

A QUANTITATIVE RISK ANALYSIS APPROACH TO A PROCESS SEQUENCE UNDER UNCERTAINTY – A CASE STUDY

David B. Johnson¹ and I. David L. Bogle

Centre for Process Systems Engineering, Dept of Chemical Engineering
University College London, Torrington Place, London, WC1E 7JE

Abstract

Process plants for the manufacture of pharmaceutical products often need to be designed and built quickly to make the most of available patent life which necessitates using uncertain or unavailable data. It is common that pilot plant equipment and data are available and new data can be generated if they are important. We present a model based approach to risk analysis to aid design for pharmaceutical processes which combines systematic modelling procedures with Hammersley sampling based uncertainty analysis and sensitivity analysis used to quantify predicted performance uncertainty and to identify key uncertainty contributions. The main contribution of the paper is the demonstration of the methodology on an industrial case study where the process flowsheet was fixed and some pilot data was available. Expected performance was improved by considering the propagation of uncertainty over the whole process. The case study results indicate the importance of considering uncertainty systematically and quantitatively. The methodology showed the opportunity to improve process performance potential through considering uncertainty systematically.

Keywords:

pharmaceutical processing; risk analysis; uncertainty; design;

1. Introduction

There is a very significant degree of uncertainty in biological and chemical pharmaceutical manufacturing processes which throws up many technical, financial and regulatory challenges (see for example Rogers and Ierapetritou, 2015). These pressures mean that process design decisions often need to be made despite the scarcity of available process knowledge and with little

¹ Current address: NEST Corporation, London

understanding of the physico-chemical phenomena characterising the process. A more structured approach to process development has the potential to save development resources, improve the quality of the final process and help ensure safety and added value in the products. In addition the efficiency of the process may be enhanced.

The regulatory authorities have required formal Quality by Design processes to systematically incorporate risk-based methods (Yu et al.) sometimes using modelling and simulation (O'Connor et al., Escotet-Espinoza et al., Gomes et al., Montes). A number of systematic methods for the generation and use of data from pilot and full scale plant have been published using design of experiments (Mandenius and Brundin) and Monte-Carlo sampling (Severson et al.).

The PSE community have developed a range of tools for the systematic analysis of uncertainty: as a two-stage stochastic programming problem for continuous processes (Pistikopoulos and Ierapetritou), for batch processes (Ierapetritou and Pistikopoulos), dynamic programming (Tsay et al.), identification-based optimisation (Wang and Baldea), multi-objective optimization (Sun and Lou), using a polynomial chaos-based approach (Shen and Braatz), chance-constrained programming (Li et al.) and optimization with probabilistic constraints (Lapteva et al.). A substantive case study using an optimization based approach was presented by Steimel and Engell. Bernardo et al. considered the role of different robustness criteria and Datskov et al. considered the problem at a later stage of the development cycle when there is operation data but it is not sufficient. Hong et al. specifically consider the challenges for biopharmaceutical manufacturing control where batch processes dominate.

Application of a systematic method for the robust optimisation of a 'here and now' operating policy under model parameter uncertainty was presented in a previous paper (Johnson and Bogle). It showed for two small examples that a more robust process can be obtained with respect to some criteria but not in others. In this paper, a much larger industrial case study is studied using a full methodology from first principles modelling through several cycles of data incorporation to produce a robust process. The process studied here, a complete integrated manufacturing process sequence, is presented using both pilot plant and production plant data. First uncertainty and sensitivity analysis techniques are employed to manage the evolving uncertainty in developing models with incoming process modelling information. Then optimisation for desired levels of prediction uncertainty reduction is used to show the required levels of input parameter uncertainty reduction. The objectives of the case study are to provide:

- a quantification of uncertainty,

- a priority list for required process knowledge
- a quantification of the effect of increased process knowledge
- analysis of optimal trade-off between the reduction in uncertainty performance criteria and that required in the uncertain parameters
- an optimal robust operating policy in the available decision variables
- the value of specific information for potentially measurable process inherent uncertainties
- a measure of tolerance to error in the optimal robust operating policy variables

The methodology is outlined in section 2. The process operations which constitute the integrated sequence are described briefly in Section 3 with further information in the supplementary material. In Section 4, the risk analysis problem is defined for a first generation of models based on the initial data available and assumptions. Iterative results of the effect of new data on the modelling effort in the risk analysis approach are summarised in Section 5. The results of the uncertainty analysis and the sensitivity analysis are discussed in section 6. Minimum reductions in input model parameter uncertainty required to achieve desired reductions in the uncertainty predicted in the important output criteria are determined in Section 7. Some conclusions on the effectiveness of the methodology are presented in section 8. All Appendices are in the Supplementary Materials.

2. A methodology for risk assessment

The methodology we proposed (Johnson and Bogle) introduces a process for the management of uncertainty associated with model representations of current knowledge. It assumes the conceptual design is already chosen, that models have been developed, there is no specific consideration of natural variability or structural uncertainty, model parameter uncertainties may be characterised by uniform, normal or other types of stable probability distributions, and the uncertainty is static in time unless they are associated with an external action to the process. Given the inevitably large amount of uncertainty the methodology aims to:

1. Reduce uncertainty by improving the current models and parameters
2. Manipulate the operating policy to improve robustness to model uncertainty by adjusting available process variables.
3. Consider process alternatives

In this paper we consider the first two stages for the pharmaceutical process case study.

The whole methodology is shown in Figure 1. The key uncertainty analysis elements are in the lower dotted box. Steps 1 to 7 are conventional process model development stages. In step 8 quantitative estimation of the uncertainties is determined by available data. Step 9 defines the stochastic system and a sampling procedure is invoked in step 10 (for more details see Johnson). The stochastic system is solved in step 11 to obtain probability distributions and distribution parameters. This is achieved by sequential simulations of the deterministic model of the complete flowsheet, steps 12 and 13. In step 14 a convergence test is employed to terminate the stochastic model solution procedure when the sample estimates of the parameters characterising the distributions (mean and variance) of the stochastic model output criteria are deemed to remain sufficiently unchanged with increasing samples. The convergence test used here is the tolerance limit on the average squared deviation measured in the distribution parameter from the estimate at the current iteration observation over a specified number of preceding iteration estimates. The results from the uncertainty analysis are compared to independent data to validate the uncertain model in step 15. In step 16 sensitivity analysis is used to estimate the ranking priority of the key stochastic inputs contributing to the uncertainty in the stochastic process output criteria. In step 17 the optimal reduction in key parameter uncertainties is determined and this determines the decisions concerning what data to collect or what new experiments to plan to improve the model uncertainty.

Figure 1.

3. Process Description

The Case Study is based on an operating process with data provided by a pharmaceutical company. The components cannot be given in their chemical form for proprietary reasons. The process objective is the production of a crystalline drug product of key component actB and of consistent purity with respect to the by-product impurities, actC and actE, from feed solids comprising the active pharmaceutical ingredient (API), actA, and a stereo-isomer impurity, actD. The considered sequence comprising of 15 stages is shown in Figure 2. The key process conditions are summarised in Table 1.

Figure 2.

Table 1. Summary of key process operating conditions for the Case Study

(LOD is the level of dampness in solids)

Stage	Key process conditions
1. Reaction	77 wt% \pm 3 wt% actA purity API feed solids and catalyst. 2 mol eq. (ratio to feed actA) reG solids. 10.4 mol eq. (ratio to feed actA) 30% aq. reH solution, controlled addition rate to maintain a constant temperature, T ₁ . Maintain a constant temperature, T ₁ \pm 1 C°, throughout entire
2. Dilution	1 volume eq. (ratio to Stage 1 reH) distilled water. 15 min agitation period.
3. Layer separation	30 min settling period. Drain heavy organic phase to parallel vessel.
4. reH destruction	0.4 mol eq. (ratio to feed actA) 6% aq. baseL solution per shot, pending litmus paper test for residual reH presence. 15 min agitation period
5. reG destruction	0.7 mol eq. (ratio to feed actA) 50% aq. baseJ solution. Agitate mixture for 120 min at a constant temperature, T ₅ .
6. Layer separation	30 min settling period. Drain heavy organic phase to original vessel.
7. pH neutralisation	0.7 mol eq. (ratio to actA feed) baseK solids 15 min agitation period.
8. Layer separation	30 min settling period. Drain heavy organic phase to parallel vessel.
9. solF distillation	Distil solF until vessel minimum stir volume is reached. 1 bar pressure and zero reflux.
10. 1st solL distillation	Add a fraction of the total solL volume:product ratio between 14 and 15. Distil a fraction of the solL. 1 bar pressure and zero reflux.
11. 2nd solL distillation	Add remaining fraction of the total solL volume:product ratio. Distil solL to achieve a final solL volume:product ratio between 7 and 8. 1 bar pressure and zero reflux.
12. Crystallisation in solL	Cool boiling mixture to 25 C° and hold for 60 min. 1 bar pressure.
13. Filtration	Vacuum filter the slurry at a constant temperature, T ₁₃ , until ~W ₁₃ % LOD is achieved.
14. Washing	Rinse the damp solids with a 2 volume:product ratio of pure solL at a constant temperature, T ₁₄ , and refilter to the prior LOD.
15. Drying	Dry with pure N ₂ at a high temperature to a low LOD value ~W ₁₅ %.

An aqueous-organic liquid phase chemical reaction takes place in Stage 1. The description for this process stage is based on the details obtained from private communication with the pharmaceutical company. The stoichiometric reactions believed to be occurring are shown in equations 1 to 4. The reaction objective is to produce the drug product, actB, from the feed. Feed solids of actA, stereo-isomer actD and reagent reG dissolve in the organic solvent, solF. Controlled addition of the aqueous reagent, reH, leads to the production of an oxidant, oxG, in a reaction between aqueous reH and dissolved reG, believed to occur at the aqueous-organic phase interface, see Figure 3. This oxidant reacts with the dissolved drug components in the organic phase. The key feed API (actA) is oxidised to the desired product (actB). Over-oxidation leads to the formation of actC from actB in a consecutive reaction. actC is the critical impurity believed to cause problems in the morphology during the crystallisation of the final product. In parallel, the feed impurity (actD) is oxidised to a secondary impurity (actE).



The following seven operations, Stages 2 to 8, provide a termination of the reaction and treatment of the residual reagents. The termination is precipitated by rapid dilution of the aqueous solution of reH with addition of distilled water followed by a period of stirring, Stage 2. An aqueous-organic layer separation is conducted after a period of settling, Stage 3. The heavy organic phase is drained to another vessel and the aqueous phase is sent to waste. Residual reH and reG are destroyed using aqueous solutions of baseI and baseJ in Stage 4 and Stage 5, respectively. The organic phase is drained back to the original vessel in another layer separation, Stage 6. An aqueous solution of baseK is charged to effect a pH neutralisation, Stage 7, followed by a final layer separation, Stage 8.

The next three operations effect a solvent exchange from solF for a crystallisation from an organic solvent, solL. solF is distilled from the vessel to a minimum concentrate, Stage 9. A fraction of a predetermined total volume of solL is added in Stage 10. Some fraction of this is removed in the Stage 10 distillation, after which the second solL fraction is added and distilled (Stage 11). This

aims to maintain a desired initial and final solL volume to product mass ratio, without violating the maximum boiling volume limit of the vessel.

The final four operations involve the purification of the product. A crystallisation from solL aims to remove actC and actE to acceptable levels in the product solids (Stage 12). This also acts to remove unreacted actA and actD. The solids are filtered and then washed in pure solL to remove residual solution containing dissolved impurities before drying (Stages 13, 14 and 15).

Figure 3.

Some important process issues concerning the product yield and final quality are indicated:

- The controlled addition of reH to the reactor is necessary to prevent a potentially strong exothermic reH reaction and maintain a constant low temperature, T_1 .
- Maintenance of a constant low temperature in the reaction helps prevent the possibility of increased impurity formation.
- The molar charges for the chemical reagents are estimated based on the initial moles of actA feed in Stage 1, and not based on measurements of the species to be quenched. The exception is Stage 4, where the complete destruction of residual reH is ensured with the utilisation of a litmus paper test, and additional base charges.
- There is a potential for drug loss in Stage 5, up to 1-2 wt% of the product yield. The presence of excess baseJ may lead to a product decomposition reaction, due to either an incorrect baseJ molar charge, or a prolonged stirring period.
- The desired values for initial and final solL volume to product mass ratios are obtained from statistical design of experiments. These ranges are believed to be important to the behaviour of the subsequent crystallisation process.
- The addition and distillation split of the total solL make up is imposed due to the maximum volume limit allowed for boiling within the given vessel. The split may be an important factor regarding the quantity of residual solF present in the Stage 12 crystallisation stream and the total process sequence time.

- A higher crystallisation temperature may lead to yield loss due to higher solubility of the product while a lower temperature may mean increased impurity content in the solids.
- Low solubility of product in solL at the controlled wash temperature, T_{14} , means that dissolution of solid product in the solL wash is assumed to be negligible.

These observations provide limitations on the operating policy and the underlying issues should provide incentives for the use of a modelling-based approach. It is important that such observations can be considered in such an approach though there may be little mechanistic understanding behind them.

4. Risk Analysis

In this section the effect of the large degree of uncertainty concerning the first generation of process models in the integrated sequence is quantified using the uncertainty and sensitivity analyses as shown in Figure 1. The effect that additional process data may bring to the models and the assumed uncertainty within the sequence is also considered.

In the hypothetical situation that data from complete process sequence runs at subsequent scale-ups becomes available, it is proposed that the methodology is used to analyse the predicted uncertainty in the current level of models with respect to this data, to determine where the models need to be developed further. This corresponds to the validation step in the systematic model development process, step 15 shown in Figure 1. Since production plant data was not available for this case study, the PPR plant data (Table 3) was used as the benchmark with which to assess the effect of the uncertainty in the predicted criteria using models and parameter uncertainty estimated or assumed to correspond to the current levels of knowledge available. The corresponding PPR plant conditions given in Table 2 were used in this assessment. Estimation of the model parameters and quantification of the parameter uncertainty is based on actual data where available.

Table 2. PPR plant run conditions

Stage	PPR plant run conditions	
1	Operation time	300 min
	Total drug feed mass	469.9 kg
	Drug feed purity	74.1 wt%
	solF	2880 kg
	30% aq. reH	1042 kg
	reH addition time	60 min
	Agitation speed	78 rpm
2	Distilled water addition	940 kg
4	6% aq. baseI solution	790 kg
5	50% aq. baseJ solution	46.9 kg
7	35% aq. baseK solution	260.0 kg
9	Minimum stir volume	500 dm ³
10/11	Total solL feed	4785 dm ³
	Total solL removed	2435 dm ³
12	Cooling rate (assumed)	1 °C min ⁻¹

Table 3. PPR plant run measurements

Stage	PPR plant run measurements	
1	X_{actA}	0.955
	actB content	71.45 wt%
	actC content	0.74 wt%
	actE content	2.88 wt%
	Post reactor crude product mass	345.8 kg
10/11	Cumulative initial solL:product ratio	14.7 dm ³ kg ⁻¹
11	Final solL:product ratio	7.2 dm ³ kg ⁻¹
	Pre-crystallisation crude product mass	325.8 kg
14	Post filtration LOD	25 %
15	Dry product mass	293.2 kg
	Post drying LOD	6 %
	actB dry content	89.4 wt%
	actC dry content	0.24 wt%
	actE dry content	1.4 wt%

4.1. Stochastic problem for the first generation model set

The first generation stochastic system to which the uncertainty and sensitivity analyses are applied is defined in this section. The main characteristics of the first generation set of deterministic models and the uncertainty inputs to the problem are summarised in Table 4.

The uncertain inputs to the stochastic system comprise of those model parameters which generate a significant response in important output criteria when individually perturbed from their nominal values. For the consecutive reaction rate constants, k_1 and k_2 , estimated simultaneously using non-linear least squares, the degree of correlation and the parameter standard deviations are estimated using a first term Taylor series expansion for the covariance matrix. For uncertain

Table 4 Summary of first generation model characteristics and stochastic problem inputs.

Stage	Operation	Main deterministic model characteristics	Uncertainty sources	Reference (Appendix B)
1	Reaction	First order reaction kinetics.	Kinetic rate parameters	Model B1
2,4,5,7	Reagent addition	Two-phase mass balance with fractional drug loss assumption due to aqueous phase solubility.	Fractional drug loss parameter from organic phase	Model B2
3,6,8	Layer separation	Two-phase mass balance with fractional organic phase loss due to imperfect phase cut.	Fractional organic phase cut loss parameter	Model B3
9,10,11	Distillation	Batch distillation assuming ideal VLE with specified reboiler duty for estimation of operation time.	VLE coefficients	Model B4
12	Crystallisation	Crystal growth kinetics based on solute saturation in a seeded batch cooling regime, with drug impurity solute concentration 'loss'.	Crystal growth rate and impurity 'loss' parameters and saturation data	Model B5
13	Filtration	Two-phase mass balance attaining a desired moisture hold-up.	Moisture hold-up and filtration rate	Model B6
14	Washing	Two-phase mass balance with moisture displacement with wash.	Moisture hold-up and displacement	Model B7
15	Drying	Two-phase mass balance attaining a desired moisture hold-up.	Moisture hold-up and drying rate	Model B8

parameters not estimated simultaneously, standard deviations are estimated from the relevant data or assumed (based on the standard deviation being a percentage of the nominal value) or for uniformly distributed uncertainty, range limits are estimated or assumed.

The failure to maintain initial (14-15) and final (7-8) desired volumes of solL solvent to expected mass of actB prior to crystallisation have been identified as an important observation with regard to the final crystal impurity content. This is given some weight in the stochastic problem. Violations of these desired operating ranges are assumed to give a greater uncertainty in parameters characterising downstream criteria believed to be related but for which there is no mechanistic understanding. The effect of this uncertainty is shown in Figure 4(a) where violation of these desired solvent to product ranges leads to an increase in the uncertainty in the impurity concentration 'loss' parameter (from solution) and in Figure 4(b) a corresponding increase in the uncertainty in endpoint key impurity content of the crystals.

Figure 4.

4.2 First generation model sequence results

A total of 431 scenarios were required to satisfy the multiple 1% mean and variance parameter error convergence criteria for both the total yield (Y_T) and endpoint key impurity content (w_{actC}). The results of the uncertainty analysis are summarised in Table 5. 5% and 95% fractiles are used to quantify the predicted uncertainty in the endpoint output and certain inter-stage criteria. If the data does not fall within the 5-95% fractiles of the predictions then the proximity values (the final column in Table 4) show how close the data is to the nearest fractile (5 or 95%), relative to the fractile width.

For example, in this case study the cumulative frequency plot in Figure 4 (a) shows the predicted distribution in endpoint total yield under uncertainty in the first generation process model sequence relative to the independent PPR plant data (the solid vertical line in the Figure). Since the plant data for this criterion does not fall within the predicted uncertainty as enclosed within the 5% and 95% fractiles (the dashed lines in the Figure), clearly some process models of the first generation model set may not be suitable for prediction of the total yield at the PPR plant scale (as would be expected). Since model parameter error has quantitatively been accounted for, an element of structural error may be suspected. The extent of the error in the prediction distribution indicated by the proximity values, are particularly large for the Stage 1 conversion (307%) and secondary impurity composition (221%) (see Table 5). It would appear that the large over prediction of the conversion in Stage 1, Figure 4 (b), contributes to the observed under prediction of the final solL to product ratio and over prediction of the Stage 1-15 total yield. A similar assessment may be made concerning the over prediction of the secondary impurity content in the final product.

Table 5. Summary of first generation model results under uncertainty

Stage	Criteria	PPR Plant data	Predicted mean	Predicted fractiles [5%, 95%]	Data proximity to fractile interval, %
1	X_{actA}	0.955	0.993	[0.987, 0.997]	307
	actB product composition, wt%	71.5	72.9	[72.1, 73.4]	49
	actC impurity composition, wt%	0.74	1.21	[0.89, 1.48]	25
	actE impurity composition, wt%	2.88	8.53	[7.45, 9.51]	221
11	Final solL:product ratio, dm ³ kg ⁻¹	7.2	6.94	[6.83, 7.04]	76
15	actB product content, wt%	89.4	90.52	[88.0, 93.5]	inside
	actC impurity content, wt%	0.24	0.33	[0.19, 0.45]	inside
	actE impurity content, wt%	1.4	2.78	[1.72, 3.70]	21
1-15	Total yield, %	84.2	90.2	[87.2, 92.4]	59

Figure 5.

Sensitivity analysis is used to estimate the key contributions to the predicted output uncertainty with regard to the propagation of uncertainty in certain inter-stage process properties and the individual

input uncertainties of the stochastic model. These indicate the relative importance of the uncertainty in the model parameters characterising the current state of knowledge in the available process models and the associated phenomena.

Contributory process sub-sequences are defined by potentially viable data measurements in this case study. It is initially assumed that only the inter-stage criterion of reaction conversion is a measurement which can be used to define the sub-sequences for the endpoint total yield. For the impurities, it is assumed that the post reaction stream is a potential inter-stage measurement. The relative contributions of the sub-sequences associated to inter-stage criteria are shown in Table 6. These estimate the fraction of the total uncertainty (quantified as the width between the 5 and 95% fractiles) in the endpoint criterion which is attributed to each specified process sub-sequence. The initial indication is that the Stage 2 to 15 sub-sequence contributes the most uncertainty to the predicted uncertainty in the total yield (82% of the final uncertainty), while the Stage 1 reaction contributes the least (18% of the final uncertainty). The model parameter uncertainties in the models for Stages 2-15 provide a much larger contribution to the uncertainty in the endpoint total yield than the Stage 1 model parameter uncertainties. To reduce the uncertainty in the total yield, the sub-sequence contributors immediately indicate that further work should be focused on the models and uncertainties assumed in Stages 2-15. This is different to implying that the models for Stages 2-15 are the main cause for the deviation in total yield from the PPR plant data. The prediction of the key impurity in Stage 1 introduces a greater uncertainty in the final product content than all the following operations. The opposite is apparent for the secondary impurity.

Table 6. First generation model sub-sequence contributions to predicted uncertainty.

Sub-sequence	Total yield	Key impurity	Secondary impurity
Stage 1	0.18	0.62	0.34
Stage 2 to 15	0.82	0.38	0.66
Endpoint	1.00	1.00	1.00

If a sample of the pre-crystallisation crude were to be available for analysis, then the Stage 2-15 sub-sequence contribution of 0.82 for total yield uncertainty could be decomposed. The new contributions to the total yield uncertainty and impurity contents, are shown in Table 7. Significant

contributions from both the Stage 2-11 (0.35) and Stage 12-15 (0.47) sub-sequences are apparent. However, Stages 2-11 do not appear to introduce any additional uncertainty to the impurity compositions predicted from Stage 1. This is reasonable considering the realistic assumptions incorporated in the first generation reagent addition and layer separation models that any drug loss is independent of the concentrations of the other drug species and that there is no loss in the distillations.

Estimated values of the coefficient of determination close to unity for total yield, key and secondary impurity contents for unranked data (0.97, 0.98 and 0.99 respectively) and ranked data (0.97, 0.96 and 0.97 respectively) indicate that the linear input parameter contributors predicted by the Sensitivity Analysis should be reliable and rank transformation of the data is not required.

Table 7. Effect of more inter-stage measurements to sub-sequence contributions.

Sub-sequence	Total yield	Key impurity	Secondary impurity
Stage 1	0.18	0.62	0.34
Stage 2 to 11	0.35	~0.00	~0.00
Stages 12-15	0.47	0.38	0.66
Endpoint	1.00	1.00	1.00

The correlation coefficients (CC) and standardised regression coefficients (SRC) over all the uncertain parameters are shown in Figure 5 (a), where the parameter index numbers are specified in Table B2 (Appendix B9, Supplementary materials). The key contributor parameters to the observed uncertainty in the endpoint yield as indicated by the CCs are k_g (0.74, index 16) then k_1 (0.38, index 1) and k_2 (-0.38, index 2), as shown in Figure 5 (a). However, the induced correlation between the product formation reaction rate constant, k_1 , and the key impurity formation reaction rate constant, k_2 , in the sampling procedure results in a false estimation of the influence of k_2 to total yield from the CCs. This is because the CCs do not measure the standardised global influence. The SRCs show that the influence of k_2 (-0.05) is actually negligible compared to k_1 (0.32), as would be expected. The parameters associated with product loss from the reagent addition and layer separation stages (indices 6 to 12) provide minor contributions (SRCs between -0.15 to -0.20). The key SRC contributors to the final impurity content are estimated as the key

impurity 'solute loss' parameter, ζ_{actC} , (0.78, index 21) then k_2 (0.77, index 2) and then the wash efficiency, η_{wash} , (-0.21, index 26), as shown in Figure 5 (b). A similar ranking is predicted for the secondary impurity but with a stronger influence on the secondary impurity 'solute loss' parameter, ζ_{actE} , (SRC of 0.96 compared to 0.32 for k_3) due to better fit of the reaction model for the parallel reaction to the bench scale data. The parameters identified as key contributors and priority do not provide any surprising results.

As may be expected the estimated key uncertain parameter contributors coincide with stages contained within the key sub-sequences. Whereas the sub-sequence contributors are useful in providing an initial idea to the key areas of the parts of the process sequence contributing uncertainty and provide a measure of the accumulation of uncertainty at specific points in the sequence based on certain (potentially measurable) inter-stage and endpoint outputs of the stochastic model, the CCs and SRCs provide a ranking of importance in the uncertain inputs which can differentiate between a large number of individual sources.

The key process sub-sequences identified and the key parameter contributors of the predicted endpoint uncertainty in the whole process sequence can be used to provide a guide to the key phenomena which are not well characterised and introduce large amounts of uncertainty. Given the identified quantitative indicators for the predicted uncertainty, indicated in Table 8, a

Table 8. List of key parameters regarding first generation model uncertainty in ascending order of priority

Total yield			
Key Stage	Key parameter	Characterised	Possible related phenomena
12	k_g	Growth kinetics	Nucleation kinetics, mixing etc. various - complex
1	k_1	Intrinsic kinetics	Intrinsic reagent kinetics, solids dissolution, mixing
2, 4, 7	u_1	Solubility loss	Mass transfer rate, equilibrium solubility
3, 6, 8	u_2	Imperfect phase cut	Phase dispersion band, drop entrainment
Key and secondary impurity content			
Key Stage	Key parameter	Characterised	Possible related phenomena
12	$\zeta_{actC}/\zeta_{actE}$	Impurity 'solute loss' rate	Various - very complex
1	k_2	Intrinsic kinetics	Intrinsic reagent kinetics, solids dissolution, mixing
14	η_{wash}	Moisture displacement	Mass transfer rate

phenomenological knowledge priority list can be inferred, based on engineering intuition. With this information the data required to progress the model development can be ascertained, either to reduce the uncertainty associated with the parameters of the current model structures if the uncertain prediction encompasses the current data, or to develop models incorporating different phenomena and/or equations. The next step of the methodology is to determine how the effect of incoming data and knowledge can improve the current predictions and what parameter uncertainty reductions are required for desired performance levels (Step 17 in Figure 1).

5. Effect of new data

With an identified model it is useful to be able to determine the effect on the prediction of the key output criteria under uncertainty with developing models based on new incoming data and observations. This aspect of the methodology can provide a driving force for the collection of certain data to progress the model-based approach, where identified to be needed. The assumption made in this study is that new data is incorporated as and when it becomes available. The availability of new data is given in Table 9, in the assumed order of incorporation. The revised models and parameter uncertainties associated with each level of incorporated knowledge are given in Appendix C (Supplementary materials). The semi-empirical reactor model is developed to incorporate further limiting phenomena to account for the key responses. The layer separation and reagent addition models are transformed from speculative assumptions towards more mechanistic type models.

Figure 6.

Figure 7

Tracking of the uncertainty (depicted by the predicted 5-95% fractiles) with sequential incorporation of knowledge into the system models, as data becomes available, is shown in Figure 6 (a) and Figure 7 (a) for endpoint yield and key impurity content predictions, respectively. The respective relevant inter-stage criteria of conversion and post reactor key impurity composition (dashed fractile lines) indicate how much how uncertainty has accumulated and propagated between the Stage 1

Table 9. Levels of incorporated knowledge with incoming data.

Knowledge level	New data/information	Model development action	Reference
0	Bench scale data, 100 gram scale.	First generation of process models.	Model B1 (Appendix B)
1	Reactor data at 90 rpm, 1000 US gallon scale.	Revised parameter fit to first generation Stage 1 model.	Model B1 (Appendix B)
2	As for knowledge level 1.	Second generation Stage 1 model, incorporating an initial rate limiting period (reagent addition/solids dissolution)	Model C1 (Appendix C)
3	Reactor data at 75, 60 rpm, 1000 US gallon	Third generation Stage 1 model, incorporating an empirical mixing effect	Model C2 (Appendix C)
4	Observations of layer separation phase cuts, 1000 US gallon scale	Second generation layer separation model (Stage 3, 6 and 8) incorporating a	Model C3 (Appendix C)
5	Drug solubility data in solF/aqueous system.	Second generation reagent addition model (Stage 2, 4, 5 and 7), incorporating	Model C4 (Appendix C)
6	Crystallisation yield data, 1000 US gallon scale.	Revised parameter fit to first generation Stage 12 model.	Model B5 (Appendix B)

and the Stage 15 predictions. Comparing the relative fractile interval widths between the Stage 1 and Stage 15 criteria it appears that while a large proportion of the uncertainty in the endpoint yield evolves after Stage 1, the uncertainty in the endpoint key impurity content is mainly due to that generated in Stage 1.

The addition of new data either leads to revised uncertainty characteristics for parameters in the same model structures (knowledge levels 1 and 6) or to new models structures with different uncertain parameters (knowledge levels 2, 3, 4 and 5), as indicated in Table 9. In the former circumstance, new sets of data are required to re-estimate the uncertainty in the existing model parameters. This does not guarantee a reduction in the uncertainty of the predicted output if there is a wide spread in the new data sets or the model structure is inadequate. Hence, there is an increase in the predicted uncertainty in endpoint yield for knowledge levels 1 and 6 in Figure 6 (a). In the latter circumstance, while the predicted endpoint uncertainty may not decrease with the addition of new data, it is hoped that the prediction of the new model is closer to the data. This is indicated from the deviations in the mean prediction from the PPR plant data that are shown in Figure 6 (b) and Figure 7 (b) for total yield and key impurity content, respectively. For knowledge levels 2, 3, 4 and 5 (see Table 8 for definitions), the total yield predictions become closer to the data as the model structures are changed, and similarly for level 3 for the key impurity content.

Consideration of these deviations with the associated uncertainties provides an indication of the quality of the model system with respect to both the spread in the predicted distribution and accuracy relative to the data. The history of the Stage 1 conversion predictions, Figure 6 (a) and (b) shows that both the proximity of the mean prediction to the data and the relative spread in the distribution are required to ascertain the quality of the model. At knowledge level 1, the mean prediction is closer to the data and under uncertainty the fractiles encompass the data. However, the large increase in the uncertainty of the prediction from level 0 to level 1 (an increase of over 800% in the fractile width) indicates a problem with the model. While the incorporation of more knowledge, in level 2, reduces the uncertainty, the prediction moves away from the data. Only at level 3, when the mixing phenomena are modelled in Stage 1, does the prediction in the reaction conversion to the data improve without such a substantial increase in the uncertainty. The uncertainty in the predicted post reactor key impurity composition and endpoint content, Figure 7 (a), are also reduced at level 3, and the prediction accuracy to the data increases, Figure 7 (b).

The change in the accumulation of the uncertainty through the sequence as knowledge is incorporated can be analysed by the comparison between the relative magnitude of the inter-stage criteria fractile widths, as illustrated in Figure 6 (a) for the conversion and total yield criteria. A clearer representation is provided with the sub-sequence contributions.

The contributions to the total yield uncertainty for Stage 1 and Stage 2-15 sub-sequences are shown in Figure 8 (a). It is no surprise that the contribution of Stage 1 to the final uncertainty becomes larger than that of Stage 2-15 at level 1, when the parameters of the bench scale Stage 1 model are fitted to the larger scale data. This is redressed with the revised model, level 2. At level 3, the contribution of Stage 1 increases again due to the uncertainty in the mixing correlation employed. With the addition of the pre-crystallisation solL to product ratio as a measured inter-stage criterion, Figure 8 (b), it is indicated that the Stage 2-11 sub-sequence becomes an increasingly minor contributor compared to the Stage 12-15 sub-sequence with the incorporation of extra layer separation and drug solubility knowledge.

The importance of the deterministic model structure on the propagation of uncertainty in the stochastic model is demonstrated by the increase in Stage 2-15 sub-sequence contribution to the total yield at knowledge level 6, as shown in Figure 8 (b). At high values of the crystal growth rate constant (k_g) the Stage 12 crystalliser model predicts that the solute concentration approaches the saturation concentration. This suppresses the effect of uncertainty in k_g to the total yield. At

knowledge level 6, the lower revised value of k_g results in a reduction in the suppression effect the model structure has on its uncertainty, despite no change in the relative level of uncertainty in k_g .

The stages associated with the key contributing model parameters, identified using CCs and SRCs, do not necessarily coincide with the key contributing sub-sequences. For knowledge levels 0, 1 and 2, the unassociated variabilities in the Stage 12 key impurity 'solute loss' parameter (ζ_{actC}) and the reaction rate constant for key impurity formation (k_2) appear to each explain a similar fraction

Figure 8.

Figure 9

of the variability in the endpoint key impurity content (SRCs of 0.78, 0.61 and 0.72 for ζ_{actC} compared to 0.77, 0.58 and 0.66 for k_2 , Table D1, Appendix D). A scatter plot for the knowledge level 0 case (Figure D3 Appendix D) does not indicate a greater relationship between either of these two inputs to the endpoint impurity composition. However, the Stage 1 sub-sequence appears to contribute a significantly greater proportion of the endpoint uncertainty than the Stage 2-15 sub-sequence as shown in Figure 9. Cumulative frequency plots (Figure D4, Appendix D), show that the magnitude of the uncertainty in the endpoint composition relative to the uncertainty in the post reactor composition is not much greater. The propagation of minor uncertain inputs in Stage 1 provide an accumulation of uncertainty which overrides the single effect of the uncertainty in ζ_{actC} in Stage 12. In this case it is important to differentiate between the key contributing sub-sequences and parameters. Focus on all the uncertainties in the Stage 1 sub-sequence would be more beneficial than on the Stage 2-15 sub-sequence, with regard to the uncertainty in the endpoint key impurity composition.

As the deterministic models are revised to accommodate different phenomena and different uncertain parameters are introduced the priority of the uncertainty contributors change. The final list of knowledge priorities at knowledge level 6 is given in Table 10. Compared to the priorities estimated at level 1 (Table 8) the main contributors to the prediction uncertainty in endpoint yield remain the crystallisation and reaction rate constants and drug-aqueous solubility parameters. Uncertainty in the layer separation parameters provide no significant contributions. Uncertainty in

Table 10. List of key parameters and knowledge priorities in the final generation of models (knowledge level 6), in ascending order of priority

Total yield			
Key Stage	Key parameter	Characterised phenomena	Possible related phenomena
12	k_g	Growth kinetics	Nucleation kinetics, mixing etc. various - complex
1	k_1	Intrinsic pseudo first order drug reaction kinetics	Intrinsic reagent-drug kinetics
2, 4, 7	σ_{sl}^*	Organic-aqueous phase drug solubility	Mass transfer rate
1	γ_1, γ_2	Rate limiting mixing case	Eddy formation and imperfect energy dissipation
Key impurity content			
Key Stage	Key parameter	Characterised phenomena	Possible related phenomena
1	k_2	Intrinsic pseudo first order drug reaction kinetics	Intrinsic reagent-drug kinetics
12	ζ_{actC}	Impurity solute 'loss' rate	Various, very complex molecular scale phenomena
14	η_{wash}	Solution displacement	Mass transfer rate
1	γ_1, γ_2	Rate limiting mixing case	Eddy formation and imperfect energy dissipation
Secondary impurity content			
1	k_3	Intrinsic pseudo first order drug reaction kinetics	Intrinsic reagent-drug kinetics
12	ζ_{actE}	Impurity solute 'loss' rate	Various, very complex molecular scale phenomena
14	η_{wash}	Solution displacement	Mass transfer rate

the Stage 1 reaction rate constants have become more important to the endpoint impurity than the Stage 12 impurity 'solute loss' parameters. At the expense of a more accurate deterministic Stage 1 model, the addition of the mixing effect (level 3) introduces minor contributions to the uncertainty in total yield and key impurity content.

We were able to show how incorporating new data can improve the robustness of the model.

6. Optimal uncertainty reduction

The key uncertain parameter contributors to the uncertainty in total yield, key and secondary impurity content predictions for knowledge levels 6 have already been identified in Table 9. The extent of the reduction in the uncertainty of these key contributors required to meet a specified reduction in the predicted output criteria can be quantified. An optimisation problem is solved for

the knowledge level 6 process model system in order to reduce the uncertainty as much as possible.

For the optimisation problem we minimise the sum of the uncertainties subject to the deterministic process stage model equations, stochastic quality constraints, decision variable bounds, and uncertainty space characterisation). The general formulation as follows:

objective function

$$\max_{\delta_{\sigma_{k_1}}, \delta_{\sigma_{k_g}}, d_{\sigma_{sl}^*}, \delta_{\sigma_{k_3}}, \delta_{\zeta_{actC}}, \delta_{\zeta_{actE}}, \delta_{\Delta\eta_{wash}}} = \sum \left[\delta_{\sigma_{k_1}}, \delta_{\sigma_{k_g}}, d_{\sigma_{sl}^*}, \delta_{\sigma_{k_3}}, \delta_{\zeta_{actC}}, \delta_{\zeta_{actE}}, \delta_{\Delta\eta_{wash}} \right]$$

subject to

deterministic process model equations models B1-9 in Appendix B (Supplementary Materials)

stochastic quality constraints

$$FW_{5\%,95\%,Y_T} \leq \alpha_{Y_T} FW'_{5\%,95\%,Y_T}$$

$$FW_{5\%,95\%,wt_{actC}} \leq \alpha_{wt_{actC}} FW'_{5\%,95\%,wt_{actC}}$$

$$FW_{5\%,95\%,wt_{actE}} \leq \alpha_{wt_{actE}} FW'_{5\%,95\%,wt_{actE}}$$

decision variable bounds

$$0 < \delta_{\sigma_{k_1}} \leq 1, \quad 0 < \delta_{\sigma_{k_g}} \leq 1, \quad 0 < \delta_{\sigma_{sl}^*} \leq 1, \quad 0 < \delta_{\sigma_{k_3}} \leq 1$$

$$0 < \delta_{\zeta_{actC}} \leq 1, \quad 0 < \delta_{\zeta_{actE}} \leq 1, \quad 0 < \delta_{\Delta\eta_{wash}} \leq 1$$

uncertainty space

$$\eta_{wash} = \left\{ \eta_{wash} \left| U \left(\mu'_{\eta_{wash}} - \delta_{\eta_{wash}} \Delta\eta_{wash}, \mu'_{\eta_{wash}} + \delta_{\eta_{wash}} \Delta\eta_{wash} \right) \right. \right\}$$

$$\Delta\eta_{wash} = \eta_{wash}^{UB'} - \mu'_{\eta_{wash}} = \mu'_{\eta_{wash}} - \eta_{wash}^{LB'}$$

$$k_1 = \left\{ k_1 \left| N \left(\mu'_{k_1}, \delta_{k_1} \sigma'_{k_1} \right) \right. \right\}, \quad k_g = \left\{ k_g \left| N \left(\mu'_{k_g}, \delta_{k_g} \sigma'_{k_g} \right) \right. \right\}$$

$$\sigma_{sl}^* = \left\{ \sigma_{sl}^* \left| N \left(\mu'_{\sigma_{sl}^*}, \delta_{\sigma_{sl}^*} \sigma'_{\sigma_{sl}^*} \right) \right. \right\}, \quad k_3 = \left\{ k_3 \left| N \left(\mu'_{k_3}, \delta_{k_3} \sigma'_{k_3} \right) \right. \right\}$$

$$\zeta_{actC} = \left\{ \zeta_{actC} \left| N \left(\mu'_{\zeta_{actC}}, \delta_{\zeta_{actC}} \sigma'_{\zeta_{actC}} \right) \right. \right\}, \quad \zeta_{actE} = \left\{ \zeta_{actE} \left| N \left(\mu'_{\zeta_{actE}}, \delta_{\zeta_{actE}} \sigma'_{\zeta_{actE}} \right) \right. \right\}$$

Stochastic inequality constraints for $\alpha\%$ reduction in the width of the predicted 5-95% fractile intervals for total yield, Y_T , key and secondary impurity content, wt_{actC} and wt_{actE} , are maintained by the minimum reduction in the standard deviations of the normal distributions of:

- the Stage 1 kinetic uncertain parameters (k_1 and k_3),
- the Stage 12 crystal growth rate parameter (k_g),
- the equilibrium drug-aqueous phase solubility parameters (σ_{sl}^*),
- the Stage 12 drug component impurity 'solute loss' parameters (ζ_{actC} and ζ_{actE}),

and by the tightening of the lower and upper bounds about the mean of the uniformly distributed Stage 14 wash efficiency parameter (η_{wash}). An equivalent reduction in the uncertainty in the Stage 1 consecutive reaction rate constant (k_2) is assumed to the reduction determined in the uncertainty of k_1 . These inputs are the key contributing uncertainties to the output uncertainty, identified from the knowledge level 6 Sensitivity Analysis (Table D1, Appendix D). It is assumed that $\alpha_{Y_T}, \alpha_{wt_{actC}}, \alpha_{wt_{actE}}$ are equal to each other.

The optimization problem was solved parametrically at different values of α to obtain a trade-off curve, Figure 10, between the required levels of total key input parameter uncertainty reduction required to meet the desired uncertainty reductions in the output criteria. It appears that the required level of parameter reduction increases sharply (objective function decreases) after a desired combined output criteria uncertainty reduction of 50% (from the original levels). For desired reductions of 70% and greater, the problem was infeasible suggesting that other non-key parameter sources of uncertainty (which were not included as reducing decision variables) have become significant.

Individual relationships between key parameter uncertainty reductions and desired output criteria uncertainty reductions are shown in Figure 11 (a) for $k_1, k_3, k_g, \sigma_{sl}^*$, and Figure 11 (b) for $\zeta_{actC}, \zeta_{actE}, \eta_{wash}$. These graphs show that the uncertainty in k_1, k_3 and k_g consistently needs to be reduced for all desired reductions in output uncertainty while reductions in uncertainty in the other considered parameters $\sigma_{sl}^*, \zeta_{actC}, \zeta_{actE}$ and η_{wash} need only be obtained for desired output uncertainty reductions of greater than 30 or 40%. Associating the key parameters to the process stages indicates where research effort should be directed for different levels of desired output criteria uncertainty reduction. For the endpoint impurity contents, reduction in the uncertainty in the reaction kinetics (k_3 and k_2 via the correlation with k_1) is more important than the crystallisation

parameter uncertainties (ζ_{actC} and ζ_{actE}) until approximately 50-60% uncertainty reduction in the impurity contents. No such distinction between parameter uncertainties can be made regarding uncertainty reduction for the total yield.

The change in the sensitivities of the key uncertain parameters to the output criteria are shown in Figure 12 regarding absolute SRC measures ($R^2 > RR^2 > 0.90$ for all cases). Regarding the total yield, Figure 12 (a) indicates that as the uncertainty in the Stage 1 reaction rate constant for the product (k_1), Stage 12 crystal growth rate constant (k_g) and the drug-aqueous solubilities (σ_{sl}^*) decrease as optimally determined (Figure 10 (a)), the SRC values measuring the contributions of the uncertainty in the Stage 1 time at which the initial rate limited period ends (t'' , defined in Model C1, Appendix C) and the reaction mixing coefficient (γ_1 , defined in Model C2, Appendix C) increase. A similar effect is shown in Figure 12 (b) regarding the endpoint key impurity content except that the Stage 1 reaction rate constant for the key impurity (k_2) and the Stage 14 wash efficiency (η_{wash}) replace k_1 and σ_{sl}^* .

Figure 10.

Figure 11.

Figure 12

These plots indicate the desired levels of uncertainty reduction in the output criteria which may be achieved before the key contributor sensitivities change such that a change in the experimental and modelling effort would become necessary to provide further uncertainty reduction in the output criteria, due to the extent of the reductions in the original key parameter uncertainties. For this case study, beyond approximately 50-60% reduction in uncertainty in the total yield and impurity content output criteria it becomes more beneficial to reduce uncertainty in different uncertain parameters (t'' and γ_1), which would require a redirection of the experimental effort since these parameters are associated with different phenomena, Stage 1 initial rate limiting period and mixing regime (and may sometimes be associated with different process stages, though not in this case).

For the case study it is shown that as more information was incorporated into the process models the predicted distributions in the output criteria do compare more favourably with the independent pilot plant data. In general, the levels of uncertainty decrease with model development iterations but not in every instance. Stream variable uncertainties may be amplified or dampened as they propagate through the sequence. This is affected by the form of the model equations. An amplifying effect in the uncertainties in the Stage 12 solute concentration and total product yield was observed for lower values of the uncertain crystallisation growth rate constant (k_g) at knowledge level 6. The incorporation of three sets of data can be identified as key to the improvement of the distribution characteristics of the total yield prediction. The first and most critical is the 1000 US gallon Stage 1 reaction data at different agitation rates leading to the development of the mixing case model (knowledge level 3). The second is the incorporation of drug aqueous-organic phase solubility data in the development of the generic aqueous reagent addition model (knowledge level 5). The third is the incorporation of larger scale crystallisation yield data (knowledge level 6). The knowledge level 3 data is also critical in the improvement and reduction in the uncertainty of the prediction for the endpoint key impurity content (given the initial laboratory model, knowledge level 0).

Sensitivity analysis showed that the relative contributions to uncertainty in the predicted process yield became approximately equivalent between the Stage 1 reaction and the downstream sequence. In the latter the contribution due to the uncertainty in the Stage 12 crystallisation grew while that due to the Stage 2-11 layer separation and solvent exchange operations receded. Uncertainty in the endpoint impurity content predictions were estimated to be mainly due to the uncertainty in the reaction model for all the knowledge level Risk Analysis iterations. In short and perhaps unsurprisingly, the key uncertain parameters (regarding the endpoint total yield and impurity contents) were associated with the intrinsic reaction kinetics and the crystallisation process, indicating the areas to which development efforts should be directed to increase the understanding and confidence in the process. For the final generation of models presented (knowledge level 6 models), optimal reduction in the uncertainties of the key uncertain parameter contributors is determined for increasing levels of desired uncertainty reduction in the output criteria. The indication is that uncertainty reduction in the reaction rate and crystal growth rate constants would be beneficial to obtain any degree of total yield and endpoint impurity content uncertainty reduction while additional reduction in the other less critical parameters only become necessary once a certain output uncertainty reduction threshold has been passed (Stage 12

crystallisation impurity 'solute loss' parameters and Stage 14 wash efficiency uncertainties become important at 50-60% output uncertainty reductions and Stage 1 transition time from initial rate limiting period and reaction mixing coefficient uncertainties become important beyond 60%).

7. Flowsheet Optimisation Under Uncertainty

Next the investigation was extended to process flowsheet optimisation under uncertainty in order to:

- optimise the economic capacity of the Base Case flowsheet through manipulation of the operating policy variables under no uncertainty in comparison to stochastic optimisation accounting for model uncertainties,
- determine the effect on optimal solutions of different characterisations of the input uncertainties with respect to an increase in the size and a change in the shape of the stochastic input distributions,
- assess the value of perfect information with regard to the potential uncertainty in the purity of the feed API,
- evaluate the maximum tolerance to error in the implementation of the Base Case optimum operating policy under uncertainty.

7.1 Nominal flowsheet optimisation problem

Optimisation of the Base Case process flowsheet is based on a profitability objective function. The profitability, P_{ty} , is defined as the revenue from the end product, less the cost of the main solvents on a basis of the total feed mass of active pharmaceutical ingredient (API) charged to the reactor (Stage 1), $F_{actA,1}$, and the total batch processing time, t_T . The units for the profitability are dollars per kilogram of API feed per hour of processing time. The values of the selling price, C_{drug} and solvent costs, C_{solF} and C_{solL} , are assumed at 2000 \$ kg^{-1} product, 5 \$ kg^{-1} solF solvent and 10 \$ kg^{-1} solL solvent, respectively.

It is assumed that six operating policy variables are available for the optimisation of the base process flowsheet, as follows:

- the Stage 1 agitation speed N_1 (rpm),
- the Stage 1 duration time, $t_{f,1}$ (min),

- the fraction of the total solL solvent used for the crystallisation (added over Stages 10 and 11), AF_{10} , added in Stage 10 with the remainder added in Stage 11,
- the fraction of the solL added in Stage 10 removed in the subsequent distillation, RF_{10} ,
- the Stage 12 linear crystallisation cooling rate, CR_{12} ($^{\circ}\text{C min}^{-1}$),
- the Stage 12 crystallisation holding period, HP_{12} (min).

The optimisation problem under no uncertainty for maximum profitability with constraint limits on the endpoint key and secondary impurity contents and the pre-crystallisation solF composition, is shown below. A fixed reboiler duty is assumed in the Stage 9, 10 and 11 batch distillation models.

objective function

$$\max_{t_{f,1}, N_1, AF_{10}, RF_{10}, CR_{12}, HP_{12}} Pfy = \left\{ \frac{Y_T}{100} \frac{1}{t_T} \left(C_{drug} - C_{solF} \frac{\sum_{s=1}^{15} F_{solF,s}}{Z_{drug,15}} - C_{solL} \frac{\sum_{s=1}^{15} F_{solL,s}}{Z_{drug,15}} \right) \right\}$$

subject to:

process model equations B1-9 in Appendix B (Supplementary Materials)

$$Y_T = \frac{Z_{drug,15}}{F_{actA,1}} \times 100\%$$

inequality constraints

$$wt_{solF,11} \leq 0.5$$

$$wt_{actC,15} \leq 0.3$$

$$wt_{actE,15} \leq 2.0$$

decision bounds

$$200 \leq t_{f,1} \leq 400$$

$$60 \leq N_1 \leq 90$$

$$0.4 \leq AF_{10} \leq 0.7$$

$$0.3 \leq RF_{10} \leq 0.7$$

$$0.5 \leq CR_{12} \leq 4$$

$$40 \leq HP_{12} \leq 120$$

where Y_T is the product component (actB) yield over the entire process sequence based on the quantity of API feed.

7.2 Stochastic flowsheet optimisation problem

The proposed optimisation problem under uncertainty aims to maximise the expected potential profitability of the Base Case process flowsheet. However, certain realisations within the uncertainty space may result in poor process performances in the endpoint impurity contents. A potential loss in profitability is modelled as the average profitability which is lost due to violation of desired upper limits on either the key or secondary endpoint impurity contents (0.3 and 2.0 wt%, respectively). Some tolerance to these violations is allowed to reduce the tendency towards overly conservative solutions. This tolerance is quantified by the incorporation of a one-sided stochastic constraint allowing an average profitability loss of up to $3.0 \text{ \$ kg}_{\text{actA}}^{-1} \text{ hr}^{-1}$.

The operating policy decisions are scenario independent, assuming the *a priori* 'here and now' mode of robust control where knowledge of particular realisations of the model parameter uncertainties is not assumed in the optimal operating policy solution. This results in six decision variables, as in the nominal optimisation given in 7.1.

To solve this problem the stochastic optimisation formulation for the base flowsheet given below is solved. The first stochastic inequality constraint tries to maintain an expected pre-crystallisation solF stream composition below 0.5 wt%. In this constraint the general continuous deviation function, f_{dev} , is replaced by the solF content at each scenario. The second stochastic inequality constraint maintains an average potential profitability loss below $3.0 \text{ \$ kg}_{\text{actA}}^{-1} \text{ hr}^{-1}$. The impurity content binary variables, β , are one if the constraint thresholds are passed and zero otherwise. A profitability loss is returned if either the key or secondary content thresholds are violated. However, the profitability loss is not incorporated into the profitability objective since it is assumed only to be a potential loss which may be rectified with further purification iterations at further expense (not included in this problem). The associated general deviation functions in the general problem, d_1 and d_2 , become respectively zero and P_{ty} in the problem below i.e. the profitability loss is not a function of the extent of the impurity content violation. The resulting continuous deviation function, f_{dev} , is defined as the potential profitability loss.

The size of the problem is reduced by limiting the uncertainty space to sensitive inputs to the output criteria, as defined in the uncertainty space characterisation in the optimisation problem below. A reduced convergence criterion of $\pm 2\%$ deviation in the output distribution parameters is permitted to reduce the number of scenarios per objective function evaluation.

objective function

$$\max_{t_{f,1}, N_1, AF_{10}, RF_{10}, CR_{12}, HP_{12}} E\{Pty\} = \frac{1}{M} \sum_{m=1}^M \left[\frac{Y_{T,m}}{100} \frac{1}{t_{T,m}} \left(C_{drug} - C_{solF} \frac{\sum_{s=1}^{15} F_{solF,s}}{Z_{drug,15,m}} - C_{solL} \frac{\sum_{s=1}^{15} F_{solL,s}}{Z_{drug,15,m}} \right) \right]$$

subject to:

process model equations B1-9 in Appendix B (Supplementary Materials)

binary variable approximations

$$\beta_{wt_{actC},15,m} = \frac{1}{2} \left[\tanh\left\{ \xi_{wt_{actC},15} (0.3 - wt_{actC,15,m}) \right\} + 1 \right] \quad \forall m = 1 \dots M$$

$$\beta_{wt_{actE},15,m} = \frac{1}{2} \left[\tanh\left\{ \xi_{wt_{actE},15} (2.0 - wt_{actE,15,m}) \right\} + 1 \right] \quad \forall m = 1 \dots M$$

stochastic inequality constraints

$$E\{wt_{solF,11}\} = \frac{1}{M} \sum_{m=1}^M wt_{solF,11,m} \leq 0.5$$

$$E\{Pty_{loss}\} = \frac{1}{M} \sum_{m=1}^M \left[(1 - \beta_{wt_{actC},15,m} \beta_{wt_{actE},15,m}) Pty_m \right] \leq 3.0$$

decision bounds

$$200 \leq t_{f,1} \leq 400, \quad 60 \leq N_1 \leq 90, \quad 0.4 \leq AF_{10} \leq 0.7, \quad 0.3 \leq RF_{10} \leq 0.7, \quad 0.5 \leq CR_{12} \leq 4.0,$$

$$40 \leq HP_{12} \leq 120$$

uncertainty space

$$k_1, k_2 = \left\{ k_1, k_2 \left| \begin{bmatrix} k_1 - \mu_{k_1} \\ k_2 - \mu_{k_2} \end{bmatrix}^T \begin{bmatrix} \sigma_{k_1}^2 & \rho \sigma_{k_1} \sigma_{k_2} \\ \rho \sigma_{k_1} \sigma_{k_2} & \sigma_{k_2}^2 \end{bmatrix} \begin{bmatrix} k_1 - \mu_{k_1} \\ k_2 - \mu_{k_2} \end{bmatrix} \right. \right\}$$

$$k_3, t''_1, \gamma_1, \gamma_2, \sigma_{st}^*, k_g, \zeta_{actA}, \zeta_{actB}, \zeta_{actC}, \zeta_{actD}, \zeta_{actE}, FR, DR = \left\{ \theta_{st} \mid N(\mu_{st}, \sigma_{st}) \right\}$$

$$\eta_{wash} = \left\{ \eta_{wash} \left| U \left(\eta_{wash}^{LB}, \eta_{wash}^{UB} \right) \right. \right\}$$

The values for the constants of the hyperbolic smoothing functions for the endpoint key and secondary impurity contents, $\xi_{wt_{actC},15}$ and $\xi_{wt_{actE},15}$, were selected to be 1000 and 150, respectively. With these values the binary approximation smoothing functions calculate zero and one for criteria values to within approximately 1% of the threshold value.

7.3 Stochastic operating policy tolerance optimisation problem

The optimisation problem aims to maximise the tolerances (uncertainty) around the previous 'here and now' optimal operating policy variables subject to a stochastic constraint forcing the expected value of the potential profitability, $E\{Pty\}$, to be at least 99% of the 'here and now' optimal solution, $E\{Pty^*\}$. The total uncertainty space is expanded to include the operating policy variables tolerances, the extents of which are the decisions, δ_d^U and δ_d^L , determined in the optimisation. The formulation for this optimisation is given here:

objective function

$$\max_{\delta_d^L, \delta_d^U} \frac{1}{D} \sum_{d=1}^D \frac{\delta_d^L (z_d^* - z_d^{LB}) + \delta_d^U (z_d^{UB} - z_d^*)}{(z_d^{UB} - z_d^{LB})} \quad \text{where } d = t_{f,1}, N_1, AF_{10}, RF_{10}, CR_{12}, HP_{12}$$

subject to:

process model equations B1-9 in Appendix B (Supplementary Materials)

binary variable approximations

$$\beta_{wt_{actC},15,m} = \frac{1}{2} \left[\tanh \left\{ \xi_{wt_{actC},15} (0.3 - wt_{actC,15,m}) \right\} + 1 \right] \quad \forall m = 1 \dots M$$

$$\beta_{wt_{actE},15,m} = \frac{1}{2} \left[\tanh \left\{ \xi_{wt_{actE},15} (2.0 - wt_{actE,15,m}) \right\} + 1 \right] \quad \forall m = 1 \dots M$$

stochastic inequality constraints

$$E\{wt_{solF,11}\} = \frac{1}{M} \sum_{m=1}^M wt_{solF,11,m} \leq 0.5$$

$$E\{Pty_{loss}\} = \frac{1}{M} \sum_{m=1}^M \left[(1 - \beta_{wt_{actC},15,m} \beta_{wt_{actE},15,m}) Pty_m \right] \leq 3.0$$

$$E\{Pty\} = \frac{1}{M} \sum_{m=1}^M \left[\frac{Y_{T,m}}{100} \frac{1}{t_{T,m}} \left(C_{drug} - C_{solF} \frac{\sum_{s=1}^{15} F_{solF,s}}{Z_{drug,15,m}} - C_{solL} \frac{\sum_{s=1}^{15} F_{solL,s}}{Z_{drug,15,m}} \right) \right] \leq 0.99 E\{Pty^*\}$$

decision bounds

$$0 < \delta_d^U \leq 1, \quad 0 < \delta_d^L \leq 1 \quad \text{where } d = t_{f,1}, N_1, AF_{10}, RF_{10}, CR_{12}, HP_{12}$$

uncertainty space

$$\Theta_{U,z_d} = \left\{ \theta_{z_d} \left| U \left(z_d^* - \delta_d^L (z_d^* - z_d^{LB}), z_d^* + \delta_d^U (z_d^{UB} - z_d^*) \right) \right. \right\}$$

$$\text{where } d = t_{f,1}, N_1, AF_{10}, RF_{10}, CR_{12}, HP_{12}$$

$$k_1, k_2 = \left\{ k_1, k_2 \left| \begin{bmatrix} k_1 - \mu_{k_1} \\ k_2 - \mu_{k_2} \end{bmatrix}^T \begin{bmatrix} \sigma_{k_1}^2 & \rho \sigma_{k_1} \sigma_{k_2} \\ \rho \sigma_{k_1} \sigma_{k_2} & \sigma_{k_2}^2 \end{bmatrix} \begin{bmatrix} k_1 - \mu_{k_1} \\ k_2 - \mu_{k_2} \end{bmatrix} \right. \right\}$$

$$k_3, t_1'', \gamma_1, \gamma_2, \sigma_{sl}^*, k_g, \zeta_{actA}, \zeta_{actB}, \zeta_{actC}, \zeta_{actD}, \zeta_{actE}, FR, DR = \left\{ \theta_{st} \left| N(\mu_{st}, \sigma_{st}) \right. \right\}$$

$$\eta_{wash} = \left\{ \eta_{wash} \left| U(\eta_{wash}^{LB}, \eta_{wash}^{UB}) \right. \right\}$$

operating policy bounds

$$z_{t_{f,1}}^{LB}, z_{t_{f,1}}^{UB} = \{200, 400\}, \quad z_{N_1}^{LB}, z_{N_1}^{UB} = \{60, 90\}, \quad z_{AF_{10}}^{LB}, z_{AF_{10}}^{UB} = \{0.4, 0.7\}, \quad z_{RF_{10}}^{LB}, z_{RF_{10}}^{UB} = \{0.3, 0.7\},$$

$$z_{CR_{12}}^{LB}, z_{CR_{12}}^{UB} = \{0.5, 4.0\}, \quad z_{HP_{12}}^{LB}, z_{HP_{12}}^{UB} = \{40, 120\}$$

The validated results of the optimisation with and without uncertainty are given in Table 11. It is immediately clear that when the optimal decisions obtained with no consideration of the uncertainty are implemented in the uncertain process, the predicted potential profitability loss (15.08 \$ kg_{actA}⁻¹ hr⁻¹) far exceeds the desired limit of 3.0 \$ kg_{actA}⁻¹ hr⁻¹. This appears to be due to low probabilities of passing the thresholds on either the key or secondary impurity contents (0.67 and 0.75, respectively), resulting in a probability of passing the loss constraint (Pr_{pass}) of only 0.51. At the expense of a reduction in the expected profitability (30.97 from 31.13 \$ kg_{actA}⁻¹ hr⁻¹ due to a lower yield), the expected potential profitability loss limit is approximately maintained in the validated results of the robust optimisation.

The optimal decisions, shown in Table 12, explain these results. The shorter Stage 1 reaction and Stage 12 crystallisation duration times determined in the uncertain optimisation, lead to a lower expected total yield (86.5%) but also restrict the formation of the impurities observed by the lower expected endpoint contents predicted (0.22 and 1.42 wt% for the key and secondary impurities, respectively). These comparisons are reflected in the cumulative frequency plots for the total yield and key impurity content, Figures 13 (a) and (b), respectively. The relative behaviours of the secondary impurity content predictions mirror those of the key impurity.

Table 11. Validated Base Case process flowsheet optimisation results under uncertainty.

Criteria	Nominal optimal operation	Uncertain optimal operation
$E\{Pty\}$ ($\$ \text{kg}_{\text{actA}}^{-1} \text{hr}^{-1}$)	31.13	30.97
$E\{Ptyloss\}$ ($\$ \text{kg}_{\text{actA}}^{-1} \text{hr}^{-1}$)	15.08	3.02
Pr_{pass}	0.514	0.894
$[E\{wt_{\text{actC}}\}, E\{wt_{\text{actE}}\}, E\{wt_{\text{solF}}\}]$ (%)	[0.28, 1.71, 0.43]	[0.22, 1.42, 0.43]
$[FW\{wt_{\text{actC}}\}, FW\{wt_{\text{actE}}\}, FW\{wt_{\text{solF}}\}]$	[0.17, 1.42, 0.002]	[0.14, 1.19, 0.002]
$E\{Y_T\}$ (%)	87.9	86.5
$FW\{Y_T\}$ (%)	4.7	5.4

Table 12. Optimal decisions for the Base Case flowsheet

Decisions	Nominal optimisation	Uncertainty optimisation
$t_{f,1}$ (min)	262	251
N_1 (rpm)	90.0	90.0
AF_{10}	0.40	0.40
RF_{10}	0.70	0.70
CR_{12} ($^{\circ}\text{C min}^{-1}$)	0.57	0.50
HP_{12} (min)	67	54

The restricted formation of the key and secondary impurity are coupled with reductions in the uncertainties of the predicted endpoints (-16% and -17% in the respective 5-95% fractile widths).

However, the uncertainty in the total yield increases by 13%. The shorter crystallisation holding period has the effect that the final Stage 12 product drug concentration is further away from the equilibrium saturation which in turn has the effect that the propagation of the uncertainties entering Stage 12 are not suppressed as much in the output variables. The optimum Stage 1 agitation speed, N_1 , and the soLL solvent exchange decisions, AF_{10} and RF_{10} , are unaffected by the incorporation of uncertainties.

The results of the Sensitivity Analysis for the optimised Base Case are given in Table 13 and Table 14. The drying time, DR_{15} , is clearly the highest ranked contributor to the uncertainty predicted in the profitability, followed by the Stage 12 crystallisation growth rate constant (k_g) and the Stage 1 reaction rate constant (k_1). For the total yield, k_g , k_1 , the aqueous-organic equilibrium drug solubilities (σ_{si}^*) and the delayed Stage 1 key impurity reaction start time, t_1'' , exhibit the strongest relationships. For the key and secondary impurity contents, the Stage 1 reaction rate constants are the clearly the strongest contributors followed by the uncertainty in the Stage 12 crystallisation impurity 'solute loss' parameters (ζ_{actC} and ζ_{actE}) and the Stage 14 wash efficiency (η_{wash}). The overwhelming importance of the uncertainty contained in the Stage 1 model to the impurity contents is corroborated by the sub-sequence contributions shown in Table 14. Implementation of the robust decisions compared to the nominal decisions did not qualitatively (nor significantly quantitatively) affect the outcomes of the Sensitivity Analysis.

Figure 13.

Table 13. SRC ranking of the key input uncertainty contributors to the Base Case flowsheet criteria.

Criteria	R ²	Key uncertainty contributors (SRC value)
Pty	0.99	DR ₁₅ (-0.768), k _g (0.405), k ₁ (0.298), FR ₁₃ (-0.184), σ_{sl}^* (-0.146), t ₁ '' (-
Y _T	0.96	k _g (0.681), k ₁ (0.465), σ_{sl}^* (-0.256), t ₁ '' (-0.231), X'' (0.108), h _{band.3.6.8}
w _{tactC}	0.99	k ₂ (0.848), ζ_{actC} (0.447), η_{wash} (-0.320), t ₁ '' (-0.240)
w _{tactE}	0.99	k ₃ (0.915), ζ_{actE} (0.331), η_{wash} (-0.186)

Table 14. Sub-sequence contributions to the uncertainty in the Base Case flowsheet.

Sub-sequence	Total yield	Key impurity	Secondary impurity
Stage 1	0.47	0.80	0.93
Stage 2 to 15	0.53	0.20	0.07
Endpoint	1.00	1.00	1.00

From these results it can be inferred that with the importance of the time aspect incorporated in the profitability criterion, it would be most beneficial to try and reduce the uncertainty in the drying operation. However, since drying appears to be an extremely difficult process to model with any degree of accuracy it may be more realistic to direct action towards improving the current models (and parameter estimations) of the crystalliser and reactor models, and in particular the kinetics.

The fact that the nominal optimal decisions, which enable the process to maintain the deterministic impurity constraints, perform so poorly when extended to the uncertain process system. This underlines the importance of the consideration of the main uncertainties in the optimisation determining these decisions. This importance has been quantified through the stochastic criteria estimated in the optimisation under uncertainty. The effect of the optimal decisions on the predicted output criteria distributions can be explained with respect to the propagation of uncertainty in the stream variables through the process sequence due to the deterministic structures of the process model Robustness Analysis of the input uncertainties

The importance of the state of knowledge of the input source uncertainties to the results of the process flowsheet optimisation under uncertainty (section 7.2) is investigated and quantified in what Kleijnen (1997) terms a Robustness Analysis. Since the problem is constrained by the potential profitability loss due to failure of some portion of the upper distribution tails of the endpoint impurity contents, the optimal solution and the corresponding decisions may be sensitive to the size and

form of the input uncertainties (the location of the means are not varied). Two variations to the Base Case problem uncertainties are considered for the re-optimisation of the process flowsheet under uncertainty,

- Case 1 considers the sensitivity to the size of the input uncertainties, where the uncertainties in the distributions of the original problem are increased by 50%,
- Case 2 considers the sensitivity to the form of the input uncertainties, where uniform distributions replace the normal distributions assumed in the original problem. The upper and lower bounds are fixed at two standard deviations (of the original normal distributions) from the mean value.

The bar chart in Figure 14 shows percent deviations in the expected values and 5-95% fractile widths of the key criteria from the Base Case. The assumption of uniform input uncertainties (Case 2) imposes very little effect compared with a 50% increase in the input uncertainties (Case 1), relative to the original normal input distributions of the Base Case is true. A deviation of only -0.3% in the expected profitability and virtually identical optimal decisions to those of the Base Case are predicted for Case 2. A deviation of -1.2% in the profitability for Case 1 is predicted, resulting from reduced expected yield due to the shorter Stage 1 reaction time (246 minutes) and Stage 12 crystallisation time (44 minutes). As expected for Case 1, the significant increases in uncertainty in the impurity contents (+36.5% and +37.1% in the fractile width of the key and secondary impurity contents, respectively) results in the distributions being shifted to the left in order to satisfy the potential profitability loss constraint and a decrease in the expected values (see Figure 13).

These observations are reflected in the cumulative frequency plots for total yield and key impurity content comparing the three input uncertainty cases, shown in Figure E1 (a) and (b) (Appendix E), respectively. The behaviour of the secondary impurity content predictions mirror those of the key impurity.

This investigation underlines the importance that the state of knowledge of the input uncertainties can have in the Uncertainty Analysis approach to process flowsheet optimisation under uncertainty. Clearly this importance can depend on whether the optimisations are concerned with averages or the tails of distributions. Within the confines of good assumptions for the input uncertainty bounds, the form of the distributions appears to be of negligible significance to the optimal solutions determined. On the other hand, a good estimate of the magnitude of the input uncertainties is

essential to the results and decisions obtained using an Uncertainty Analysis approach to process flowsheet optimisation. This is particularly true when the tails of output distributions are important.

7.4 The importance of process input specification

The concept of the value of perfect information (VPI) is applied to the Base Case flowsheet with regard to potential feed purity knowledge of the API, $p_{f,1}$. The VPI is defined below (and approximated with the second equation) as the expected gain in the potential profitability when using an informed ‘wait and see’ optimal approach as opposed to the uninformed ‘here and now’ approach, with penalisation of violations in an acceptable profitability loss constraint in the latter. A linear function in the extent of violation in the loss constraint penalises the value of information expected with the ‘here and now’ approach, VI_{here} .

$$\begin{aligned}
 VPI &= E_{\theta_{pc}} \{VI_{\text{wait}} - VI_{\text{here}}\} \\
 &= E_{p_{f,1}} \left\{ \left(E_{\theta_{st}, \theta_{dt}} \{Pty_{\text{wait}}\} \right)_{m'} - \right. \\
 &\quad \left. \left[\left(E_{\theta_{st}, \theta_{dt}, p_{f,1}} \{Pty_{\text{here}}\} \right)_{m'} - \beta_{Pty_{\text{loss, here, } m'}} \left(\left(E_{\theta_{st}, \theta_{dt}, p_{f,1}} \{Pty_{\text{loss here}}\} \right)_{m'} - 3.0 \right) \right] \right\}
 \end{aligned}$$

where m' are the observations in $p_{f,1}$ space. Under potential uncertainty in the feed purity of 74 ± 3 wt%, the VPI with knowledge of this uncertainty is a profitability of $11.46 \text{ \$ kg}_{\text{actA}}^{-1} \text{ hr}^{-1}$. It is clear from the increasing value of the VPI with increasing feed purity, Figure 15 (a), that at low values of $p_{f,1}$ (below 72.5 wt%) the main contribution to the VPI is incurred. As indicated in Figure 15 (b) the VI_{here} is greatly reduced for these values of feed purity. The relatively large violation of the profitability loss constraint is shown by the dashed line representing the VI_{here} without any penalisation of the constraint violation. A shallow optimum in the value of the feed purity (~ 74 wt%) is observed in the VI_{wait} solutions, Figure 15 (b). This could be useful knowledge if the feed purity can be more closely specified. At lower feed purities optimal solutions are constrained by the increased propensity for formation of the secondary impurity in the Stage 1 reaction leading to high values in the endpoint secondary impurity content. At higher feed purities the greater amount of product formed results in increasing proportions of the key impurity.

The importance of certain *a priori* knowledge to the process has been quantified in the form of the VPI. The results show that this value can be very sensitive within certain ranges of the available information.

Figure 14

Figure 15

7.5 Base case operating policy tolerance optimisation results

The solution of the stochastic operating policy tolerance optimisation problem (section 7.3) gives an optimum value for the average tolerance of 0.088, shown in Table 15. This is the maximum average fraction of the space of all the operating policy variables (defined by the upper and lower policy bounds) within which a feasible solution is permitted under error or uncertainty in the implementation of the ‘here and now’ optimal policy. As shown in Table 15, the solution is tightly constrained by the stochastic constraints on the expected potential profitability (30.66 \$ kg_{actA}⁻¹ hr⁻¹, at 99% of the ‘here and now’ optimal solution), the expected potential profitability loss (3.00 \$ kg_{actA}⁻¹ hr⁻¹) and the pre-crystallisation solF content (0.50 wt%).

The resulting tolerance limits around the ‘here and now’ optimal operating policy variables (given previously in Table 11) on the operating policy are given in Table 1. The tolerances permitted on

Table 15. Results for tolerance optimisation of the Base Case flowsheet operating policy.

Objective	Average tolerance	0.088
Stochastic constraints	$E\{Pty\} (\$ \text{ kg}_{actA}^{-1} \text{ hr}^{-1})$	30.66
	$E\{Ptyloss\} (\$ \text{ kg}_{actA}^{-1} \text{ hr}^{-1})$	3.00
	$E\{wt_{solF}\} (\%)$	0.49

Table 16. Tolerance limits for Base Case optimal operating policy.

	$t_{f,1}$ (min)	N_1 (rpm)	AF ₁₀	RF ₁₀	CR ₁₂ (°C min ⁻¹)	HP ₁₂ (min)
50	267	90.0	0.41	0.70	0.52	64
0.5	235	89.5	0.40	0.68	0.50	43

the reaction time ($t_{r,1}$) and the crystallisation time (HP_{12}) indicate that the expected profitability criteria is relatively insensitive to the total yield due to a strong dependence on the overall process time to which the available optimisation operating policy variables provide a relatively small contribution. Virtually no tolerance is acceptable in any of the other operating policy variables, hence the relatively low value for the average tolerance. The values of the optimal decisions (δ_a^U and δ_a^L) are given in Table E5 (Appendix E).

7.6 Discussion

The use of a multiscenario stochastic optimisation approach for the optimisation of integrated flowsheets under model parameter uncertainty is demonstrated with respect to the set of models comprising the Base Case process flowsheet of the Case Study. Its application allowed the selection of 'here and now' operating policy decisions to optimise and manage certain aspects of the distributions of the uncertain output predictions. An expected profitability objective function is optimised within a threshold on an expected profitability loss due to failure in endpoint impurity contents. Under no uncertainty, the nominal optimal decisions result in poor behaviour of the endpoint impurity content distributions and a high potential profitability loss when the quantified model parameter uncertainty is considered in Uncertainty Analysis. This highlights the importance of the consideration of uncertainty in the process optimisation.

It is also shown that the state of knowledge of the model uncertainties is important with respect to the magnitude but not so much to the characteristic distribution. The value of perfect information in potentially uncertain process stream inputs is considered with respect to the feed purity. It was found that below a certain value in the feed purity (72.5%), perfect knowledge had a very significant impact and that a shallow optimum value existed (~74%).

The key uncertainties under the optimal policy decisions are identified in the Sensitivity Analysis with the conclusion that realistic efforts to reduce the current levels of uncertainty in the profitability and yield should be primarily directed towards improving the confidence in the crystallisation kinetics followed by the intrinsic rate constant for the product reaction. The ability to provide a better prediction for the time taken for the drying is the key factor in reducing the uncertainty in the profitability due to its dependency on the total batch process time. However, the provision of a useful mechanistic model for drying may be unrealistic considering the complexity of the physical phenomena associated with drying. To reduce the uncertainty predicted in the endpoint impurity contents the analysis (not unexpectedly) strongly indicates that efforts aimed at reducing the

uncertainty in the Stage 1 intrinsic reaction kinetics would be the most rewarding. This information, based on systematic and quantitative procedures, provides focus to the actions which may be invested towards the development of more reliable models for the uncertainty management.

A further stochastic optimisation determined the maximum uncertainty allowed around the optimal 'here and now' operating policy values of the Base Case flowsheet. The upper and lower tolerance limits are constrained to meet a small relaxation of the 'here and now' profitability together with the original stochastic constraints. This quantifies the relaxation permitted to the optimal operating policy actions determined in association to the second management response.

8. CONCLUSIONS

The risk analysis methodology has been applied to a comprehensive case study comprising of an integrated sequence of 15 process operations. The integration of the proposed Risk Analysis methods with model development iterations as more process information becomes available has been demonstrated.

The Risk Analysis methods permit the quantification and tracking of the combined influence of parameter uncertainties contained in the entire sequence of process models as they are developed in the systematic model development procedures with the progression of process development. The sensitivity analysis methods allow the efficient estimation of the key uncertain parameters of the stochastic system and the internal sub-sequence contributions from the results of the uncertainty analysis. The information obtained may be used to ascertain levels of uncertainty and help focus data collection and modelling effort towards those parts of the process sequence in which the uncertainty has the greatest influence on the output.

The analysis allows us to apply a multiscenario stochastic optimisation approach for the optimisation of integrated flowsheets under model parameter uncertainty. Its application allowed the selection of 'here and now' operating policy decisions to optimise and manage certain aspects of the distributions of the uncertain output predictions. An expected profitability objective function is optimised within a threshold on an expected profitability loss due to failure in endpoint impurity contents. Under no uncertainty, the nominal optimal decisions result in poor behaviour of the endpoint impurity content distributions and a high potential profitability loss when the quantified

model parameter uncertainty is considered in the uncertainty analysis. This highlights the importance of the consideration of uncertainty in the process optimisation.

Having determined the most robust set of conditions it would now be possible to compare alternative flowsheets to enable selection of the best flowsheet according to these different criteria under their respective model uncertainties. This will be the topic of a future paper.

The approach presented here gives a structured way to explore the uncertainty arising from inaccurate or unavailable process data during process development. It produces guidance as to which measurements would best to reduce the uncertainty. The approach was demonstrated on a process for which pilot scale data was available. The approach is applicable to any process sequence such as those found in the pharmaceutical industry and would also be applicable for processes with recycle.

Acknowledgements

The authors acknowledge financial support for DBJ from the Engineering and Physical Sciences Research Council in the UK, Pharmacia Corp. and the Centre for Process Systems Engineering. The authors are also very grateful for the provision of data by Pharmacia Corp and the time of their staff.

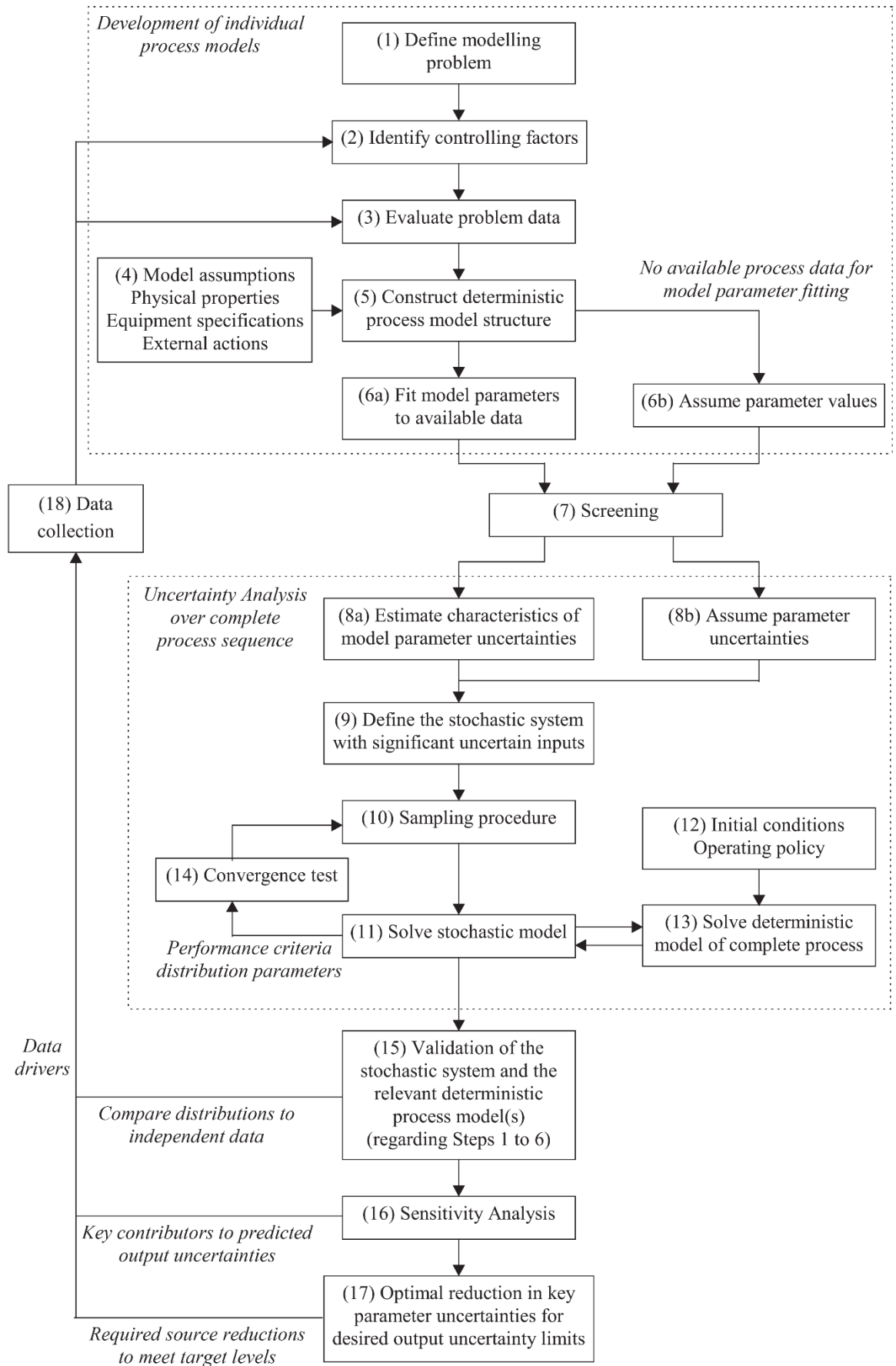
References

- Bernardo, F., Saraiva, P. and Pistikopoulos, E. (2003). Process design under uncertainty: Robustness criteria and value of information. *Computer Aided Chemical Engineering*, pp.175-208.
- Boukouvala, F., Muzzio, F. and Ierapetritou, M. (2011). Dynamic Data-Driven Modeling of Pharmaceutical Processes. *Industrial & Engineering Chemistry Research*, 50(11), pp.6743-6754.
- Datskov, I., Ostrovsky, G., Achenie, L. and Volin, Y. (2006). Process optimization under uncertainty when there is not enough process data at the operation stage. *Optimization and Engineering*, 7(3), pp.249-276.
- Diangelakis, N., Burnak, B., Katz, J. and Pistikopoulos, E. (2017). Process design and control optimization: A simultaneous approach by multi-parametric programming. *AIChE Journal*, 63(11), pp.4827-4846.
- Escotet-Espinoza, M., Singh R., Sen M., and O'Connor T. (2015). Flowsheet Models Modernize Pharmaceutical Manufacturing Design and Risk Assessment. *Pharmaceutical Technology*, 39(4), pp.34-42, 86.

- Floudas, C. and Gümüş, Z. (2001). Global Optimization in Design under Uncertainty: Feasibility Test and Flexibility Index Problems. *Industrial & Engineering Chemistry Research*, 40(20), pp.4267-4282.
- Gomes, J., Chopda, V. and Rathore, A. (2014). Integrating systems analysis and control for implementing process analytical technology in bioprocess development. *Journal of Chemical Technology & Biotechnology*, 90(4), pp.583-589.
- Hong, M., Severson, K., Jiang, M., Lu, A., Love, J. and Braatz, R. (2018). Challenges and opportunities in biopharmaceutical manufacturing control. *Computers & Chemical Engineering*, 110, pp.106-114.
- Ierapetritou, M. and Pistikopoulos, E. (1996). Batch Plant Design and Operations under Uncertainty. *Industrial & Engineering Chemistry Research*, 35(3), pp.772-787.
- Johnson D. (2003) Integrated design under uncertainty for pharmaceutical processes. PhD thesis, University of London.
- Johnson D. and Bogle I.D.L. (2006) Handling uncertainty in the design and development of chemical processes. *Reliable Computing*. *Reliable Computing* (2006) **12**: 409–426 DOI: 10.1007/s11155-006-9012-7
- Lapteva, T., Ziyatdinov, N., Ostrovskii, G. and Pervukhin, D. (2010). Optimization of chemical technology processes under probabilistic constraints. *Theoretical Foundations of Chemical Engineering*, 44(5), pp.651-659.
- Li, P., Arellano-Garcia, H. and Wozny, G. (2008). Chance constrained programming approach to process optimization under uncertainty. *Computers & Chemical Engineering*, 32(1-2), pp.25-45.
- Mandenius, C. and Brundin, A. (2008). Bioprocess optimization using design-of-experiments methodology. *Biotechnology Progress*, 24(6), pp.1191-1203.
- Montes, R. (2011). Variation Transmission Model for Setting Acceptance Criteria in a Multi-staged Pharmaceutical Manufacturing Process. *AAPS PharmSciTech*, 13(1), pp.193-201.
- O'Connor, T., Yang, X., Tian, G., Chatterjee, S. and Lee, S. (2017). Quality risk management for pharmaceutical manufacturing. *Predictive Modeling of Pharmaceutical Unit Operations*, pp.15-37.
- Pistikopoulos, E. and Ierapetritou, M. (1995). Novel approach for optimal process design under uncertainty. *Computers & Chemical Engineering*, 19(10), pp.1089-1110.
- Rogers, A. and Ierapetritou, M. (2015). Challenges and opportunities in modeling pharmaceutical manufacturing processes. *Computers & Chemical Engineering*, 81, pp.32-39.
- Rogers, A. and Ierapetritou, M. (2016). Mathematical Tools for the Quantitative Definition of a Design Space. *Methods in Pharmacology and Toxicology*, pp.225-279.
- Rooney, W. and Biegler, L. (2003). Optimal process design with model parameter uncertainty and process variability. *AIChE Journal*, 49(2), pp.438-449.

- Sen, M., Rogers, A., Singh, R., Chaudhury, A., John, J., Ierapetritou, M. and Ramachandran, R. (2013). Flowsheet optimization of an integrated continuous purification-processing pharmaceutical manufacturing operation. *Chemical Engineering Science*, 102, pp.56-66.
- Shen, D. and Braatz, R. (2016). Polynomial chaos-based robust design of systems with probabilistic uncertainties. *AIChE Journal*, 62(9), pp.3310-3318.
- Severson, K., VanAntwerp, J., Natarajan, V., Antoniou, C., Thömmes, J. and Braatz, R. (2015). Elastic net with Monte Carlo sampling for data-based modeling in biopharmaceutical manufacturing facilities. *Computers & Chemical Engineering*, 80, pp.30-36.
- Steimel, J. and Engell, S. (2016). Optimization-based support for process design under uncertainty: A case study. *AIChE Journal*, 62(9), pp.3404-3419.
- Sun, L. and Lou, H. (2008). A Strategy for Multi-objective Optimization under Uncertainty in Chemical Process Design. *Chinese Journal of Chemical Engineering*, 16(1), pp.39-42.
- Tsay, C., Pattison, R. and Baldea, M. (2017). A Dynamic Optimization Approach to Probabilistic Process Design under Uncertainty. *Industrial & Engineering Chemistry Research*, 56(30), pp.8606-8621.
- von Stosch, M., Hamelink, J. and Oliveira, R. (2016). Hybrid modeling as a QbD/PAT tool in process development: an industrial E. coli case study. *Bioprocess and Biosystems Engineering*, 39(5), pp.773-784.
- Wang, S. and Baldea, M. (2014). Identification-based optimization of dynamical systems under uncertainty. *Computers & Chemical Engineering*, 64, pp.138-152.
- Wang, Z., Escotet-Espinoza, M. and Ierapetritou, M. (2017). Process analysis and optimization of continuous pharmaceutical manufacturing using flowsheet models. *Computers & Chemical Engineering*, 107, pp.77-91.
- Yu, L., Amidon, G., Khan, M., Hoag, S., Polli, J., Raju, G. and Woodcock, J. (2014). Understanding Pharmaceutical Quality by Design. *The AAPS Journal*, 16(4), pp.771-783.

Figure 1. Schematic for the systematic model development incorporating the Risk Analysis approach under uncertainty



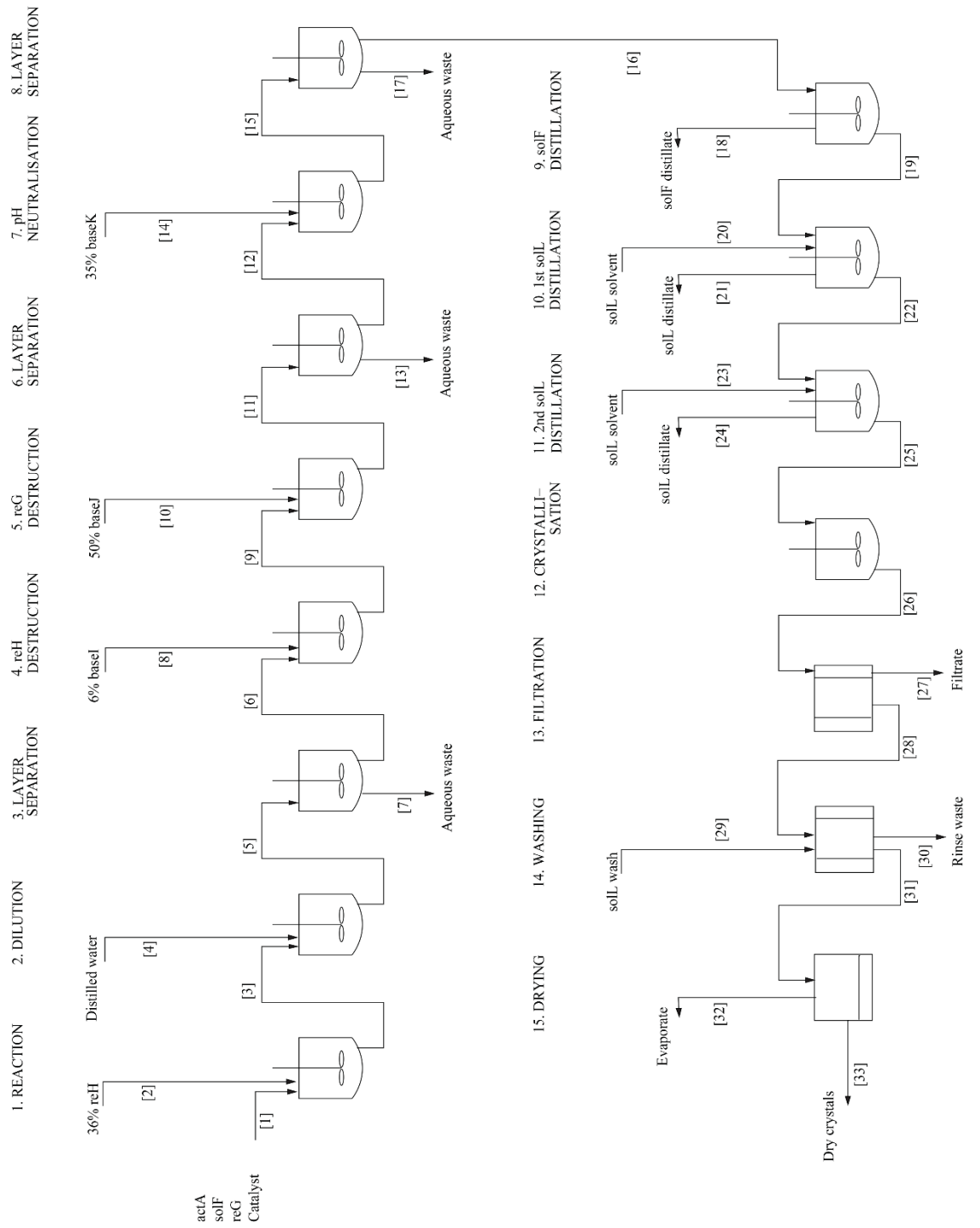


Figure.2. Process flow diagram for Case Study – 15 stages

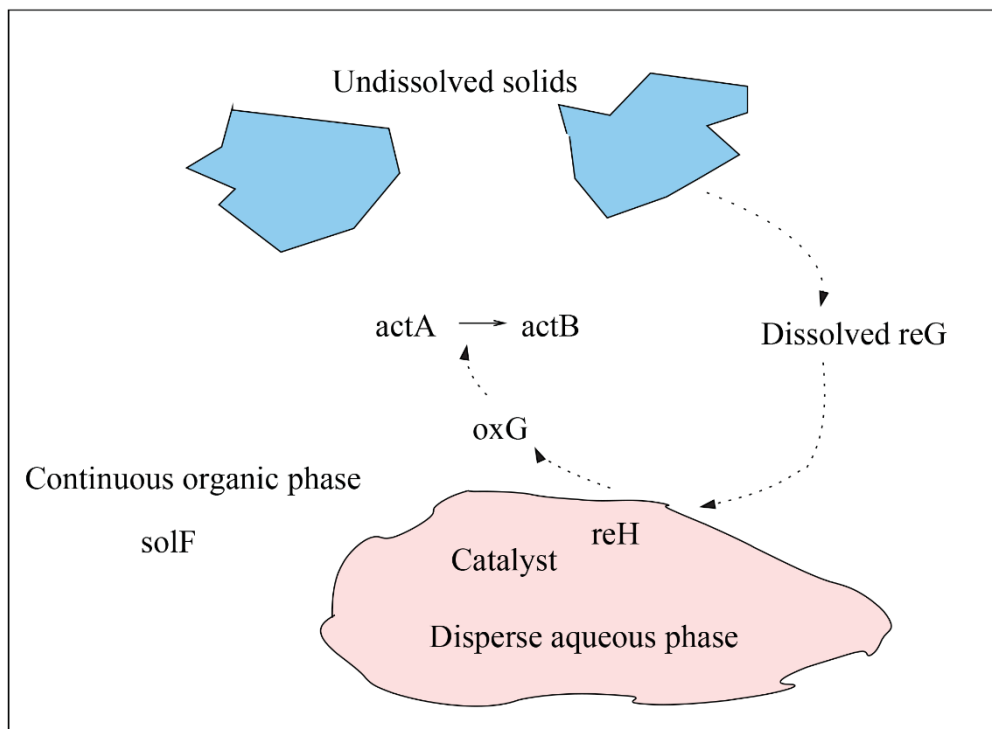


Figure 3. Stage 1 Bisphasic chemical reaction

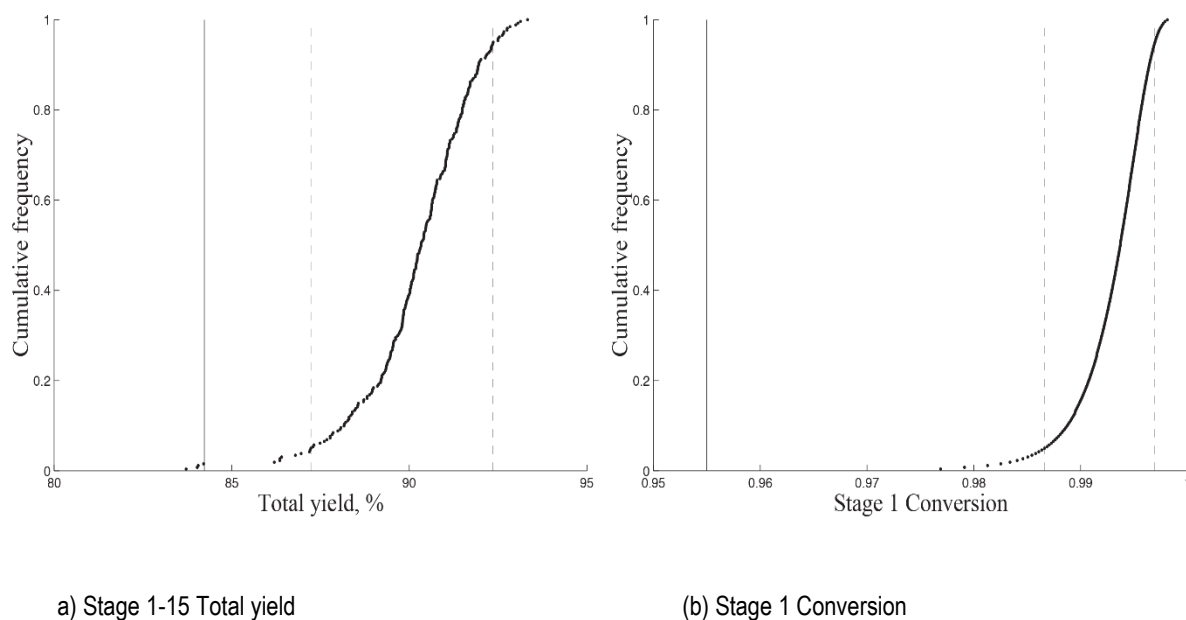
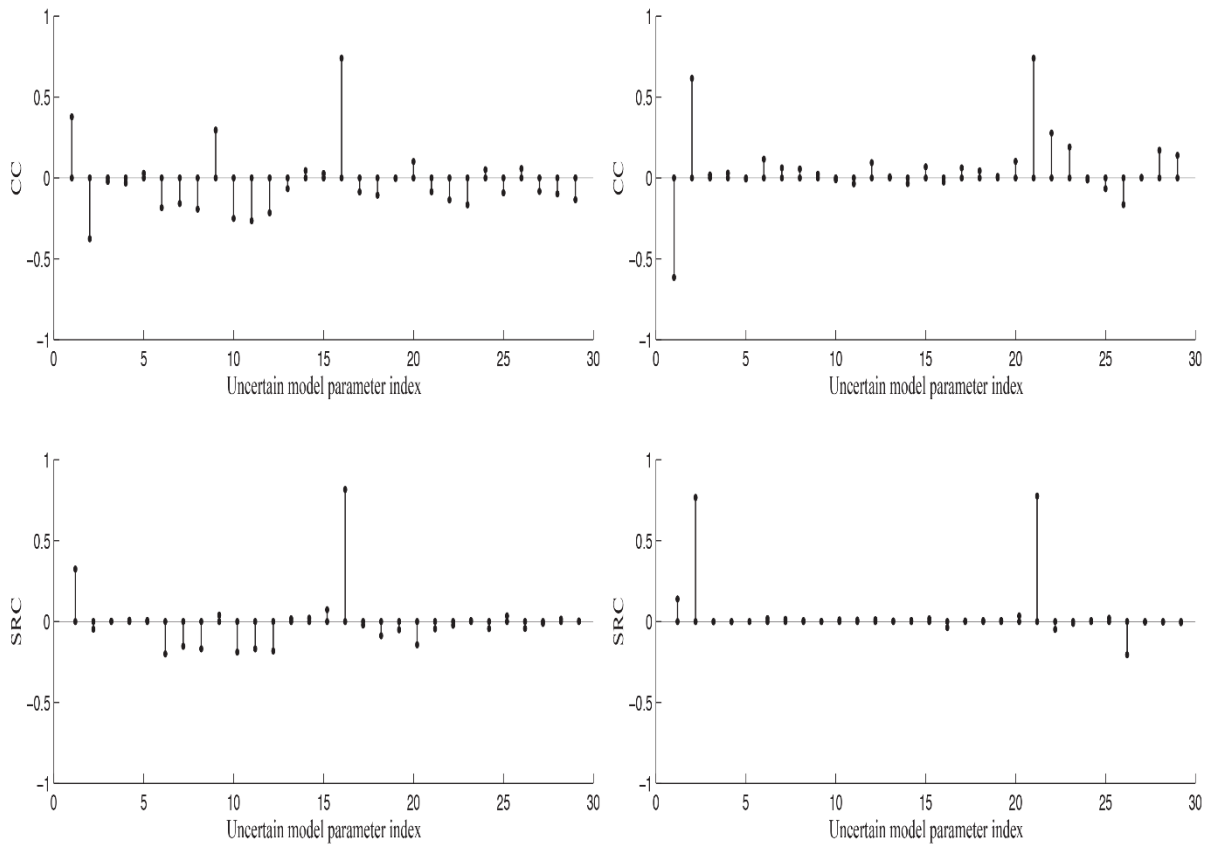


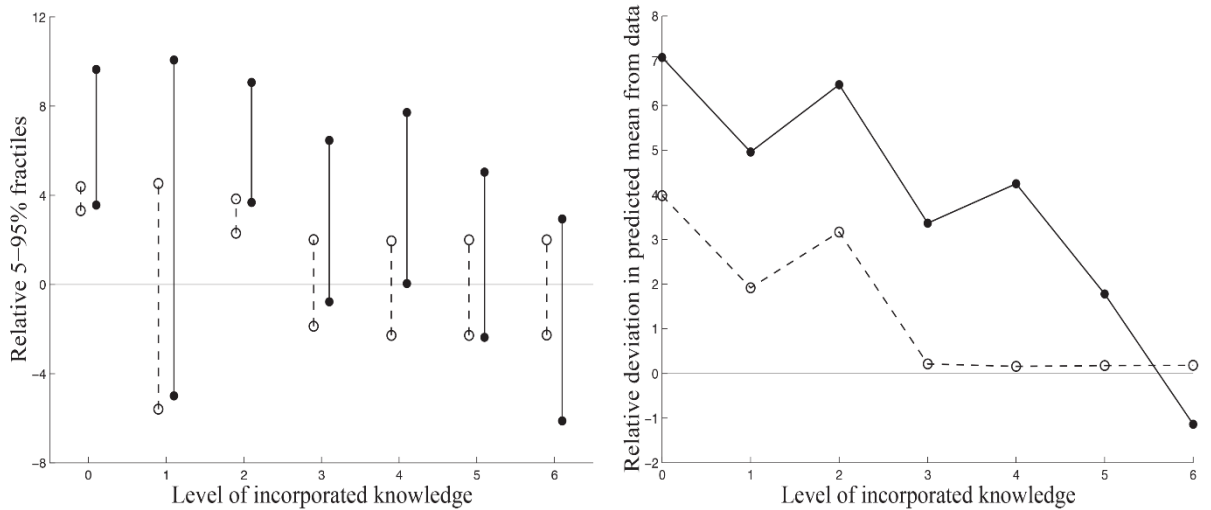
Figure 4. Cumulative frequency plots for the first generation model set predictions under uncertainty Key: • = predicted results, — = PPR data, --- = 5%, 95% fractiles



(a) Total yield

(b) Key impurity content

Figure 5. Uncertain parameter contributor measures for endpoint criteria

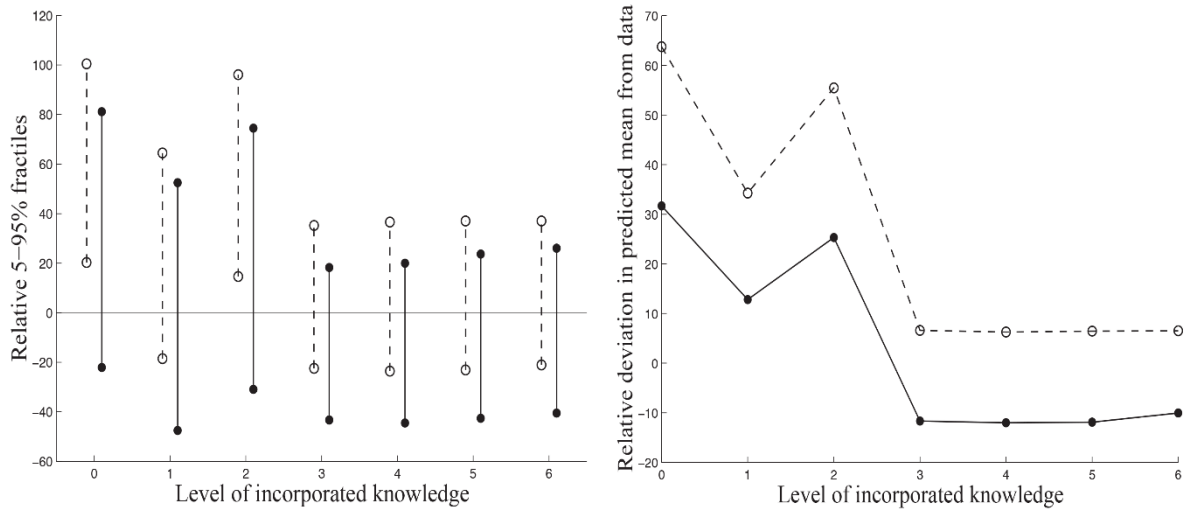


(a) Predicted 5-95% fractiles relative prediction to the plant data

(b) Percent deviation of mean relative to the plant data

Figure 6. Effect of knowledge incorporation to total yield and conversion predictions

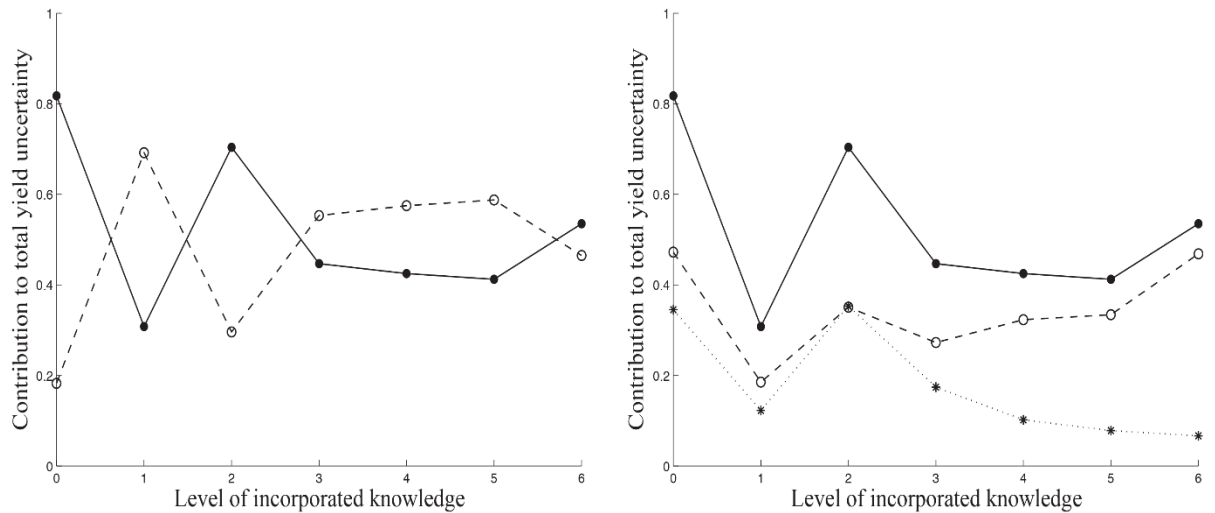
Key: ● = Total yield, ○ = Conversion



(a) Predicted 5-95% fractiles relative prediction to the plant data (b) Percent deviation of mean relative to the plant data

Figure 7. Effect of knowledge incorporation to key impurity composition predictions

Key: ● = Dry crystal key impurity content, ○ = Post reaction crude key impurity composition



(a) Two sub-sequences

(b) Additional sub-sequence

Key: ○ = Stage 1 contribution,
● = Stage 2-15 contribution

Key: ● = Old Stage 2-15 contribution,
* = New Stage 2-11 contribution,
○ = New Stage 12-15 contribution

Figure 8. Effect of knowledge incorporation to the endpoint key impurity uncertainty

Key: ○ = Stage 1 contribution, ● = Stage 2-15 contribution

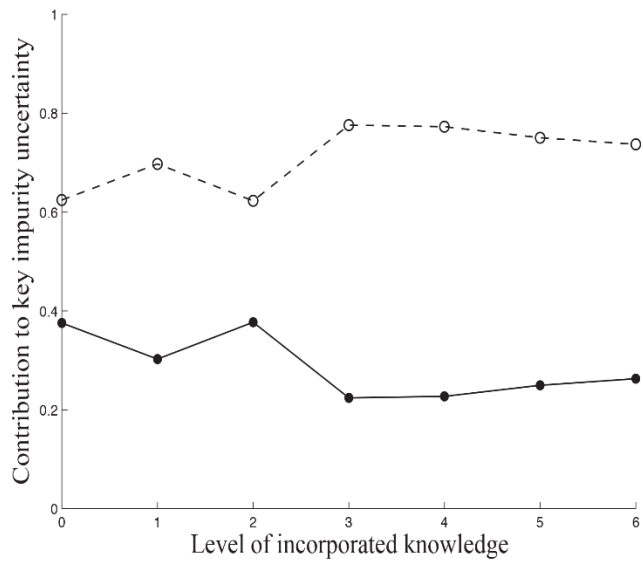


Figure 9. Effect of knowledge incorporation to the endpoint key impurity uncertainty

Key: o = Stage 1 contribution, • = stage 2-15 contribution

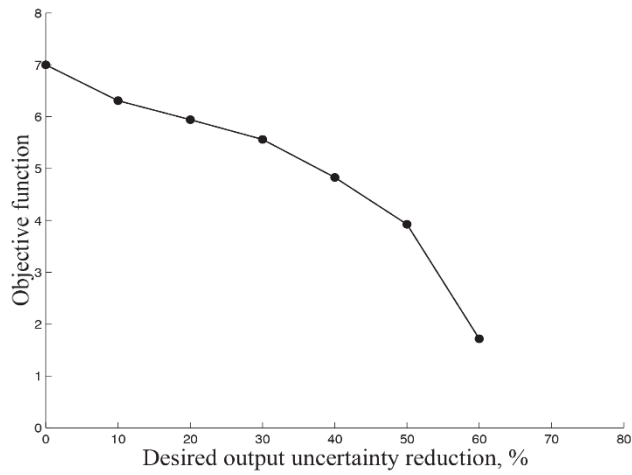
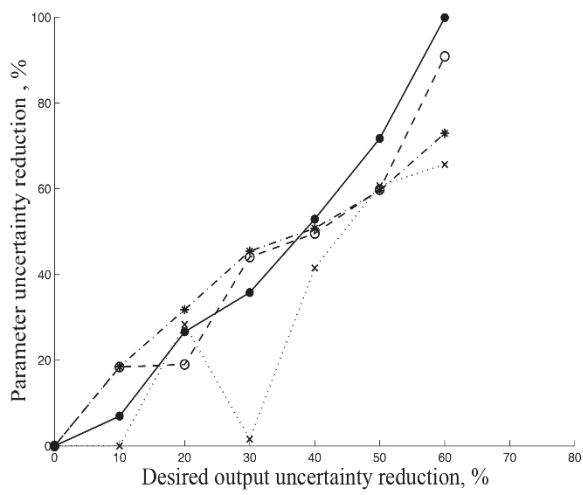
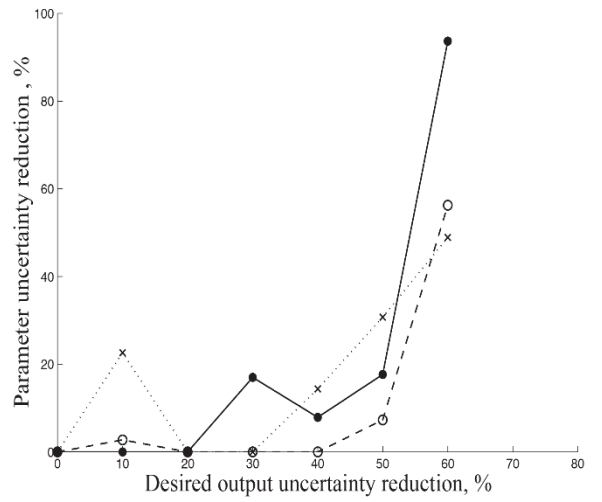


Figure 10. Objective function value, $\Sigma\delta_p$, measuring total required levels of input parameter reduction against desired levels of output criteria uncertainty reduction.



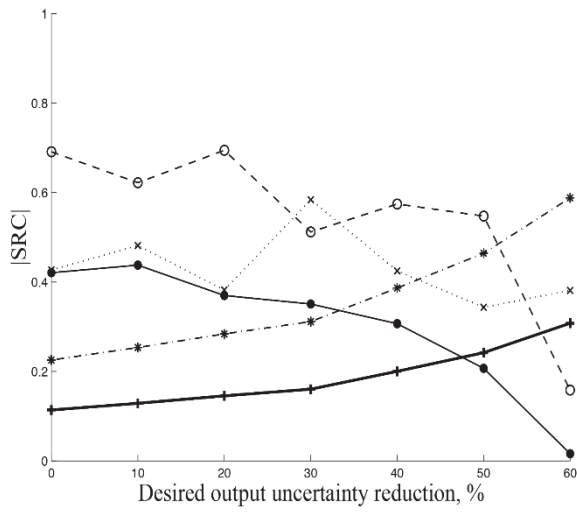
(a) Key: —●= k_1 , ---○= k_g , -●-●= k_3 , ●●●□= α_{sl}



(b) Key: —●= σ_{actC} , ---○= σ_{actE} , ●●●□= α_{sl}

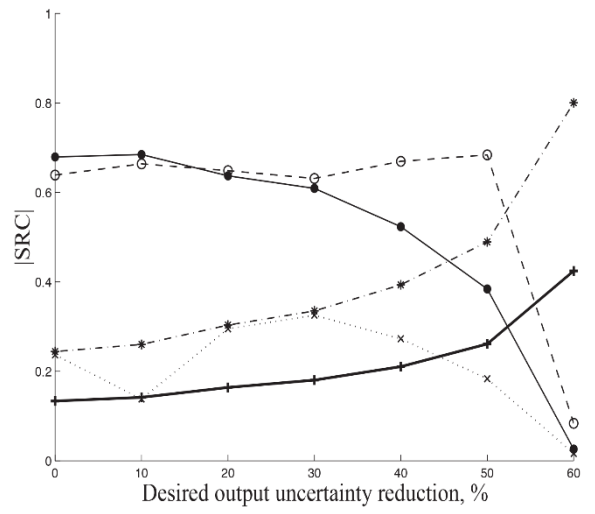
ρ_{wash}

Figure 11. Optimal degree of key input parameter uncertainty reduction (from original uncertain values) required to meet desired levels of output criteria uncertainty reduction.



(a) Total yield

Key: —● = k_1 , ---○ = k_g ,
× = σ_{sl}^* , -.-.* = t'' , —+ = γ_1 .



(b) Key impurity content

Key: —● = k_2 ,
 ---○ = ζ_{actC} ,× = η_{wash} , -.-.* = t'' , —+ = γ_1 .

γ_1 .

Figure 12. Change in absolute SRC sensitivities between output criteria and key uncertain parameters with optimal uncertainty reductions

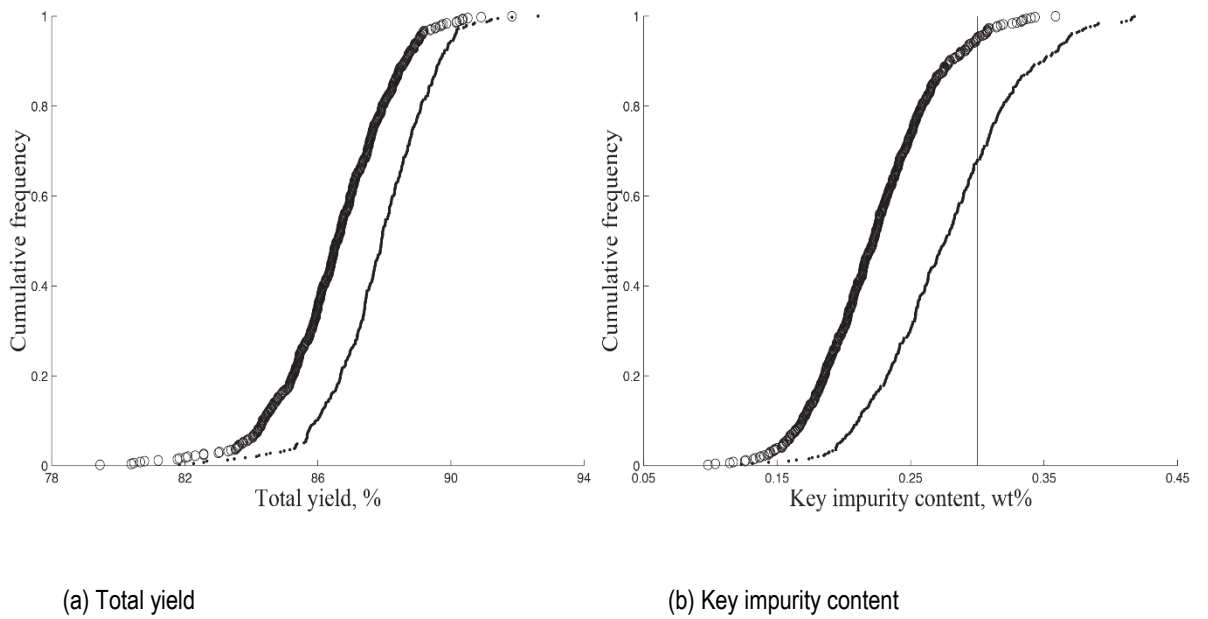


Figure 13. Cumulative frequency plots for validated results for optimisation with and without uncertainty consideration, Case Study II

Key: ● = nominal optimisation, ○ = optimisation under uncertainty, — = PPR data

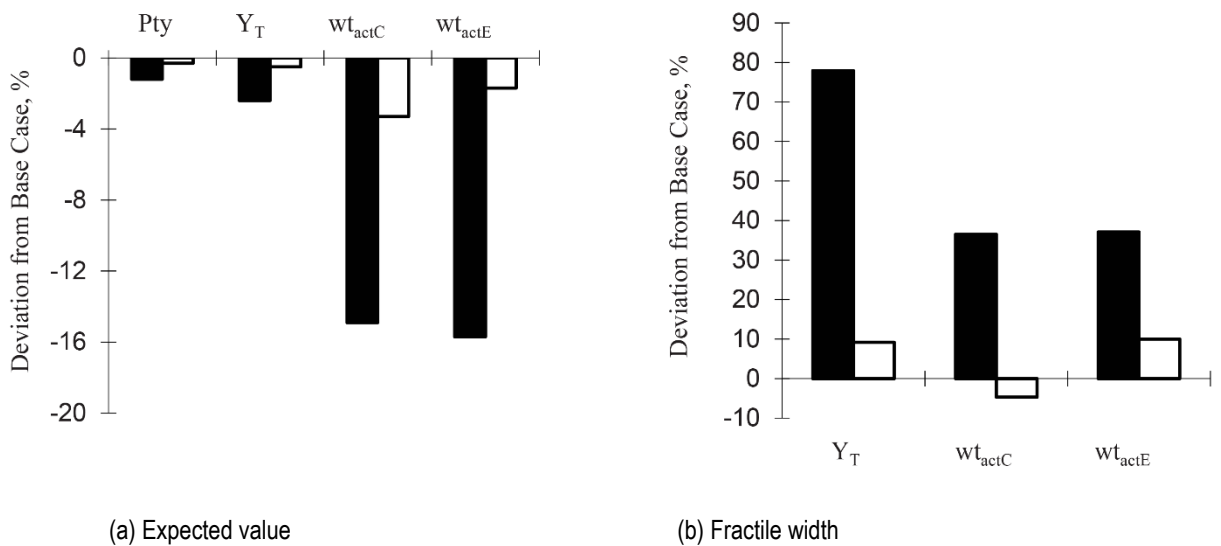
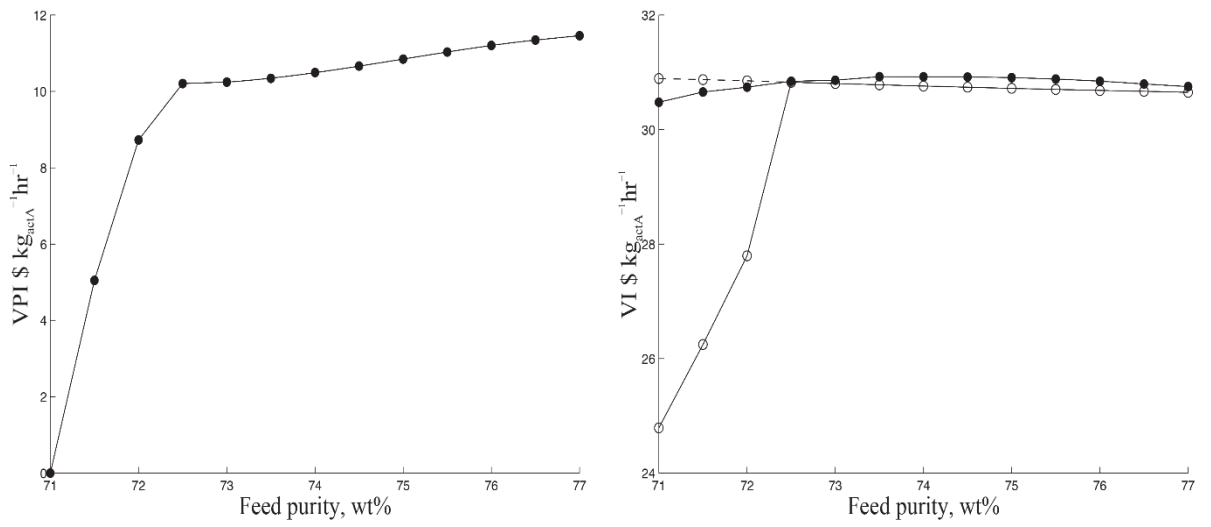


Figure 14. Deviation in optimal results with input uncertainty variations, from the Base Case

Key: unfilled = Case 1, filled = Case 2



(a) Value of Perfect Information

= V_{Iwait} , 0 = V_{Ihere} ,

(b) Value of Information

Key: ●

---- = unpenalised V_{Here}

Figure 15. Value of feed purity information to potential profitability