Appendix B

B. CASE STUDY II: FIRST GENERATION PROCESS MODELS

The deterministic models, associated modelling assumptions and uncertain parameter characterisations for the processes comprising the complete process sequence investigated in Case Study II are stated in Appendix B. In addition, an expression is given which is used to account for additional levels of parameter uncertainty due to the violation of desired ratios or ranges in inter-stage state variable measurements. Some additional Uncertainty Analysis results for Case Study II are presented.

B.1 Reaction model

The available bench scale data for the reaction comprises concentration-time profiles of each drug species, reG and reH. The following experimental results and analysis are obtained from private communication with a pharmaceutical company. These are discussed to understand the assumptions in the model. With this data it is assumed an intrinsic kinetic model for the reaction process can be developed. Since the fates of aqueous reagent, reH, solid organic reagent, reG, and the resulting oxidant, oxG, are complex and not well understood, it is not possible to develop a rigorous kinetic model which accounts for these species. Under the conditions used oxG does not appear to be a limiting factor in the kinetics of the drug reactions. Simplified pseudo-first order reaction kinetics based on the stoichiometry shown in Equations B1 and B2, are assumed in the organic solvent phase. The first generation model for the Stage 1 reaction is given in Model B1 and the process diagram is shown in Figure B1.

$$actA_{org} \rightarrow actB_{org} \rightarrow actC_{org} \quad \text{(B1)}$$

$$actD_{org} \rightarrow actE_{org} \quad \text{(B2)}$$

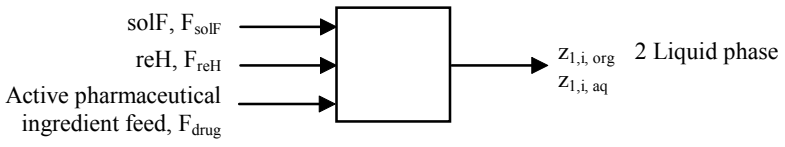


Figure B1. Stage 1 reaction, Case Study II.

Feed specification,

$$f_{drug,i} = \frac{p_{f,i}}{100} F_{drug}$$

$$F_{reH} = molratio_{F_{reH}} RMM_{reH} \frac{f_{drug,1}}{RMM_{drug,1}} \frac{100}{wtaq_{reH}}$$

$$V_{reH} = \frac{F_{reH}}{\rho_{reH}}$$

Drug component mole balance,

$$\frac{dm_{1,org}}{dt} = -k_1 m_{1,org}$$

$$\frac{dm_{2,org}}{dt} = k_1 m_{1,org} - k_2 m_{2,org}$$

$$\frac{dm_{3,org}}{dt} = k_2 m_{2,org}$$

$$\frac{dm_{4,org}}{dt} = 0 \Big|_{t_o,t'} = -k_3 m_{4,org} \Big|_{t',t_f}$$

$$\frac{dm_{5,org}}{dt} = 0 \Big|_{t_o,t'} = k_3 m_{4,org} \Big|_{t',t_f}$$

Solvent balance,

$$z_{1,6,org} = F_{solF}$$

$$z_{1,7,org} = 0$$

$$z_{1,8,org} = F_{reH}$$

End point criteria,

$$z_{1,i,org} = \frac{RMM_i m_{i,org}}{1000} \qquad for \quad i = 1..5$$

$$X_1 = \frac{m_{o,1,org} - m_{1,org}}{m_{o,1,org}}$$

$$comp_i = \frac{z_{1,i,org}}{\sum_i^5 z_{1,i,org}} \times 100 \qquad for \quad i = 1..5$$

Initial conditions,

$$m_{o,i} = \frac{f_{drug,i}}{RMM_{drug,i}} \times 1000 \qquad for \quad i = 1..5 \qquad \text{(Model B1)}$$

where

comp = composition by weight, wt%

f = component feed mass, kg

F = total stream feed mass, kg

k_1 = first order rate constant for main reaction, min^{-1}

k_2 = first order rate constant for consecutive reaction, min^{-1}

k_3 = first order rate constant for sub-reaction, min^{-1}

m = moles

molratio = mole ratio of reagent to moles of active pharmaceutical ingredient in feed

p_f = purity of active pharmaceutical ingredient feed, wt%

RMM = relative molecular mass (in aqueous streams, refers to RMM of solute compound)

V = volume, m^3

wtaq = reagent strength in aqueous solution, wt%

X = conversion

z = mass, kg

ρ = density, kg m^{-3}

subscripts

aq = aqueous phase

i = component species {actA, actB, actC, actD, actE, solF, solL, aq}

o = initial condition

org = organic phase

t_o = reaction starting time (zero), min

t' = time at which sub-reaction starts, min

t_f = total time of Stage 1 operation, min

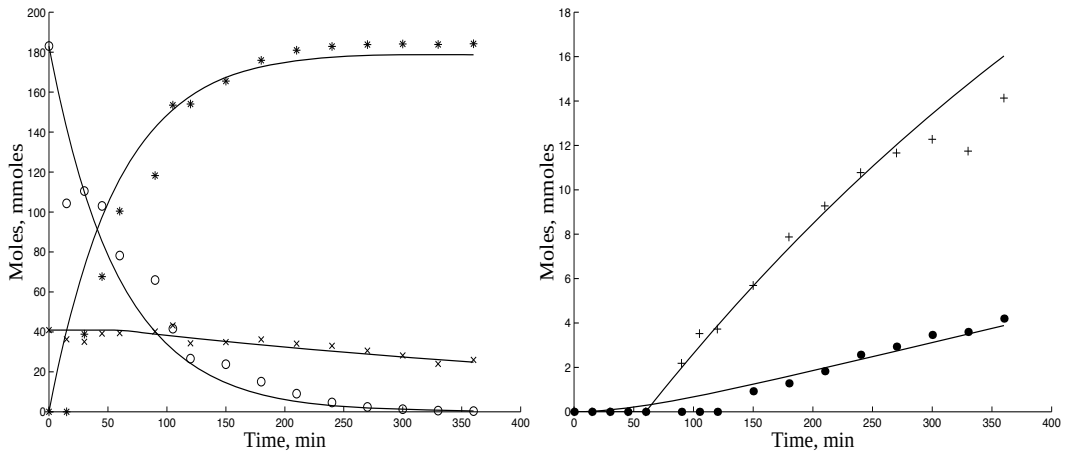
Possible degrees of freedom include F_{drug}, $molratio_{reH}$, $wtaq_{reH}$ and t_f. The main assumptions made in this model include:

- pseudo first order stoichiometry and elementary reaction kinetics (oxG in excess throughout),
- observation of intrinsic kinetics in the assumed reactions (perfect mixing throughout),
- the sub-reaction for the secondary impurity (actE) starts after 60 minutes,
- no feed solids dissolution effects,
- instantaneous addition of reH feed,
- reG and reH species are not explicitly modelled, since their fate is not understood,
- isothermal operation and no other limiting heat transfer effects,
- no mass transfer of drug species from the organic phase to the aqueous phase.

The fitted parameter values for the bench scale model are given in Table B1. Since the predicted drug component profiles exhibit a reasonably good fit to the bench scale data, see Figure B2, the assumptions of pseudo-first order kinetics and perfect mixing appear satisfactory for this system.

Table B1. Parameters for the bench scale Stage 1 model, Case Study II.

Fitted model parameter values	Imposed laboratory conditions
$k_1 = 0.0169$ min^{-1}	$t_f = 360$ min
$k_2 = 7.14 \times 10^{-5}$ min^{-1}	$p_{f,i} = [81.7, 0, 0, 18.3, 0]$ wt%
$k_3 = 1.66 \times 10^{-3}$ min^{-1}	$F_{drug} = 0.100$ kg
$t' = 60$ min	$molratio_{reH} = 10.4$
	$wtaq_{reH} = 30\%$



(a) Key to data points: o = actA, * = actB, x = actD. (b) Key to data points: • = actC, + = actE.

Figure B2. Bench scale drug profile predictions for the first generation Stage 1 model, Case Study II.

B.2 Reagent addition model

A mass balance comprises the first generation model used to describe the addition of aqueous reagent operations. The first generation model for the Stage 2, 4, 5 and 7 reagent addition operations is given in Model B2 and the process diagram is shown in Figure B3.

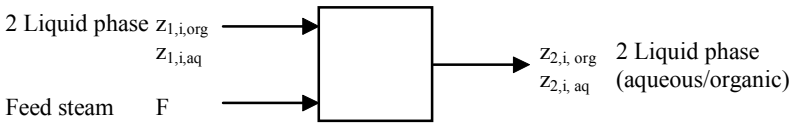


Figure B3. Reagent Addition, Case Study II.

Feed specification,

$$F = V_F \rho_F \qquad \text{for distilled water}$$

or

$$F = molratio_F \, RMM_F \frac{m_{o,1}}{RMM_1} \frac{100}{wtaq_F} \qquad \text{for aqueous solution}$$

Drug component balance,

$$z_{2,i,org} = \left(1 - u_1\right) z_{1,i,org} \qquad for \quad i = 1..5$$

$$z_{2,i,aq} = \left(u_1 z_{1,i,org}\right) + z_{1,i,aq} \qquad for \quad i = 1..5$$

Solvent balance

$$z_{2,i,org} = z_{1,i,org} \qquad for \quad i = 6..8$$

$$z_{2,8,aq} = z_{1,8,aq} + F$$

$$z_{2,i,aq} = z_{1,i,aq} \qquad for \quad i = 6,7 \qquad \text{(Model B2)}$$

where u_1 is a parameter representing the fraction of drugs in the organic phase of the input stream, which are soluble in the aqueous phase and the component species, i, are {actA, actB, actC, actD, actE, solF, solL, aq}. An additional parameter, u_o, not explicitly expressed in Model B2, is used specifically for the Stage 2 dilution model to represent the time in minutes, after the desired Stage 1 termination point ($t_{f,1}$, whereupon the diluent is added), that passes before the reaction has fully terminated. Degrees of freedom could be either V_F for the addition of distilled water or $molratio_F$ and $wtaq_F$ for the addition of an aqueous reagent solution. The main assumptions of this model include:

- instantaneous addition of feed stream, F,
- instantaneous reactions, no mixing effects,
- no limiting heat transfer effects,
- any mass transfer of drugs to the aqueous phase, $z_{2,i,aq}$ due to solubility, is represented by parameter u_1, and is assumed instantaneous and the same proportion for each drug species,
- solubility of solF in the aqueous phase, $z_{2,i,aq}$, is unimportant and is assumed to be zero.

B.3 Layer separation model

A simple mass balance comprises the first generation model used to describe layer separations. The first generation model for the Stage 3, 6 and 8 layer separation operations is given in Model B3 and the process diagram is shown in Figure B4.

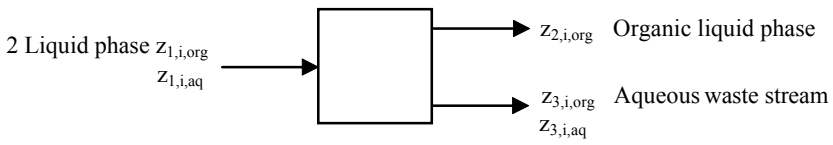


Figure B4. Layer separation, Case Study II.

Component mass balance

$$z_{2,i,org} = (1 - u_2) z_{1,i,org} \qquad for \quad i = 1..8$$

$$z_{3,i,org} = u_2 z_{1,i,org} \qquad for \quad i = 1..8$$

$$z_{3,i,aq} = z_{1,i,aq} \qquad for \quad i = 1..8 \qquad \text{(Model B3)}$$

where u_2 is a parameter representing the fraction of organic phase of the input stream, which is an undesired cut in the aqueous waste stream phase and the component species, i, are {actA, actB, actC, actD, actE, solF, solL, aq}. There are no degrees of freedom in this model. The main assumptions made in this model include:

- light aqueous phase (reH) is disperse, heavy organic phase (solF) is continuous,
- no mixing or time-dependent effects,
- no aqueous phase hold-up in the output organic stream, z_2, so that the efficiency of subsequent chemical destruction or solvent exchange operations is maintained (i.e. instead a small amount of organic phase loss, $z_{3,i,org}$, is incurred in the aqueous phase cut, z_3),
- any organic phase hold up (within a dispersion band) retained in the aqueous waste stream, z_3, is characterised by the parameter fraction u_2,
- no drug solubility in the aqueous phase.

B.4 Batch distillation model

The model used here is a batch distillation from a reboiler, with an energy balance to evaluate the vapour flowrate to the top product. A total condenser is not explicitly modelled but its operation is assumed. The first generation model for the Stage 9, 10 and 11 solvent exchange operations is given in Model B4 and the process diagram is shown in Figure B5.

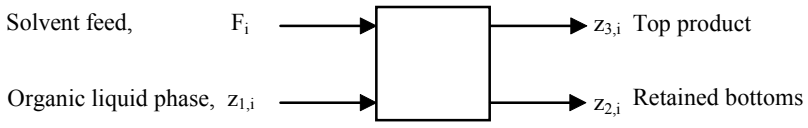


Figure B5. Batch distillation, Case Study II.

Component mole balance,

$$\frac{dm_{L,i}}{dt} = -y_i V_{flow} \qquad for \quad i = 1..8$$

$$z_{2,i} = m_{L,i} RMM_i \qquad for \quad i = 1..8$$

$$z_{3,i} = \left(F_i \frac{comp_{F,i}}{100} \right) - z_{2,i} \qquad for \quad i = 1..8$$

Energy balance

$$\frac{d\left(M_L H_L\right)}{dt} = -H_V V_{flow} + Q_r$$

Vapour-liquid equilibrium

$$y_i = x_i K_i \qquad for \quad i = 1..8$$

$$K_i = \frac{p_{o,i}}{P} \qquad for \quad i = 1..8$$

Total liquid mole balance,

$$M_L = \sum_{i=1}^{8} m_{L,i}$$

Product volume,

$$V_{top} = \sum_{i=1}^{8} z_{3,i} \Big/ \sum_{i=1}^{8} \frac{y_i}{\rho_i}$$

$$V_{bot} = \frac{\left(\sum_{i=6}^{8} z_{2,i}\right) + \left(\sum_{i=6}^{8} z_{2,i}\right)\left(1 - \sum_{i=6}^{8} x_i grad_i\right)}{\sum_{i=6}^{8} \frac{x_i}{\rho_i}}$$

Enthalpy,

$$H_L = \sum_{i=1}^{8} x_i Cp_{L,i}$$

$$H_V = \sum_{i=1}^{8} x_i \left(Cp_{L,i} + dH_{pVo,i}\right)$$

Liquid mole fraction,

$$m_{L,i} = x_i M_L \qquad for \quad i = 1..8$$

Normalisation equations,

$$\sum_{i=1}^{8} y_i = 1$$

$$\sum_{i=1}^{8} x_i = 1$$

Initial conditions,

$$m_{o,i} = \left(F_i \frac{wt\%_{feed,i}}{100} + z_{1,i} \right) \frac{1}{RMM_i} \qquad for \quad i = 1..8 \qquad \text{(Model B4)}$$

where

Cp	=	pure component heat capacity, J/kmol/K
dH_{pVo}	=	pure component heat of vaporisation, J/kmol
F	=	mass solvent feed, kg
H	=	enthalpy, J/kmol
K	=	vapour liquid equilibrium K-value
m	=	moles, kmol
M	=	total moles
p_o	=	pure component pressure, Pa
P	=	total pressure, Pa
Q_r	=	reboiler duty J/h
V	=	volume, m^3
V_{flow}	=	vapour flowrate, kmol/hr
x	=	liquid phase mole fraction
y	=	vapour phase mole fraction
z	=	mass, kg

subscripts

bot	=	bottom product
i	=	component species {actA, actB, actC, actD, actE, solF, solL, aq}
F	=	solvent feed stream
L	=	liquid phase
top	=	top product
V	=	vapour phase

The assumptions made in this model include:

- distillation time is an important variable otherwise a constant vaporisation rate model could be used,
- the condenser is not explicitly modelled (not assumed to be a limiting factor), but total condensation is assumed in the equation for top volume (V_{top}),
- the physical properties, dH_{pVo} and Cp_L, for the drug components are unknown, so the properties for dioctylphthalate (ChemCAD V database, Chemstations, Inc., USA), $C_{12}H_{38}O_4$ (RMM = 390), are assumed due to a similarity in RMM (its properties predict no vaporisation under the range of operating conditions considered in this case study),
- the physical property methods used are: ideal VLE from K-value model, the generic physical property equations for pure component vapour pressure, heat of vaporisation for pure liquid components and pure liquid component heat capacities, as specified in the physical property library of the ChemCAD V simulation software,
- operation at zero reflux and 1 bar pressure,
- the assumption that an estimation for the maximum available reboiler duty per hour is available,
- the equation for bottom volume (V_{bot}) is derived from an assumed linear function between drug solute concentration and solution density when solF is the only solvent present (the bottom volume estimation is only required for the solF single solvent case), where the gradient ($grad_i$) is assumed to be 0.5 for solF (i=6) and the intercept is the pure solvent density (ρ).

Possible degrees of freedom in this model could include the reboiler duty, Q_R, and the initial quantity of mixture, z_1, and pure solvent feed, F.

B.5 Cooling batch crystallisation model

The cooling batch crystalliser model used in this case study incorporates conventional growth kinetics for the product drug component, in which the method of moments is used to solve the population balance, Hulbert and Katz (1964). The first generation model for the Stage 12 crystallisation is given in Model B5 and the process diagram is shown in Figure B6.

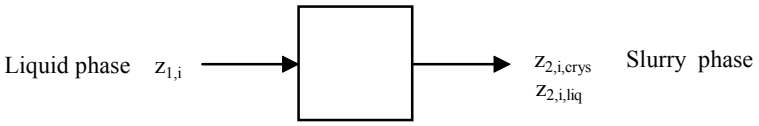


Figure B6. Crystalliser, Case Study II.

Seeded moments,

$$\frac{dN_s}{dt} = 0$$

$$\frac{dL_s}{dt} = GN_s$$

$$\frac{dA_s}{dt} = 2GL_s$$

$$\frac{dV_s}{dt} = 3GA_s$$

Growth kinetics,

$$G = k_g \Delta c_{2,2,liq}{}^{g}$$

Solute concentration balance,

$$\Delta c_{2,2,liq} = c_{2,2,liq} - c_{2,2,liq}\left(T\right)^{*}$$

$$c_{2,2,liq} = c_{o,2,2,liq} - Z_{pc}$$

$$Z_{pc} = V\rho_c f_v$$

Component mass balance,

$$z_{2,2,crys} = Z_{pc} z_{2,7,liq}$$

$$z_{2,2,liq} = z_{1,2} - z_{2,2,crys}$$

$$z_{2,i,crys} = z_{2,7,liq}\left(c_{o,2,i,liq} - c_{2,i,liq}\right) \qquad for \quad i = 1,3,4,5$$

$$z_{2,i,liq} = z_{1,i} - z_{2,i,crys} \qquad for \quad i = 1,3,4,5$$

$$z_{2,i,liq} = z_{1,i} \qquad for \quad i = 6,7,8$$

$$z_{2,i,crys} = 0 \qquad for \quad i = 6,7,8$$

Impurity growth,

$$\frac{dc_{i,liq}}{dt} = \zeta_i c_{2,i,liq} \qquad for \quad i = 1,3,4,5$$

Initial conditions,

$$N_{o,s} = \frac{Z_{o,s}}{f_v \rho_c L_{o,s}^3 z_{1,7,liq}}$$

$$L_{o,s}$$

$$A_{o,s} = \frac{Z_{o,s} F}{\rho_c L_{o,s} z_{1,7,liq}}$$

$$V_{o,s} = \frac{Z_{o,s}}{\rho_c z_{1,7,liq}}$$

$$N_{o,n} = L_{o,n} = A_{o,n} = V_{o,n} = 0$$

$$c_{o,2,i,liq} = \frac{z_{1,i}}{z_{1,7}} \qquad for \quad i = 1,3,4,5$$

Operating policy,

$$\frac{dT}{dt} = CR \Big|_{t_o,t'} = 0 \Big|_{t',t_f}$$

$$t_f = t' + HT \qquad \text{(Model B5)}$$

where

A = total crystal surface area, $m^2\ kg^{-1}$ solvent

c = solute concentration, kg kg^{-1} solvent

c^* = equilibrium solubility solute concentration, kg kg^{-1} solvent

CR = cooling rate, $^oC\ min^{-1}$

f_v = volumetric shape factor

F = overall shape factor

g = kinetic order of growth

G = growth rate, m min^{-1}

HT = holding time, min

k_g = kinetic growth rate constant, m min^{-1} (kg kg^{-1} solvent)$^{1/g}$

L = total crystal length, m kg^{-1} solvent

N = number of crystals, kg^{-1} solvent

t = time, min

t' = time at which holding temperature is achieved, min

t_f = termination time, min

T = temperature, ^oC

V = total crystal volume, $m^3\ kg^{-1}$ solvent

Z = stream mass, kg kg^{-1} solvent

ρ = crystal density, kg m^{-3}

ζ = first order rate constant for loss of impurity concentration, min^{-1}

subscripts

crys = solid phase

i = component species {actA, actB, actC, actD, actE, solF, solL, aq}

liq = liquid phase

o = initial value

pc = product crystal

s = seeds

Possible degrees of freedom for the batch crystallisation model could include: CR, HT, $L_{o,s}$, $Z_{o,s}$. The assumptions in the crystalliser model include:

- seeded operation, with $Z_{o,s}$ kg of crystals of size $L_{o,s}$, and no nucleation,
- a power law function is suitable to describe growth kinetics for the crystallisation of the product drug,
- since a lower holding temperature is believed to lead to increased crystal impurity content, but no data or mechanistic knowledge is available, the holding temperature is not considered a degree of freedom,
- due to the lack of understanding regarding the drug impurity effects, their crystalline presence is explained using first order solute loss functions of liquid phase drug impurity concentration (as opposed to an alternative assumption of linear impurity concentration loss which is not sensitive to changes in the initial value), and independent to temperature (the holding temperature remains constant),
- since data is not available concerning the presence of crystalline impurities other than the drug components (i.e. reG, solF, solL) no characterisation for these effects is portrayed in the model, although the presence of solF in the pre-crystallisation stream is considered an important criterion in this case study,

- estimation of growth rate constant, k_g, is based on solubility data for an alternative high relative molecular mass organic compound in pure solL solvent (based on three temperature data points and fitted with a 2nd order polynomial, RMM = 354, Crossfire Beilstein Database, Beilstein Chemiedaten und Software GmbH) since solubility data for the drug is not available in this study,
- the limitation in process understanding precludes crystal size distribution (CSD) prediction,
- an assumed fixed value for growth rate order (g = 1.2) due to lack of profile data points,
- perfect cooling control at a constant rate and associated heat transfer effects are not limiting,
- size independent growth,
- perfect spheres assumed for overall shape factor (F) and volumetric shape factor (f_v),
- the initial mixture is at the composition boiling point predicted by the ideal VLE batch distillation model with the physical properties of dioctylphthalate (ChemCAD V) used to represent the unknown drug properties.

B.6 Filtration model

The first generation filtration operation is described with a simple mass balance. The lack of available data precludes the use of conventional filtration/centrifugation models found in chemical engineering literature. The first generation model for the Stage 13 washing operation is given in Model B6 and the process diagram is shown in Figure B7.

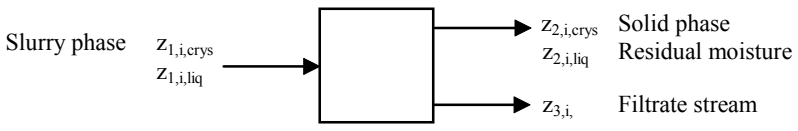


Figure B7. Filtration unit, Case Study II.

Component mass balance,

$$z_{2,i,crys} = z_{1,i,crys} \qquad for \quad i = 1..8$$

$$LOD = \frac{\sum_{i=1}^{8} z_{2,i,liq}}{\sum_{i=1}^{8} z_{2,i,liq} + \sum_{i=1}^{8} z_{2,i,crys}} \times 100\%$$

$$\frac{z_{1,i,liq}}{\sum_{i=1}^{8} z_{1,i,liq}} = \frac{z_{2,i,liq}}{\sum_{i=1}^{8} z_{2,i,liq}} = \frac{z_{3,i,liq}}{\sum_{i=1}^{8} z_{3,i,liq}} \qquad for \quad i = 1..8$$

Batch filtration time,

$$t_f = FR \sum_{i=1}^{8} z_{1,i,crys} \qquad \text{(Model B6)}$$

where

FR = filtration rate, min kg^{-1} solids

i = component species {actA, actB, actC, actD, actE, solF, solL, aq}

LOD = level of dampness in solids, %

t_f = operation time, min

No degrees of freedom are associated with this model, since not enough information is available to justify a model which relates an operating policy to performance. The assumptions for the filtration model include:

- no change in the slurry liquid composition such that the composition of drugs in entrained in the damp solids, $z_{2,i,liq}$ is the same as in the filtrate, $z_{3,i,liq}$,
- no change in dry solids composition or mass,
- a pre-determined desired value of the level of dampness (LOD) is achieved,
- a fixed processing rate per mass of solids, independent of scale, LOD and CSD.

B.7 Washing model

The first generation washing model consists of a mass balance with displacement of residual moisture with wash solvent. The first generation model for the Stage 14 washing operation is given in Model B7 and the process diagram is shown in Figure B8.

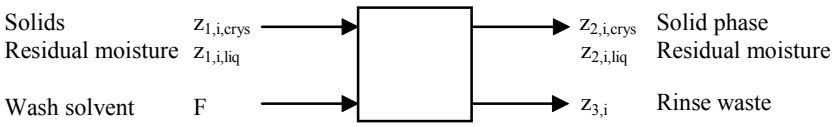


Figure B8. Washing unit, Case Study II.

Non-wash solvent component mass balance,

$$LOD = \frac{\sum_{i=1}^{8} z_{2,i,liq}}{\sum_{i=1}^{8} z_{2,i,liq} + \sum_{i=1}^{8} z_{2,i,crys}} \times 100\%$$

$$z_{2,i,crys} = z_{1,i,crys} \qquad for \quad i = 1..8$$

$$z_{2,i,liq} = \left(1 - \eta_{wash}\right) z_{1,i,liq} \qquad for \quad i = 1..6,8$$

$$z_{3,i} = \eta_{wash} z_{1,i,liq} \qquad for \quad i = 1..6,8$$

Wash solvent component (solL, i = 7) mass balance,

$$z_{2,7,liq} = \left(1 - \eta_{wash}\right) z_{1,7,liq} + \left(\eta_{wash} \sum_{i=1, i \neq 7}^{8} z_{1,i,liq} \right)$$

$$z_{3,7} = \left(\eta_{wash} z_{1,7,liq}\right) + \left(F - \eta_{wash} \sum_{i=1, i \neq 7}^{8} z_{1,i,liq} \right) \qquad \text{(Model B7)}$$

where

F = mass of wash solvent feed, kg

LOD = level of dampness in solids, %

t_f = operation time, min

η_{wash} = wash efficiency, representing the split fraction of initial residual moisture which is replaced with pure wash solvent

and subscript i represents the component species {actA, actB, actC, actD, actE, solF, solL, aq}. No degrees of freedom are associated with this model, since not enough information is available to justify a model which relates an operating policy to performance. Assumptions for the washing model include:

- a final LOD equal to the initial LOD is achieved,
- a fractional displacement of the initial residual moisture ($z_{1,i,liq}$) with pure wash solvent (F) represented with an assumed wash efficiency (η_{wash}),
- negligible dissolution of crystalline drug components in the pure solL wash solvent (operated at ambient temperature),
- the composition of the displaced residual moisture is equal to the composition of the initial moisture,
- no change in dry solids composition or mass,
- a fixed wash time, independent of scale.

B.8 Dryer model

Mass and heat transfer effects are likely to be complex and a lack of data and general understanding of the drying process permits only a simple mass balance model based on an efficiency measure in drying rate. The first generation model for the Stage 15 drying operation is given in Model B8 and the process diagram is shown in Figure B9.

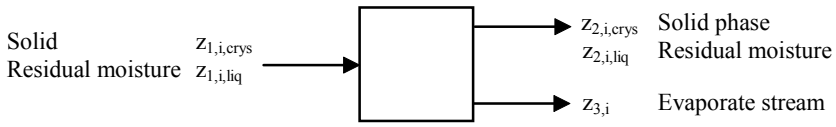


Figure B9. Drying unit, Case Study II.

Drug component mass balance,

$$z_{2,i,crys} = z_{1,i,crys} \qquad for \quad i = 1..8$$

$$z_{2,i,liq} = z_{1,i,liq} \qquad for \quad i = 1..5$$

Solvent component mass balance

$$LOD = \frac{\sum_{i=1}^{8} z_{2,i,liq}}{\sum_{i=1}^{8} z_{2,i,liq} + \sum_{i=1}^{8} z_{2,i,crys}} \times 100\%$$

$$\sum_{i=6}^{8} z_{2,i,liq} + \sum_{i=1}^{5} z_{2,i,liq} = \left(\frac{\sum_{i=1}^{8} z_{2,i,crys}}{\frac{100}{LOD} - 1} \right)$$

$$\frac{z_{2,i,liq}}{z_{1,i,liq}} = \frac{\sum_{i=6}^{8} z_{2,i,liq}}{\sum_{i=6}^{8} z_{1,i,liq}} \qquad for \quad i = 6...8$$

$$z_{3,i} = z_{1,i,liq} - z_{2,i,liq} \qquad for \quad i = 6...8$$

$$p_{drycrys} = \frac{z_{2,i,crys} + z_{2,i,liq}}{\sum_{i=1}^{5} z_{2,i,crys} + \sum_{i=1}^{5} z_{2,i,liq}} 100\% \qquad for \quad i = 1...5$$

$$t_f = DR \sum_{i=1}^{8} z_{1,i,crys} \qquad \text{(Model B8)}$$

where

DR = drying rate, min kg^{-1} solids

LOD = level of dampness in solids, %

$p_{drycrys}$ = purity of final crystals, dry weight percent % (excluding solvent moisture)

and subscript i represents the component species {actA, actB, actC, actD, actE, solF, solL, aq}. No degrees of freedom are associated with this model, since not enough information is available to justify a model which relates an operating policy to performance. Assumptions for the drying model include:

- no change in dry solids composition or mass,
- any drug components in the initial residual moisture ($z_{1,i,liq}$) is retained and does not leave in the evaporate ($z_{3,i}$),
- a pre-determined desired value of the LOD is achieved,
- a fixed processing rate per mass of solids (DR), independent of scale, LOD, CSD and ambient temperature.

B.9 Uncertainty in first generation models

Table B2. Uncertainty characterisation in the parameters of the first generation models, Case Study II.

Stage	Parameter (stochastic model parameter index)	Normal distribution, N(μ, σ) Uniform distribution, U(min, max)
1	k_1 (1), k_2 (2)	$M\begin{pmatrix} k_1^* = 1.69\times10^{-2} \\ k_2^* = 7.14\times10^{-5} \end{pmatrix}, \hat{V} = \begin{bmatrix} 2.27\times10^{-6} & -1.80\times10^{-8} \\ -1.80\times10^{-8} & 1.48\times10^{-10} \end{bmatrix}$
	k_3 (3)	N(1.67×10^{-3}, 1.51×10^{-4})
	t′ (4)	N(60, 10% nominal)
2	u_o (5)	U(3, 6)
	u_1 (6)	U(0, 0.01)
3	u_2 (7)	U(0, 0.01)
4	u_1 (8)	U(0, 0.01)
5	u_1 (9)	U(0, 0.01)
6	u_2 (10)	U(0, 0.01)
7	u_1 (11)	U(0, 0.01)
8	u_2 (12)	U(0, 0.01)
9	$v_{p,solF,A}$ (13)	N(101.6, 0.25% nominal)
	$v_{p,solF,C}$ (14)	N(-12.2, 0.25% nominal)
	$grad_{solF}$ (15)	N(0.49, 10% nominal)
10	$v_{p,solF,A}$ (13)	N(101.6, 0.25% nominal)
	$v_{p,solF,C}$ (14)	N(-12.2, 0.25% nominal)
	$grad_{solF}$ (15)	N(0.49, 10% nominal)
11	$v_{p,solF,A}$ (13)	N(101.6, 0.25% nominal)
	$v_{p,solF,C}$ (14)	N(-12.2, 0.25% nominal)
	$grad_{solF}$ (15)	N(0.49, 10% nominal)
12	k_g (16)	N(6.61×10^{-5}, 2.31×10^{-5})
	C_{21oC}^* (17)	N(0.0056, 5% nominal)
	C_{30oC}^* (18)	N(0.0110, 5% nominal)
	C_{78oC}^* (19)	N(0.2530, 5% nominal)
	ζ_{actA} (20)	N(0.0025, $\sigma_{\zeta_{actA}}$) $\sigma_{\zeta_{actA}} = f\{ratio(solL:product)\}$
	ζ_{actC} (21)	N(0.0058, $\sigma_{\zeta_{actC}}$) $\sigma_{\zeta_{actC}} = f\{ratio(solL:product)\}$
	ζ_{actD} (22)	N(0.0021, $\sigma_{\zeta_{actD}}$) $\sigma_{\zeta_{actD}} = f\{ratio(solL:product)\}$
	ζ_{actE} (23)	N(0.0034, $\sigma_{\zeta_{actE}}$) $\sigma_{\zeta_{actE}} = f\{ratio(solL:product)\}$
13	FR (24)	N(0.5, 10% nominal)
	LOD (25)	N(25, 10% nominal)
14	η_{wash} (26)	U(0, 1)
	LOD (27)	N(25, 10% nominal)
15	DR (28)	N(2.0, 10% nominal)
	LOD (29)	N(6, 10% nominal)

B.10 Uncertainty from violation of predetermined operating ranges

To accommodate the lack of understanding and mechanistic knowledge regarding possible consequences due to deviations from desired operating conditions obtained from design of experiment analyses, an extra degree of uncertainty is incorporated. This introduces a form of the stochastic system model where the uncertainty is dynamic, dependent on future decisions or knowledge. In these instances it is assumed that a deviation of a particular measured criterion from the desired value or range results in an increase in the prior uncertainty for a parameter characterising the possible consequence,

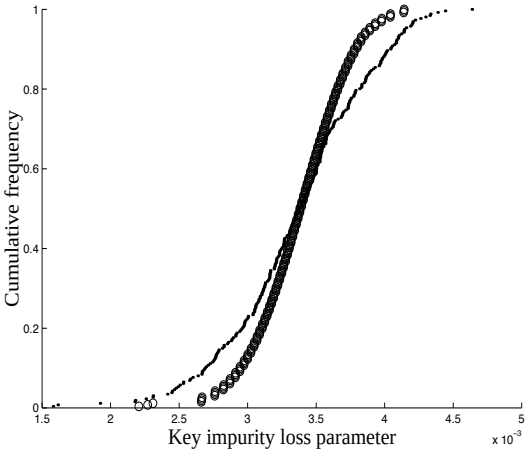
$$\sigma_{\theta,i} = \begin{cases} \sigma_{o,\theta} & \\ \quad if & \left| Q_{j,i} - \dfrac{Q_j^{UB} + Q_j^{LB}}{2} \right| \le \dfrac{Q_j^{UB} - Q_j^{LB}}{2} \\ \sigma_{o,\theta} & \times \left(1 + \sum_j^J g_j \left(\left| Q_{j,i} - \dfrac{Q_j^{UB} + Q_j^{LB}}{2} \right| - \dfrac{Q_j^{UB} - Q_j^{LB}}{2} \right) \right) \\ \quad if & \left| Q_{j,i} - \dfrac{Q_j^{UB} + Q_j^{LB}}{2} \right| \ge \dfrac{Q_j^{UB} - Q_j^{LB}}{2} \end{cases} \quad for \quad \begin{matrix} \theta = 1 \ldots \Theta \\ j = 1 \ldots J \\ i = 1 \ldots I \end{matrix} \qquad \text{(B3)}$$

where $\sigma_{\theta,i}$ is the standard deviation used to generate the ith parameter scenario θ_i, $\sigma_{o,\theta}$ is the standard deviation of uncertain parameter θ used if no range violation occurs, Q_j is the value of the jth criterion and g is the factor by which $\sigma_{\theta,i}$ increases linearly from $\sigma_{o,\theta}$ with deviation of Q_j outside the desired criterion range Q_j^{UB} and Q_j^{LB}. In a conservative assumption, the standard deviation for a given uncertain parameter increases additively for deviations from multiple criteria ranges (which may be associated with the uncertain parameter in question), as shown in Equation B3. In the event of a criterion deviation outside the desired range, the uncertain parameter scenario is re-sampled from the newly characterised distribution.

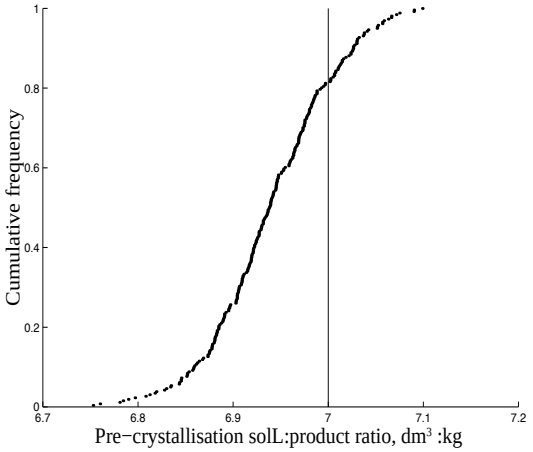
For the violation of the total initial and pre-crystallisation desired solL solvent volume to product mass ratio operating ranges (14-15 and 7-8, respectively) it is assumed that the uncertainty (standard deviation) in the Stage 12 crystallisation parameters characterising the crystal impurity content, ζ_i, increase linearly at a rate of unity with extent of the (additive) deviations from the limits of the initial and final solvent ratio ranges.

B.11 Uncertainty Analysis results for the first generation of models, Case Study II.

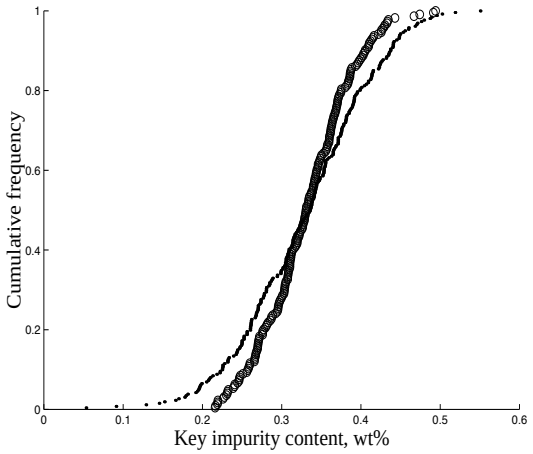
Figure B10 shows the quantitative effect of employing the expression for additional uncertainty (Equation B3) due to the violation of the desired initial and final range in the predicted solvent to product ratio (latter shown in Figure B10 (b)) on the crystallisation key impurity 'solute loss' parameter (ζ_{actC}), Figure B10 (a), and on the endpoint impurity content (wt_{actC}), Figure B10 (c).



(a) Key impurity 'solute loss' parameter.
Key: • = with additional uncertainty,
o = without additional uncertainty.



(b) Pre-crystallisation solL solvent volume to product mass ratio.



(c) Key impurity content.
Key: • = with additional uncertainty,
o = without additional uncertainty.

Figure B10. Effect of additional uncertainty in crystallisation 'solute loss' parameter for the key impurity due to violation of desired solL solvent volume to product mass operating range, Case Study II.

Figure B11 shows the arrival of sampling convergence in the evolution of the mean and variance parameters for the predicted total yield. The effect of inducing rank correlation using Iman and Conover (1982) technique, as expressed in Figure B12, is discussed. The parameter regression covariance matrix for the knowledge level 0 Stage 1 reaction rate constants, k_1 and k_2, was determined assuming linearisation around the optimal estimates from which the associated correlation matrix is obtained (as stated in Section 5.4),

$$\hat{V} = \begin{bmatrix} 2.27 \times 10^{-6} & -1.80 \times 10^{-8} \\ -1.80 \times 10^{-8} & 1.48 \times 10^{-10} \end{bmatrix}, \hat{C} = \begin{bmatrix} 1.0 & -0.9820 \\ -0.9820 & 1.0 \end{bmatrix} \tag{B4}$$

$\hat{C}$ is set as the desired correlation matrix for the sample generated k_1 and k_2 vectors. A matrix K, obtained from an independently generated Hammersley sequence sample, has a correlation matrix E,

$$E = \begin{bmatrix} 1.0 & -0.0024 \\ -0.0024 & 1.0 \end{bmatrix} \tag{B5}$$

Since E is close to the identity matrix (the correction for K^* is not required) and the correlation and rank correlation (E^* and E^*_{rk}) matrices of K^* are close to each other,

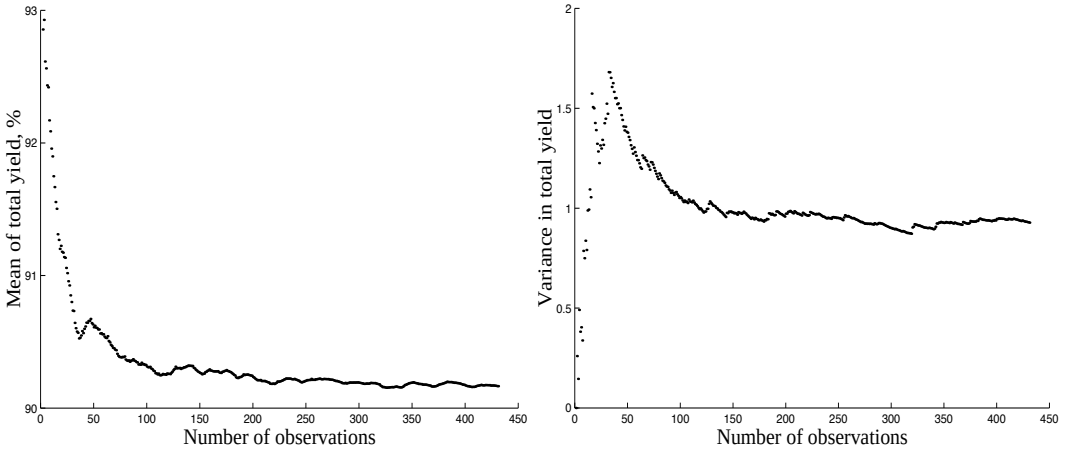
$$E^* = \begin{bmatrix} 1.0 & -0.9821 \\ -0.9821 & 1.0 \end{bmatrix}, \quad E^*_{rk} = \begin{bmatrix} 1.0 & -0.9829 \\ -0.9829 & 1.0 \end{bmatrix} \tag{B6}$$

then the desired rank correlation can be induced into k_1 and k_2 sample vectors by rearranging the elements according to the rank order of K^*. The resulting sample correlation matrix for the 431 sample of k_1 and k_2 is,

$$\hat{C}_s = \begin{bmatrix} 1.0 & -0.9866 \\ -0.9866 & 1.0 \end{bmatrix} \tag{B6}$$

which is close to the desired correlation matrix, $\hat{C}$. Figure B12 (a) shows the contrast of the rearranged observations of the independent Hammersley sample matrix for a desired correlation of -0.9820 (dots) to the unit hyper-cube sample matrix (circles), before inversion over the standard normal cumulative distribution. X^* is expressed for the normally distributed k_1 and k_2 parameters with the same desired correlation in Figure B12 (b). The circles represent the distributed observations before the induced correlation and the dots represent the observations after the rearrangement. The 95% confidence regions assume linearisation of the model about the optimal parameter estimates (see Section 5.4). The inducement of the correlation in the sample appears to be reasonable compared to the 95% confidence

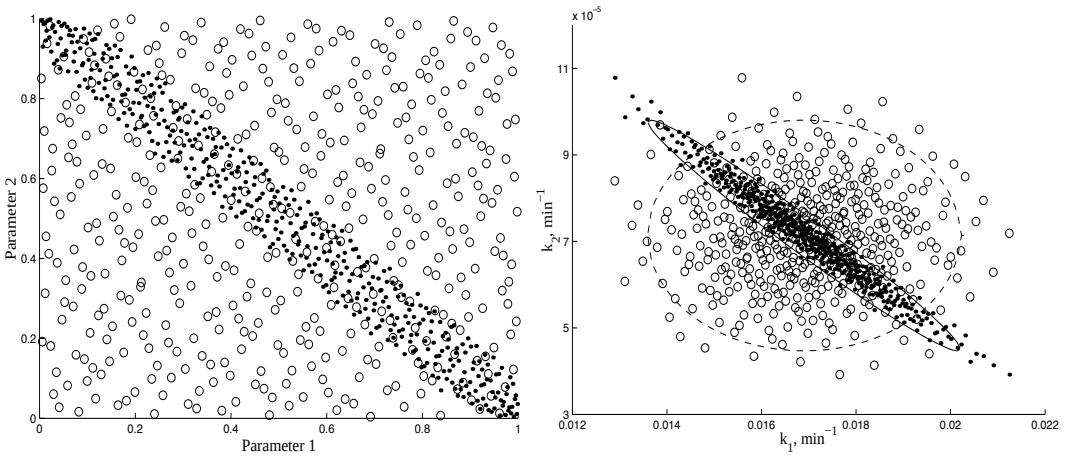
region for the correlated parameters (the solid ellipse). The scatter plots in Figure B13 shows that the linear Sensitivity Analysis measures (CC and SRC) based on unranked data are adequate measures of the key uncertain parameter contributions for the performance criteria shown.



(a) Mean.

(b) Variance.

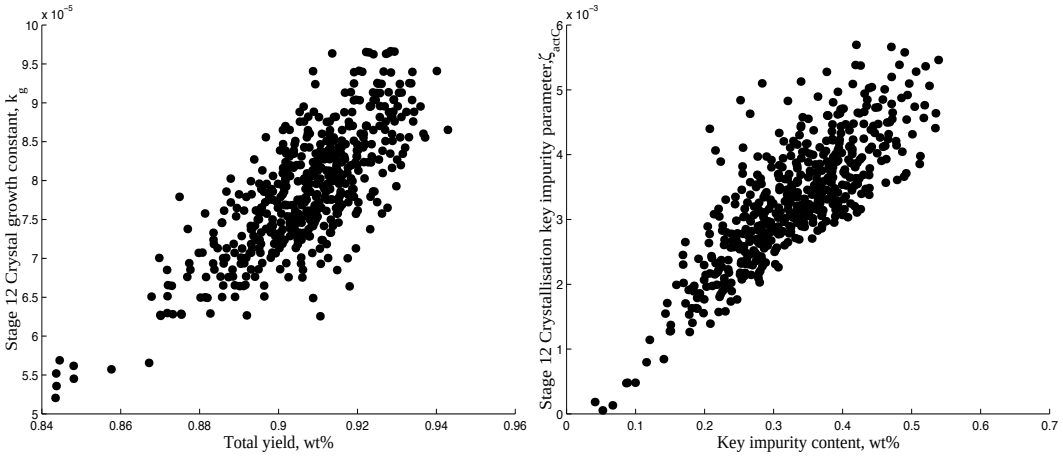
Figure B11. Evolution of distribution parameter estimates with sample observations for the total yield of the first generation of models (knowledge level 0), Case Study II.



(a) The first two dimensions of the unit hyper-cube sample (before inversion over the standard normal distribution) Key: o = initial Hammersley sample, • = rearranged sample.

(b) Normally distributed k_1 and k_2 reaction rate constants. Key: o = initial Hammersley sample, • = rearranged sample, --- = 95% confidence region for uncorrelated sample, — = 95% region for rank correlated sample.

Figure B12. Scatter plots showing the effect of induced rank correlation (ρ = -0.98) in the Stage 1 rate constant parameters for the first generation of models (knowledge level 0), Case Study II.



(a) Stage 12 crystal growth constant and total yield (SRC = 0.82).

(b) Stage 12 crystallisation key impurity 'solute loss' parameter and key impurity content (SRC = 0.78).

Figure B13. Scatter plots of the key uncertain parameters with endpoint criteria for the first generation of models (knowledge level 0), Case Study II.