Homoscedastic and Heteroscedastic response in Antoine's Equation Optimal Designs

Design and Analysis of Experiments (2021), 14^{th} *October 2021*

de la Calle-Arroyo, Carlos - Universidad de Castilla-La Mancha, Escuela de Ingeniería Industrial
carlos.callearroyo@uclm.es
López-Fidalgo, Jesús - Universidad de Navarra, DATAI
Rodríguez-Aragón, Licesio - Universidad de Castilla-La Mancha, Escuela de Ingeniería Industrial

Abstract

Vapor pressure is a temperature-dependent characteristic of pure liquids, and also of their mixtures. This thermodynamic property can be characterized through a wide range of models. Antoine's equation stands out among them for its simplicity and precision. Its parameters are estimated via maximum likelihood with experimental data. Once the parameters of the equation have been estimated, vapor pressures between known values of the curve can be interpolated. Other physical properties such as heat of vaporization can be predicted as well.

The probability distribution of a physical phenomenon is often hard to know in advance, as it depends on the phenomenon itself as well as the procedures to carry on the experiments and the measurements. Hence, assuming a probability distribution for such events has to be done with caution, as it affects the Fisher Information Matrix and consequently the optimal designs. This work presents D-, Ds-, A- and I-optimal designs to estimate the unknown parameters of the Antoine's equation as accurately as possible for homoscedastic and heteroscedastic normal distribution of the response, with the characteristic objectives of the different criteria.

An online tool to calculate Antoine's optimal designs for the criteria included in this work has been developed.

Antoine's Equation

The Antoine's Equation is a class of semi-empirical correlations describing the relation between vapor pressure, p, and temperature, T, for pure components (Wisniak, 2001). The usual homoscedastic variance, constant *absolute error*, was considered initially

$$P(T) = \eta(T) + \varepsilon = 10^{a-\frac{b}{c+T}} + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma_o^2),$$

and the heteroscedastic model with constant *relative error*, as suggested by Brozena et al. (2016), was considered later on

$$P(T) = \eta(T) + \varepsilon = 10^{a-\frac{b}{c+T}} + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma_e^2 \eta(T)^2).$$

being $\theta = (a, b, c)^t$ the unknown parameters of the model related with enthalpy and entropy of vaporization.

Optimal Experimental Design

The aim of the Optimal Experimental Design (OED) is to define the best experiment, regarding a certain optimality criterion. Different optimality criteria have been use in this work. Since the model is non-linear for some of the parameters, the designs calculated numerically are locally optimal. For this sake, the example of water in liquid state has been considered, giving the design space $\mathcal{X} = [T_{min}, T_{max}] = [1°C, 100°C]$, and nominal values $\theta_0 = (8.07131, 1730.63, 233.426)^t$, Dortmund Data Bank.

D-Optimality for Antoine's Equation

D-Optimality is a criterion with the goal of minimizing the determinant of the inverse of the information matrix, which translates to minimizing the volume of the parameters confidence ellipsoid.

A handful of theoretical results have been proved in de la Calle-Arroyo et al. (2021). These results lead to the analytical expressions for the D-optimal designs, both for the homoscedastic and the heteroscedastic model. The D-optimal designs for the homoscedastic and heteroscedastic model for liquid water have, respectively, the following expressions

$$\xi^\star_{oD} = \begin{Bmatrix} 44.90 & 83.20 & 100 \\ 1/3 & 1/3 & 1/3 \end{Bmatrix}, \quad \xi^\star_{eD} = \begin{Bmatrix} 1 & 41.87 & 100 \\ 1/3 & 1/3 & 1/3 \end{Bmatrix}.$$

The efficiency of the heteroscedastic optimal design for the homoscedastic model is $\text{eff}_{oD}(\xi^\star_{eD}) = 25.4\%$, while for the analogous it is $\text{eff}_{eD}(\xi^\star_{oD}) = 18.7\%$.

The nominal values are the best guesses for the values of the parameters. There can be uncertainty about the accuracy of these specifications. A *sensitivity analysis* can be performed to estimate the loss of efficiency in case the guess is not precise.

Both models are quasi-linear on a, and hence the optimal designs do not depend on the value of this parameter. Figure 1 shows that, in the homoscedastic model, it is better to either underestimate or overestimate both parameters at the same time than to underestimate one and overestimate the other. The heteroscedastic model is not dependent on the value of b, and is very robust to a misspecification of the parameter c. It is better to understimate this parameter, although the loss of efficiency is very small on either case.

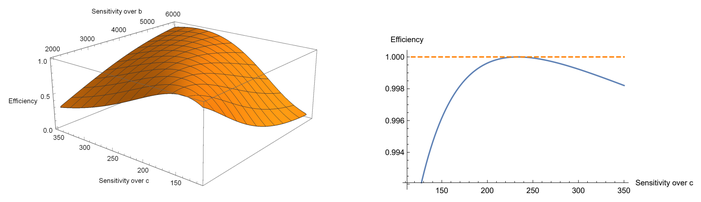

Figure 1: Efficiency for a misspecification of the parameters of the models, homoscedastic (left) and heteroscedastic (right).

D_s-Optimality

The D_s-Optimality is a criterion that, with the same philosophy as the D-Optimality, looks for the best design in order to estimate a subset of s parameters of θ.

Table 1: D_s-optimal designs for the six different subsets of the parameters. In **bold** are the weights greater than $1/3$. The first part of the table gives the optimal designs for the homoscedastic model while the second part includes those for the heteroscedastic model.

Criterion	D_s-optimal designs			$\text{eff}_{eD_s}(\cdot)$
$\xi^\star_{oD_a}$	33.11 (**0.561**)	83.90 (0.302)	100 (0.138)	2.6%
$\xi^\star_{oD_b}$	33.13 (**0.582**)	83.90 (0.289)	100 (0.129)	2.8%
$\xi^\star_{oD_c}$	33.14 (**0.606**)	83.90 (0.274)	100 (0.119)	2.3%
$\xi^\star_{oD_{ab}}$	42.25 (**0.487**)	79.82 (**0.334**)	100 (0.178)	10.7%
$\xi^\star_{oD_{ac}}$	42.29 (**0.491**)	79.66 (**0.340**)	100 (0.169)	9.7%
$\xi^\star_{oD_{bc}}$	42.33 (**0.493**)	79.53 (**0.347**)	100 (0.161)	8.9%
				$\text{eff}_{oD_s}(\cdot)$
$\xi^\star_{eD_a}$	1 (0.210)	41.75 (**0.492**)	100 (0.297)	3.3%
$\xi^\star_{eD_b}$	1 (0.229)	41.77 (**0.500**)	100 (0.271)	3%
$\xi^\star_{eD_c}$	1 (0.250)	41.75 (**0.500**)	100 (0.250)	3.5%
$\xi^\star_{eD_{ab}}$	1 (0.256)	44.65 (**0.338**)	100 (**0.406**)	13.6%
$\xi^\star_{eD_{ac}}$	1 (0.295)	43.31 (**0.333**)	100 (**0.372**)	12.8%
$\xi^\star_{eD_{bc}}$	1 (**0.333**)	41.89 (**0.333**)	100 (**0.333**)	12.8%

Table 1 features D_s-optimal designs for the six different parameters subset, for both the homoscedastic and heteroscedastic model. Each of column represent a pair of point and (weight), whereas the last column indicates the efficiency of the design in case of a wrong assumption over the heteroscedasticity or homoscedasticity of response.

Note that the support points remain the same for each of the three designs estimating a single parameter, while there are slight variances on the weight of these points.

A-Optimality

The A-optimality criterion minimizes the mean of the estimates of the parameters, a, b and c. For the two response variances, the respective A-optimal designs are

$$\xi^\star_{oA} = \begin{Bmatrix} 33.13 & 83.73 & 100 \\ 0.582 & 0.289 & 0.129 \end{Bmatrix}, \quad \xi^\star_{eA} = \begin{Bmatrix} 1 & 41.77 & 100 \\ 0.229 & 0.500 & 0.271 \end{Bmatrix}, \tag{1}$$

The efficiency of the heteroscedastic optimal design for the homoscedastic model is $\text{eff}_{oA}(\xi^\star_{eA}) = 2.8\%$. Meanwhile, the efficiency of the homoscedastic design for the heteroscedastic model is $\text{eff}_{eA}(\xi^\star_{oA}) = 3\%$. In both cases, once again, we see very low efficiencies when comparing optimal designs for the different models. Figure 2 shows the sensitivity function of the A-optimal design for both models. It can be seen that the Equivalence Theorem holds for both of them.

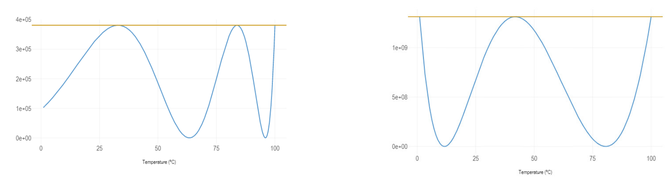

Figure 2: Sensitivity function for the A-optimal designs, homoscedastic (left) and heteroscedastic (right).

I-Optimality

The I-optimality criterion aims to minimize the variance over a region of interest. In these models, it is of special relevance as precise estimation near the boiling point is important.

Two probability distributions were used, the uniform distribution, $\mathcal{U}$, with different regions of interest near the vaporization point, and the triangular distribution, $\mathcal{T}$, with vertex at 100ºC with the same regions.

In Table 2 we can see the different I-optimal designs, considering the probability distributions mentioned above, for both the homoscedastic and heteroscedastic model. Again, in the last column the efficiency of the design for the other model is included. For these designs, while the homoscedastic design for the heteroscedastic model has a reasonable effiency, around 60%, the efficiency loss of the analogous case is very strong.

Table 2: I-optimal designs for different choices of $\mu(T)$. In **bold** the weights greater than $1/3$.

Criterion	I-optimal designs			$\text{eff}_{eI}(\cdot)$
$\xi^\star_{oI_{\mathcal{U}(70,100)}}$	52.46 (0.168)	83.80 (**0.578**)	100 (0.253)	61.7%
$\xi^\star_{oI_{\mathcal{U}(60,100)}}$	52.58 (0.280)	83.06 (**0.506**)	100 (0.214)	57.6%
$\xi^\star_{oI_{\mathcal{T}(70,100)}}$	49.71 (0.108)	85.72 (**0.560**)	100 (0.332)	65.6%
$\xi^\star_{oI_{\mathcal{T}(60,100)}}$	51.18 (0.175)	84.54 (**0.538**)	100 (0.287)	61.5%
				$\text{eff}_{oI}(\cdot)$
$\xi^\star_{eI_{\mathcal{U}(70,100)}}$	1 (0.023)	62.07 (**0.444**)	100 (**0.533**)	1%
$\xi^\star_{eI_{\mathcal{U}(60,100)}}$	1 (0.026)	58.72 (**0.512**)	100 (**0.462**)	1%
$\xi^\star_{eI_{\mathcal{T}(70,100)}}$	1 (0.019)	62.70 (**0.335**)	100 (**0.646**)	0.9%
$\xi^\star_{eI_{\mathcal{T}(60,100)}}$	1 (0.023)	60.43 (**0.399**)	100 (**0.578**)	0.9%

Shiny App

All the designs and efficiencies presented in this work, as well as the sensitivity function figures, have been calculated using an online tool developed for the use of the practitioners. The software uses the analytical expressions, when available, and an implementation in R of the cocktail algorithm to calculate the optimal designs. It is an intuitive app that allows optimal design calculation with a minimal set up, choosing the response of the model, the design space and best guesses of parameters, with some substances already available to compute. The application can be found at `https://kezrael.shinyapps.io/AntoineOptimal/` or accessed through the QR code.

Figure 3: QR code for the Shiny Application

Conclusions

This work presents locally optimal designs for the Antoine Equation for the homoscedastic and heteroscedatic models. Efficiencies for these models have been compared, showing strong differences between optimal designs for each situation. Therefore is very important to obtain information about the behavior of the variance of the response in order to obtain robust designs.

With the Shiny Application $D-$, D_s-, $A-$ and $I-$ optimal designs for water in the range $\mathcal{X} = [1, 100]$ºC have been obtained. Other compounds, or a manual setting of the parameters, are available for the practitioner to choose from and calculate optimal designs for their particular substance.

Acknowledgements

This work was sponsored by Ministerio de Economía y Competitividad MTM2016-80539-C2-1-R and by Consejería de Educación, Cultura y Deportes of Junta de Comunidades de Castilla-La Mancha and Fondo Europeo de Desarrollo Regional SBPLY/17/180501/000380.

References

A. Brozena, C. E. Davidson, B. Schlindler, and D. E. Tevault. Vapor pressure data analysis and statistics. Technical Report ECBC-TR-1422, Edgewood Chemical Biological Center, U.S. Army RDECOM, Aberdeen Proving Ground, MD, December 2016. URL `https://apps.dtic.mil/dtic/tr/fulltext/u2/1022530.pdf`.

C. de la Calle-Arroyo, J. López-Fidalgo, and L. J. Rodríguez-Aragón. Optimal designs for antoine equation. *Chemometrics and Intelligent Laboratory Systems*, 214:104334, 2021. doi: DOI:10.1016/j.chemolab.2021.104334. URL `https://doi.org/10.1016/j.chemolab.2021.104334`.

Dortmund Data Bank, 2020. URL `www.ddbst.com`.

J. Wisniak. Historical development of the vapor pressure equation from dalton to antoine. *Journal of Phase Equilibria*, 22: 622–630, 2001.