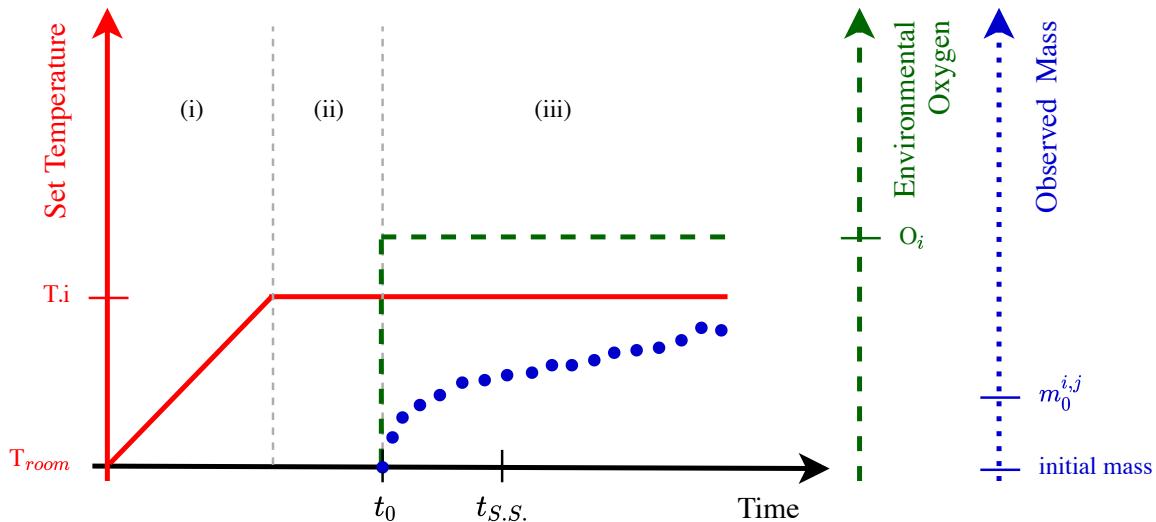


Figure 2.1. (i) is a temperature ramp segment under reducing conditions (ii) being an isothermal hold under reducing conditions (iii) an isothermal hold under oxidizing conditions. Experimental mass-time data is represented by filled, blue circles and the dotted vertical axis, the content of oxygen in the experimental environment is represented by dashed green lines and vertical axis, and the temperature

With the definition of the reaction rate coefficient comes the complication that the rate

of a material's oxidation mass–time behaviour is not only temporally dependent (as described by Equation 2.1 for isothermal conditions) but also dependent on the temperature at which the reaction is occurring. This relationship is commonly represented by the Arrhenius relationship

$$k_p = A \exp^{\frac{-E_a}{RT}}$$

where the steady-state oxidation behaviour is now also defined in terms of a thermal activation energy (E_a) and a pre-exponential factor (A , also interpreted as the frequency), R being the universal gas constant with a value of $8.314 \text{ J mol}^{-1} \text{ K}^{-1}$. This relationship, between the reaction rate coefficient and the temperature of a material, describes that the temporal rate at which a material will change in mass is limited by a fundamental material constant (E_a), which is required to be overcome as an activation energy for a particular fraction of the A frequency of oxidation reactions to successfully oxidize the material. With some rearranging this relationship can also be interpreted in a linear form of

$$\ln(k_p) = \frac{-E_a}{RT} + \ln(A) \quad (2.6)$$

With a progression from the rate law to inform the general nature of an oxidation mechanism, to the reaction rate to quantify the temporal progression of a material's mass change, and concluding with the thermal nature of the material's oxidation rate, Equations 2.1, 2.3, and 2.6 are used by OKiDA to quantify a material's oxidation kinetics by deducing the values of a number of associated parameters.

CHAPTER 3

The Structure of OKiDA

As discussed in Chapter 2, an approach to quantitatively describe the oxidation kinetics of a material is the definition and evaluation of normalized parameters. If the numerical value of the parameters are known, oxidation behaviour can be correlated between differing materials and even characterized in terms of their thermal dependence. In Chapter 2 these parameters were defined to help describe the different relationships between the change in mass of a material with both time and temperature as it experiences a reaction such as oxidation. It is the rearrangement of such relationships into their respective linear forms that forms the framework of the experimental data analysis performed by OKiDA. It is the intention of this chapter to describe the general procedure used by OKiDA to perform such a qualitative analysis, which includes calculation of these oxidation parameters.

When provided with data appropriate to each oxidation kinetics relation in Equations 2.1, 2.3, and 2.6, OKiDA exploits the similarity between the form of the oxidation kinetics rela-

tions and the general equation of a linear line between two variables x and y

$$y = kx + b \quad (3.1)$$

Comparing each of the oxidation kinetics relations to Equation 3.1, it is clear that the oxidation kinetics parameters for not only the rate–law and reaction–rate (n , k_p , and m_o), but also the thermal dependence of the reaction rate (E_a and A), are easily relatable to a line’s slope and its intersection with the independent variable axis (y–axis in Equation 3.1’s case). It is therefore utilized within the structure of OKiDA that if a set of mass–time data are supplied (by experiments, first principle predictions, etc.), the data can be transformed into the appropriate forms ($\ln(m) - \ln(t)$ for the rate law, $m - t^{1/2}$ for the reaction rate, and $\ln(k_p) - T^{-1}$ for the Arrhenius relation) for their trend to be regressed to fit the respective linear relationships finding the best possible description of the material’s behaviour with the use of adjustable parameters (slope and y–intercept). In this way it is clear by comparing the respective kinetics relations to Equation 3.1, that the parameters of interest can be easily determined from the most statistically sound regression parameters. As a result of the linear form of the logarithmic mass–time relation for the rate law for example, if data is recorded on a material’s mass as it changes with time while subject to oxidizing conditions at a particular temperature, their relation can be linearly regressed according to the logarithmic relation in Equation 2.1 where the oxidation kinetics parameter called the rate law coefficient (n) is clearly deducible from the fitted slope coefficient. This same reasoning can be applied to the reaction rate relationship in Equation 2.3 and Arrhenius relationship in Equation 2.6 to produce

$$n = \left(\frac{d \ln(\Delta m)}{d \ln(t)} \right)^{-1} \quad (3.2)$$

$$k_p = \left(\frac{d \Delta m}{d t^{1/2}} \right)^n \quad (3.3)$$

$$E_a = -R \left(\frac{d \ln(k_p)}{dT^{-1}} \right) \quad (3.4)$$

It should be clear why the n and k_p parameters are called *coefficients*, as they represent a fitted parameter in a respective regression model. From a trio of linearly regressed models, all five parameters can be associated with the general fitted parameters m and b that are used to describe a unique, best-approximated relationship between two variables y and x as in Equation 3.1. However a linear regression is exploited for the purpose of quantifying a material's oxidation kinetics, it is important to consider not only the implications and limitations of the many different approaches to linear regression, but also their relation to the experimental design and the relation between the available data-sets. The following discussions in this chapter, along with those related in Chapter 4 on the different approaches to linear regression, are inherently related to the execution and statistical precision of regression, the selection of data for analysis by OKiDA and the associated choice of *regression method* as illustrated in Figure 3.1.

For a collection of data which does not consist of multiple data-sets at any temperature (simply referred to as *repeats* where r_i , the number of repeats at temperature i , holds a value of unity across all m temperatures), utilizing linear regression to determine the rate law and/or reaction rate parameters is as straightforward as described in Chapter 3. In summary, the parameter/s of isothermal oxidation kinetics are defined as a form of the tuneable regression parameters of the underlying relationship (Equations 2.1, 2.3, and 2.6). A benefit of such an approach is that the act of linear regression can also provide estimates on the precision of each fitted parameter's value. Equation 3.5 depicts the form of a linear model constructed for each data-set, none which share a common temperature indexed uniquely by i . However, limited by the highest level of random selection/independence, the capabilities for statistical inference from such an experimental design are severely limited. Without sampling the material- or experimental uncertainty-space beyond a single data-set, it would be inappropriate to deduce the behaviour of other samples of material from the single regression.

$$y_i = k_i x + b_i \quad (3.5)$$

where : $i = 1 \dots m$

To improve the level of statistical inference, repeat data-sets at identical temperatures may be commonly considered by an analysis (j represents the repeat identifier where $r_i > 1$). By analyzing the oxidation behaviour from a number of randomly sampled data-sets at repeat/identical experimental conditions, confidence is improved for projecting the conclusions drawn beyond the local data-sets. However, despite the statistical advantage that the consideration of repeat data-sets introduces, a complication is also realized in the interpretation of repeat data-sets. Instead of a one-to-one temperature-parameter relationship, the collection of repeat data-sets represents numerous individual values associated with the modelled regression parameters (each with their respective uncertainty) at the related temperature. This multi-valued scenario is illustrated in Figure 3.1 (where r_1 data-sets are shown for the first temperature where $i = 1$ and r_m data-sets are considered for the m^{th} temperature) and is also the case in the example data (where three repeat experiments exist at the first and second temperatures as in $r_1 = 3 = r_2$, and two data-sets are included at the third temperature, $r_3 = 2$). Therefore, questions are posed on how the repeat data-sets should ultimately represent the single temperature-parameter relationship required by the Arrhenius Equation 2.6 (a multi-valued function for the Arrhenius is non-physical). Such an environment is managed by double subscripting (i, j) the regression parameters, as portrayed in the following form of a general linear model, to keep track of the accounting of the individual analysis of repeated data-sets and also to distinguish such an analysis on the data-set level from that on the temperature level (where any repeated parameter values/data-sets are reduced by some means)

$$y_{i,j} = k_{i,j} x + b_{i,j} \quad (3.6)$$

where : $i = 1 \dots m$

$$j = 1 \dots r_i$$

To aid in the following discussions it may be fruitful to formulate a concise representation of any *data-set* which the oxidation kinetics analysis of OKiDA can be analyzed from. As such, such a representation can be easily referred to and may help the discussions of different forms of data be visualized. To begin, a general form of a mass-time data-set may be embodied by

$$\begin{bmatrix} T.i & u_i \Delta t & m_{i,j,t_{u_i}} \end{bmatrix} \quad (3.7)$$

$$\text{where : } i = 1 \dots m$$

$$j = 1 \dots r_i$$

$$u_i = 1 \dots z_i$$

$$t_{u_i} = u_i \Delta t$$

where $T.i$ represents a columnar vector consisting of identifiers of the isothermal temperature at which the data-set is affiliated, $m_{i,j,t_{u_i}}$ describes a matrix of mass data for each time increment (t_{u_i}) within each data-set (j) at each temperature (i) (either rank three or four depending on if there are repeats where $r_i \geq 1$), and where u_i is matrix of indices for each data point within each data-set spanning from 1 to the number of data points collected for each data-set represented by z_i (equal to the time span divided by the duration between each measurement Δt). In the simplistic case of a data-set which does not share a common temperature with another (where $r_i = 1$), the general form of the data-set 3.7 can be expanded as in

$$\begin{bmatrix} T.i & 1 \Delta t & m_{i,t_1} \\ T.i & 2 \Delta t & m_{i,t_2} \\ \vdots & \vdots & \vdots \\ T.i & z_i \Delta t & m_{i,t_z} \end{bmatrix} \quad (3.8)$$

The expansion of the non-repeated data-set shows the case where the index j is not needed, by which a linear regression such as Equation 3.5 and its parameters are associated with.

On the other hand, the general form data-set 3.7 can also shown to be expanded for the more universal case of repeat data-sets at a particular temperature $T.i$ as in

$$\begin{bmatrix} T.i & 1 \Delta t & m_{i,1,t_1} & m_{i,2,t_1} & \dots & m_{i,r_i,t_1} \\ T.i & 2 \Delta t & m_{i,1,t_2} & m_{i,2,t_2} & \dots & m_{i,r_i,t_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ T.i & z_i \Delta t & m_{i,1,t_z} & m_{i,2,t_z} & \dots & m_{i,r_i,t_z} \end{bmatrix} \quad (3.9)$$

Coupled with the second column of time, *repeat data-sets* are clearly defined in this form by any additional mass columns beyond the third column (where $j = 1$). It should be clear that the universal data-set, as expanded in the example form of a data series 3.9, reduces to that shown in data-set 3.8 if at $T.i$ there exists no repetition in the data (where no j index is needed). It is also clear that with repeated data-sets at any temperature, the dimension (and related complexity) of an analysis is increased. Whether it be through data reduction, parameter reduction, or a more considerate modelling routine, it is the choice of procedure for considering repeat parameters that is the primary complication when utilizing repeat data.

Figure 3.1 shows a summary of how OKiDA manages the complex environment parameters of a material's oxidation kinetics from mass-time data which may be provided across a series of temperatures (represented by $T.i$ from $T.1$ to $T.m$), each possibly with their own number of replicates (r_i). The first thing to note from Figure 3.1, is that despite the possible presence of repeat data-sets at any number of temperatures ($r_i > 1$ for any i), OKiDA first computes a respective regression model for every data-set's rate law and/or reaction rate by OLS regression in the general form of Equation 3.6. This means that each mass column at a temperature $T.i$, such as in the series of data-sets 3.9, is regressed according to the appropriate rate law and/or reaction rate relationship to produce individual regression models equal in number to the total number of data-sets provided to OKiDA. Impartial to the presence of repeat data-sets, this initial, *sweeping* regression is performed prior to handling any repeated

measurements for three reasons. Firstly, individual analysis of each data-set is valuable to compare the behaviour and similarities between repeat data-sets. In this way, an outlier data-set can be identified early and can therefore be possibly negated in a subsequent running of OKiDA if justified. Obtaining regression parameters on the data-set level (specific to each repeat) is also required by a number of techniques when eventually considering the relation between each data-set and constructing unique parameters for each temperature. Lastly, the immediate OLS regression of each data-set directly produces the final regression parameters for non-repeated data-sets according to Equation 3.5. In other words, data-sets which are not repeated ($r_i = 1$), the double indexed regression parameters of Equation 3.6 reduce to the final model of temperature specific parameters in Equation 3.5. OKiDA does this first sweeping, repeat-careless regression analysis using a custom-built function called *createFit*. The reason for generating and using the *createFit* function is that it amalgamates the stock *fitlm* function's computation of regression models, uncertainty analyses, and residual matrices. For each data-set there are therefore regression coefficients and constants appropriately calculated, along with the respective uncertainty of each (standard deviation and confidence interval). OKiDA constructs plots of each regression model overlaid with the respective data-set, along with an annotation of the fitted equation. A figure consisting of three regression residual analysis plots is also produced for each data-set's regression. The residual analysis consists of a residuals-predictor plot, a lagged residuals plot, and a q-q plot for the user to scrutinize the appropriateness and function of the fitted model. In this individual regression, none of the complicated and numerous user-specified treatments of repeat data-sets, those of which are discussed in Chapter 4, are required.

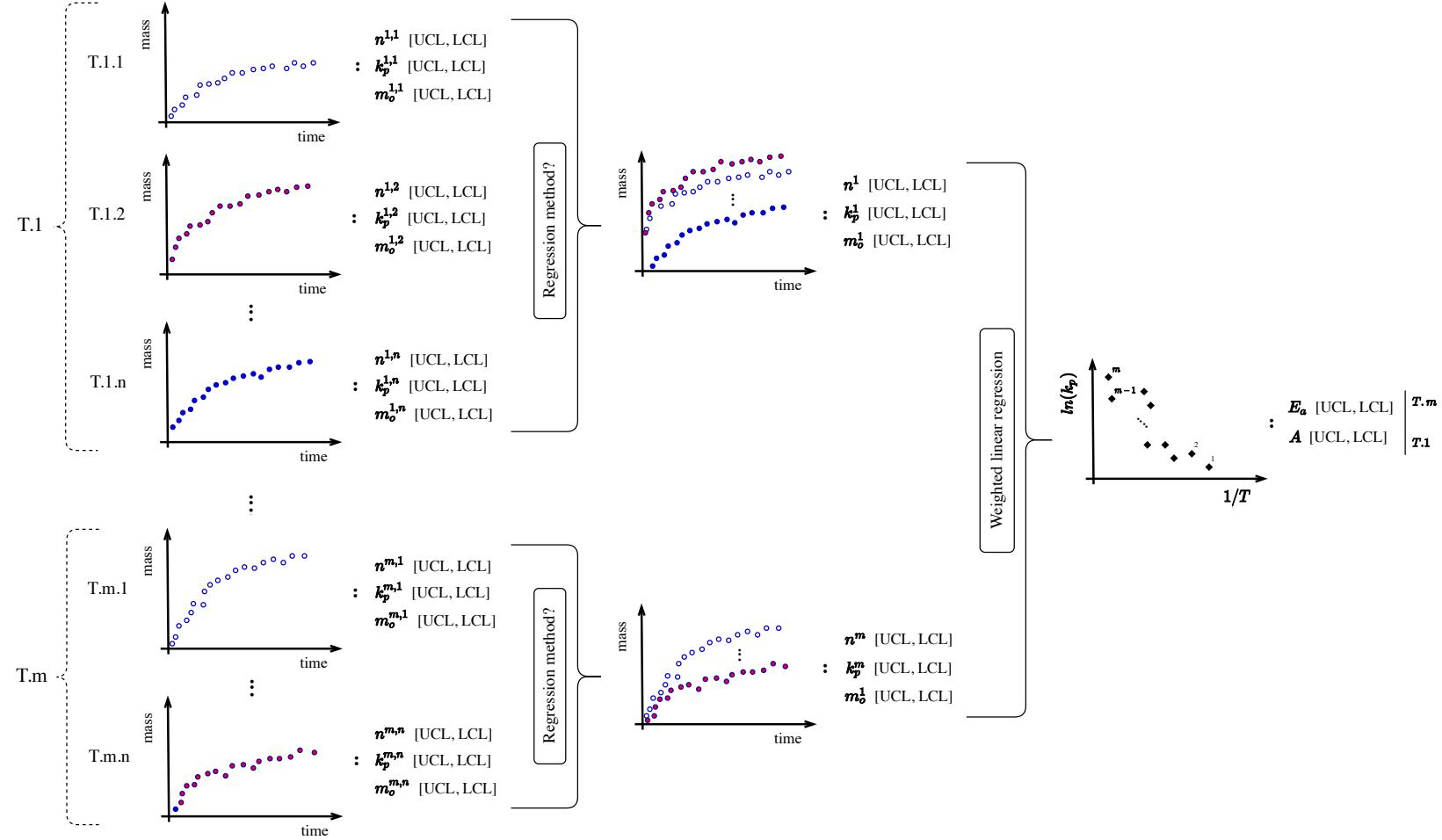


Figure 3.1. A high-level schematic of how OKiDA handles input data-sets. Where input data may consist of repeat measurements at the same temperature ($T.i$, where $i = 1 \dots m$), OLS linear regression may be performed on each data-set indexed by j (as part of a unique data-set identifier $T.i.j$) to quantify the rate law ($n^{i,j}$) and/or reaction rate ($k_p^{i,j}$ and $m_o^{i,j}$). The data-sets that share a common temperature (those with the same i index) are then considered by a specific, user-specified analysis to generate final n_i , k_p^i , and m_o^i values (and their confidence bounds) for each temperature. After the construction of temperature defined reaction rates (k_p^i), their Arrhenius behaviour can then be examined by a linear regression weighted by the confidence in each fitted k_p^i value.

The overarching characterization of a material's oxidation kinetics is its reaction rate's thermal behaviour as described by the Arrhenius relation in Equation 2.6. This is depicted in Figure 3.1, where, as such, the analysis by OKiDA then considers temperatures at which repeated data-sets (distinguished by the dual index i, j) are provided by attempting to reduce, aggregate, or somehow concede their individual data and/or analysis to finally construct unique, representative parameters at each temperature (distinguished only by an index i) according to a user defined treatment. The choice of this parameter/data reduction procedure (from a series indexed by j to a single value i) can be complex and is discussed in more detail in Chapter 4. In either case, regression plots are generated with their fitted equation by OKiDA, along with the same trio of associated residual plots as discussed prior. The opportunity to print the value of the calculated oxidation kinetics parameters and their (95 %) confidence bounds in MATLAB's command window is also certainly available with OKiDA.

With the coupling of the data-set defined temperatures with the calculated reaction rate coefficients, the thermal dependence of the material's oxidation behaviour can be explored by OKiDA as described previously by regression according to Equation 2.6. Independent of the user-defined treatment of repeat data-sets in the construction of representative k_p^i values, the material's fitted Arrhenius relationship and the calculation of the associated oxidation parameters (E_a and A) is performed by Weighted Least Squares (WLS) with the knowledge of each regression points' unique error estimated in the prior step. This is why a sound approach to the repeat data-sets' treatment in calculating the final, representative reaction rate regression at each temperature is important.

CHAPTER 4

Handling Repeated Data–Sets in Regression

One of the strengths of OKiDA is its flexibility in handling experimental designs with repeated measures. As described in Chapter 3, the user-defined treatment of repeated data–sets is an important consideration in the use and interpretation of an analysis by OKiDA. This section aims to describe the options available with the use of OKiDA for generating temperature unique rate law and/or reaction rate parameters, as shown in Figure 3.1. Each following data–treatment option is defined by a long–form name in its section title, along with an associated, italicized *user–prompt* code in brackets that the user should specify (according to the instructions in Chapter 5) if they wish to employ that specific method.

4.1 Independent Regression (*Ind*)

The *Ind* regression option is one of the simplest approaches for handling repeated data–sets within a regression. The main inspiration for the *Ind* method is that, for each temperature that repeat data–sets exist, it exploits the individual regression (be that regarding the rate law or reaction rate of Equations 3.2 or 3.3 respectively) of each mass–time column of

data series 3.9 to generate single, temperature defined parameters. Specifically, using each individual regression, the mean of each of Equation 3.6's associated regression parameter/s (which share a common i index, differing only by j) is computed and simply taken as the final parameter value/s of Equation 3.5. In this way, the *Ind* data treatment *reduces* repeat regression models (Equation 3.5) to a single representation of all of them at that temperature. No data aggregation is performed, and instead, an aggregation of the conclusions from each analysis is conducted. OKiDA then calculates the uncertainty in this mean from the standard deviation of each of the individual, constituent, repeated data-sets' respective regression parameters. The calculations employed by the *Ind* data treatment to reduce a series of repeated data-set's isothermal parameters and their uncertainty (variance) is formally shown respectively as follows

$$Q_i = \bar{Q}_{i,j} \\ = \frac{\sum_{j=1}^{r_i} Q_{i,j}}{r_i} \quad (4.1)$$

$$s_i^2 = \frac{\sum_{j=1}^{r_i} (Q_{i,j} - Q_i)^2}{r_i - 1} \quad (4.2)$$

where Q represents any oxidation kinetics parameter from the linear relations for rate law or the reaction rate (may therefore represent n , k_p , and/or m_0). The parameter r_i is the sample size of repeated data-sets at each temperature (which itself is indexed by i), while $s_{i,j}^2$ is the variance of $Q_{i,j}$ from the individual OLS regression of each repeated data-set in Equation 3.6.

Using the *Ind* method of reducing a series of repeat parameter estimates from their own individual regression, by averaging their values to generate a parameter representative of their common temperature, a pseudo-fitted model can be constructed to resemble that of Equation 3.5. OKiDA then constructs a three-plot figure for the residual analysis of this pseudo-constructed model with respect to all of the (repeat) data at the associated temper-

ature.

It should be clear that for repeat data–sets, no consideration of the individual data–set’s regression parameter/s’ error is made in the final (mean) regression parameter/s at each temperature. As such, a disadvantage of this approach is the extreme reduction in statistical degrees of freedom when calculating the confidence in the final regression parameter/s. A discrepancy is therefore obvious: the sample size of a non-repeated data–set at a particular temperature (the number of all data points minus one: $\sum_{j=1}^{r_i}[n_j - 1]$) would frequently be much larger than that of the final regression parameter/s of repeat data–sets (the number of repeats minus one: $r_i - 1$). The natural advantage of increasing confidence in an inferred estimate by having repeated data–sets is nullified by the fact that the final confidence of the averaged (mean) parameter only considers the number of repeat data–sets and not the number of data–points within each which is likely far larger and contribute to calculating confidence levels on the individual repeated parameter estimates which are eventually not even considered.

4.2 Independent Regression with Pooled Variance (*Ind Pooled*)

The *Ind Pooled* regression option is similiar to the *Ind* option described in Section 4.1; where repeat data–sets exist, the mean of a data series 3.9’s repeated regression parameters (equal in number to the number of columns minus 2, and equal to r_i) is taken at each temperature according to Equation 4.1 in order to construct a pseudo-fitted regression model by the reduction of repeat regression parameters. A residual analysis is likewise then conducted on the pseudo-fitted model of Equation 3.5 in a manner identical to the *Ind* treatment with a three–plot figure. Characteristic to the *Ind Pooled* data treatment, however, is the technique for computing the uncertainty in the value of such a reduced (mean) parameter. Unlike *Ind*, this approach utilizes the uncertainty of each repeat’s regression parameter ($s_{i,j}^2$), which were

already determined when initially fitting the individual data-sets by OLS, to construct an estimate of the uncertainty of the mean parameter value which was reduced from repeat data-sets according to Equation 4.1. By pooling the variance of each appropriate repeat regression parameter as in the following (where $s_{p(i)}^2$ is the pooled variance of the associated regression parameter at temperature i), a more encompassing confidence estimate can be constructed for the mean regression parameters

$$\begin{aligned} s_i^2 &= s_p^2 \\ &= \frac{\sum_{j=1}^{r_i} (n_j - 1)s_{i,j}^2}{\sum_{j=1}^{r_i} (n_j - 1)} \end{aligned} \tag{4.3}$$

This approach, by capturing the significant sample size associated with each data-set, is therefore highly considerate. The *Ind Pooled* approach respects the uncertainty in each repeat's regression parameters used to calculate the mean value of the parameter.

4.3 Independent Regression with Propagated Variance (*Ind Prop*)

A third technique for estimating the uncertainty in reduced regression parameters of repeat data-sets is available with *Ind Prop*. Identical to Section 4.1's and 4.2's *Ind* and *Ind Pooled* OLS regression of individual data-sets, this approach also relies on the reduction of repeat parameters by calculating their mean according to Equation 4.1. Likewise, a pseudo-fitted model is constructed by OKiDA, and an analysis of the model's residuals is generated in a three-plot figure. Similar to *Ind Pooled*, this approach also utilizes the variance of the repeat regression parameters ($s_{i,j}^2$) to generate an estimate for the uncertainty in the repeat parameter's mean, however *Ind Prop* does so by propagating their values. Founded on standard error propagation rules, the error propagation for a mean value (calculated in the

4.4. Regression on Data Averaged Between–Group (*BGAvg*)

form of Equation 4.1 for any parameter n_i , k_p^i , and/or m_0) can be calculated according to

$$s_i^2 = \sqrt{\sum_{j=1}^{r_i} (s_{i,j}^2)^2 \frac{1}{r_i}} \quad (4.4)$$

This approach, similiar to *Ind Pooled*, is also highly considerate as it captures the significant sample size associated with each data-set. The *Ind Prop* approach respects the uncertainty in each repeat's regression parameters used to calculate the mean value of the parameter.

4.4 Regression on Data Averaged Between–Group (*BGAvg*)

Rather than reducing the repeated model parameters that are generated from the initial regression of individual repeat data-sets themselves (as in *Ind*, *Ind Pooled*, and *Ind Prop*), *BGAvg* presents the option of obtaining the unique regression parameters of Equation 3.5 by aggregating the data between each repeated group before performing any sort of regression and/or oxidation kinetics analysis. This technique therefore does not utilize the individual Equation 3.6 regression of each repeat data-sets which OKiDA initially computes by OLS (the r_i regressions of the individual mass columns in the general data series 3.9 go unused). For each collection of repeat data-sets, the *BGAvg* technique instead couples the repeat mass-time data-sets with respect to time and subsequently computes the mean of the repeat data-sets' mass data for each time interval according to

$$\begin{aligned} m_{i,t} &= \bar{m}_{i,j,t} \\ &= \frac{\sum_{j=1}^{r_i} m_{i,j,t}}{r_i} \end{aligned} \quad (4.5)$$

where the individual mass data for each repeat at each time is represented by $m_{i,j,t}$ (where there are r_i data points at each time, equivalent to the number of repeats at temperature

$T.i$), whereas the mean mass data at each time is represented by $m_{i,t}$. In other words, the mass-time data as in the example data series 3.9 is considered row by row, where the mean of each row's mass is calculated ($m_{i,t}$) and associated with each time interval ($t_{u_i} = u_i \Delta t$), resulting in a form analogous to that shown in data-set 3.8 by the coordinate $[t_{u_i}, m_{i,t}]$. In a way, the repeated series within the example data series 3.9 (the additional mass columns) are aggregated to a mass-time vector as in the data-set 3.8.

Upon obtaining the aggregated (mean) data-set, regression by OLS is performed on the mean data-set in order to generate a model representative of the particular temperature's behaviour (whether that be the rate law or the reaction rate relationship in Equations 2.1 and 2.3 respectively) from which the associated parameters (and their uncertainty) are obtained according to Equation 3.5. This technique is therefore given the identifier *BWAvg* to describe that the aggregation is performed between-groups and should not be confused with the subsequent regression on the aggregated data-set that is therefore performed within-group. OKiDA then produces a figure of the reduced data-set overlaid with the OLS regressed model and its fitted equation. Comparing the model to both the reduced data-set and the entire collection of repeat data-sets, OKiDA then also generates two, three-plot residual analysis figures to be interpreted by the user.

Benefitting from the reduction of data itself, rather than the fitted parameters, it is clear that the *BGAvg* treatment of repeat data improves the minimal degrees of freedom considered by the *Ind* approach. Unlike the *Ind Pooled* and *Ind Prop* techniques, which consider the sample size to be the entire collection of isothermal repeat data, *BGAvg* interprets the degrees of freedom to be only that of a single data-set's sample size. In this way, the degrees of freedom of the *BGAvg* approach is somewhere between the small value used by *Ind* and those assumed by *Ind Pooled* and *Ind Prop*. With the use of OLS to generate a regression model for the mean data-set, it is assumed that the error in each mean data point is identical. No changes in the aggregate data's uncertainty are considered as the regression moves through different positions of time (rows in data-set 3.8).

In this way however, *BGAvg* does not utilize all of the information afforded by a series of repeated data–sets (such as insights into its variance). One of the biggest uncertainties regarding this approach is the question and scope of inference that the constructed model can answer. As a result of a *BGAvg* model being constructed from aggregated data, the highest level of (pseudo–) independence is the average data–set, not the sample itself. As such, statistical inference, only being applicable on the highest level of sampling independence, is limited to answering a question about the overall, average trend between the regressed variables, not on the nested level of the sample where the data is not dependent. It is also noted that the assumption of independence is not robust, as the average that each data point represents essentially transfers the relationship between the repeated data–sets prior to being aggregated to one now between each temporal data point.

4.5 Weighted Regression on Data Averaged Between–Group (*WtBGAvg*)

As introduced as an oversight of the *BGAvg* approach, between–group aggregation of repeat data–sets may also provide insight into the uncertainty of each aggregated data point from the distribution of the repeat data points at each time from which the mean data–set was constructed (from the distribution of mass data within each row in the example form of a data series 3.9). As described previously, the *BGAvg* treatment of data constructs a such a data–set by taking the mean of the repeated data at each recorded time interval, and is there also afforded with such information. However, with the use of OLS, *BGAvg* does not incorporate such information into the computation of a regression model and instead assumes homoskedasticity in the data. However, with the knowledge of each aggregated data point’s uncertainty, Weighted Least Squares (WLS) can be used to construct a model sensitive to the confidence in each mean data point. This is the only difference between *BGAvg* and

$WtBGAvg$, where instead of OLS, WLS is utilized to regress within-group a between-group mean data-set of repeat data-sets. The weights associated with each aggregated data point of $WtBGAvg$ are computed as the inverse of the standard deviations of the repeated data points at each respective time interval from which the mean (reduced) data-set was generated. It should be noted that despite the improved confidence afforded by considering the repeated-experiment's variance in the weight of each aggregated data point, the primary inference problems are equally as present as in the $BGAvg$ approach.

4.6 Regression on All Data (*All*)

The simplest approach to handling repeat data-sets is to consider the entire collection of data together as a single, disaggregated data-set. In this way, the entire data-set is non-singular, as it consists of data from both the same experiment and from differing experiments, therefore having multiple recorded mass values at the same time interval. This is summarized by understanding that the considered data-set (at each temperature) may be represented by the example form of a data series 3.9 with a series of data points at the coordinates $[t_{u_i}, m_{i,j,t_{u_i}}]$. Here the time series is shared identically between each repeated data-set (highlighted by the fact that the timestep u_i is not indexed by the repeat index j) however many data points exist at each (one from each repeated data-set), as j spans from 1 to r_i .

At a first consideration, this approach is clearly the one most closely related to a standard linear regression. No data transformations, aggregations, or reductions are performed prior to the *All* analysis, and a regression is performed on each temperature's entire series of raw data-sets that were specified by the user. As a result however, this approach demands the intrinsic strict assumptions associated with basic linear regression. In other words, the *All* approach can be considered a non-nested OLS regression. By considering such implications, it becomes clear that, by definition, a collection of all of the data at a temperature will not meet the independence criteria if repeated data-sets are considered where a nested, multi-level structure of data is present. In other words, heteroskedasticity will be evident in

4.6. Regression on All Data (*All*)

the nested levels of a temperature's amalgamated data-sets. The relationship between some of the data which share a common data-set (a common j index) introduces a multi-level problem when blindly considering all of the data together with the *All* approach. As a result of the inter-dependence between large quantities of the amalgamated *All* data-set (such as those having same j index), there may be trends introduced with the presence of nested groups on the between-group level. Taking the influence of nested data to an extreme, the effects from a series of repeated data-sets may be conceptually most evident if considered in the case where they are aggregated within-group but regressed between group (called here the *BWAvg* but not included in OKiDA because of its clear difference in analysis question than that concerned by oxidation kinetics). Because there is no reason to believe that those introduced trends may not be related or even proportional to the phenomenological behaviour/question of interest to a specific experiment's oxidation kinetics, a regression not accounting for (and negating) nested-levels within each temperature's entire collection of data (the j sample specific level and the i temperature level) may be severely influenced by competing or atleast uncharacterized trends. Instead of answering the question of 'what is the relation between mass-time of a sample coupon of X material?', the *All* approach risks a change to a different question, 'what is the relationship between mass-time between sample coupons of X material?'.

Another view in this regard is that, considering the clear dependence in the entire collection of data, the degrees of freedom should be reasonably less than the entire sample size. To account for the dependence between each repeated data-set's data, one would expect that the degrees of freedom be reduced beyond the total number of dependent data points collected between each repeat. In the *All* approach however, no reduction in the degrees of freedom is employed and the cumulative sample size across all of the temperature's data-sets is taken in the measure of the confidence.

4.7 Hierarchical Linear Modelling (*HLM*)

When providing multiple data-sets at the same temperature to OKiDA for analysis, the associated regressional modelling is susceptible to whatever trends exist between the known and underlying unknown or unconsidered variables depending on how one presents the final analysis with the data. As has been discussed, there are many different ways that a series of repeated data-sets can be handled. The *Ind* approach considers the repeat data-sets to a first approximation by finding the mean of their individual regressional models (the parameters specifically) and estimating the confidence in them simply according to the spread between each of the constituent regressional parameters. Beyond this, *Ind Pooled* and *Ind Prop* responsibly account for the estimated variance in each data-sets fitted model by either pooling or propagating their variances to the mean. On the other hand, the between-group mean of the data can be calculated to form a single, aggregated data-set. A regression technique accounting for the known spread in each aggregated (mean) data-set, and one assuming a constant variance in each aggregated data point, are defined by the *WtBGAvg* and *BGAvg* respectively. The within-group mean of the data can also be regressed, however by definition this has to be regressed between-group and thus strays from the research question of OKiDA and is therefore not considered. There is also the possibility of blindly regressing all of the data together with the *All* treatment, however this is simultaneously susceptible to both between- and within-group trends and is also therefore unlikely to be appropriate in most cases.

A hierarchical linear model is likely the most appropriate approach to considering a series of repeat data-sets in such an analysis. Hierarchical Linear Modelling (HLM) allows for an entire data-set to be considered in a regression, despite consisting of nested groups and their within- and between-group trends, while allowing for a specific level of the multi-level data to be regressed upon all while considering the possible synergetic effects from the other levels. In the case of such an oxidation analysis, the nested, within-group level of an individual sample coupon is of interest to the overarching regression relationships, and any between-group

4.7. Hierarchical Linear Modelling (*HLM*)

trends are considered experiment and/or coupon specific noise. There considerable justification for the use of a HLM approach, yet it is most recommended that the user explores for themselves the residuals and relative differences in regressed models, oxidation parameters, and specifically their confidence bounds to satisfy their own use of a particular data treatment. Resultantly, OKiDA also allows for the user to specify the presence or lack-thereof of a relationship between the fitted slope and y-intercept in an HLM model.

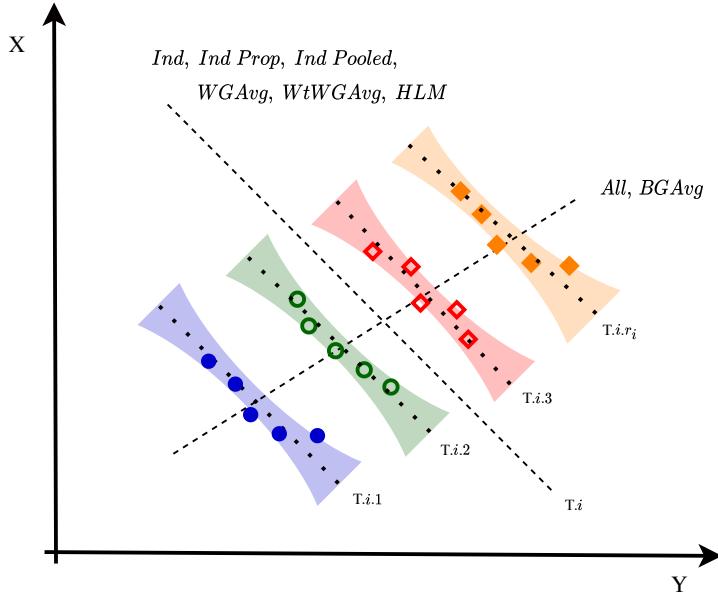


Figure 4.1. A high-level illustration of the handling of regressing repeat data-sets of two arbitrary variables X and Y. The main objective is to consider the repeat data-sets to improve the estimation of the final regression parameters and their confidence bounds. In this way, the individual data-sets and potentially their own regressional models ($T.i.j$) are reduced in some fashion (to be chosen by the user and described in the following sections) to a single regressional model representing $T.i$. Individual data-sets are differentiated by their colour and marker shape (filled circles, annular circles, annular diamonds, and filled diamonds), along with being labelled with their individual indexes $T.i.j$. These individual data-sets are graphical representations of each mass-time coupled column in the example form of the example form of a data series 3.9. The individual regression of each data-set (irregardless of the data treatment option chosen by the user) is represented by the dotted lines, differentiated between each other and affiliated with each data-set by the colour of their confidence bounds. The *All* data treatment is distinctly unlike all other available approaches to handling repeat data-sets in OKiDA. By ignoring all within-group dependencies in a collection of repeated data, *All* is susceptible to answering a between-group question, changing the focus of the statistical inference of the regression analysis away from the phenomenological question of interest to oxidation kinetics of a single material sample. This is exemplified in this illustrative case, where the fitted regression (dashed, black, labelled line) using the *All* approach results in a completely different trend being constructed than the other methods (the other dashed, black, labelled line). The differences between the rest of the available options (*Ind*, *Ind Prop*, *Ind Pooled*, and *HLM*) is not depicted for the sake of clarity and also because some of the predominant differences are in the calculated confidence bounds which are not depicted also for clarity.

CHAPTER 5

How to Use OKiDA

The use of OKiDA is straightforward once the user provides data that is in an acceptable form and alters to a small section of the code itself called the *user-defined inputs*. Upon fulfilling these requirements, which describe the user’s experimental environment, the use of OKiDA is conveyed to the user through a series of prompts within MATLAB’s command window each time the program is run. The user must respond to these user-prompts by typing a specific command into the command window (written between quotations as in: “command”) during associated pauses in the performance of the algorithm. A high-level flowchart of the experience that a user may expect when running OKiDA is shown at the end of this section in Figure 5.11. The following discussion will refer to this user-experience flowchart as a map when describing how to use OKiDA, coupled with screen-shots of the potential outputs from the program.

The first requirement of the user is to ensure that the data that they wish to analyze is in a proper format. The best way to ensure data-readability by OKiDA is to provide data-sets that mimic those provided with the source-code (at <https://github.com/DanHallatt/OKiDA>).

In short, individual comma-separated values files populated with isothermal mass and corresponding time should be provided, where the data is organized into three columns like

<i>Temperature (°C)</i>	<i>Time (min)</i>	<i>Mass (mg or mg/cm²)</i>
#	#	#
#	#	#
#	#	#
...

where the mass is normalized to start at a value of zero to transform the column precisely into *mass change* (each raw mass recording is subtracted by the initial mass of the sample). The data-files must be named according to the following naming convention, where i represents an index of the data-set's temperature (i increasing incrementally from 1 to m) and j is an index for a potentially repeated data-set at a particular temperature (j increasing incrementally by 1 from 1 to r_i)

$$T.i.j \text{ Isothermal Section.csv}$$

After saving each data-set of interest in a folder accessible by the OKiDA program (generally within the same folder is a fail-safe location), the user must address a series of user-defined inputs within the code which are specified for the particular data-set and experimental environment. A screen shot of the actual code that must be altered by the user is shown in Figure 5.1 as the user-defined inputs. Each variable within the code of OKiDA that the user must address is described as follows:

ExpCodes

The first thing that must be specified is a series of experiment identifiers that represent **each data-set**. This variable represents a (row) vector which stores this information, to be populated with comma-separated sequential experiment identifiers according to the convention $i.j$. The particular order/sequence of recording

these indices is easiest if increasing in temperature, starting at the index $i = 1$, and looping in j for each repeated data-set such as the following (just as in the example data where $m = 3$, $r_1 = r_2 = 3$, and $r_m = 2$)

$$[1.1, 1.2, \dots, 1.r_1, \dots, m.1, m.2, \dots, m.r_m]$$

SA

The surface area (in units of squared millimetres) of the sample coupon for each experiment associated with **each data-set** is recorded in this comma-separated vector. This row vector is to be specified in the sequential order of increasing temperature ($i = 1\dots m$) with loops in j for each repeated data-set as in (where ‘S.A.’ simply represents the value of the surface area)

$$[\text{S.A.}_{1,1}, \text{S.A.}_{1,2}, \dots, \text{S.A.}_{1,r_1}, \dots, \text{S.A.}_{m,1}, \text{S.A.}_{m,2}, \dots, \text{S.A.}_{m,r_m}]$$

Temperature

The isothermal temperature (in units of Kelvin) associated with **each data-set** is recorded in this comma-separated vector. This quantity is illustrated in Figure 2.1 which depicts the presence of such a quantity (represented by T_i) within the context of a likely experimental prescription designed observe a material’s oxidation kinetics. This row vector is to be specified in the sequential order of increasing temperature ($i = 1\dots m$) where each repeated data-set (sharing an i value) are defined by the same temperature by definition as in (where ‘T’ simply represents the value of the temperature)

$$[\text{T}_{1,1}, \text{T}_{1,2}, \dots, \text{T}_{1,r_1}, \dots, \text{T}_{m,1}, \text{T}_{m,2}, \dots, \text{T}_{m,r_m}]$$

OxygenFrac

The fraction of oxygen (%) in the experimental environment associated with **each data-set**, is recorded in this comma-separated vector. This quantity is illustrated in Figure 2.1 which depicts the presence of such a quantity (represented

by O_i) within the context of a likely experimental prescription designed observe a material’s oxidation kinetics. This row vector is to be specified in the sequential order of increasing temperature ($i = 1 \dots m$) with loops in j representing each repeated data-set as in (where ‘O’ simply represents the value of the oxygen fraction)

$$[O_{1,1}, O_{1,2}, \dots, O_{1,r_1}, \dots, O_{m,1}, O_{m,2}, \dots, O_{m,r_m}]$$

DayExpPerformed

The date (‘dd-mm-yy’) that the data in **each data-set** was experimentally obtained is recorded in this semi-colon-separated vector. This columnar vector is also to be specified in the sequential order of increasing temperature ($i = 1 \dots m$) with loops in j representing each repeated data-set.

TimeOfDayExpPerformed

The 24-hour time of day (‘hh:mm’) that the data in **each data-set** was experimentally obtained is recorded in this semi-colon-separated vector. This columnar vector is also to be populated in the sequential order of increasing temperature ($i = 1 \dots r$) with loops in j representing each repeated data-set.

SteadyStateStart &

SteadyStateStartInd

The location within the data-set where the analysis will begin must also be specified for each temperature (each i). In this way, the user is specifying a certain amount of time, in the form of a row in the .csv file (not explicitly the *time*, but the row corresponding to a particular time), that must pass to ensure that any initial transient behaviour is avoided and that OKiDA therefore only considers the steady-state regime of the material’s oxidation behaviour. The reasoning behind this consideration was discussed in Section 2, where Figure 2.1 illustrates the location of such a time (represented by $t_{S>S}$) in the reference frame of all the

stages of an oxidation experiment. The user is to state a number of the first rows in each temperature's data-sets which a steady-state is likely to continue after in the given data. Two vectors called *SteadyStateStart* and *SteadyStateStartInd* should be populated with the location of the first data to be analyzed (the new t_0) at **each temperature** and for **each data-set**, respectively, as in the following (where 'S.S.' simply represents the start of the data to ensure steady-state behaviour in each of the .csv data-sets)

$$[\text{S.S.}_{i=1}, \text{S.S.}_{i=2}, \dots, \text{S.S.}_{i=m}]$$

$$[\text{S.S.}_{1,1}, \text{S.S.}_{1,2}, \dots, \text{S.S.}_{1,r_1}, \dots, \text{S.S.}_{m,1}, \text{S.S.}_{m,2}, \dots, \text{S.S.}_{m,r_m}]$$

The length of *SteadyStateStart* will thus will be m , whereas *SteadyStateStartInd* will be $\sum_{i=1}^m r_i$ in length (where r_i is the number of repeated data-sets at each temperature indexed by i).

ExcelDataStart

Similiar to *SteadyStateStart* & *SteadyStateStartInd*, where instead of representing the row location of the data-file to *guarantee* steady-state behaviour afterwards, *ExcelDataStart* simply identifies any initial rows of text (usually describing the experimental details) of **each data-set** which are sometimes output from a TGA device that shall not be included in the analysis. It is assumed that the extent of this *cut-off* in each .csv file is the same for each temperature, as . The comma-separated values in this row vector are to be specified in the sequential order of increasing temperature ($i = 1\dots m$) with loops in j for each repeated data-sets as in (where 'E.S.' simply represents the value of the start of the data in each of the the Excel .csv data-sets)

$$[\text{E.S.}_{1,1}, \text{E.S.}_{1,2}, \dots, \text{E.S.}_{1,r_1}, \dots, \text{E.S.}_{m,1}, \text{E.S.}_{m,2}, \dots, \text{E.S.}_{m,r_m}]$$

RepeatTestAtTemp

The number of repeat data-sets at **each temperature** (also known as r_i) are specified by the user in this variable. The comma-separated contents of the *RepeatTestAtTemp* the (row) vector correspond to increasing temperatures (increasing i index from 1 to m) as in

$$[r_{i=1}, r_{i=2}, \dots, r_{i=m}]$$

DataPointTypes

The vector *DataPointTypes* represents different shapes and colours of data points used by OKiDA to represent different data-sets and/or temperatures throughout the performance of the code, and can be adjusted if the user sees fit. One important point to note is that the number of elements in the *DataPointTypes* vector may be greater than but not less than the number of data-sets analyzed (repeats included). As such, *DataPointTypes* must be extended to potentially include repeated data point types if necessary for large numbers of data-sets. The stock, sequential (from left to right) preferences for data-point markers are as follows

1. blue circle
2. red upward-pointing triangle
3. green square
4. magenta diamond
5. black downward-pointing triangle
6. cyan left-pointing triangle
7. yellow right-pointing triangle

or in MATLAB

$$["ob", "r", "sg", "dm", "kv", "<c", ">y"]$$

IsMassDensity

This allows the user to account for mass data that is not normalized by the sample coupon's surface area. In the instance when a data-set is in the form of mass-density (units of g/cm²), this is reflected in the *IsMassDensity* vector being populated with values of '0'. This is the case for every data-set in the available example data. If the mass data in a particular data-set is not yet normalized to area (in units of grams), the user should reflect this in the comma-separated vector by changing its respective value to '1'. The binary data-type identifiers in the *IsMassDensity* row vector are to be populated in the sequential order of increasing temperature ($i = 1 \dots r$) with loops in j for each repeated data-sets such as, for example (from the example data-set in mass-density and the stock OKiDA code)

[0, 0, 0, 0, 0, 0, 0, 0]

```

% -----
% -----
% ----- **** ----- **** ----- **** ----- **** ----- **** ----- **** -----
% ----- * USER DEFINED INPUTS                               USER DEFINED INPUTS *
% ----- **** ----- **** ----- **** ----- **** ----- **** ----- **** -----
% -----
% ExpCodes = [1.1,1.2,1.3,2.1,2.2,2.3,3.1,3.2]; % The test's codes are: [i,j] where i designates the tem
SA = [2.594946,2.586856,2.469012,2.586856,2.537856,2.519378,2.68956,2.856472];% Surface area ('SA') of
SteadyStateStart = [500,700,500]; % Set value of time (UNITS ARE NOT TIME, BUT ACTUALLY EXCEL ROWS) wh
SteadyStateStartInd = [500,500,500,700,700,700,500,500]; % Set value of time (UNITS ARE NOT TIME, BUT
Temperature = [1473.15,1473.15,1473.15,1623.15,1623.15,1623.15,1673.15,1673.15]; % Temperature codes c
OxygenFrac = [20,20,20,20,20,20,20,20]; % The oxygen fraction (%) in each of the measurement's environ
RepeatTestsAtTemp = [3,3,2]; % The number of repeat experiments ran at each temp. (1st column is T.1.#
ExcelDataStart = [34,37,37]; % What row each temperature data should be started from in the .csv file.
DataPointTypes = ["ob","^r","sg","dm","vk","<c",">y"]; % Defining the hierarchy of preferred data symbol
IsMassDensity = [0,0,0,0,0,0,0,0]; % Data that is manually corrected for Al2O3 (buoyancy) is already i
DayExpPerformed = ['26-Jan-2019';'27-Jan-2019';'29-Jan-2019';'30-Jan-2019';'30-Jan-2019';'31-Jan-2019'
TimeOfDayExpPerformed = {'00:00';'04:03';'04:32';'16:52';'01:45';'21:49';'05:15';'15:40';'02:45'}; % C

% ----- **** ----- **** ----- **** ----- **** ----- **** ----- **** -----
% ----- **** ----- **** ----- **** ----- **** ----- **** ----- **** -----
% -----

```

Figure 5.1. The *user-defined inputs* section of OKiDA. The contents of each variable in this section must be considered and appropriately specified for the user's particular experimental environment.

After addressing and specifying the values of each vector in the user-defined inputs section of the source code, OKiDA can be executed by pressing the *run* button in the MATLAB script. The resulting outputs and user-prompts which follow are illustrated in Figure 5.11. The first event that the user will experience is a prompt asking if they want to calculate the

rate law coefficient for the first temperature which was specified in the user-defined input vector called *Temperature*. The user responds by typing “N” or “Y” (including the quotation marks) into the command window of MATLAB and pressing ‘enter’ on their keyboard. If the user wishes to study the rate law of the material at the temperature indexed by $i = 1$, following entering “Y” into the command window, OKiDA will loop through each of the repeat data-sets available (corresponding to the temperature $T.1.j$ where $j = 1...m_1$) and perform independent OLS linear regression according to Equation 2.1 to not only produce plots of the regression but also the residuals between the data and the fitted linear model. By generating the residual plots of the fit, OKiDA stresses the importance of scrutinizing the functional validity of the calculated linear model with each respective, individual data-set. The first result that may be output from OKiDA is therefore this rate law regression fit and its residuals for the very first data-set $T.1.1$ at the 1st temperature. As explained in Section 4, the oxidation kinetics of such a regression fit (Equation 2.1) is most interested in the calculated coefficient of the regression’s slope with the aim of deducing the rate law coefficient known as $n_{i,1}$ from it according to Equation 3.2.

Before performing a similar rate law analysis of any other data-set, OKiDA then asks the user to specify, from an explicitly stated list within the command window of MATLAB, which regression method the user would like to employ when eventually reducing the repeated rate law analyses (represented by the series of calculated $n_{1,j}$) to a single rate law analysis (represented by n_i) for each temperature. This process is illustrated in Figure 3.1, where a series of three repeated measurements at $T.1$ (with the series $n_{1,j}$ being possibly calculated from the 1st temperature’s repeated data-sets, and where $j = 1...3$ is depicted to illustrate the case of the example data) are reduced to a single analysis representing that temperature (n_1). The method of regression on an entire temperatures’ series of data-sets is to be chosen by the user as it is highly dependent on the experimental design as long as it is from the list of options described in Section 4. Just like every user-prompt from OKiDA, specifying the regression method here is achieved within the command window, where the user types an appropriate command in quotations (such as “HLM”) and presses ‘enter’.

Before OKiDA can proceed to investigate any repeated data-sets at the first temperature (where $j > 1$), or even other temperatures in the case where no repeat data-sets were supplied at T.1, the reaction rate analysis is also initiated in the same manner. The introduction of the reaction rate analysis to the user is done so in a similar way as the rate law perviously described; OKiDA asks the user if they want to calculate the reaction rate of the first temperature, where, if confirmed by “Y” being entered into the command window, OKiDA then also asks for the regression method at the temperature-level to be specified from the same list as that offered for the rate law and as described in Section 4. It should be noted that naturally the analysis of the reaction rate is not dependent on the user answering “Y” to OKiDA performing a rate law analysis, however an eventual prediction of the Arrhenius activation energy of the oxidation process is dependent on the reaction rate analysis being performed. The choice of regression method is also independent between the rate law and reaction rate calculations of OKiDA, hence why the user is asked in both instances for their preference. It is noted however that the author does not find any reasonable circumstance when a different regression method will be appropriate between the two analyses. The goal and output of the very first rate law and/or reaction rate analysis of the first data-set are identified in Figure 5.11 as $n_{i,1}$ and $k_p^{i,1}$ respectively. Additionally, and presumably upon determining the rate law coefficient of the material’s behaviour, OKiDA asks the user to specify the rate law coefficient of the material (that which specifies the reaction rate relationship which the data is modelled in the subsequent analysis). Upon the analysis of rate law and/or reaction rate being performed on the first data-set, OKiDA continues to loop through each repeat data-set at each temperature and generates their respective regression plots and the associated plots of each model’s residuals. The output from the analysis of subsequent repeated data-sets is identified by $n_{i,j>1}$ and $k_p^{i,j>1}$ as in Figure 5.11. An example of the output regression plot from modelling on the level of each data-set is shown for the reaction rate in Figure 5.2, whereas that of the residuals is shown in Figure 5.3. The outputs from rate law regression at the data-set level will produce similar figures.

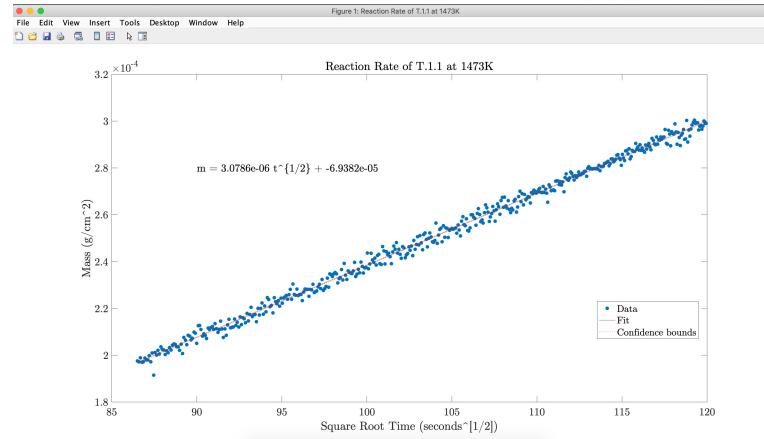


Figure 5.2. An example figure output from OKiDA showing a regression plot of an individual data-set. In the particular case of this figure, the reaction rate regression plot of T.1.1 from the example data is shown. The fitted line is shown as a red, solid line, and the fitted equation is displayed at a location specified by the user within the code of OKiDA (most-likely after performing the analysis once prior [and thus obtaining this figure once before] to determine an appropriate placement of the equation). Confidence bounds on the fitted line are also included as red, dotted lines shoulder the model line to represent 95 % of future repeat measurements. This type of output can be expected from any output from OKiDA corresponding to the superscript 'T' in Figure 5.11.

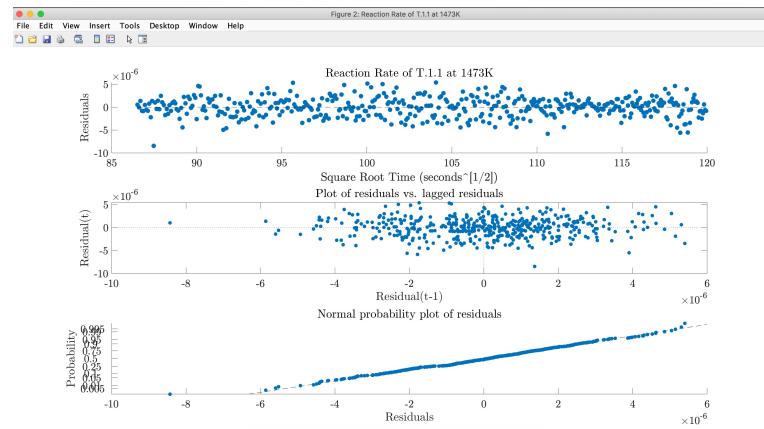


Figure 5.3. An example figure output from OKiDA showing plots of the residuals to assess the functional validity of a particular regression of an individual data-set. In the particular case of this figure, the residuals of a regression for the reaction rate of T.1.1 from the example data is shown. This type of output can be expected from any output from OKiDA corresponding to the superscript 'I' in Figure 5.11.

Upon completing regression of each individual, repeated data-set, OKiDA then obeys the repeat data-handling technique that was specified earlier by the user in order to construct

estimates of the respective oxidation kinetics analysis (whether that be the rate law and/or reaction rate, as specified by the user) in generating unique oxidation parameters representing each temperature. Figure 3.1 illustrates the concept of reducing the analysis of multiple (repeated) data-sets at the same temperature, to produce unique oxidation parameters for the given temperature, done so according to the details of any of the techniques in Section 4. Figure 5.11 shows that at this time there may be a couple of outputs expected. Firstly, for each temperature, the representative isothermal rate law and/or reaction rate analysis generates a plot of the regressed linear model which the associated oxidation parameters n_i and/or k_p^i and m_o^i can be deduced from according to Equations 2.1 and/or 2.3. Figure 5.4 is an example of such a final figure for the reaction rate of the example data at 1473 K. The rate law analysis will also produce a similarly formatted figure. As stressed, the specific statistical approach that is employed by OKiDA to obtain such representative parameters of the temperatures with repeated data-sets is the one specified by the user around the time of analyzing the very first data-set as shown in both Figures 3.1 and 5.11. Additionally, a plot of all the data-set's raw mass-time data is produced for each temperature. This plot distinguishes data from each data-set by both colour and marker shape and also illustrates the location at which the user specified the analysis to begin (the quasi- t_0) in the user-defined input vectors called *SteadyStateStart* and *SteadyStateStartInd*. This value should be chosen from a best guess and iteratively evaluated/adjusted with the help of the residual plots and this raw-data plot for its acceptable position as a ‘safe bet’ to ensure steady-state behaviour of the data considered by OKiDA. Figure 5.5 is an example of the raw-data plot produced by OKiDA from the example data at 1473 K.

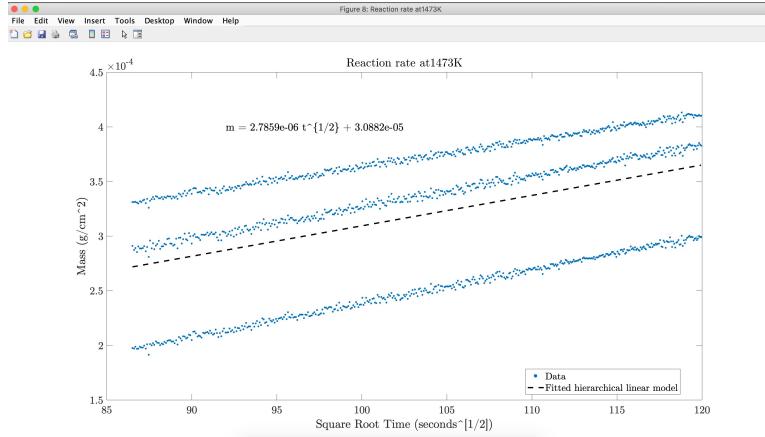


Figure 5.4. An example figure output from OKiDA showing the regression plot of a collection of repeated data-sets at a single temperature. In the particular case of this figure, the reaction rate regression of T.1 was generated by HLM from the example data is shown, which has three repeat data-sets at 1473 K. The fitted line is shown as a black, dotted line, and the fitted equation is displayed at a location specified by the user within the code of OKiDA (most-likely after performing the analysis once prior [and thus obtaining this figure once before] to determine an appropriate placement of the equation). This type of output can be expected from any output from OKiDA corresponding to the superscript 'II' in Figure 5.11. The details of the regression method of repeated data-sets is dependent on the regression method specified by the user.

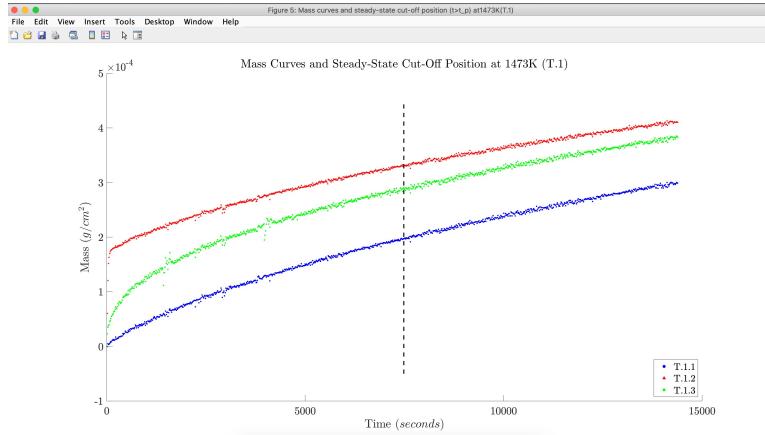


Figure 5.5. An example figure output from OKiDA showing a plot of all of the raw data-sets provided to OKiDA that are repeats at a specific temperature (at T.*i*). In the particular case of this figure and the example data, three repeated data-sets ($j = 1, 2, 3$) at 1473 K (where $i = 1$) were included in the data ($r_1 = 3$). The chosen location of steady-state cut-off is also displayed for the user to visualize the domain of the data being analyzed by OKiDA (all data to the right of the vertical dotted black line). Data is organized by both colour and shape according to their repeat index j . This type of output can be expected from any output from OKiDA corresponding to the superscript 'III' in Figure 5.11.

With the repeat–considerate plots of the rate law and/or reaction rate regression modelling being produced, OKiDA then asks the user if they would like the particular analysis’ parameters displayed in the command window. If the rate law was calculated, the rate law coefficient and its 95% confidence bounds may displayed as n_i [UCL, LCL]. The case is the same for reaction rate’s k_p^i [UCL, LCL] and m_o^i [UCL, LCL]. An example of this output from the example data is shown in Figure 5.6, which demonstrates the expected form of the calculated values of the reaction rate parameters and their confidence bounds at 1473 K, obtained in the example by using a user–specified “HLM” technique with correlated slope and intercept regression method.

```

Do you want to display reaction rate (Kp & Mo) data at 1473K (T.1)? ("Y"/"N")"Y"
The following is the calculated rate law coefficient (Kp) for the temperature:
ans =
1x3 table
    Reaction Rate Coefficient (Kp)      Upper Bound 95%      Lower Bound 95%
    _____          _____          _____
    7.761e-12          9.5972e-12          6.1196e-12

The following is the calculated reaction rate constant (Mo) for the temperature
ans =
1x3 table
    Reaction Rate Constant (Mo)      Upper Bound 95%      Lower Bound 95%
    _____          _____          _____
    3.0882e-05          0.00011954          -5.7774e-05

```

Figure 5.6. An example of printed text output in the command window from OKiDA showing values of the fitted parameters of a regression. This specific figure shows the requested parameter values of the reaction rate (k_p & m_o) at a temperature of 1473 K (T.1), however similiar output of any regression parameters can be expected from any output from OKiDA corresponding to the superscript ‘IV’ in Figure 5.11.

Immediately following the reaction rate analysis of all data–sets at a particular temperature (following each analysis indexed by $j = r_i$), OKiDA automatically calculates the theoretical transient–state component of the temperature’s data–set according to the idealized relation of Equation 2.5. The comparison and legitimacy of the calculated transient–state of each data–set is illustrated in a plot as shown in the example of Figure 5.7, produced from the example data at 1473 K. The calculation of the transient–state behaviour is another way in which the user–defined temporal position of steady–state (within the variables

SteadyStateStart and *SteadyStateStartInd*) can be checked. By Equation 2.5's definition of the calculated transient data, the user should ensure that the transient–curve flattens to a constant value of mass before the displayed steady–state cut–off. If such a case is found, as is the case in Figure 5.7, the user can be more confident that kinetics examined by OKiDA are not influenced by an unaccounted for transient phenomena.

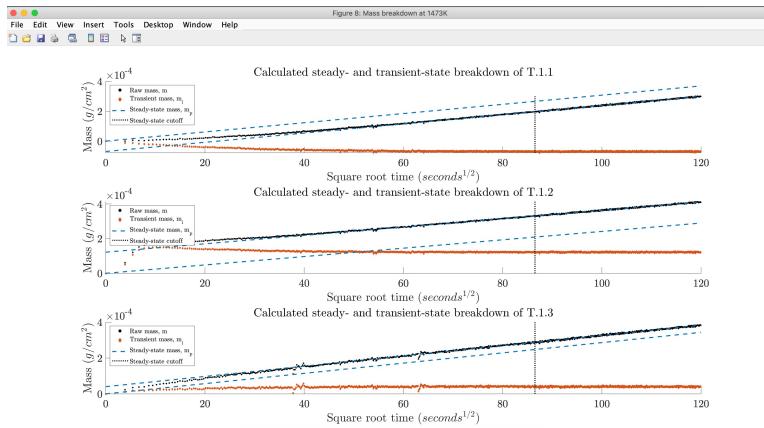


Figure 5.7. An example figure output from OKiDA showing a plot of the steady–state and transient–state calculated breakdown of the oxidation behaviour of all of the raw data–sets provided to OKiDA that are repeats at a specific temperature (at $T.i$). In the particular case of this figure and the example data, three repeated data–sets were included in the data at 1473 K (where $i = 1$), so three plots are included here for T.1.1, T.1.2, and T.1.3. The black circles represent the oxidation data of each data–set considered by OKiDA, the red diamonds represent those calculated from the black circles to be the transient–state data, and the two blue, dashed lines are the calculated steady–state behaviours (one tangent to the raw–data due to an offset of an initial transient mass change and one intersecting the origin). A vertical, dotted, black line is also included to illustrate the position of steady–state cut–off, that of which is defined by the user within OKiDA's user–defined inputs prior to running the program. This type of output can be expected from any output from OKiDA corresponding to the superscript 'V' in Figure 5.11.

Thanks to the user specifying the date and time of day in the user–defined input vectors *DayExpPerformed* and *TimeOfDayExpPerformed* (respectively), OKiDA then automatically constructs an plot of the residuals of each repeat–considerate regression model (both the rate law and reaction rate if chosen) against the environmental factors. An assessment of the experimental environment and the environmental functional validity of each model is thus able to be easily performed by evaluating trends in the pair of produced figures. Figure 5.8 shows an example of OKiDA's environmentally plotted residuals of each of the example

data-set's reaction rate regression.

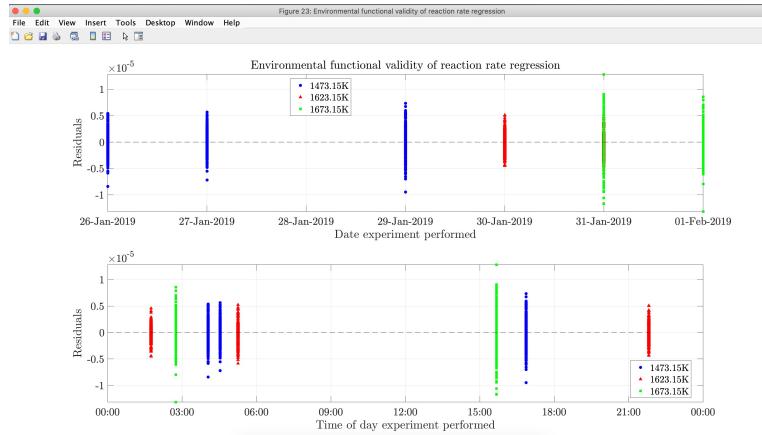


Figure 5.8. An example figure output from OKiDA showing plots of the residuals according to environmental factors for an assessment of a particular regression's environmental functional validity. In the particular case of this figure and the example data, three temperatures ($m = 3$ in Figure 3.1) are spanned by the data (at 1473, 1623, and 1673 K, i.e.: at T.1, T.2, and T.3). Data is organized by both colour and shape according to their temperature and trends in the variability are assessed between the date and time of day. This type of output can be expected from any output from OKiDA corresponding to the superscript 'VI' in Figure 5.11.

The final stage of OKiDA's operation is to provide the possibility of assessing Equation 2.6's thermal Arrhenius behaviour of the reaction rate coefficient ($k_p[T]$) as introduced in Section 2. As a product of calculating reaction rates that span a particular temperature domain from the series of data-sets, OKiDA asks the user if they would like to calculate the Arrhenius parameters E_a and A from a linear regression of each temperature's representative reaction rate coefficient according to the relationship in Equation 2.6. OKiDA performs this linear regression by WLS in a consideration of the calculated confidence bounds associated with each regressed k_p^i value. The value of these confidence bounds, those of the reaction rate coefficient (k_p^i [UCL, LCL]), are highly dependent on the method of regression chosen to reduce repeated data-sets, and thus, so are the weights associated with each point regressed in the Arrhenius analysis. Figures 5.9 and 5.10 are examples from the entire collection of example data-sets (from 1473 to 1673 K) of what the automatically generated Arrhenius fitted plots and the model's residuals may resemble (assuming the user responds to the Arrhenius prompt by typing "Y" and hitting 'enter'). OKiDA concludes by asking the user if they

would like to display the fitted Arrhenius parameters of the material (E_a and A), similar to the conclusion of the rate law (for n_i) and reaction rate (for k_p^i and m_o^i) analysis.

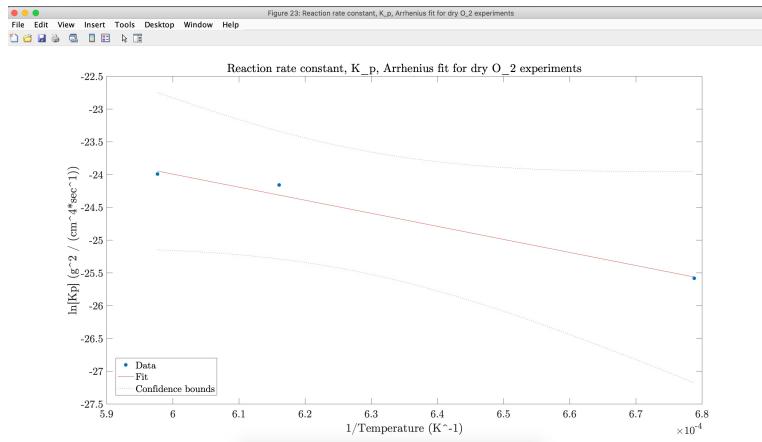


Figure 5.9. An example figure output from OKiDA showing a plot of the Arrhenius regression, fitted by weighted least squares. In the particular case of this figure and the example data, three temperatures ($m = 3$ in Figure 3.1) were spanned by the data (at 1473, 1623, and 1673 K, i.e.: at T.1, T.2, and T.3). The blue, filled circles represent the calculated reaction rate coefficients at each temperature (k_p^i), whereas the red, solid line is the fitted regression line, and the dotted, red curves are the 95% confidence bounds of the linearly fitted model. The weighted regression of the Arrhenius plot considers the confidence (specifically, the inverse of the standard deviation) of each data point (k_p^i [UCL, LCL]) and adjusts the significance of each accordingly. This type of output can be expected from any output from OKiDA corresponding to the superscript ‘VII’ in Figure 5.11.

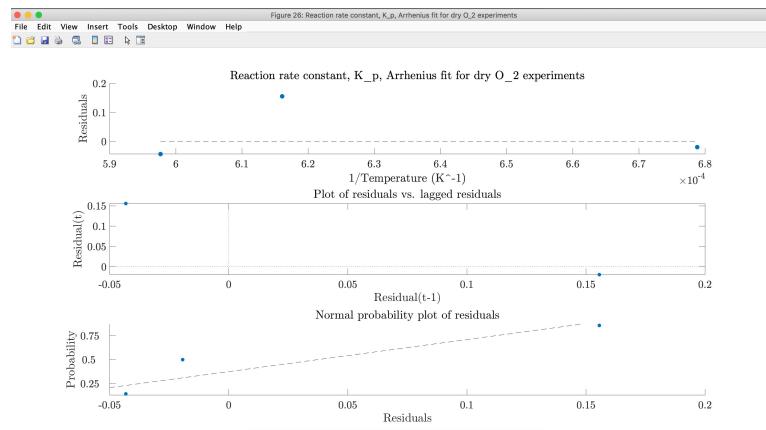


Figure 5.10. An example figure output from OKiDA showing plots of the residuals to assess the functional validity of the Arrhenius regression. In the particular case of this figure and the example data, three temperatures ($m = 3$ in Figure 3.1) were spanned by the data (at 1473, 1623, and 1673 K, and represented by T.1, T.2, and T.3). The blue, filled circles represent the calculated reaction rate coefficients at each temperature (k_p^i), whereas the red, solid line is the fitted regression line, and the dotted, red curves are the 95% confidence bounds of the linearly fitted model. This type of output can be expected from any output from OKiDA corresponding to the superscript ‘VII’ in Figure 5.11.

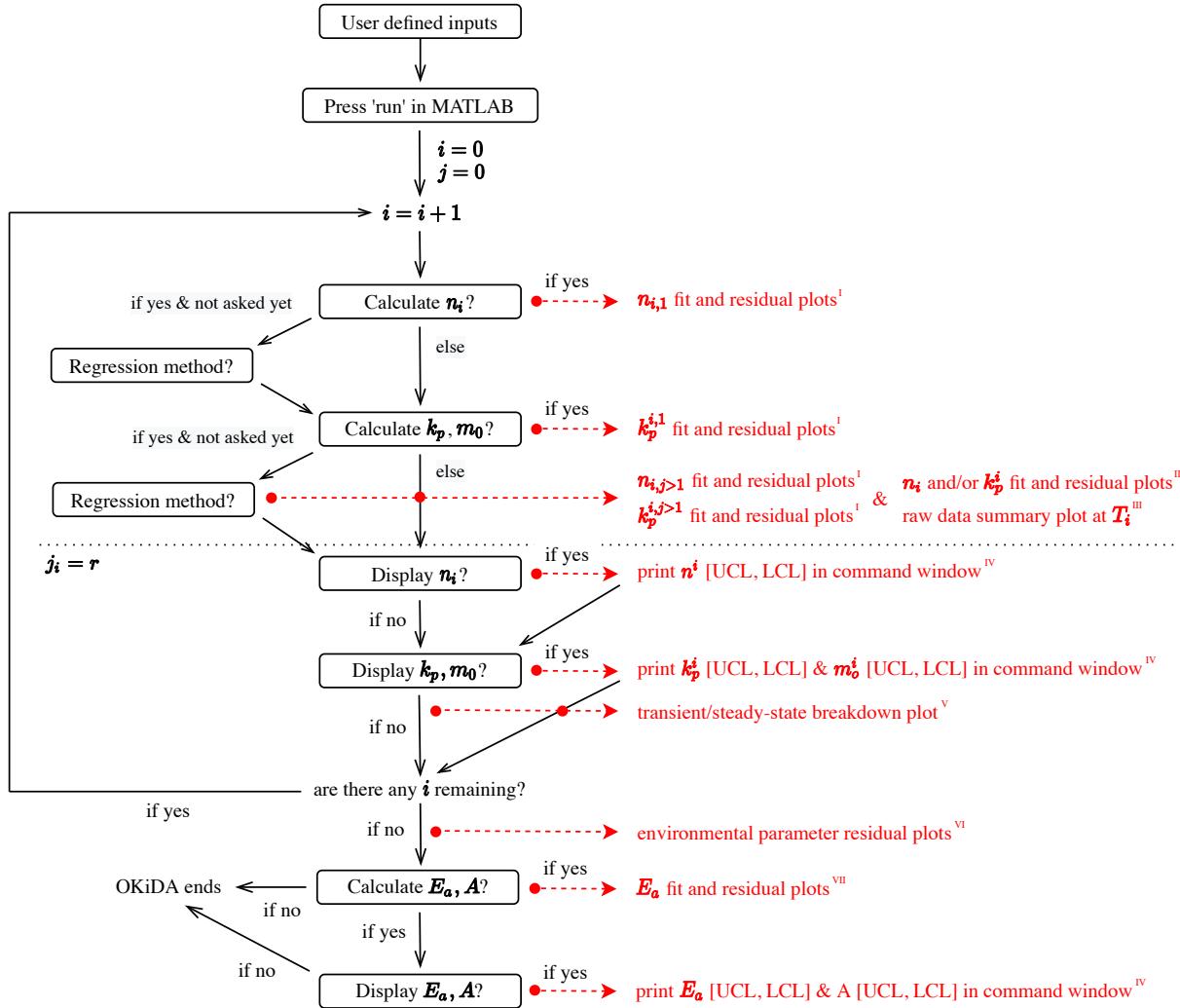


Figure 5.11. A flowchart illustrating the expected experience of a user of OKiDA. Time through the program starts at the top after the user has modified the header of OKiDA’s code with the *User defined inputs* and runs the constructed code in MATLAB. Instances when the user must provide input to OKiDA are represented by events within rectangular boxes. As explained, with the exception of the user defined inputs, the command window describes the possible options that can be chosen at each instance and prompts the user for the input command (to be entered between quotation marks as “*command*”). Possible outputs (dependant on the choice of analysis) are represented by red text and are arrived at at a time represented roughly by red, filled circles with extending red, dotted arrows. Common outputs from the analysis are graphs or printed values from calculations, the former shown as their mean and respective upper and lower 95% confidence bound values (*UCL* and *LCL*). In most of the instances, the analysis output from OKiDA is dependent on the choice of associated regression method. Further explanation of the significance and requirement for a user-specified regression method are discussed in Section 4 and also partially illustrated in Figure 3.1.