

Parameter estimation

Oliwer Sliczniuk^{*a,*}, Pekka Oinas^a and Francesco Corona^a

^aAalto University, School of Chemical Engineering, Espoo, 02150, Finland

ARTICLE INFO

Keywords:

Supercritical extraction

Parameter estimation

Mathematical modelling

ABSTRACT

Given a system of partial differential equations, $F(t, x, \dot{x}, p, u) = 0$, where x represents state variables, p are the parameters, and u are control variables, the process model is simultaneously solved for both x_i and a set of sensitivity functions, dx_i/dp_j , overall times t . These sensitivity functions measure the influence of the parameter change on the model's output. As an example, the supercritical extraction process is presented. The impact of mass flow rate, pressure, and inlet temperature on the model's output is discussed. The sensitivity analysis results prove that the considered variables can affect the extraction's yield and be used as control variables in optimization problems. Moreover, the local sensitivity analysis results are analyzed from a phenomenological point of view to enhance understanding of the process model.

1. Introduction

The extraction of natural substances from solid materials and liquids with solvents have been a popular subject of research and development in the last years. Supercritical fluids have multiple applications in an extraction process due to the pressure-dependent dissolving power and both gas- and liquid-like properties (for example fluid-like density and gas-like diffusivity). Among different supercritical fluids, the supercritical CO_2 is one of the most popular due to it is nontoxic, non-flammable and non-corrosive properties. The critical point of CO_2 is relatively low (73.8 bar and $31^\circ C$), compare to other fluids. The supercritical extraction with CO_2 become attractive alternative, to replace traditional extraction techniques.

The supercritical extraction process is considered unconventional due to the lack of formulation of one general model, properties that strongly depend on working conditions and its unsteady nature. Understanding the physics behind the phenomena and the proper mathematical description is necessary to predict the process outcome. Several extraction models have been developed to characterize different mass transfer mechanisms and equilibrium relationships. The extraction of valuable compounds from fixed bed of biomass can be described by one of many mathematical models as presented by Huang et al. [1]. The selection the extraction model is not arbitrary and should be based on the knowledge of phenomena occurring in the operational unit. Each model has its own assumptions and describe different mass transfer mechanisms and equilibrium relationships.

Reverchon [2] developed a model, which considers an oil as a single component and assumes that the extraction process is controlled by internal-mass transfer resistance. As a result of these assumptions, the external mass transfer was neglected. The original model of Reverchon does not consider the influence of axial dispersion and does not take


into account the change of density and flow rate along the bed.

Based on analogy to heat transfer, Reverchon et al. [3] suggested a hot ball model, where the extraction process is treated analogously to a process in which a hot ball is cooled in a uniform medium the process of cooling down a hot ball in a uniform medium. The hot ball model is used to describe an extraction process from solid particles which contains small quantities of solute so that the solubility is not a limiting factor.

Sovova [4] presented the Broken and Intact Cell model, which assumes the presence of two solid phases, the broken and intact cells. The results of applying Sovova's model allow distinguishing three phases of the extraction process (constant extraction rate, falling extraction rate, and diffusion-controlled periods) directly at the modelling stage. presented the Broken and Intact Cell model. The BIC model describes a system where the outer surfaces of particles have been mechanically interrupted. The solute from the broken cells is easily accessible for the solvent. The extraction of easily accessible solute is fast and directly controlled by its diffusion and convection in the solvent. However, the rest of the solute is less accessible because it is closed in the particle's core or intact cells. This solute slowly diffuses through the walls of a cell due to high mass transfer resistance. In this model, the extraction process is divided into three periods controlled by different mass transfer mechanisms: Constant extraction rate (CER) period, Falling extraction rate (FER) period and Diffusion-controlled (DC) period.

As in many engineering problems, mathematical modelling is crucial to mimic real-life objects, but it requires knowing multiple parameter values. One way to better understand the phenomena occurring in the process is to investigate the influences of various parameters on the model's output. One of the methods that can be applied is sensitivity analysis. Among different types of sensitivity analyses, there are two main groups: a local and global sensitivity analysis. The local method involves taking the total derivative of the output with respect to an input parameter at a single point. The derivative can be estimated directly by the automatic

*Corresponding author

 oliwer.sliczniuk@aalto.fi (O. Sliczniuk*)

ORCID(s): 0000-0003-2593-5956 (O. Sliczniuk*); 0000-0002-0183-5558 (P. Oinas); 0000-0002-3615-1359 (F. Corona)

differentiation. The total derivatives form an additional set of equations that is solved simultaneously with the original system. An alternative approach to the direct method is the Adjoint Sensitivity Analysis method. In this method, the original model is integrated to obtain the full trajectory of the state. Then the adjoint equation is solved backwards in time. The adjoint operates backwards in the sense that it determines a gradient with respect to input from a gradient with respect to output as discussed by Errico [5]. The adjoint method is faster and requires the least memory than the local method. However, not all ODEs are stable under backwards integration. Moreover, both methods belong to one-at-a-time (OAT) methods, as they allow an investigation of the influence of only one parameter at a time. One way to determine the model response to "interactions" between parameters and to observe the non-linear behaviour of the system is to use higher-order derivatives. An alternative method is to apply a global sensitivity analysis, which utilises information about parameter uncertainty. A sensitivity analysis is considered global when all the parameters vary simultaneously. According to Razavi and Gupta [6], the main problem of generalising local sensitivity measures to represent "global" properties is the lack of a unique definition of global sensitivity. These methods allow selecting the most influential parameters, which can be used for optimisation (the finding is a set of control variables, which requires the least effort to affect a model's output), model reduction (removing the least influential parameters), or parameter fitting. In this work, the example of supercritical extraction is used to show the practical application of sensitivity analysis. Some researchers working with the supercritical fluids extraction process examined the influence of the parameters on the model's output. Most of them perform a parametrisation of the process model and solve it multiple times for different combinations of parameters. The literature review is presented below. A sensitivity analysis is helpful to identify parameters that have the greatest effect on the model's output. The sensitivity analysis results can be used to find regions in the space of input factors for which the model output is maximum. Another application is model reduction - for example, by identifying and removing redundant parts of the model structure. In the case of parameter fitting with a large number of parameters, non-sensitive parameters can be excluded from the fitting procedure, reducing the computational cost. A sensitivity analysis is used to increase understanding of the relationships between input and output variables in a model.

Fiori et al. [7] presented how a model uncertainty affects a supercritical extraction process. Authors performed a local sensitivity analysis taken at the mean values of the parameters, determined during the parameter estimation. Their analysis shows the importance of two parameters: the particle diameter and the internal mass transfer coefficient. The authors found that the supercritical extraction process is almost completely insensitive during the first phase of the process (controlled by the solute's solubility in the supercritical solvent) for a change of these parameters.

However, the authors claim that the model's sensitivity to parameters becomes high during the second phase of the process (controlled by internal diffusion).

Rabi et al. [8] analysed how the different parameters affect the model's output. The parameters are dimensionless numbers (e.g., mass transfer Biot, Péclet and Damköhler numbers), which are assumed to take one of few discrete values. The simulations are performed for each value of parameters and compared with each other. The authors found that the Péclet number can strongly affect the extraction efficiency, so the diffusive transport should not be neglected. The authors investigated effect of the Damköhler number on the process. The authors observed that higher intra-particle diffusion accelerated solid particles depletion so that an optimal extraction time can then be looked for. Furthermore, the authors proved that others investigated parameters like partition coefficient and bed porosity influence the extraction yield.

Glover and LeVan [9] presented a local sensitivity analysis for an adsorption process. Considering the similarities between extraction and adsorption processes, it is interesting to see how the same technique can be applied to different cases. The authors examine the sensitivity of the breakthrough of adsorption beds to system parameters. The impact of mass and energy transfer effects and adsorbent layer thicknesses are determined by calculating the derivatives of the outlet concentration and outlet temperature. The authors applied a Newton-based method along with the sensitivities to determine the optimal layering of an adsorption bed.

In this work, the process model represents supercritical extraction and consists of a set partial differential equations, $F(t, x, \dot{x}, p, u) = 0$, where x represents state variables, p are the parameters, and u are control variables. The process model is simultaneously solved for both x_i and a set of sensitivity functions, dx_i/dp_j , overall times t . The impact of mass flow rate, pressure, and inlet temperature on the model's output is analyzed. In this work, the sensitivity analysis results are used to prove if the considered variables affect the extraction process and can be used as decision variables in optimization problems. The novelty is the application of the local sensitivity analysis in the case of supercritical extraction. In this work, a distributed-parameter model, based on Reverchon [2] and Vargas et al. [10], is used to describe a fluid-solid extraction process of 'carqueja' oil from 'Baccharis Trimeria' leaves with CO_2 as a solvent. The process model consists of three partial differential equations: two mass balance equations and one heat balance equation. To enhance understanding of the process model, the local sensitivity analysis is used to investigate the effect of parameters on the model's output. The presented sensitivity analysis results are restricted to parameters related to operating conditions only. Furthermore, the sensitivity analysis is used to prove which of the analysed variables can be used as decision variables in optimisation problems.

2. Materials and methods

The extraction process is assumed to operate in a semi-continuous mode in a cylindrical vessel. The solvent is firstly brought to super-critical condition, pumped through a fixed-bed, of finely chopped biomass, where the solute is extracted from the biomass. Then, the solvent and solute from the extractor are separated in a flush drum, and the extract is collected. The flow rate (F_{in}) and inlet temperature (T_{in}) of the extractor's feed can be measured and manipulated. The pressure (P) in the vessel can be measured and manipulated, while the outlet temperature (T_{out}) can only be measured. A simplified flow diagram is depicted in Figure 1.

The solid-fluid extraction process is assumed to operate in a semi-continuous mode in a cylindrical vessel. In the extractor, essential oils are removed from biomass (for example, carqueja seeds) by supercritical carbon dioxide. As the solvent flows continuously through the bed of porous particles, the CO_2 molecules diffuse into the pores and adsorb on the particle surface to form an external fluid film around the solid particles. Then the solute dissolves and diffuses into the solvent in the pores and eventually into the bulk. The solvent-solute mixture leaves the extractor and undergoes depressurization. The CO_2 is released to a gaseous state, and the solute precipitates. In the vapour-liquid separator, the gaseous solvent is vent off while the precipitated product is collected.

In Figure 1, an instrumental set-up is presented. The flow rate (F_{in}) and the temperature (T_{in}) of the solvent entering the extractor are measured and can be adjusted using a pump and a heat exchanger, respectively. Moreover, the extractor's pressure (P) is assumed to be measurable and adjustable using a back-pressure regulator valve. In addition, the temperature of the outlet stream T_{out} and the temperature profile of the fixed-bed can be measured.

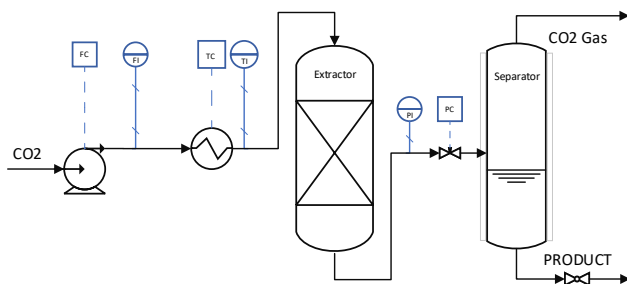


Figure 1: Process flow diagram for solid-fluid extraction process

2.1. Extraction model

The main components of the model are three one-dimensional differential equations related to mass balances (solid and fluid phase) and heat balance. As in all batch processes, our system has only one steady-state. The steady-state occurs only if the solute is entirely taken out of the biomass. The extraction from a fixed bed is an unsteady process, implying that the solute concentrations in the solid and

solvent phases vary over time and space. The convection and the diffusion of the mass and the heat are considered only in the axial direction. The radial distribution of such properties like a velocity is assumed to be flat in that direction. The assumption of the flat velocity profile allows considering a lack of back mixing. The pressure drop in the bed is assumed to be negligible. Under such conditions, the fluid flow can be considered a plug flow. Since the solute is assumed to be a single pseudo-component, there is only one mass balance equation for each phase (Eq. s 1 and 3). On the other hand, there is only one heat balance (Eq. 8) because the presence of a single pseudo-phase is assumed. Moreover, the process model does not consider the change of particle size or the void fraction during the process. The particles are assumed to be uniformly distributed along the bed and to have the same size.

A distributed-parameter model, based on Reverchon [2], is used to describe the solid-fluid extraction process. The process model consists of three partial differential equations: two mass balance equations relative to the concentration of solute in the fixed solid phase and the mobile fluid phase, and the heat balance equation relative to the temperature of the fluid phase. The amount of solute in the solvent is considered negligible. Therefore, the fluid phase can be described as pseudo-homogenous, and its properties are assumed to be the same as the solvent.

The movement of the mobile pseudo-homogeneous phase (Eq. 1) is considered only in the axial direction. The properties of the system in the radial direction is assumed to be uniform. In addition, it is considered that the boundary layer adjacent to the inner wall of the extractor does not exist. Therefore, the velocity profile is constant across any cross-section of the extractor perpendicular to the axial direction. As a result, the plug flow model can be introduced. The particle size distribution and the void fraction of the solid phase are assumed to be uniform in space and remain constant in time. Moreover, the pressure is considered to be constant along the fixed bed. The mass balance for the fluid phase consists of convection, diffusion, and kinetic terms.

Considering the solid phase to be fixed, the convection and diffusion terms in the corresponding mass balance are both assumed to be negligible or absent. Therefore, the mass balance for the solid phase (Eq. 3) consists of the kinetic term only.

As for the mass transfer from the solid to the fluid phase, a kinetic term based on two-film theory for the solute is considered. The mass transfer kinetic (Eq. 4) consists of the overall diffusion coefficient and the concentration gradient, which acts as a driving force for the process. A single pseudo-component is used to represent the extract collectively.

The heat balance (Eq. 8) consists of the convective and diffusive terms. It follows the assumption of a pseudo-homogeneous phase, which properties are the mean between fluid and solid phases. We consider no heat loss through the wall, and there is no heat generation in the system. Therefore,

the temperature of the extractor can be changed only by manipulating the temperature of the inlet stream $T_{In}(t)$.

For the sake of clarity of the process model, different colors have been used in the equations to indicate: **control variables**, **state variables**, **variables** and **parameters**.

2.1.1. Mass balance for the fluid phase

The fluid phase is a mobile phase, which means that the fluids mass balance (Eq. 1) is made of convection, diffusion, and mass transfer kinetic terms.

$$\frac{\partial c_f(t, z)}{\partial t} = - \underbrace{\frac{1}{\epsilon A \rho_f} \frac{F(t)}{[T(t, z), P(t)]} \frac{\partial c_f(t, z)}{\partial z}}_{\text{Convective term}} \quad (1)$$

$$+ \underbrace{D_e^M [T(t, z), P(t), F(t)] \frac{\partial^2 c_f(t, z)}{\partial z^2}}_{\text{Diffusive term}} + \underbrace{\frac{1 - \epsilon}{\epsilon} r_e(t, z)}_{\text{Kinetic term}} \quad (2)$$

where $c_f(t, z)$, $c_s(t, z)$, $T(t, z)$ correspond to concentration of solute in the fluid phase, concentration of solute in the solid phase and the temperature, respectively. $r_e(t, z)$ is a mass transfer kinetic term. $F(t)$ is the mass flow rate, $P(t)$ is the pressure, ϵ is the void fraction of the bed, A is the extractor's cross section, $\rho(T(t, z), P(t))$ is the fluid's density, ρ_s is the solids density, $D_e^M(T(t, z), P(t), F(t))$ is the axial mass diffusion coefficient.

2.1.2. Mass balance for the solid phase

The solid phase is a stationary phase, which indicates a lack of convection and diffusion terms in the mass balance (Eq. 3). Therefore, the only term present in this equation is the kinetic term (defined as presented in equation 4), which links solid and fluid phases.

$$\frac{\partial c_s(t, z)}{\partial t} = \underbrace{r_e(t, z)}_{\text{Kinetics}} \quad (3)$$

2.1.3. Kinetic term

As the solvent flows through the bed, it forms a film around the porous solid particle and occupies its pores. the CO_2 molecules diffuse into the pores and adsorb on the particle surface to for an external fluid film around the solid particles through the solvent-solid matrix interactions. Assuming that the mean free path of the molecule is much smaller than the pore diameter, the effect of Knudsen diffusion is small and can be neglected. In this work, the molecular diffusion is assumed. The dissolved solute diffuses from the particle's core through the solid-fluid interface, the pore, and the film into the bulk. The graphical representation of the mass transfer mechanism is shown in Figure 2. The mean solute concentration in the solid phase is denoted as c_s . At the solid-fluid interface, the equilibrium concentrations are given as c_s^* and c_p^* , respectively for solid and fluid phases. The concentration of the solutes in the fluid phase in the centre of the pore is denoted as c_p . As the solute diffuses through the pore, its concentration changes and reaches c_{pf} at the end the opening of the pore. The solute diffuses through the film around the particle and reaches a

concentration in the bulk c_f . It can be assumed that the two-film theory describes the solid-fluid interface inside the pore. The overall mass transfer coefficient can be introduced if the relation between the solute concentration in one phase and its equilibrium concentration is known.

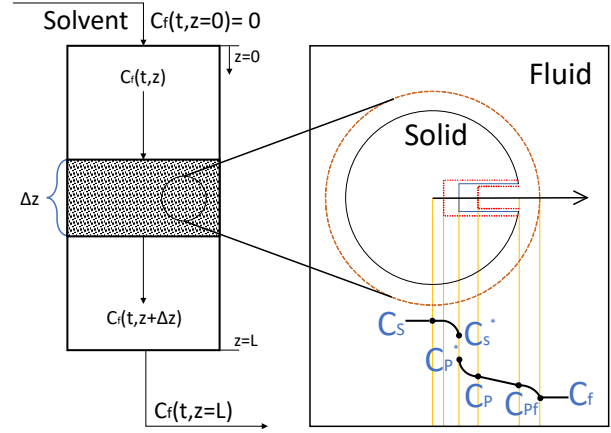


Figure 2: The extraction mechanism

Bulley et al. [11] suggest a process where the driving force for extraction is given by the difference between concentration of the solute in the bulk, c_f , and in the centre of the pore, c_p^* . The concentration c_p^* is in equilibrium with c_s according to an equilibrium relationship. The rate of extraction is thus $r_e(c_f - c_p^*(c_s))$.

On the other hand, Reverchon [2] proposes a driving force given by the difference between c_s and c_p^* . c_p^* is determined by an equilibrium relationship with c_f and the extraction rate is $r_e(c_s - c_p^*(c_f))$.

The kinetic term, denoted as equation 4, corresponds to a solute transfer rate between both phases. This term is made of the overall diffusion coefficient and the concentration gradient (between one phase and its equilibrium concentration), which acts as a driving force for the process. The mass transfer kinetic (Eq. 4) describes a diffusion rate of the solute from solid particles to the solvent. The kinetic term consists of the overall diffusion coefficient and the concentration gradient, which acts as a driving force for the process. Based on the work of Bulley et al. [11], Reverchon [2] proposed to define the kinetic terms as

$$r_e(t, z) = \frac{D_i(T(t, z))}{\mu l^2} (q(t, z) - q^*(t, z)) \quad (4)$$

where $D_i(T(t, z))$ corresponds to the overall diffusion coefficient and $q^*(t, z)$ is a concentration at the solid-fluid interface (which according to the internal resistance model is supposed to be at equilibrium with the fluid phase). $q^*(t, z)$ can be calculated based on the equilibrium relationship (Eq. 5).

The kinetic term of the model takes into account the properties of the biomass like a sphericity (through the shape coefficient μ equal to 1/3 for slabs and 3/5 for spheres), a

characteristic dimension of particles (through the parameter $l = r/3$ where r is the mean particle radius), and the solid density ρ_s . The fixed bed is characterized by a void fraction ϵ .

The kinetic term of the model takes into account properties of the solid particles like a sphericity (through the shape coefficient μ), a characteristic dimension of particles (through the parameter $l = r/3$ where r is the mean particle radius), and the bulk density ρ_s . The fixed bed is characterized by a void fraction ϵ .

According to Bulley et al. [11], a linear equilibrium relationship (Eq. 5) can be used to find an equilibrium concentration of the solute in the fluid phase $c_f^*(t, z)$ is based on concentration of the solute in the solid phase $c_s(t, z)$

$$c(t, z) = k_p(T(t, z), P(t))q^*(t, z) \quad (5)$$

The volumetric partition coefficient $k_p(T(t, z), P(t))$ behaves as an equilibrium constant between the solute concentration in one phase and the corresponding equilibrium concentration at the solid-fluid interphase. The $k_p(T(t, z), P(t))$ can be expressed as mass partition factor $k_m(T(t, z))$ according to Spiro and Kandiah [12]. According to Spiro and Kandiah [12], the term $k_p(T(t, z), P(t))$ can be expressed as the function of mass partition factor $k_m(T(t, z))$

$$k_m(T(t, z)) = \frac{k_p(T(t, z), P(t))\rho_s}{\rho(T(t, z), P(t))} \quad (6)$$

Equation 7 is the full representation of the kinetic term.

$$r_e(t, z) = -\frac{D_e(T(t, z))}{\mu l^2} \left(q(t, z) - \frac{\rho_s}{k_m(T(t, z))\rho(T(t, z), P(t))} c(t, z) \right) \quad (7)$$

2.1.4. Heat balance

The heat balance equation (Eq. 8) was developed, assuming the existence of a pseudo-homogeneous phase, which properties are the mean between fluid and solid phases (the amount of solute is considered small enough not to affect the overall heat balance). Equation 8 contains the convection and the diffusion terms. It is considered that there is no heat loss through the wall, and there is no heat generation in the system. The temperature of the extractor can be changed only by increasing the temperature of the inlet stream. The pseudo-homogeneous phase is assumed to flow only in the axial direction. The numerator of the factor in front of the convection term of the heat equation contains only the specific heat of the fluid $C_p(T(t, z), P(t))$ because the solid phase is stationary. Therefore, this factor can be understood as the fraction of the fluid's total heat through convection. On the other hand, the axial heat diffusion is calculated based on the definition of thermal diffusivity for the fluid, as explained in the appendix.

The heat balance (Eq. 8), is based on Srinivasan and Depcik [13], and consists of convective and diffusive terms. It follows the assumption of a pseudo-homogeneous phase, which properties are the mean between fluid and solid phases. It is considered that there is no heat loss through the walls, and there is no heat generation in the system.

The temperature of the extractor can be changed only by manipulating the temperature of the inlet stream $T_{Inlet}(t)$.

$$\frac{\partial T(t, z)}{\partial t} = \underbrace{\frac{F(t)C_p(T(t, z), P(t))}{A[(1 - \epsilon)\rho(T(t, z), P(t))C_p(T(t, z), P(t)) + \epsilon\rho_s C_{ps}]} \frac{\partial T(t, z)}{\partial z}}_{\text{Convection}} + \underbrace{D_e^T(T(t, z), P(t)) \frac{\partial^2 T(t, z)}{\partial z^2}}_{\text{Diffusion}} \quad (8)$$

where $D_e^M(T(t, z), P(t), F(t))$ is the axial mass diffusion coefficient, $C_p(T(t, z), P(t))$ is the fluid's specific heat, C_{ps} is the specific heat of the solid phase, $D_e^T(T(t, z), P(t), F(t))$ is the axial heat diffusion coefficient.

2.1.5. Extraction yield

The efficiency of the process (the yield) is calculated according to equation 9. The yield is defined as the ratio between a difference of initial solute's mass in the solid phase m_0 , and an actual amount of solute in the same phase m , compared to the initial amount of the solute in the fixed bed m_0 . The yield can also be calculated based on the change of solutes concentration in the solid phase if the volume of the extractor V is constant. An output function $y(t)$ returns a yield curve that represents the fraction of the solute's mass over time. The yield is obtained by function $g(q(t, z))$. Function $g(q(t, z))$ corresponds to a 'measurement device' that allows to calculate the yield based on the state variable $q(t, z)$. The efficiency of the process (the yield) is calculated according to Eq. 9, and is based on the change of solute's mass in the solid phase. It is defined as ratio between the difference of its initial value m_0 and the actual value $m(t, z)$, compared to the initial one. The yield can also be calculated based on the change of solutes concentration in the solid phase if the volume of the extractor V is constant.

$$y(t) = g(q(t, z)) = \frac{(m_0 - m(q(t, z)))}{m_0} = \frac{V(q_0 - q(t, z))}{Vq_0} \quad (9)$$

where $q(t, z)$ is concentration of solute in a solid phase and a temperature, respectively. m_0 is a initial mass of the solute in the solid phase, q_0 is a initial concentration of the solute in the solid phase, V is an extractor volume.

2.1.6. Dependencies

The Peng-Robinson equation of state is used to describe the thermodynamics of the fluid phase. The thermal properties of the pseudo-homogeneous phase are assumed to vary along the extractor, though always remaining equal to the corresponding mean values between the fluid and solid phase.

In the process model, properties of the supercritical fluid such as the density $\rho(T(t, z), P(t))$, the specific heat $C_p(T(t, z), P(t))$, the thermal conductivity $k^T(T(t, z), P(t))$, or the viscosity $\eta(T(t, z), P(t))$ are calculated based on the local temperatures along the fixed bed and the pressure. The Peng-Robinson equation of state is used to find values of a $\rho(T(t, z), P(t))$ and $C_p(T(t, z), P(t))$.

The inter-particle diffusion coefficient $D_e^M(T(t, z), P(t), F(t))$ is calculated from the empirical correlation, as the function of the Reynolds number $R_e(T(t, z), P(t), F(t))$ and the Peclet's number $P_e(T(t, z), P(t), F(t))$. The values of the internal diffusion $D_i(T(t, z))$ and the partition coefficient $k_m(T(t, z))$ come from the laboratory experiments. The expressions and the detailed information of the parameter fitting can be found in the appendix.

2.1.7. State-space representation

It is assumed that the solvent is free of solute at the entrance of the extractor and that all the solid particles have the same initial solute content q_0 . Moreover, it is considered that the initial temperature of the extractor in every place is equal to T_0 . Therefore, the initial conditions employed in the simulation are:

$$c(t = 0, z) = 0$$

$$q(t = 0, z) = q_0$$

$$T(t = 0, z) = T_0$$

The process model can be written in a general form:

$$\begin{bmatrix} \frac{\partial c(t, z)}{\partial t} \\ \frac{\partial q(t, z)}{\partial t} \\ \frac{\partial T(t, z)}{\partial t} \end{bmatrix} = \begin{bmatrix} \phi_1(c(t, z), q(t, z), T(t, z); \theta) \\ \phi_2(c(t, z), q(t, z), T(t, z); \theta) \\ \phi_3(c(t, z), q(t, z), T(t, z); \theta) \end{bmatrix} = \phi(t, z; \theta) = \frac{\partial \chi(t, z)}{\partial t} \quad (11)$$

where θ is a set of parameters present in the model, ϕ is a set of functions that correspond to state equations of the model, and χ is the state-space model.

The method of lines is used to transform the process model equations into a set of ODEs denoted as $G(x(t); p)$. The partial derivatives in z -direction are computed using a first-order and second-order finite difference approximation. The backward finite difference is used to approximate first-order derivative, while the central difference scheme is used to approximate second-order derivative. The length of the fixed bed is divided into N_z equally distributed points in z -direction. Each function ϕ_i is transformed to a corresponding set of N_z discretized equations denoted as $G_{i \times N_z + 1}$ to $G_{(i+1) \times N_z}$, where i corresponds to the process model equation. The state-space model $\chi(t, z)$ after the discretization is represented by $\dot{x}(t)$ (Eq. ??).

$$\dot{x}(t) = \frac{dx(t)}{dt} = \begin{bmatrix} \frac{dc_{f,1}(t)}{dt} \\ \vdots \\ \frac{dc_{f,N_z}(t)}{dt} \\ \frac{dc_{s,1}(t)}{dt} \\ \vdots \\ \frac{dc_{s,N_z}(t)}{dt} \\ \frac{dT_1(t)}{dt} \\ \vdots \\ \frac{dT_{N_z}(t)}{dt} \end{bmatrix} = \begin{bmatrix} G_1(x(t), q(t), T(t); p) \\ \vdots \\ G_{N_z}(c(t), q(t), T(t); p) \\ G_{N_z+1}(c(t), q(t), T(t); p) \\ \vdots \\ G_{2N_z}(c(t), q(t), T(t); p) \\ G_{2N_z+1}(c(t), q(t), T(t); p) \\ \vdots \\ G_{3N_z}(c(t), q(t), T(t); p) \end{bmatrix} = \underbrace{\begin{bmatrix} G_1(x(t), q(t), T(t); p) \\ \vdots \\ G_{3N_z}(c(t), q(t), T(t); p) \end{bmatrix}}_{G(x(t); p)}$$

where $x \in \mathbb{R}^{N_x=3N_z}$ and $p \in \mathbb{R}^{N_p=N_\theta+N_u}$, N_θ is the number of model parameters, N_u is the number of control variables.

In a state-space sense, the state variables of the system are the local concentrations of solute in the fluid and solid phases ($c(t, z)$ and $q(t, z)$, respectively), and the local temperature of the pseudo-homogeneous phase ($T(t, z)$). The controllable input variables are the mass flow-rate and temperature of the solvent in the feed ($F_{in}(t) = F(t)$ and $T_{in}(t) = T(t, z = 0)$, respectively) and the pressure in the extractor ($P(t, z) = P(t)$). We also assume that extraction yield can be modelled as a function of a known initial mass of solute in the solid phase and it can be measured after the separator ($Y(t)$). The system is controllable by manipulating the flow-rate and temperature of CO2 in the feed, and the pressure in the extractor.

3. Bayes theorem

As discussed by Himmelblau [14], the Bayesian approach to estimation make use of prior information. Such prior knowledge can come from theoretical considerations, from the results of previous experiments, or from assumptions by the experimenter. Typically, a Bayesian approach assumes a prior probability distribution of an unknown parameter θ in some parameter space θ . The distribution is updated by using Bayes' rule to obtain the posterior probability distribution.

Consider a set of events or outcomes, A_1, A_2, \dots, A_n , and some other event B . Bayes' theorem states that the probability that event A_i will occur, given that event B has already occurred, which will denoted by $P\{A_i|B\}$, is equal to the product of the probability that A_i will occur regardless of whether B will take place and the probability that B will occur, given that A_i has already taken place, divided by the probability of the occurrence of B :

$$P\{A_i|B\} = \frac{P\{B|A_i\}P\{A_i\}}{P\{B\}} \quad (12)$$

Further, if all events comprising the set $\{A_i\}$ are included in A_1, A_2, \dots, A_n , then

$$P\{A_i|B\} = \frac{P\{B|A_i\}P\{A_i\}}{\sum_{i=1}^n P\{B|A_i\}P\{A_i\}} \quad (13)$$

We can interpret these symbols as follows:

1. $P\{A_i\}$ is a measure of our degree of belief that event A_i will occur or that hypothesis A_i is true prior to the acquisition of additional evidence that may alter the measure. $P\{A_i\}$ is denoted the *prior probability*.
2. $P\{A_i|B\}$ is a measure of our degree of belief that event A_i will occur or that hypothesis A_i is true, given additional evidence B pertinent to the hypothesis. $P\{A_i|B\}$ is termed the *posterior probability*.
3. $P\{B|A_i\}$ denotes the likelihood that event B will occur, given that event A_i is true. $P\{B|A_i\}$ is a conditional probability, interpreted in the Bayesian framework as a likelihood, $L(A_i|B)$.

For continuous variable, Bayes' theorem can be more conveniently expressed in terms of the probability density function rather than the probabilities themselves. Equation 13 can be expressed in terms of a set of observed values of the random variable X , the experimental data \mathbf{x} and unknown parameter θ as

$$p(\theta|X = \mathbf{x}) = p(\theta|\mathbf{x}) = \frac{L(\theta|\mathbf{x})p(\theta)}{\int_{-\infty}^{+\infty} L(\theta|\mathbf{x})p(\theta)d\theta} \quad (14)$$

where

$p(\theta|\mathbf{x})$ = the posterior probability density function for θ ; it includes knowledge of the possible values of θ gained from the experimental data \mathbf{x}

$p(\theta)$ = the prior probability density function for θ (before the experiment in which \mathbf{x} was observed)

$L(\theta|\mathbf{x}) = p(\mathbf{x}|\theta)$ = the probability density function termed the likelihood function of θ given \mathbf{x}

The denominator in Equation 14 is a normalizing factor chosen so that the integration of the posterior distribution is unit, i.e., $\int_{-\infty}^{+\infty} p(\theta|\mathbf{x})d\theta = 1$. By taking into account the law of total probability, the denominator can be written as

$$\int_{-\infty}^{+\infty} p(\mathbf{x}|\theta)p(\theta)d\theta = p(\mathbf{x}) \quad (15)$$

If the prior distribution is a uniform distribution, that is the prior distribution is a constant, the Equation 14 reduces to

$$p(\theta|\mathbf{x}) = \frac{L(\theta|\mathbf{x})}{\int_{-\infty}^{+\infty} L(\theta|\mathbf{x})d\theta} \quad (16)$$

If the prior knowledge concerning a postulated event or hypothesis is poor, the posterior probability is largely or entirely determined by the likelihood function, that is, by the additional accumulated evidence for which the likelihood function acts as a mathematical expression. If prior knowledge outweighs recent evidence, however, then the posterior probability is determined almost solely by the prior probability.

In the application of tests and the design of experiments, certain definitions and rules concerning probability are needed and are listed below.

1. It follows from the frequency theory of probability that: $0 \leq P \leq 1$
2. If the probability of occurrence of one event A depends on whether or not event B has occurred, the two events are termed *dependent*; if the probability of occurrence of event A does not depend on the occurrence of B , or the reverse, the two events are *independent*.
3. **Addition Rule** If A_1, A_2, \dots, A_n are mutually exclusive events, i.e., cannot occur at the same time, the probability of occurrence of just one of the events is equal to the sum of the probabilities of each A_i :

$$P(A_1 \text{ or } A_2 \dots \text{ or } A_n) = \sum_{i=1}^n P(A_i) \quad (17)$$

Very often we let

$$\sum_{i=1}^n P(A_i) = 1 \quad (18)$$

Also, if each event is equi-probable so that $P(A_i) = q$,

$$\sum_{i=1}^n q = nq = 1 \quad \text{or} \quad q = \frac{1}{n} = P(A_i) \quad (19)$$

In set theory, mutually exclusive event have no points in common. The union of the sets which represents the set of all elements that belong to

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n) \quad (20)$$

4. **Multiplication Rule** If A and B are *independent* event

$$P(A \text{ and } B) = P(A)P(B) \quad (21)$$

In set theory the intersection of A and B is the set of all elements that belong to A and B :

$$P(A \cap B) = P(A)P(B) \quad (22)$$

If the A and B are *dependet* events,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (23)$$

where the symbol $P(A|B)$ means "probability of A given B ". As a corollary,

$$P(A \cap B) = P(B)P(A|B) \quad (24a)$$

$$= P(A)P(B|A) \quad (24b)$$

Two kind of probabilities enter Equation 24a (or Equation 24b): the absolute probability of event B (or A) irrespective of whether or not A (or B) has occurred, and that the conditional probability of event A (or B) computed on the assumption that B (or B) has occurred. Equation 22 or 23 is special case of Equation 24a or 24b, because if the event are independent $P(A|B) = P(A)$. For the case of many event, Equation 22 can be expanded to

$$\begin{aligned} P(A_1 \text{ and } A_2 \text{ and } \dots \text{ and } A_n) &= P(A_1) \cdot P(A_2) \cdot \dots \cdot P(A_n) \\ &= \prod_{i=1}^n P(A_i) \end{aligned} \quad (25)$$

5. Another useful relationship for event which are not mutually exclusive is

$$P(A) + P(B) - P(A \cap B) = P(A \cup B) \quad (26)$$

4. Parameter estimation

Conceptually, the unobservable error $\epsilon(t)$ is added to the deterministic model output, $y(t)$, to give the observable dependent variable $Y(t)$, for example the experimental process output. For discrete observations:

$$Y(t_i) = y(t_i) + \epsilon(t_i) \quad (27)$$

and for continuous variables:

$$Y(t) = y(t) + \epsilon(t) \quad (28)$$

The objective in parameter estimation is to obtain the "best" estimate of θ based on the continuous observations $Y(t)$ or the discrete observations $Y(t_i)$. We assume that $x(t)$ is given a priori (for example as a solution of process model) but θ is to be estimated over the time interval $0 \leq t \leq t_n$ from discrete observation described by the following relation:

$$Y(t_i) = \mathbf{h}(t_i)y(t_i) + \epsilon(t_i) \quad 1 \leq i \leq n \quad (29)$$

where $Y(t_i)$ is an $n \times 1$ column vector, $\mathbf{h}(t_i)$ is an $n \times v$ matrix given a priori, and $\epsilon(t_i)$ is an $n \times 1$ column vector ("noise" vector) whose elements are the unobservable errors. The above equation can be written in general form as:

$$Y(t_i) = y(\theta, t_i) + \epsilon(t_i) \quad (30)$$

where y is the output model, which is a function of θ and time t_i .

Because of the difficulty of obtaining analytical solutions to the deterministic process model, experiments have been arranged whereby the vector of derivatives $dY(t_i)/dt$ is measured rather than $Y(t_i)$ itself. In such cases, it is assumed that the unobservable error is added to the deterministic derivative $dy(\theta, t_i)/dt$ as follows:

$$\frac{dY(t_i)}{dt} = \frac{dy(\theta, t_i)}{dt} + \epsilon(t_i) \quad (31)$$

4.1. Least squares estimation

A least squares parameter estimation does not require prior knowledge of the distribution of unobservable errors, yield unbiased estimates, that is, $\mathbb{E}\{\hat{Y}(t) - y(t)\} = 0$, and results in the minimum variance among all linear unbiased estimators. If the observations Y for the model responses are continuous functions of time from $t = 0$ to $t = t_f$, the Markov (or "rigorous least squares") criterion is to minimize:

$$\phi = \frac{1}{2} \int_0^{t_f} [Y(t) - y(\theta, t)]^T \Gamma^{-1} [Y(t) - y(\theta, t)] dt \quad (32)$$

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \phi \quad (33)$$

where Γ is the covariance matrix (or perhaps a matrix of appropriate weights), $\hat{\theta}$ is the parameter estimate, Θ is the parameter space, and ϕ is the time integrated value of the error squared ("integral squared error"). If the observations are made at discrete instants of time, t_i , $i = 1, 2, \dots, n$, the Markov criterion is to minimize:

$$\phi = \frac{1}{2} \sum_{i=1}^n [Y(t_i) - y(\theta, t_i)]^T \Gamma^{-1} [Y(t_i) - y(\theta, t_i)] \quad (34)$$

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \phi \quad (35)$$

If Γ is a diagonal matrix, ϕ becomes a "weighted least squares" criterion; if $\Gamma = \sigma^2 \mathbf{I}$ where σ is a standard deviation, ϕ is the "ordinary least squares" criterion.

4.2. Weighted sum of squares - Cumulative data

One characteristic feature of certain special experiments is the use of the same batch of materials for the entire series of measurements. If unobservable error in the first observation is designated as ϵ_1 , the second error observation ϵ'_2 includes ϵ_1 plus a random component introduced side from ϵ_1 , or $\epsilon'_2 = \epsilon_1 + \epsilon_2$. The error in the third observation is $\epsilon'_3 = \epsilon_1 + \epsilon_2 + \epsilon_3$, and so forth. Mandel [15] distinguished between the usually assumed type of independent measuring error in the dependent variable and a 'cumulative' or interval error in which each new observation includes the error of the previous observations. Cumulative errors, arising because of the fluctuations as a function of time in the process itself due to small changes in operating, are not independent - only the differences in measurement from one period to the next are independent. Thus, if we consider two model:

Model A - independent error:

$$Y(t_i) = y(\theta, t_i) + \epsilon_i \quad i = 1, 2, \dots, n$$

Model B - cumulative data:

$$Y(t_j) - Y(t_{j-1}) = y(\theta, t_j) - y(\theta, t_{j-1}) + \epsilon_j, \quad j = 1, 2, \dots, n$$

or

$$Y(t_i) = \sum_{j=1}^i (Y(t_j) - Y(t_{j-1})) = \sum_{j=1}^i (y(\theta, t_j) - y(\theta, t_{j-1})) + \sum_{j=1}^i \epsilon_j$$

in which ϵ_i are independent random variables with $\mathbb{E}\{\epsilon_i\} = 0$ and $\text{Var}\{\epsilon_i\} = \sigma_1^2$, a constant. The ϵ_j are also random independent variables because they represent differences such as $\epsilon'_3 - \epsilon'_2 = (\epsilon_1 + \epsilon_2 + \epsilon_3) - (\epsilon_1 + \epsilon_2) = \epsilon_3$. We assume that $\mathbb{E}\{\epsilon_j\} = 0$ and $\text{Var}\{\epsilon_j\} = \sigma_1^2 (y_j - y_{j-1})$; i.e., the variance of ϵ_j can be a function of the test interval. The experimental point tend to stay on one side of the regression line of the best fit for Model B.

The best unbiased estimate of the Model B is obtained by minimizing the weighted sum of squares:

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \sum_{j=1}^n \frac{[(Y(t_j) - Y(t_{j-1})) - (y(\theta, t_j) - y(\theta, t_{j-1}))]^2}{y(\theta, t_j) - y(\theta, t_{j-1})}$$

4.3. Maximum likelihood estimation

Maximum likelihood estimation (MLE) is a method of estimating the parameters of an assumed probability distribution, given some observed data. This is achieved by maximizing a likelihood function so that, under the assumed statistical model, the observed data is most probable. The MLE has the desirable characteristics of asymptotic efficiency and normality. Each time it has been associated with the (joint) normal distribution because of mathematical convenience. Consider the joint probability density function (the likelihood function) $p(\theta|y)$ for random sample $y = y(t_1), y(t_2), \dots, y(t_n)$, parameter θ . If a maximum of this function over all choices of θ can be found, the estimates so obtained are maximum likelihood estimates. The conditions at the maximum can be evolved incorporating prior information as follows.

The posterior probability density $p(\theta|y)$ can be expressed as the ratio of two probability densities if we make use of the analogue for continuous variables of Equation 23:

$$p(\theta|y(t_n), \dots, y(t_1)) = \frac{p(\theta, y(t_n), \dots, y(t_1))}{p(y(t_n), \dots, y(t_1))} \quad (36)$$

The numerator of the right-hand side of Equation 36 using Equation 24a becomes

$$p(\theta, y(t_n), \dots, y(t_1)) = p(y(t_n)|\theta, y(t_{n-1}), \dots, y(t_1)) \cdot p(\theta, y(t_{n-1}), \dots, y(t_1)) \quad (37)$$

These operations can be continued repetitively until we get

$$p(\theta, y(t_n), \dots, y(t_1)) = p(\theta) \prod_{i=1}^n p(y(t_i)|\theta, y(t_{i-1}), \dots, y(t_1)) \quad (38)$$

Examination of Equation 30 shows that $Y(t_i)$ depends only on t_i , θ and $\epsilon(t_i)$ and is not conditioned by any previous measurement. Consequently, we can write

$$p(y(t_i)|\theta, y(t_{i-1}), \dots, y(t_1)) = p(y(t_i)|\theta) \quad (39)$$

provided Equation 30 is observed as a constraint. The desired joint conditional probability function is thus

$$p(\theta|y(t_n), \dots, y(t_1)) = \frac{p(\theta) \prod_{i=1}^n p(y(t_i)|\theta)}{p(y(t_n), \dots, y(t_1))} \quad (40)$$

We can get rid of the evidence term $p(y(t_n), \dots, y(t_1))$ because it's constant with respect to the maximization. Moreover, if we are lacking a prior distribution over the quantity we want to estimate, then $p(\theta)$ can be omitted. In such a case:

$$p(\theta|y(t_n), \dots, y(t_1)) = \prod_{i=1}^n p(y(t_i)|\theta) = \prod_{i=1}^n L(\theta|y(t_i)) \quad (41)$$

By collecting a random sample of values of the random variable y and selecting the values of θ that maximize the likelihood function $L(\theta|y)$, a function described in chapter 3 in connection with Bayes' theorem. Such estimators, $\hat{\theta}$, are known as maximum likelihood estimators. In effect, the methods selects those values of θ that are at least as likely to generate the observed sample as any other set of values of the parameters if the probability density of the random variable y were to be extensively simulated through use of the probability density $p(y|\theta)$. In making a maximum likelihood estimate, we assume that the form of the probability density (only the θ need be determined) and that all possible values of θ are equally likely before experimentation.

The likelihood function for the parameters based on several observations is the product of the individual functions if the observations are independent.

$$L(\theta|y(t_n), \dots, y(t_1)) = \prod_{i=1}^n L(\theta|y(t_i)) = p(y(t_1)|\theta) p(y(t_2)|\theta) \dots p(y(t_n)|\theta) \quad (42)$$

In choosing as estimates of θ the values that maximize L for the given values $(y(t_i))$, it turns out that it is more convenient to work with the $\ln L$ than with L itself:

$$\ln L = \ln p(y(t_1)|\theta) + \ln p(y(t_2)|\theta) + \dots + \ln p(y(t_n)|\theta) = \sum_{i=1}^n \ln p(y(t_i); \theta) \quad (43)$$

By assuming that the conditional distribution of \bar{Y}_i , given y_i , is normal, then we form the likelihood function based on the probability density:

$$p(\theta, \sigma|y(t_n), \dots, y(t_1)) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{1}{2\sigma^2} (Y(t_i) - y(\theta, t_i))^2 \right] \\ L(\theta, \sigma|y(t_n), \dots, y(t_1)) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{1}{2\sigma^2} (Y(t_i) - y(\theta, t_i))^2 \right] \quad (44)$$

where σ is the variance

By taking the natural logarithm of the Equation 44, the final form of the objective function can be obtained:

$$\ln L = -n \ln \sqrt{2\pi} - \frac{n}{2} \ln \sigma - \frac{\sum_{i=1}^n [Y(t_i) - y(\theta, t_i)]^2}{2\sigma^2} \quad (45)$$

The parameter estimation problem can be formulated as follow:

$$\hat{\theta}_{MLE} = \arg \max_{\sigma, \theta \in \Theta} \ln L = \arg \max_{\sigma, \theta \in \Theta} p(\theta|y) \\ \text{subject to} \quad \dot{x} = f(t, x, \theta) \\ \quad \quad \quad \dot{\theta} = 0 \\ \quad \quad \quad y = g(x) \quad (46)$$

Based on the first order optimality condition, the $\ln L$ can be maximized with respect to the vector θ by equating to zero the partial derivatives of $\ln L$ with respect to each of the parameters:

$$\frac{\partial \ln L}{\partial \theta} = \frac{\partial \sum_{i=1}^n \ln p(y(t_i)|\theta)}{\partial \theta} = 0 \quad (47)$$

Solution of Equations 47 yield the desired estimates $\hat{\theta}$. For some models, these equations can be explicitly solved for $\hat{\theta}$ but in general no closed-form solution to the maximization problem is known or available, and an MLE can only be found via numerical optimization.

By carrying out this operation, under fairly unrestricted conditions, it can be shown that as n approaches infinity the maximum likelihood estimates have the desirable asymptotic properties:

- (1) $\lim_{n \rightarrow \infty} \mathbb{E}\{\hat{\theta}\} = \theta$
- (2) $\left[\sqrt{n}(\hat{\theta} - \theta) \right]$ is normally distributed

and for the case of two parameters:

$$(3) \lim_{n \rightarrow \infty} \text{Var}\{\hat{\theta}\} = \frac{1}{n} \left[\mathbb{E} \left\{ \left(\frac{\partial \ln p}{\partial \theta} \right)^2 \right\} \right]^{-1} \frac{1}{1 - \rho_{\hat{\theta}}^2}$$

where $\rho_{\hat{\theta}}^2$ is the matrix of correlation coefficients.

Maximum likelihood estimates are not necessarily unbiased. Maximum likelihood estimates, however are efficient and, hence, consistent estimates. Furthermore, where a sufficient estimate can be obtained, the maximum likelihood method will obtain it. Finally, if $\hat{\theta}$ is a maximum likelihood estimator of θ , the $f(\hat{\theta})$ is a maximum likelihood estimator of $f(\theta)$, a function of θ .

As already discussed, the prior information about the estimated parameters can be incorporated in to the maximum likelihood estimator to form a maximum a posteriori probability (MAP) estimate. MAP can be used to obtain a point estimate of an unobserved quantity on the basis of empirical data. MAP estimation can therefore be seen as a regularization of maximum likelihood estimation. The relationship between MAP and MLE can be obtained from the Bayes' rule:

$$\begin{aligned}\hat{\theta}_{\text{MAP}} &= \arg \max_{\sigma, \theta \in \Theta} p(\theta|y) \\ &= \arg \max_{\sigma, \theta \in \Theta} \frac{p(y|\theta)p(\theta)}{p(y)} \\ &= \arg \max_{\sigma, \theta \in \Theta} p(y|\theta)p(\theta) \\ &= \arg \max_{\sigma, \theta \in \Theta} \ln(p(y|\theta)p(\theta)) \\ &= \arg \max_{\sigma, \theta \in \Theta} \ln p(y|\theta) + \ln p(\theta)\end{aligned}\quad (48)$$

We assume the normal distribution of random samples, with identical variance and the prior distribution of θ given by $\mathcal{N}(\theta^0, \sigma_\theta^2)$. The function to be maximized can be written as

$$\begin{aligned}&\arg \max_{\sigma, \theta \in \Theta} \left[\ln \prod_{i=1}^n \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(Y(t_i) - y(\theta, t_i))^2}{2\sigma^2}} + \ln \prod_{j=0}^p \frac{1}{\sigma_\theta \sqrt{2\pi}} e^{-\frac{(\theta - \theta^0)^2}{2\sigma_\theta^2}} \right] \\ &= \arg \max_{\sigma, \theta \in \Theta} \left[-\sum_{i=1}^n \frac{(Y(t_i) - y(\theta, t_i))^2}{2\sigma^2} - \sum_{j=0}^p \frac{(\theta - \theta^0)^2}{2\sigma_\theta^2} \right] \\ &= \arg \min_{\sigma, \theta \in \Theta} \frac{1}{2\sigma^2} \left[\sum_{i=1}^n (Y(t_i) - y(\theta, t_i))^2 + \frac{\sigma^2}{\sigma_\theta^2} \sum_{j=0}^p (\theta - \theta^0)^2 \right] \\ &= \arg \min_{\sigma, \theta \in \Theta} \left[\sum_{i=1}^n (Y(t_i) - y(\theta, t_i))^2 + \lambda_{L2} \sum_{j=0}^p (\theta - \theta^0)^2 \right]\end{aligned}\quad (49)$$

In Equation 49 some of the constants (with respect to θ) were dropped and factored to simplify the expression. We can adjust the amount of regularization we want by changing λ_{L2} . Equivalently, we can adjust how much we want to weight the priors carry on the coefficients (θ). If we have a very small variance (large λ_{L2}) then the coefficients will be very close to 0; if we have a large variance (small λ_{L2}) then the coefficients will not be affected much (similar to as if we didn't have any regularization).

Let's assume that the prior distribution of parameter θ is given by the Laplace distribution. The likelihood function can be defined as follow:

$$\begin{aligned}&\arg \max_{\sigma, \theta \in \Theta} \left[\ln \prod_{i=1}^n \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(Y(t_i) - y(\theta, t_i))^2}{2\sigma^2}} + \ln \prod_{j=0}^p \frac{1}{2b} e^{-\frac{|\theta - \theta^0|}{\sigma_\theta}} \right] \\ &= \arg \max_{\sigma, \theta \in \Theta} \left[-\sum_{i=1}^n \frac{(Y(t_i) - y(\theta, t_i))^2}{2\sigma^2} - \sum_{j=0}^p \frac{|\theta - \theta^0|}{\sigma_\theta} \right] \\ &= \arg \min_{\sigma, \theta \in \Theta} \frac{1}{2\sigma^2} \left[\sum_{i=1}^n (Y(t_i) - y(\theta, t_i))^2 + \frac{\sigma^2}{\sigma_\theta} \sum_{j=0}^p |\theta - \theta^0| \right] \\ &= \arg \min_{\sigma, \theta \in \Theta} \left[\sum_{i=1}^n (Y(t_i) - y(\theta, t_i))^2 + \lambda_{L1} \sum_{j=0}^p |\theta - \theta^0| \right]\end{aligned}\quad (50)$$

The Laplacean prior act as $L1$ -type of regularization on the optimization problem. Instead of preventing any of the coefficients from being too large (due to the squaring),

$L1$ promotes sparsity. Due to the sparsity some of the parameters can be zero out, and eventually removed from the model. $L1$ is more robust to outliers than $L2$.

If the cost function is given by Equations 49 or 50, then the optimization problem can be formulated as follow:

$$\begin{aligned}\hat{\theta}_{\text{MAP}} &= \arg \max_{\sigma, \theta \in \Theta} \ln p(y|\theta) + \ln p(\theta) \\ \text{subject to} &\quad \dot{x} = f(t, x, \theta) \\ &\quad \dot{\theta} = 0 \\ &\quad y = g(x)\end{aligned}\quad (51)$$

If the cost function to be maximized is given by Equation 48 and constrained by Equation 30, then the method of Lagrange multipliers can be used to convert a constrained problem into a form such that the derivative test of an unconstrained problem can be applied. The Lagrange function (L^*) can be defined as $\ln(L)$ plus the Lagrangian multipliers $\lambda(t_i)$ times the constraint function

$$\begin{aligned}L^* &\equiv \ln p(\theta) + \sum_{i=1}^n \{ \ln p(y(t_i)|\theta) \\ &\quad + \lambda^T \cdot [Y(t_i) - y(\theta, t_i) - \epsilon(t_i)] \\ &\quad - \ln [p(y)]\end{aligned}\quad (52)$$

By assumption of the relation of Equation 30

$$p(y(t_i)|\theta, y_0) = p(\epsilon(t_i))\quad (53)$$

and specially

$$p(\epsilon(t_i)) = \frac{1}{(2\pi)^{v/2} |\Gamma(t_i)|^{1/2}} \exp \left[-\frac{1}{2} \epsilon^T(t_i) (\Gamma(t_i))^{-1} \epsilon(t_i) \right]\quad (54)$$

where $|\Gamma|$ is the det Γ , and Γ is the covariance matrix of ϵ , i.e., of the response.

After Equation 53 is substituted into Equation 52, and L^* can be differentiated with respect to each of the estimates and $\epsilon(t_i)$. **Based on the first order optimality conditions**, the resulting expression is equated to zero:

$$\frac{\partial L^*}{\partial \theta} = \frac{\partial}{\partial \theta} \ln p(\theta) - \sum_{i=1}^n \lambda^T(t_i) \frac{\partial}{\partial \theta} y(\theta, t_i) = 0\quad (55a)$$

$$\frac{\partial L^*}{\partial \epsilon(t_i)} = \frac{\partial}{\partial \epsilon(t_i)} \ln p(\epsilon(t_i)) - \lambda^T(t_i) = 0^T\quad (55b)$$

Substitution of Equation 54 into the last Equation of 55 makes it possible to solve for $\lambda(t_i)$:

$$\begin{aligned}\lambda(t_i) &= -(\Gamma(t_i))^{-1} \epsilon(t_i) \\ &= -(\Gamma(t_i))^{-1} [Y(t_i) - y(\theta, t_i)]\end{aligned}\quad (56)$$

and to eliminate $\lambda(t_i)$ from the first equation of 55.

For convenience we shall define a new column vector θ^* in which all the elements of θ are arranged as follows:

$$\theta^* = [\theta_{1,1} \quad \theta_{1,2} \quad \dots \quad \theta_{1,v} \quad \theta_{2,1} \quad \dots \quad \theta_{2,v} \quad \dots \quad \theta_{v,v}]^T\quad (57)$$

Finally, we assume that θ^* are distributed by a joint normal distribution and that the prior distribution are, respectively

$$p(\theta^*) = \frac{1}{(2\pi)^{v/2} |\Omega_{\theta^*}|^{1/2}} \exp \left[-\frac{1}{2} (\theta^* - \theta^{*(0)})^T \Omega_{\theta^*}^{-1} (\theta^* - \theta^{*(0)}) \right]\quad (58a)$$

where Ω 's are the respective covariance matrices for θ^* and y_0 , and the subscript (0) designates the prior estimated of θ^* and y_0 . If we assume that θ^* and y_0 are independent

Introduction of the prior distributions, Equations 58, plus the expression for $\lambda(t_i)$ into the first two equations of 55 gives the final equations from which the estimators of θ^* and y_0 can be obtained:

$$-(\hat{\theta}^* - \theta^{*(0)})^T \Omega_{\theta^*}^{-1} + \sum_{i=1}^n [Y(t_i) - y(\hat{\theta}, t_i)]^T \Gamma^{-1}(t_i) \cdot \frac{\partial}{\partial \theta^*} y(\hat{\theta}, t_i) = 0^T \quad (59a)$$

where the overlay caret denotes estimated parameters.

Note that under the assumption that the elements of Ω are essentially infinite (prior knowledge is diffuse), the elements of Ω^{-1} are zero, the equations for the maximum likelihood estimates coincide with those for the least squares estimates, and the same calculations for precision in the estimates apply.

5. Results

6. Conclusions

References

- [1] Zhen Huang, Xiao han Shi, and Wei juan Jiang. Theoretical models for supercritical fluid extraction. Journal of Chromatography A, 1250: 2–26, aug 2012. doi: 10.1016/j.chroma.2012.04.032.
- [2] E. Reverchon. Mathematical modeling of supercritical extraction of sage oil. AIChE Journal, 42(6):1765–1771, jun 1996. doi: 10.1002/aic.690420627.
- [3] Ernesto Reverchon, Giorgio Donsi, and Libero Sesti Osseo. Modeling of supercritical fluid extraction from herbaceous matrices. Industrial & Engineering Chemistry Research, 32(11):2721–2726, nov 1993. doi: 10.1021/ie00023a039.
- [4] H. Sovova. Rate of the vegetable oil extraction with supercritical co₂. modelling of extraction curves. Chemical Engineering Science, 49 (3):409–414, 1994. doi: 10.1016/0009-2509(94)87012-8.
- [5] Ronald M. Errico. What is an adjoint model? Bulletin of the American Meteorological Society, 78(11):2577–2591, nov 1997. doi: 10.1175/1520-0477(1997)078<2577:wiaam>2.0.co;2.
- [6] Saman Razavi and Hoshin V. Gupta. What do we mean by sensitivity analysis? the need for comprehensive characterization of “global” sensitivity in earth and environmental systems models. Water Resources Research, 51(5):3070–3092, may 2015. doi: 10.1002/2014wr016527.
- [7] L. Fiori, D. Calcagno, and P. Costa. Sensitivity analysis and operative conditions of a supercritical fluid extractor. The Journal of Supercritical Fluids, 41(1):31–42, may 2007. doi: 10.1016/j.supflu.2006.09.005.
- [8] José A. Rabi, Fernando Lima Caneppele, Rogers Ribeiro, and Gustavo C. Dacanal. Scale-up of extraction processes: Dimensionless modeling and virtualization via lattice boltzmann method. Journal of Food Process Engineering, 43(1), sep 2019. doi: 10.1111/jfpe.13244.
- [9] T. Grant Glover and M. Douglas LeVan. Sensitivity analysis of adsorption bed behavior: Examination of pulse inputs and layered-bed optimization. Chemical Engineering Science, 63(8):2086–2098, apr 2008. doi: 10.1016/j.ces.2008.01.001.
- [10] R. M. F. Vargas, E. Cassel, G. M. F. Gomes, L. G. S. Longhi, L. Atti-Serafini, and A. C. Atti-Santos. Supercritical extraction of carqueja essential oil: experiments and modeling. Brazilian Journal of Chemical Engineering, 23(3):375–382, sep 2006. doi: 10.1590/s0104-66322006000300011.
- [11] N. R. Bulley, M. Fattori, A. Meisen, and L. Moyls. Supercritical fluid extraction of vegetable oil seeds. Journal of the American Oil Chemists' Society, 61(8):1362–1365, aug 1984. doi: 10.1007/bf02542243.
- [12] M. Spiro and M. Kandiah. Extraction of ginger rhizome: partition constants and other equilibrium properties in organic solvents and in supercritical carbon dioxide. International Journal of Food Science & Technology, 25(5):566–575, jun 2007. doi: 10.1111/j.1365-2621.1990.tb01116.x.
- [13] A. Srinivasan and C. Depcik. One-dimensional pseudo-homogeneous packed-bed reactor modeling: I. chemical species equation and effective diffusivity. Chemical Engineering & Technology, 36(1):22–32, dec 2012. doi: 10.1002/ceat.201200458.
- [14] David Mautner Himmelblau. Process analysis by statistical methods, 1970. [by] David M. Himmelblau., Includes bibliographical references.
- [15] John Mandel. Fitting a straight line to certain types of cumulative data. Journal of the American Statistical Association, 52(280):552–566, dec 1957. doi: 10.1080/01621459.1957.10501413.

A. Appendix

Table 1
Notation

Symbol	Description	Unit
A	cross-section	m^2
c	concentration in fluid phase	$kg\ m^{-3}$
Cp	specific heat of the fluid	$J\ mol^{-1}\ K^{-1}$
Cp_s	specific heat of the solid	$J\ mol^{-1}\ K^{-1}$
D_e^M	axial mass diffusion coefficient	$m^2\ s^{-1}$
D_e^T	axial heat diffusion coefficient	$m^2\ s^{-1}$
Di	internal diffusion coefficient	$m^2\ s^{-1}$
dp	particle diameter	m
$F(t)$	mass flow-rate	$kg\ s^{-1}$
km	partition coefficient	$[-]$
k^T	thermal conductivity	$W\ m^{-1}\ K^{-1}$
l	characteristic dimension	m
L	total length of the bed	m
m	mass of the oil in solid phase	kg
m_0	initial mass of the oil in solid phase	kg
M_{CO_2}	molecular mass of CO2	$mol\ kg^{-1}$
Np	number of model parameters and control variables	$[-]$
N_θ	number of model parameters	$[-]$
Nu	number of control variables	$[-]$
Nz	number of grid points in z-direction	$[-]$
p	vector of model parameters and control variables	$[-]$
$P(t)$	pressure	bar
Pe	Peclet's number	$[-]$
q	concentration in solid phase	$kg\ m^{-3}$
R	gas constant	$J\ K^{-1}\ mol^{-1}$
Re	Reynolds number	$[-]$
t	time	s
T	temperature	K
T_0	initial temperature	K
V	volume of the extractor	m^3
y	yield	$[-]$
z	length	m
Z	compressibility factor	$[-]$
ϵ	void fraction	$[-]$
ρ	density of the fluid	$kg\ m^{-3}$
ρ_s	solid density	$kg\ m^{-3}$
μ	shape coefficient	$[-]$
θ	vector of model parameters	$[-]$
η	viscosity	cP
Subscript		
0	initial conditions	$[-]$
*	equilibrium conditions	$[-]$