Chapter 1

Thesis overview and literature survey

This Thesis is concerned with the development of technologies for product quality monitoring in the batch manufacturing of high value added goods. Two kinds of products are considered: those whose "quality" is determined by chemical/physical characteristics (e.g., viscosity, concentration, …), and those where surface properties (e.g. texture, roughness, …) define "quality". Two main issues are investigated: *i*) the development of a strategy to design of soft sensors for the online estimation of product quality and the realtime prediction of batch length in batch chemical processes; and *ii*) the development of a strategy to design of automatic systems for surface characterization in the manufacturing of hardware devices. Tools from multivariate statistical analysis (namely, projection to latent subspaces) are used to develop the proposed technologies.

In this Chapter, after an outline of the aims of the Thesis, the concepts of quality and statistical quality monitoring are briefly reviewed. Then, a survey will follow on the use of multivariate statistical tools for statistical process control, with particular reference to batch processes, for which several challenges are still open for investigation. A roadmap to the reading of the Thesis will conclude the Chapter.

1.1 Aim of the project

Ensuring the conformance of the final product to a predetermined standard is of vital importance in high value added manufacturing in order to achieve the success in today's increasing competitiveness of the global market. However, satisfying the requirements of the customers and meeting reproducibility and high quality of the final product is particularly difficult in most processes. Furthermore, most of the manufacturing processes are inherently multivariate, and quality itself is the multivariate expression of a plurality of indices that are related to process, possibly subject to visual features, and sometimes to personal judgement as well. The aim of this project is the development of multivariate statistical tools that enable to monitor the product quality in batch manufacturing systems in a systematic manner, in such a way as to analyze quality through the information embedded in process data or in images of the product. The proposed techniques are applied to different case studies:

- the development of a strategy to design multivariate statistical soft sensors for the estimation of the product quality and for the prediction of the batch length in batch processes;
- the development of a strategy to design an automatic method for the monitoring of the surface quality of a product through multiresolution and multivariate image analysis.

The systems for the realtime estimation of product quality and for the realtime prediction of the batch length are applied to the case of a real-world industrial process for the production of resins by batch polymerization. This case study demonstrates that the proposed techniques are effective strategies to help the online adjustment of the process recipe when the quality deviates from the nominal conditions and before the final product is affected. Furthermore, these are a valid support for the organization of the production and for the scheduling of the use of the equipment and the coordination of the labour resources.

The novel methodologies developed for the automatic characterization of the surface quality by image analysis are applied to the case of the surface monitoring in the after-photolithography inspections that are carried out in the manufacturing of integrated circuits. In detail, a fully automatic system for the assessment of the surface characteristics of a semiconductor is developed to perform the monitoring of both the surface roughness and the surface patterns.

To sum up, the main contributions of the PhD project are:

- the development of innovative technologies for the online estimation of the product quality in batch processes;
- the non-conventional application of latent variables subspace methods for the prediction of the length of batch processes;
- the development of new methodologies for the multiresolution and multivariate systematic monitoring of the product quality from images of manufactured products.

1.2 Introduction to quality and statistical quality monitoring

The quality movement traces its roots back to the late 13th century, when European craftsman began organizing into “guilds”, responsible for suggesting strict rules on the product and service quality, for adopting inspection committees, and for promoting special marks for flawless goods. Later, the industrial revolution followed this example. However, it was only after World War II that the idea of the “total quality” was introduced, and the notion of “inspection” extended to process technology improvement. Nowadays, “quality” embraces the entire organization of a company and, in the increasing competition of the global market, it is of critical importance that every process can manufacture high quality products with maximum yield. Meeting quality requirements is especially difficult when products consist of large numbers of components, or when processes consist of dozens, even hundreds, of individual steps (Seborg *et al.*, 2004). For example, batch processes for chemical manufacturing and microelectronic fabrication are carried out through a series of operating steps, where quality in each stage is strictly related to the quality of the other stages and heavily influence the final product quality. This results in the need of quality-oriented technologies. On October 1st, 2008, during the meeting on the "Future of quality" of the American Society for Quality (Milwaukee, WI, USA), it was pinpointed that the 21st century technologies are one of the key forces that will shape the future of the quality (`http://www.asq.org/index.html`). This PhD Thesis inserts in this scenario, developing automatic techniques for the realtime quality assessment in the high value added productions.

The concept of quality is still not completely defined. In the common sense, quality is the degree of excellence of a product, a process, or a service. From the engineering point of view, quality is assumed to be a measurement of the conformance to a required standard, to guarantee high performances in terms of reliability, serviceability, durability, etc… (Montgomery, 2005). Namely, the purpose of quality is not only to force a product or a process to respond to predetermined features in order to reach a target or a nominal value in terms of physical, sensory, or time-oriented characteristics (quality of design), but also to improve the product and the process performances in order to reduce the defectiveness, the scraps, the costumer complaints, the rates of waste and of rework (quality of conformance). Therefore, the aim of quality monitoring is not only to monitor the quality of design, but also the quality of conformance (Montgomery and Runger, 2003). In summary, quality is inversely proportional to variability.

Since the variability is an inconsistency that introduces unevenness and determines the major sources of poor quality, the improvement of quality can be reached through the decrease of the variability in products and processes. To reduce the variability, one of the most effective tools is the systematic use of statistics. In his pioneering work, Shewhart (1931) showed how the fundamental steps of the engineering quality control (i.e.: specification of the process goals; fabrication of in-spec products; and tests on the fabricated devices) can be traced by statistical quality control (SQC). SQC fixes (statistical) limits on the state of the production, and improves the uniformity of the quality, assessing the agreement of the product/process to an optimal reference. SQC has gained increasing interest both by the research community and by the industrial one (Hare, 2003).

It should be acknowledged that quality is a synopsis of multiform attributes, depending on a composite combination of related parameters, which are often not accessible by common instrumentation hardware, sometimes not even measurable or quantifiable. Otherwise stated, quality is an inherently multivariable attribute. Furthermore, quality is often related to the values of all the process variables that can be measured during the product manufacturing. On this basis, classical SQC has moved a step forward to statistical process control (SPC) (Geladi and Kowalski, 1986; Wold *et al.*, 1987; MacGregor *et al.*, 1991; Jackson, 1991). SPC unveils the multivariate nature of a system and, furthermore, it can relate the quality parameters to the conditions in which the production process is carried out (Kresta *et al.*, 1991; MacGregor *et al.*, 1991).

1.3 Multivariate statistical techniques for process monitoring

Generally speaking, SPC is a field of technology expansion, whose philosophy is to supervise the process performances over time for emphasizing the anomalous events leading to the degradation of the quality specifications (Kresta *et al.*, 1991; Romagnoli and Palazoglu, 2006). Therefore, the goal of SPC is the quick and reliable detection of the existence, the amplitude and the time of occurrence of the changes that cause a process or a quality feature to deviate from a prescribed standard in the manufacturing of a product. SPC supports this task (MacGregor *et al.*, 1991; Kourti and MacGregor, 1995; Seborg *et al.*, 2004) and facilitates to quantify the probability in observing a process behaviour that does not conform to the expected one (Nomikos and MacGregor, 1994; Flores-Cerrillo and MacGregor, 2002 end 2003; García-Muñoz *et al.*, 2003). Consequently, SPC not only provides underlying information on the state of a plant or of a product, but also assists the operators and the process engineers to remedy a process abnormality (fault[1]). The results are safer operations, downtime minimization, yield maximization, quality improvement, and reduced manufacturing costs (Chiang *et al.*, 2001; Edgar, 2004).

Since in the industrial practice every process exhibits some variability regardless how well it is designed, operated, and instrumented, it is important to discriminate between the common cause (natural and random) variability, which is a cumulative outcome of a series of unavoidable phenomena, and the abnormal (non-random) variability triggered by assignable causes, such as process changes, faulty conditions, errors, etc… The common cause variability is a sort of “background noise” that should operate with only “chance causes of variation” (Montgomery, 2005). This allow processes/products to stay in a state of statistical control. Unfortunately, other kinds of variability may occasionally be present in the output of a process, arising form improperly maintained (or controlled) machinery, operator errors, defective raw materials, unavoidable events, etc… The assignable causes lead to unacceptable levels of process performances or product defectiveness, and determine an out-of-control state. SPC helps investigating what does not work in a process and assists in undertaking the corrective actions before non-conforming products are manufactured. Therefore, monitoring is not only understanding the status of the process, but also the possibility of controlling the product quality. Direct inspection of the quality is usually impractical or, at least, delays the discovery of the abnormal process conditions, because the appearance of the defects in the

[1] A fault is an unpermitted deviation in a system (i.e.: process changes, disturbances, problems to sensors or actuators), which is often not handled adequately by process controllers.

final product takes time. However, information about the quality is encoded in the process variables, which are often measured online, frequently and in an automatic fashion, thus enabling the refinement of the measure information and the inference of the product quality (Kresta *et al.*, 1994; Çinar *et al.*, 2003). In this way one can examine both the process performance and the product quality, ensuring repeatability, stability and the capability of the process to operate with little variability around an assigned target (i.e., the nominal conditions). Accordingly, SPC is a powerful tool to achieve process stability and improving process capability (Montgomery and Runger, 2003).

Traditional monitoring methods consist of limit sensing and discrepancy detection (Chiang *et al.*, 2001). The limit sensing raises an alarm if the state of the observed system crosses predetermined thresholds, while the discrepancy detection raises an alarm depending on model accuracy. The limit sensing imposes some limits to the observations of every process variable, but ignores the relation of each variable with the other ones (i.e., it is univariate).

To detect the departures from a prescribed state of statistical control, control charts can be used. Their use is entrusted because they are proven techniques for improving productivity, are effective in defect avoidance, prevent unnecessary process adjustments, and provide diagnostic and process capability information. In statistical terms, the control charts are hypothesis testing techniques[2] that verify if a process/product is in a state of statistical control. The in-statistical-control condition is the null hypothesis[3] to be proved. The null hypothesis is verified, with a certain degree of uncertainty (level of confidence or significance) when the status of the observed phenomenon stays in proximity of the nominal conditions. Being the nominal conditions identified by the process average conditions, and the amplitude of the confidence limits identified by the common cause variability, moving the limits farther from the average conditions (rising the degree of uncertainty) decreases the risk of type I error[4] (false alarm), and increases the chance of type II error[5] (scarce sensitivity).

The procedure suggested by Kourti (2003) for statistical process control develops through:

- selection of the most representative observations (process data) from an historical database to the purpose of the model building. The selected observation should identify the so-called normal operating conditions (NOC) ;
- pre-treating of the input data to facilitate the statistical analysis;

---

[2] The statistical hypothesis testing is a methodology to make statistical decisions based on experimental data, almost always made rejecting, or failing to reject a null hypothesis.

[3] The null hypothesis is a statement about a plausible scenario which may explain a given set of data and is presumed to be sufficient unless statistical evidence. The null hypothesis is tested to determine whether the data provide sufficient reasons to pursue some alternative hypotheses.

[4] The type I error (or α-error, or false positive) is rejecting a correct null hypothesis, i.e. a false alarm. It occurs every time an out-of-control state is called by the monitoring charts when there is no assignable cause.

[5] The type II error (or β-error, or false negative) is failing to reject a null hypothesis when it is false, i.e. an inadequate sensitivity. This is the risk that a point may still fall within the confidence limits of the monitoring charts when the status is really out of control.

- model calibration;
- checking the "observability" of the model, to test the efficiency of the monitoring model through a validatory procedure;
- checking the performances of the monitoring model in the diagnosis of the special causes that affect a process or a product and determine a detriment of the quality or a loss of process performances.

In typical industrial scenarios, hundreds, if not thousand of process data are available every few seconds, being collected online from process computers and stored in the supervision systems (Nomikos and MacGregor, 1995a; Nomikos, 1996). These data are characterized by spatial correlation (i.e. relations among variables) and serial correlation (i.e. relations among measurement of the same variable taken at different times or locations). Spatial correlation is due to the fact that several process variables are usually sampled throughout the process, and the response to a certain assignable cause affects several process variables. This means that the process variability is usually restricted to a much lower dimension than the one related to the number of variables collected in a process. The process data are serially correlated, as well, because of the relatively small sampling intervals. Furthermore, missing data and noise are often present. The need to handle correlation, noise, and missing data and the requirement to keep the dimensionality of highly correlated data to a reasonably low level calls for the calibration of multivariate statistical models, such as principal component analysis (PCA) and projection to latent structures (PLS, or partial least squares regression). PCA and PLS are data-driven methodologies with computationally non-expensive input-output model structures (Kresta *et al.*, 1994; Cinar *et al.*, 2003), whose frame is a typical black-box representation that derives from the historical data collected during experiments or industrial practice. For the purpose of SPC, PCA and PLS can be used to analyze process data, and to develop inferential models or statistical process control schemes (MacGregor *et al.*, 1991). Both PCA and PLS extract the most important, systematic information hidden into process data, usually assembled in bidimensional (2D) matrices (observations×variables), and compress it through algebraic concepts, in such a way that the information is found in the correlation pattern rather than in the individual variables' signals (Eriksson *et al.*, 2001). Hence, massive volumes of highly collinear and noisy variables can be examined by projecting them onto a subspace made of few fictitious variables, called principal components (PCs) or latent variables (LVs), which explain the direction of maximum variability of the data and contain the greatest part of the relevant information embedded into data. Therefore, both methods are concerned with explaining the variance and covariance structure of a dataset through linear combinations (i.e.: PCs and LVs) of the original ones. This is the reason why PCA and PLS models are linear correlative representations, but not causal models. Note that PCA and PLS have slightly different meanings. In particular, if the case is interpreting and modelling one block of data (e.g., process data), PCA is the proper solution (Jackson, 1991; MacGregor *et al.*, 1991;

Kourti and MacGregor, 1995). If it is necessary to investigate the relationship between two groups of data (e.g., process variables and quality variables) to solve a regression problem, the proper method is PLS, which can estimate or predict some response variables from a collection of predictor variables (Geladi and Kowalski, 1986; Höskuldsson, 1988; Kresta *et al.*, 1991; Burnham *et al.*, 1999; Wold *et al.*, 2001). In summary, the former method maximizes the variance captured from the input data, while the latter maximizes the covariance between the predictor variables and the predicted ones. Although in this Thesis the main interest is in process engineering applications of multivariate statistical methods, several applications of these techniques are reported in the most diverse fields. An incomplete excerpt of some recent applications outside the process engineering community is reported in Table 1.1.

Table 1.1 *Topics of recent papers on applications of multivariate statistical methods in non-process engineering areas.*

Reference	Area	Topic
Dokker and Devis (2007)	biology	sunflower and maize root cell structure study
Giri *et al.* (2006)	biology	examination of the metabolism of nut alkaloids in mice
Škrbić and Onjia (2007)	biology	detection of microelement content of wheat
Viñasa *et al.* (2007)	geology	volcano surveillance
Harrison *et al*. (2006)	medicine	texture analysis of non-Hodgkin lymphoma
Lee *et al.* (2008b)	medicine	citoxicity of substances for cancer treatment
Tan *et al.* (2005)	medicine	persistence of pollutants in adipose tissue
Whelehan (2006)	medicine	detection of ovarian cancer by proteomic profiles:
Übeyli (2007)	medicine	automated diagnostic system for breast cancer
Giordani *et al.* (2008)	energy	electronic-nose for bio-diesel sources identification
Kirdar *et al.* (2008)	bioprocessing	supporting key activities for bioprocessing
Trendafilova (2008)	mechanics	vibration based damage detection in aircrafts wings
Durante *et al.* (2006)	food processing	fragrance sensing and taste estimation
Apetrei *et al.* (2007)	food processing	fragrance sensing and taste estimation
Marín *et al.* (2007)	food processing	fragrance sensing and taste estimation
Arvisenet *et al.* (2008)	food processing	fragrance sensing and taste estimation
Clément *et al.* (2008)	food processing	fragrance sensing and taste estimation
ElMasry *et al.*, (2008)	food processing	defects detection
Quevedo *et al.* (2002)	food processing	food classification and characterization
Doneski *et al.* (2008)	food processing	food classification and characterization
Schievano *et al.* (2008)	food processing	food classification and characterization
Viggiani *et al.* (2008)	food processing	food classification and characterization
Liu *et al.* (2008)	food processing	evaluation of aging or maturity
Qiao *et al.* (2007)	food processing	quality survey
Kim and Choi (2007)	image analysis	face recognition
Liu *et al.* (2007b)	image analysis	mineral processing
Liu *et al.* (2005)	image analysis	wood manufacturing

Finally, multivariate statistical techniques can be extremely useful in the analysis of data from non-conventional sensors (e.g., cameras) and are applied to the field of image analysis as multivariate image analysis (MIA; Geladi, 1995), either in some of the classic fields of chemical engineering, such as plastic material processing (Liu and MacGregor, 2005), steel industry (Bharati *et al.*, 2004; Liu *et al.*, 2007a), and furnaces flames control (Szatvanyi *et al.*, 2006), or in other applications for high value added productions, namely wood manufacturing (Bharati *et al.*, 2003), snack-food statistical quality monitoring and control (Yu and MacGregor, 2003; Yu *et al.*, 2003) and food processing and packaging (Du and Sun, 2004; Brosnan and Sun, 2004; Du and Sun, 2008).

Because batch manufacturing is the main focus of this project, in the following subsections a survey on how the SPC is applied to batch processes is presented.

1.3.1 Multivariate statistical process control for batch processes

Batch and semi-batch processes are used to manufacture high value added goods, such as specialty chemicals and biochemicals, polymers, composites, pharmaceuticals, and materials for food, agriculture or microelectronics. With respect to their continuous counterpart, batch processes can accommodate multiple products in the same production facility, are flexible, easy to set up, and relatively simple to carry out, because the processing recipe usually evolves through a sequence of elementary steps performed in a assigned order to yield relatively small volumes of product with specified quality. Furthermore, for a batch process to be set up, it is often sufficient to have limited fundamental knowledge of the underlying process mechanisms.

Although the batch manufacturing of a product is performed according to a given recipe, the product quality may show great variability, if no corrective actions are taken, and it is often difficult to manufacture multiple consistent products in accordance to strict requirements. In many instances, to meet the quality specification, only the batch duration is adjusted. Sometimes, the operating recipe can be corrected in real time in addition.

There are several reasons that make batch monitoring and control an hard task (Seborg *et al.*, 2004): the time varying characteristic of batch processes; their nonlinear and irreversible behaviour; lack of adequate mechanistic and fundamental models; lack of online sensors, sensor inaccuracy and infrequent sampling of quality indices; existence of constrains; unmeasured disturbances (i.e.: operators errors, fouling, impurities of raw materials, etc…).

The data routinely obtained online from batch processes are not only multivariate in nature, but also nonlinear, highly auto-correlated and cross-correlated[6], and time varying. The time variation implies that a new dimension should be taken into account in the data, i.e. the time. Namely, the data from batch processes can be collected in three-dimensional (3D) matrices (observations×variables×time) that hold both the variation between batches and the variation in time within a batch. PCA and PLS models are linear correlative models, which are valid when the correlation structure of the data remains unchanged in time. However, the correlation structure of the data usually changes during a batch run (Kourti, 2003). Moreover,

[6] The auto-correlation identifies repeating pattern during time or along the space in a periodic signal. The cross-correlation is a measure of similarity between signals.

it changes not only within a batch, but also between batches, due to process changes, plant maintenance, sensor drifts, seasonal effects, etc… For this reason the multivariate statistical techniques evolved to embody not only the multivariable and correlative structure of the data, but also the nonlinearity and the time-varying nature of the batch data.

To face the problem of time variation and change in the correlation structure of the data several methods have been suggested. Basically, four classes of approaches are highlighted in the literature:

- nonlinear multivariate statistical methods, which are the traditional multivariate statistical techniques modified in a nonlinear manner and tailored for the nonlinear nature of the input data and the nonlinear correlation structure of the data;
- multiway models, in which time is considered as an additional dimension of the data and the variability during time evolution can be assessed;
- multiphase models, which split the data in series of segments in which a steady correlation structure of the data is preserved;
- preliminary treatment of the data, in such a way as to rectify the inputs to a multivariate statistical method, either by decomposing the data signals in different resolution scales (e.g. through wavelets transform), or by de-correlating the dataset through auto-regressive moving average (ARMA) models or state space modelling.

In the following sub-sections, the main characteristics and the limits of the abovementioned four classes of multivariate statistical methodologies are overviewed.

1.3.1.1 Nonlinear multivariate models

Nonlinear multivariate statistical techniques were developed to overcome the problem of nonlinearity of the input data and of the nonlinear correlation structure of the data. The key strategy is to alter the algorithm of the PCA and the PLS to include the nonlinearity in the model, either through imposing nonlinear relation between variables (Wold *et al.*, 1989; Baffi *et al.*, 1999a), or through a neural network framework (Baffi *et al.*, 1999b; Doymaz *et al.*, 2003; Zhao *et al.*, 2006b). The search for the right nonlinear structure of the model can be very demanding.

1.3.1.2 Multiway multivariate models

When batch processes have to be examined and the third dimension (i.e.: time) is present in the data, the most popular multivariate statistical strategy is multiway SPC (Nomikos and MacGregor, 1994). Multiway PCA (MPCA) and multiway PLS (MPLS) are statistically and algorithmically consistent with PCA and PLS, respectively. In fact, MPCA and MPLS are equivalent to perform respectively PCA and PLS on an augmented 2D matrix derived by unfolding the 3D matrix.

In the so-called batch-wise unfolding (BWU) method, the data are spread out in a 2D matrix that considers the data time order (Wise and Gallagher, 1996), putting side-by-side the time slices of the original 3D matrix. Simple pre-treatment of the input data (i.e., mean-centring[7]) can remove the major nonlinearity of the variables (Nomikos and MacGregor, 1995b). The result is that BWU-MPCA and BWU-PLS summarize the variability of the data with respect to both the variables and their time evolution (Kourti and MacGregor, 1995). Accordingly, the cross-correlation between variables is explained together with the auto-correlation within each variable. Namely, the entire history of the batch is taken into account and the batch dynamics is properly represented into the model. This is an effective approach for a batch-to-batch monitoring strategy, but some problems arise in the realtime monitoring during a batch run. In fact, not only the BWU approach starts to work well only by the time that at least the 10% of the batch history is already available (Nomikos and MacGregor, 1995b), but also it has two main drawbacks: *i*) the batch processes to be monitored must all have the same length, and *ii*) the *entire* history of a batch should be available *during* the batch evolution in order to be able to complete the 2D process data matrix.

To solve the latter problem, Nomikos and MacGregor (1995a) suggested to fill the incomplete matrix under the hypothesis that either the future unknown observations conform to the mean reference conditions, or the current deviation from the mean variables' trajectory remain unchanged for the rest of the batch duration.

The problem of uneven batch duration is very demanding. Using the BWU-MPCA or BWU-MPLS requires effective methods for the alignment and synchronization of the variables time trajectories, by stretching or shrinking the batch run to the length of a reference one. The most popular methods for the synchronization of the variables profiles are dynamic time warping (Kassidas *et al.*, 1999) and indicator variable (Westerhuis *et al.*, 1999). The latter method uses a monotonic variable as a batch maturity index, so that it is possible to align the batches, being the indicator variable an index of the percentage of batch completion. Otherwise, an indicator variable is not always available among the data. On the other hand, the dynamic time warping is a signal synchronization technique based on a pattern matching scheme of couples of trajectories, expanding or compressing a variable profile to match a reference one. Despite some attempts to streamline the computational burden (Kaistha and Moore, 2001; Ündey *et al.*, 2002), the warping requires a very expensive algorithm structure and only few online applications of synchronizing strategies have been reported (Fransson and Folestad, 2006; Srinivasan and Qian, 2005 and 2007). Additionally, the synchronization is not always practicable, because it often entails the interpolation of the existing data in fictitious time points that can alter the auto- and cross-correlation structure of the data (Kourti, 2003).

---

[7] Mean-centring is a pre-treating procedure operated subtract the mean of each variable to the actual value.

Alternative MPCA or MPLS strategies were developed. One such approach refers to a different unfolding methodology of the 3D data structure, i.e. the so-called variable-wise unfolding (VWU). VWU (Wold *et al.*, 1987) spreads out the batch data in 2D matrices that preserve the direction of the variables, but do not consider the data time order. Variable-wise unfolded matrices are constituted by putting the horizontal slices of the original 3D matrix (i.e. observations) in vertical position one underneath the other. Using this procedure, neither estimating the future unknown part of the batch, nor synchronizing the batches are necessary. This results in easier online application than BWU approach, because filling the incomplete matrix with fictitious observations and aligning variables profiles of uneven length would introduce a certain degree of arbitrariness. On the contrary, VWU has the disadvantages that: *i*) it does not consider the time order, so the dynamics of the batch is lost, and the auto-correlation of the variables' signals is not considered, and *ii*) the correlation structure is forced to be constant during the entire batch (Kourti, 2003). Accordingly, the issue in the VWU scheme is to take into account the dynamics of the process, the data auto-correlation, and the change of cross-correlation during time.

The dynamics of a process can be included into a VWU framework assuming an autoregressive (AR) structure. An AR model regresses the present (or future) values of a variable through a linear combination of the values of the same variable at the previous time instants. This is completely consistent with the fact that in dynamic processes the current state depends on the past time points (Ku *et al.*, 1995). This idea can be easily integrated into the VWU scheme by putting side-by-side the VWU data with the lagged version of the variables' time signals in the so-called dynamic PCA (DPCA) and dynamic PLS (DPLS) procedures. Lu *et al.* (2005b) introduced a dynamic structure to compute the dynamic effect both within a batch and between consecutive batches. In general, DPCA and DPLS are straightforward methods to take into account the process dynamics, and the result is a much more limited correlation of the system (Chen and Liu, 2002). However, the issue of the data nonlinearity and the change in the correlation structure for the VWU approach are still present.

1.3.1.3 Multiple multivariate models

Multiple model approaches based on a BWU strategy are: *i*) the local models (one model per sampling instant; Rännar *et al.*, 1998); *ii*) the evolving models (one model for every sampling instants and all the past sampling instants; Louwerse *et al.*, 2000; Ramaker *et al.*, 2005); and *iii*) the moving window models (models for a limited part of the batch, the current sampling instant and few past observations; Lennox *et al.*, 2001; Lee *et al.*, 2004). The abovementioned multiple model approaches do not necessitate the filling of the incomplete data matrix with future observations. However, they require the synchronization of the batches, and involve a very large number of models, that is not always feasible.

The alternative is splitting the process into a sequence of approximately linear segments (Kourti, 2003), following a multi-model structure based on the VWU-MPCA (or MPLS) analysis. At first, the need for phase division was introduced for the monitoring of multiple operating modes in continuous processes (Hwang and Han, 1999), but it revealed to be a viable and efficient solution for batch processes, too (Ündey and Çinar, 2002; Ündey *et al.*, 2003a; Camacho and Picò, 2008a and 2008b). Therefore, more than one model is derived for a batch, each one for a different phase within the batch (Zhao *et al.*, 2006a). The multiple phase modeling attenuates the problems related to the nonlinearity, and tracks the changes of correlation between variables during the batch. Camacho and Picò (2006a and 2006b), Lu *et al.* (2004a, 2004b and 2004c), Lu and Gao (2005a and 2006), Zhao *et al.* (2007a; 2007b) and Yao and Gao (2009) have designed different strategies for the automatic phase detection and switching.

1.3.1.4 Preliminary data treatment for multivariate statistical methods

The preliminary treatment of the multivariate input data can be performed through: *i*) multi-resolution methodologies of decomposition of the input signals on different frequency scales, or through *ii*) ARMA models and state space modelling to remove the correlation between data.

The latter methods intend to erase any correlation on the latent space of the PCs (or LVs). Indeed, the multivariate statistical representations usually show high degree of auto-correlation of the PCs and the LVs. This determines high rate of false alarms in SPC systems. Furthermore, the filtering of the PCs (or LVs) with ARMA models can remove the auto-correlation. However, the univariate ARMA approach may not be sufficient for clearing the correlation, as demonstrated by Xie *et al.* (2006). Furthermore, the faults magnitude and time signatures of a process may be distorted by the ARMA filtering action (Lieftucht *et al.*, 2006), so a Kalman innovation or state space models result to be preferable (Table 1.2) to better represent the multivariate case (Ljung, 1999).

Table 1.2 *Some papers on the methods for the data linearization based on Kalman innovations and state space models.*

Paper	Topic
Xie *et al.* (2006)	Kalman innovation
Lieftucht *et al.* (2006)	Kalman innovation
Shi and MacGregor (2000)	state space models
Li and Qin (2001)	state space models
Treasure *et al.* (2004)	state space models
Lee and Dorsey (2004)	state space models

In order to de-correlate the variables and to extract the deterministic features of a signal the wavelet transform can be used. The wavelet transformation produces a rectification of a signal for any aperiodic, noisy, intermittent and transient signal, examining it in both the time and frequency domain (Addison, 2002). Mathematically speaking, the wavelet transform is a convolution of a wavelet function with a signal, which converts the signal in a more amenable way (Addison, 2002). In fact, the transformed version of the signal is filtered in such a way as to result more easily manageable (linear and stable white noise) by multivariate statistical techniques, making it suitable to work with data that are typically non-stationary and represent the cumulative effect of many underlying phenomena, each operating at a different scale, such as in batch processes (Kosanovich and Piovoso, 1997). In this way, the contributions of different scales of resolution are detected for all the events whose behaviour change over time and frequency. Once the signal is decomposed in different scales of resolution, the multivariate statistical model can be built both in the domain of the frequency (through the approximations and the details of the signal) and in the time domain (reconstructing the filtered version of the signal). Usually, one model is built for each decomposition scale (Bakshi, 1998; Bakshi *et al.*, 2001; Yoon and MacGregor, 2004; Lee *et al.*, 2005b; Maulud *et al.*, 2006; Chang *et al.*, 2006), and considers only the most interesting scales to the purpose of the monitoring, either by denoising the signal (Shao *et al.*, 1999) or by removing the higher frequencies to avoid the effects of the process drifts or the seasonal fluctuations (Teppola and Minkkinen, 2000). Moreover, these techniques are very useful for an unambiguous fault detection (Misra *et al.*, 2002) and isolation (Reis *et al.*, 2008).

1.3.2 Multivariate image analysis

In recent years, some attractive industrial applications involve the use of non-conventional and non-invasive sensors, such as cameras, for product quality characterization. Images are 2D light intensity mapping of a 3D scene, and are characterized by several challenging issues:

- high dimensionality of the space, because images are may not only be monochromatic representations on gray levels, but may also have several channels of transmission (e.g.: RGB[8] images, hyperspectyral images, etc…);
- multivariate nature, because an image is an aggregation of a wide plurality of pixels[9];
- different characteristics in different scales of resolution;
- high spatial correlation, because of the effect of neighbourhood of the pixels;
- non-linearity, because of the physical structure of the object that is represented in the image;
- combination of spatial and spectral information;
- presence of noise, a random fluctuation of the light intensity that is an artefact of the signal.

---

[8] RGB is a representation of the colours from an additive model derived by the primary colours red, blue and green.

[9] In digital imaging, the pixel is the smallest piece of information of an image arranged in a 2D grid.

Multivariate statistical methods are ideal techniques to deal with the high dimensionality of the images and their inherent multivariate nature. Accordingly, multivariate image analysis (MIA) gained increasing interest (Geladi and Grahn, 1996) for both inferential modeling and statistical process control. MIA is a set of multivariate statistical techniques that allow to analyze images in a reduced dimension space rather than in the image space (Kourti, 2005). The aim of this approach is to extract subtle multivariate information from the image, in a different way from the usual digital image processing where the image is enhanced in such a way that its features become visible. Note that the problems of spatial correlation (correlation between pixel), neighbourhood, nonlinearity and noise can be faced analogously to what was suggested in the Section 1.3.1. Indeed, nonlinear models, as well as multiway, multi-model and multiresolution approaches, can be extremely useful and well tailored to the purpose of the image inspection. In fact, to a certain extent, it is possible to associate the concepts of neighbourhood and spatial correlation with the ones of process dynamics, auto- and cross-correlation, and the concept of spatial nonlinearity to the one of temporal nonlinearity. Moreover, images combine spectral (in terms of both light intensity and colour) and spatial information. In the literature, the use of multi-resolution MIA is often suggested (Liu and MacGregor, 2007; Bortolacci *et al.*, 2006), where the spectral information are properly studied by MIA classical approach, while the wavelet transform (Mallat, 1989; Ruttimann *et al.*, 1998) is adopted to grasp the spatial information. Furthermore, the spatial information can be assessed including the study of the textural features of the inspected image (Salari and Ling, 1995; Tessier *et al.*, 2007). In this way, effective frameworks are developed through image analysis for the task of either quality monitoring and control (Yu and MacGregor, 2003; Yu *et al.*, 2003; Borah *et al.*, 2007), or quality classification (Bharati *et al.*, 2004), or quality prediction (Tessier *et al.*, 2006).

1.4 Thesis overview

As was mentioned earlier, the two main topics of this Thesis are the design for multivariate statistical techniques for: *i*) the realtime product quality estimation and length prediction in batch chemical processes, and *ii*) product quality monitoring through image analysis in batch manufacturing. The challenges of both topics are presented and discussed in the following.

1.4.1 Realtime quality estimation and length prediction in batch processes

In principle, the operation of a batch process is easy, because the processing usually evolves through a "recipe", i.e. a series of elementary steps (e.g.: charge; mix; heat-up/cool; react; discharge) that can be easily carried out even without supervision if the production facility is outfitted with a fairly large degree of automation. However, it is often the case that batch plants are poorly instrumented and automated, and may require intervention by the operating personnel to provide online adjustments of the operating recipe with midcourse corrections to avoid the production of off-specification products. In fact, if the instantaneous product quality is not found to track a specified trajectory, the processing recipe must be adjusted in real time (possibly several times during a batch), and the batch is kept running until the end-point quality meets the specification. Unfortunately, most of the batch processes are run in an open-loop fashion with respect to product quality control, because information about product quality is not available online, but is obtained offline from laboratory assays of few product samples. To contain the laboratory-related expenses (in terms of: need of dedicated personnel, consumption of chemicals, use of analysis equipment, etc…) only few product samples are taken during the course of a batch and sent to the lab for analysis. Even so, in a typical industrial scenario where several productions are run in parallel, 15,000-20,000 samples may need to be taken and analyzed each year, which add up to an important fraction of the total product cost. Because of the lack of real time information on the product quality, it may be difficult to promptly detect quality shifts and to counteract them by adjusting the operating recipe accordingly. Therefore, significant drifts on the quality profiles may be experienced before any intervention can be done on the batch. The net result is that the recipe adjustments are delayed, the total length of the batch is increased, and the economic performance of the process is further penalized.

In this context, two typical challenges need to be addressed by a monitoring system in the production of specialty chemicals: the real time estimation of the instantaneous quality of the product, and the real time estimation of the length of the batch (or the length of any production stage within the batch). In fact, the performance of a batch process could be highly improved if accurate and frequent information on the product quality were available. Software sensors (also called virtual sensors or inferential estimators) are powerful tools for this task. They are able to reconstruct online the estimate of "primary" quality variables from the measurements of some "secondary" process variables (typically, temperatures, flow rates, pressures, valve openings), by using a model to relate the secondary variables to the primary ones. These issues are faced in this Thesis with reference to a real-world industrial case study, i.e. a batch process for the production of resins by polymerization.

It is well known that developing a first-principles model to accurately describe the chemistry, mixing and heat, mass and energy transfer phenomena occurring in a batch process (e.g.: polymerization; crystallization; etc…) requires a very significant effort. Several designed experiments may be needed to identify the most representative set of equations and all the related parameters. Furthermore, if the plant is a multi-purpose one, this effort must be replicated for all the products obtained in the same facility. Finally, the resulting first-principles soft sensor may be computationally very demanding for online use.

Multivariate statistical soft sensors may overcome these difficulties (Kresta *et al.*, 1994; Chen *et al.*, 1998; Neogi and Schlags, 1998; Chen and Wang, 2000; Kano *et al.*, 2003; Kamohara *et al.*, 2004; Zamprogna *et al.*, 2004; Lin *et al.*, 2007; Kano and Nakakagawa, 2008; Gunther *et al.*, 2009). This class of inferential estimators does not require to develop extra information on the process in terms of mechanistic equations or values assigned to physical parameters. Rather, they extract and exploit the information already embedded in the data as these data become available in real time from the measurement sensors. Very often, a multivariate statistical method, i.e. PLS, can be exploited to design a soft sensor for the online estimation of quality properties. Several studies about the online estimation of product quality through multivariate statistical techniques are available for continuous polymerization processes. Most of the literature on the application of multivariate statistical methods to batch polymerization processes is related to the prediction of the end-point product quality only, or to batch classification, or is limited to simulation studies, as can be seen in Table 1.3.

Table 1.3 *Literature review on the estimation of the product quality in polymerization processes: papers and topics.*

Reference	Processing	Problem	Data
Russel *et al.* (1998)	continuous	realtime estimation	industrial
Komulainen *et al.* (2004)	continuous	realtime estimation	industrial
Lee *et al.* (2004)	continuous	realtime estimation	industrial
Lu *et al.* (2004b)	continuous	realtime estimation	industrial
Warne *et al.* (2004)	continuous	realtime estimation	industrial
Kim *et al.* (2005)	continuous	realtime estimation	industrial
Aguado *et al.* (2006)	continuous	realtime estimation	industrial
Sharmin *et al.* (2006)	continuous	realtime estimation	industrial
Zhang and Dudzic (2006)	continuous	realtime estimation	industrial
Zhao *et al.* (2006a)	continuous	realtime estimation	industrial
Yabuki and MacGregor (1997)	batch	end-point estimation	industrial
Kaitsha and Moore (2001)	batch	end-point estimation	industrial
Flores-Cerrillo and MacGregor (2004)	batch	end-point estimation	industrial
Ündey *et al.* (2004)	batch	end-point estimation	industrial
Zhao *et al.* (2008b)	batch	end-point estimation	industrial
Nomikos and MacGregor (1995)	batch	realtime estimation	simulation
Rännar *et al.* (1998)	batch	realtime estimation	simulation
Chen and Liu (2002)	batch	realtime estimation	simulation
Ündey *et al.* (2003a)	batch	realtime estimation	simulation
Ündey *et al.* (2003b)	batch	realtime estimation	simulation
Zhang and Lennox (2004)	batch	realtime estimation	simulation
Lu and Gao (2005)	batch	realtime estimation	simulation
Camacho and Picò (2006)	batch	realtime estimation	simulation
Doan and Scrinivasan (2008)	batch	realtime estimation	simulation
Zhao *et al.* (2008a)	batch	realtime estimation	simulation

Very few papers present industrial applications of multivariate statistical software sensors for the realtime estimation of the product quality for industrial batch processes (Marjanovic *et al.*, 2006; Chiang and Colegrove, 2007). In this PhD Thesis, multivariate statistical techniques are proposed to provide the online estimation of product quality in batch industrial polymerization processes.

There are several specialty productions for which the total batch length is not known *a priori*, nor is it the length and the number of the processing stages within the batch. Knowing in advance the processing time is useful for several reasons. In fed-batch processes, for example, fresh raw material and catalysts should be loaded into the process vessels at a convenient time instant to adjust the batch run in real time. The ability to estimate in real time this instant (which may change from batch to batch) can result in savings both in the number of quality measurements to be processed by the laboratory and in the required total processing time (Marjanovic *et al.*, 2006). On a different perspective, realtime estimation of the total length of the batch can be very useful for production planning, scheduling of equipment use, as well as to coordinate the operating labor resources. For these reasons, the non-conventional use of multivariate statistical techniques for the realtime prediction of the batch length is suggested and discussed in the Thesis.

The abovementioned multivariate statistical techniques are applied and implemented to an industrial case study of batch polymerization for the production of resins. This process is monitored online through a fairly large number of process measurements. Several challenging features are present in this case study:

- process measurements are noisy, auto-correlated and cross-correlated;
- quality measurements are available offline from lab assays, but are scarce, delayed with respect to the sampling instant and unevenly spaced in time (a case which is rarely considered in literature);
- the batches evolve through a nominal recipe, which is subject to several online adjustments made by the plant personnel depending on the actual evolution of the batch, as it is monitored by the offline quality measurements, and their personal judgment;
- the process is poorly automated;
- the batch length exhibits a large variability.

All of these features make each batch hardly reproducible, and the online quality estimation a challenge.

1.4.2 Multivariate statistical quality monitoring through image analysis

There is a class of products whose quality is not related to chemical or physical properties, but to surface properties (like roughness, pattern, colour, texture, and the like). For these products, quality is assessed by the analysis of an image of the manufactured device. For example in semiconductor manufacturing image analysis is used for quality monitoring, but only for the task of measuring the most important physical parameters of the manufactured device, despite several other key features of the semiconductor which determine the device quality are hidden and remain unmeasured. In particular, image inspections are used in photolithography. Photolithography is a process that selectively removes parts from a thin film using light, so that a geometric pattern can be transferred (often from a mask) to a light sensitive chemical (the resist) deposited on a substrate. This process is used during the fabrication of integrated circuits (IC) as well as in many other micro-fabrication processes (e.g., micro-compressors in mechanics: Waits *et al.*, 2005; in biotechnology applications: Lee *et al.*, 2008a). In particular, a microelectronics manufacturing process comprises an extensive sequence of complex semi-batch processes (Helbert and Daou, 2001), among which photolithography is referred to as one of the most important (Blais *et al.*, 2001). In fact, photolithography: *i*) recurs up to 35 times for a given device; *ii*) defines the wafer critical dimension (CD) and the other most influencing parameters; and *iii*) affects all the successive processing phases (e.g., the doping) and the interconnection between different segments of the device. From an economical point of view, the lithography is responsible for about 60% of the processing time and 35-40% of the total cost of the IC fabrication (Blais *et al.*, 2001). As a consequence, it is quite clear that monitoring the product quality during photolithography through a fast, sensitive, and reliable system is highly advocated.

Although considerable effort has been dedicated to define technologies and procedures to meet the requirements on the product quality (Guldi, 2004; Yaakobovitz *et al.*, 2007), automatic process control has not yet been implemented on a large scale in semiconductor manufacturing, and the industrial practice is often carried out empirically with relatively little understanding of the underlying physics and chemistry (Edgar *et al.*, 2000), or through run-to-run control strategies (Zhang *et al.*, 2007 and 2008). Statistical process control techniques, too, are sometimes adopted (Edgar *et al.*, 2000; Yue *et al.*, 2000; Waldo, 2001) in order to monitor the variability of the process, to detect the abnormal conditions, and to identify the cause for a perceived anomaly.

Currently, the most advanced monitoring strategies exploit hardware and software devices for both signal filtering and image processing (Rao, 1996; Lee, 2001). For instance, the use of scanning electron microscopy (SEM) images is common for the measurement of the physical parameters of a device (Knight *et al.*, 2006) such as the CD (Constantoudis *et al.*, 2003; Patsis *et al.*, 2003). However, the typical inspecting tools focus on inline optical metrology systems measuring the CD of the pattern and its variability; only the most sophisticated instruments also determine the edge height and the side-wall angle (SWA; El Chemali *et al.*, 2004). Several important quality features like the line edge roughness (LER), the edge surface smoothness, the actual shape of an edge (and its variability) are still rather resilient to effective, fast and low-cost monitoring technologies. Only recently some researchers (e.g., Zhang *et al.*, 2007; Yaakobovitz *et al.*, 2007; Khan *et al.*, 2008) have suggested procedures to start tackling some of the above issues.

Thus, the demand of satisfying the multiple requirements of wafer fabrication and the dynamics of a quickly changeable microelectronics market call for new and more powerful monitoring tools. The quality of the manufacturing could be greatly improved if fast and more meaningful information were retrieved in a reliable fashion. For this reason, an innovative methodology is presented to inspect the surface of a product. In particular, the main components of the proposed quality monitoring strategy are:

- sensitive filtering pre-treatment, to denoise the image signal removing the artifacts (i.e., the non-systematic fluctuations of the image light intensity) without affecting the featured parts and their peculiar characteristics (i.e., the real surface roughness);
- tailored multivariate statistical monitoring models, based on a principal component analysis approach, which extract the information content on surface roughness and patterned shape.

In particular, the analysis is performed by PCA on different scales of resolutions. Innovative modifications of the PCA model are proposed to analyze both the surface roughness and the shape of the patterned surface. The effectiveness of the proposed approach is tested in the case of semiconductor surface SEM images after the photolithography process, but the approaches are general and can be applied also to inspect a product through different types of images or different phases of the same production systems, or through different types of processes.

1.4.3 Thesis roadmap

Chapter 2 overviews the mathematical and statistical background of the methods adopted in this Thesis, i.e. multivariate statistical models and multiresolution techniques. In particular, PCA and PLS are presented, and the issue of both data pre-treatment and model enhancement are discussed. Finally, multiresolution methodologies are recalled.

Chapter 3 describes the industrial process under study (i.e. production of resins by batch polymerization). Details on the plant and on the production recipe are provided. The industrial system of supervision is briefly presented.

Chapter 4 show how to design a multivariate statistical estimators of the product quality for the processes under study. Different architectures of the soft sensor are presented, and improvements of the estimation performance are proposed by including a multiphase structure and dynamic information on the process.

The problem of the prediction of the batch length is the topic of Chapter 5, in which the effectiveness of time-evolving methods is demonstrated.

In Chapter 6, the industrial implementation of prototypes of the abovementioned soft sensors is briefly described.

Chapter 7 deals with the development of a fully automatic monitoring systems for the characterization of the surface of high value added products by means of multiresolution and multivariate image analysis. Reference is made to the manufacturing of integrated circuits. A prototype interface for photolithography monitoring is also presented.

Final remarks conclude the Thesis.