Why Cross-Terms Between Temperature and RH are Not Useful

Kenneth C. Waterman, Ph.D.
FreeThink Technologies, Inc.

SUMMARY: *Proposals to add more terms to the modified Arrhenius fitting (with an exponential dependence on reciprocal temperature and water activity) used with ASAPprime® run the risk of over-fitting limited data sets. Over-fitting of accelerated data is particularly dangerous compared to common DOE optimizations since the optimized parameters are used in extrapolated fitting, where small errors can exaggerate behavior. The scientific basis for the moisture-modified Arrhenius equation is reviewed. Any added fitting term also has the disadvantage of reducing the fitting degrees of freedom. Reasons why an apparent cross-term between temperature and RH is observed include carrying out stability studies in packaging rather than open and not following an isoconversion process. The current ASAPprime® method has been shown to be accurate in virtually all systems examined to date (where no phase boundary was crossed).*

BACKGROUND

A common suggestion since the inception of the Accelerated Stability Assessment Program (ASAP) in 2001 (and with introduction of ASAP*prime*® in 2012) is to add another fitting parameter corresponding to a temperature (T) times relative humidity (RH). This can be viewed as part of the more general potential of fitting experimental stability data to functions including more parameters such as the following (only the first three terms are actually used in ASAP*prime*® fitting):

$$lnk = lnA - \frac{E_a}{RT} + B(RH) + C(RH)(T) + DT^2 + E(RH)^2 + \cdots \tag{1}$$

Here, k is the isoconversion rate, A is the collision frequency, E_a is the activation energy, R is the gas constant, B is humidity sensitivity term, C and D are additional fitting parameters. Even more broadly, one could consider both exponential terms (as shown in Eq. 1), and non-exponential terms:

$$k = Ae^{-\frac{E_a}{RT}+B(RH)+CT(RH)} + DT(RH) + ET^2 + F(RH)^2 + \cdots \tag{2}$$

The logic of this general idea is that the fitting will be better with more terms and provide a better optimization. This thinking is in line with analyses of complex systems with multivariate approaches such as principle component analysis (PCA) and factorial analysis as used with design of experiment (DOE) optimizations. With a number of commercial software systems available for such fitting, an optimum mathematical set of coefficients could be determined based on the data. This approach can be very effective for optimizations within a design space independent of the underlying science and "true" behavior. It can also be dangerous to use these optimization processes for projections outside the data range itself. As we show below, when extrapolating outside of the data design space, such analyses can provide very poor predictions, especially for small data sets. Ultimately, any added term will improve fitting, but provide poorer predictions if the terms are not accurately describing the underlying science. In addition, with each added term, there is a corresponding loss of a degree of freedom. This makes the points each have greater leverage and results in greater error bars for any predictions.

SCIENCE OF TEMPERATURE AND RELATIVE HUMIDITY SENSITIVITY

The origin of temperature and relative humidity sensitivity for solid-state chemical kinetics has been described in the previous white papers "Science of Temperature Impact on Degradation Rates" and "Science of Humidity Impact on Degradation Rates in Solids". As discussed in those papers, temperature sensitivity in solid-state (first in the absence of any moisture) is due to a combination of the impact of temperature on the proportion of molecules with adequate energy to overcome an activation barrier for reaction (reflected in the activation energy for the reaction, $E_{a\ rxn}$), and the change in diffusion with temperature (reflected in activation energy for diffusion, $E_{a\ dif}$). The two activation energy terms appear as an additive factor with a functional dependence that is equivalent to the original Arrhenius equation (Eq. 3).

$$lnk = lnA - \frac{E_{a\ rxn}+E_{a\ dif}}{RT} = lnA - \frac{E'_a}{RT} \tag{3}$$

With moisture present, there is an exponential increase in the amount of water present as a function of RH in amorphous regions of solids. Humidity sensitivity on reactivity is attributed to the shift in the amount of active dissolved as a function of the increased amount of moisture condensed as a function of RH.

Three potential factors need to be considered with respect to a potential T-RH cross term: (1) Whether the degree of moisture condensation alters the state of the reactive material thereby changing its temperature sensitivity (activation energy); (2) whether the moisture condensation as a function of RH is temperature dependent; and (3) whether the solubility of the active as a function of temperature needs to be accounted for. We examine each of these potential sources of T and RH interactions below:

(1) ***Moisture Condensation Changing Reaction Activation Energy:*** In general, the amount of material dissolved as a function of moisture condensation will be well below the saturation level. The amount of water condensed even at high RH conditions is generally less than fifty weight percent. For an active to completely dissolve, this is equivalent to a solubility of about 1000 mg/mL, and is very unlikely to be hit by most materials. Because of this, any increase in

condensation will result in a corresponding increase in the amount dissolved as a function of RH; i.e., more material is dissolved (saturated solution) rather than dissolved material becoming more dilute. This means that all the dissolved material is chemically equivalent and will have the same activation energy (temperature dependence) independent of RH; there will simply be more present at higher RH conditions. This factor would not therefore be expected to have any cross-terms for T and RH.

(2) *Change in Moisture Condensation as a Function of Temperature*: Water condensation is always a balance between enthalpy (which favors condensation) and entropy (which favors evaporation). As temperature increases, the entropy term makes more water go into the gas phase. This is true, however, whether there is a solid present or not. Since RH is corrected against the saturated vapor pressure of water at each temperature, condensation as a function of temperature follows the same behavior. When temperature increases, more water goes into the vapor phase; however, more water is in the vapor phase at equilibrium at each RH due to the increase in the saturated vapor pressure. The greater drive for evaporation is exactly balanced by the greater amount of water in the vapor phase. In the end, this means that there is very little impact of temperature on the condensation (sorption isotherm) for amorphous solids.

(3) *Change in Solubility as a Function of Temperature*: The solubilities of the majority of organic solids in water have been found to increase exponentially with temperature (°C).[1] This exponential behavior in °C is very close to exponential in the reciprocal of absolute temperature, as shown in Figure 1.

Figure 1 Aqueous solubility change as a function of temperature for an average organic material reflected in an Arrhenius plot (based on ref. 1).

$y = -3576.3x + 15.161$
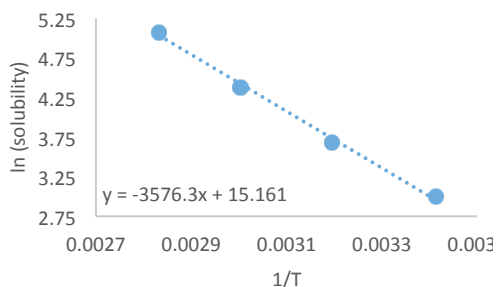
Since this exponential dependency in reciprocal of T is the same functional form as the Arrhenius equation, the overall result is that the temperature impact of solubility adds to the apparent activation energy in a third component term (average solubility change with temperature corresponds to an activation energy of 7 kcal/mol).

Each of the three component terms for exponential dependency on reciprocal temperature (reaction barrier, diffusion rate changes with temperature in solids, solubility changes with temperature) has no dependency on RH. Likewise, the exponential dependency on RH has only temperature terms that separate out in the Arrhenius form: no term requires the two factors be coupled. This all supports a separation of terms in solid-state, stability kinetics between those exponentially dependent on 1/T and those exponentially dependent on RH.

A NOTE ON PACKAGED PRODUCTS

It is important to note that solids in packaging will not face a constant internal environment with respect to RH even when the external conditions remain constant.[2] Moreover, there is a temperature dependence on the change in RH as a function of time due to changes in packaging permeability as a function of temperature. For this reason, when stability modeling experiments are done, they require that the conditions be open to the controlled environment. This also means that there will be an apparent cross term in modeling experiments when the solid is seeing a changing environment due to packaging. Instead, one can determine the explicit T and RH dependence of any degradation using open conditions, then calculate the RH as a function of time inside the package using well-established calculations to determine the shelf-life in packaging (as done with ASAP*prime*®). In fact, because reaction rates can depend significantly on RH, ambient extrapolations to the end of shelf-life (e.g., 6-12 months out to 24 months) can be less accurate than using ASAP*prime*® extrapolation from higher temperature to room temperature, which explicitly takes into account the changing RH as a function of time within the package.

A NOTE ON ISOCONVERSION

Another reason that some stability studies on solids appear to show a cross-term between T and RH is that the kinetics of the degradation is not simple, something common in >50% of solids examined to date. With complicated kinetics, any simple fitting method will give a rate constant that depends on the extent of reaction. With a range of T and RH conditions, this varying extent of reaction can lead to poor fitting to the modified Arrhenius equation. A key to enable the use of the modified Arrhenius equation is the use of isoconversion. Isoconversion changes the focus on stability studies from amount of degradant at a specific time, to the time at each condition to hit the failure point (specification limit).[3] ASAPprime is specifically designed to work with the isoconversion principle. By using isoconversion, we are not bound by the kinetic model or an assumption of linearity.

OVERFITTING AND ITS RISKS

ADDED CROSS-TERMS FOR ASAP

One risk inherent with the ready availability of data fitting software (e.g., Matlab, R, SAS) is a tendency for "overfitting". Overfitting is the use of fitting functions that are overly complex compared to the underlying science. In some sense, overfitting is a violation of Occam's Razor[4] (Solomonoff's theory of inductive inference[5-6]) which states that shorter computable theories have more weight when calculating the probability of the next observation, using all computable theories which perfectly describe previous observations. In other words, Occam's Razor states that while there are an infinite number of possible fittings that can made to a data set, the simplest one that fits the data is preferable. One way to visualize the challenge of overfitting can be seen in a simple example with four points shown in Figure 2. As can be seen, the data can be fit to a linear or higher order functions, with the latter giving progressively better fit. In fact, higher functions will invariably give better fitting which makes using the quality of the fit a poor indicator of the predictive capability of these functions. In this case, with a third order function, there are no degrees of freedom, so the fit will have to be perfect (R^2 is 1.0). As can be seen in this example, the difference between the functions is small within the data set itself, emblematic of the situation for most DOE optimizations. In other words, within the design space, higher order functions may not be more accurate, but there is often only small differences. In the extrapolated area, the differences become very significant. For example, at double the highest value in the data set (at 30), the predicted values for the first, second and third order functions are 65, 94 and 229, respectively. As this example shows, the underlying correct function (if linear for example), will give very different extrapolated predictions than other functions which fit the data more exactly.

Figure 2   Example of the risk of overfitting.   In this case, four points are fit with a line (blue), a second order polynomial (red) and a third order polynomial (green).   The fit becomes progressively better, as expected.   When extrapolating the data, the difference between the fits becomes significant.
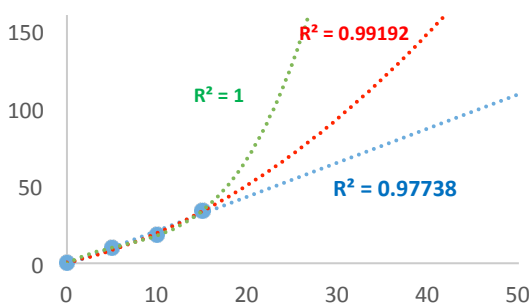
LOOKING FOR CROSS-TERMS

We have looked at a number of to determine whether the RH dependence is truly independent of T, and likewise whether the T dependence is independent of RH. Another way to state this is whether the B term is the same at different fixed temperatures, and whether the E_a term is the same at different fixed RH values. The question to consider is not whether the terms are perfectly independent, but rather, are they the same within the error bars of the experimental measurements. As stated above, these examinations need to be done open to the environment in each case, and be based on finding the failure point (isoconversion). Because of the limited data points typically studied (and the error bars involved), to date, studies appear consistent with there being no obvious additional dependency. Looking at an example product (a tablet) studied at FreeThink, isoconversion times and error bars are shown in Table 1. Figure 3 shows the data at two fixed RH conditions with the corresponding activation energies (activation energies and error bars were calculated using ASAP*prime*®). Figure 4 plots the data at three temperatures showing similar with respect to the error bars humidity effects (B-terms) across the range (error bars were calculated using ASAP*prime*®).

Table 1   Results of a typical stability study of a drug product studied at FreeThink.

T (°C)	RH	Isoconversion time (days)	Std. dev.
50	78	66.9	21.0
61	52	61.7	18.2
61	81	13.1	2.1
70	29	90.2	35.4
70	52	14.1	2.2
70	78	2.1	0.9
80	11	37.3	6.4
80	59	1.9	0.3

Figure 3 Using data from Table 1 at two fixed RH conditions, plot shows the similar slopes (activation energies and error bars calculated using ASAP*prime*®)
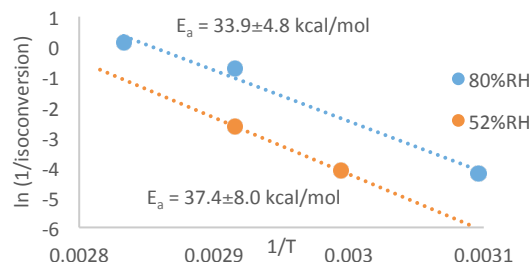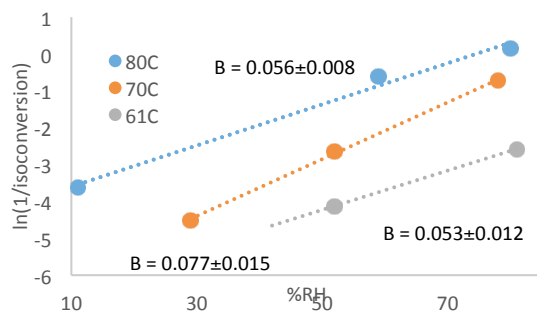

Figure 4 Using data from Table 1 at three fixed T conditions, plot shows the similar slopes (error bars calculated using ASAP*prime*®)

ADDED CROSS-TERMS FOR ASAP

This type of analysis has been conducted on many products. With the noise intrinsic in these experiments, the fitting shows that the current model without a cross-term is appropriate for all examples studied to date. In these examples, adding a cross-term would over-fit the data and result in poorer predictions. Ultimately, the predicting ability of ASAPprime has been borne out with numerous products. While we cannot rule out the presence of subtle added terms, the number of data points typically used and the variability observed make use of more complex fitting schemes of dubious benefit.

CONCLUSION

Proposals to add more terms, such as a cross-term between T and RH, to the moisture-modified Arrhenius fitting (with its exponential dependence on 1/T and RH) used with ASAP*prime*® have the risk of over-fitting limited data sets. Over-fitting of accelerated data is particular dangerous compared to common DOE optimizations since the optimized parameters are used in extrapolated fitting outside the design space, where small errors can exaggerate behavior. The scientific basis for the current fitting is reviewed. Each added fitting term also has the disadvantage of reducing the fitting degrees of freedom. The current method has been shown to be accurate in virtually all systems examined to date, with the exception of cases of phase boundaries crossed in the ASAP design space.

1. Black S; Muller F. On the effect of temperature on aqueous solubility of organic solids. Org. Process Res. Dev. 2010, 14, 661-665.
2. Waterman KC; MacDonald BC. Package selection for moisture protection for solid, oral drug products. J. Pharm. Sci. 2010, 99, 4437-4452.
3. Waterman KC; Swanson JT; Lippold BL. A scientific and statistical analysis of accelerated aging for pharmaceuticals. Part 1: Accuracy of fitting methods. J. Pharm. Sci. 2014, 103, 3000-3006.
4. Soklakov AN. Occam's Razor as a formal basis for a physical theory. Foundations of Physics Letters, 2002, 1-29.
5. Solomonoff R. A formal theory of inductive inference, Part I. Information and Control, 1964, 7, 1–22.
6. Solomonoff R. A formal theory of inductive inference, Part II. Information and Control, 1964, 7, 224–254.