INTERNATIONAL JOURNAL OF ADAPTIVE CONTROL AND SIGNAL PROCESSING
Int. J. Adapt. Control Signal Process. 2003; **00**:1–15

Probabilistic Advisory Systems for Data-Intensive Applications

A. Quinn[1] and P. Ettler[2] and L. Jirsa[3] and I. Nagy[4] and P. Nedoma[3]

[1] *University of Dublin, Trinity College, Dublin 2, Ireland, e-mail: aquinn@tcd.ie*
[2] *COMPUREG Plzeň s.r.o., P.O. Box 334, 306 34 Plzeň, Czech Republic, e-mail: ettler@compureg.cz*
[3] *ÚTIA AV ČR, P.O. Box 18, 182 08 Prague, Czech Republic, e-mail: school@utia.cas.cz*
[4] *Faculty of Transportation Sciences CTU, 110 00 Prague 1, Konviktská 20, Czech Republic, e-mail: nagy@utia.cas.cz*

SUMMARY

Real-world, multidimensional, dynamic, non-linear processes typically exhibit many distinct modes of operation. Mixtures of dynamic models improve greatly on traditional one-component linear models in this context. Improved prediction then points the way to effective adaptive control design. This paper presents the experience gained under the EU Project, ProDaCTool, in designing and implementing advisory systems, based on dynamic mixtures, in diverse domains: urban traffic regulation, therapy recommendations in nuclear medicine, and operator support for metal-strip rolling mills. Efficient, recursive estimation of the dynamic mixtures from archive data is accomplished using the Quasi-Bayes algorithm, implemented with dedicated software developed within ProDaCTool. The advisory systems are designed using the probabilistic control design technique presented in the previous paper. Highly encouraging prediction and performance enhancements are reported for the applications considered.

KEYWORDS: mixture of dynamic models; Bayesian inference; control design; advisory system; applications in transport, medicine, manufacturing.

1. INTRODUCTION

In many complex processes, the control loop is closed by human operators who react to system performance by adjusting controllable parameters. The conjecture underlying the treatment of the *process-operator loop* in this paper is that study of the loop will often reveal a countable number of distinct modes of operation. These modes reflect not just the variable state of the system but, notably, of the operator. The latter is subject to variations of performance and skill, such as might arise with personnel shifts, *etc.* [2]. The approach taken here is to represent each of these modes via a component in a mixture of dynamic models (typically referred to as a *dynamic mixture model* in this paper) [4, 5]. Each component is a probability model for data-generation under an appropriate—usually multidimensional and highly flexible—regression structure [5]. Identification of this multiple model from archive or off-line data is the starting point for the control design strategy presented in [5], and is the one exploited throughout this paper. The second ingredient is a probability function expressing an ideal for the controlled

Contract/grant sponsor: Project ProDaCTool IST-99-12058, AVČR S1075102

system performance. Since, in all applications presented in this paper, the loop is closed via the operator, reference is therefore made to *advisory systems* and their design, rather than to control design *per se*.

In Section 2, the key elements in the advisory design strategy [5] are summarized. In Section 3, urban traffic flow measurements are predicted using various single-component regressions and their mixtures. In Section 4, an advisory system for nuclear medicine dosimetry is developed, using an archive of patient data. In Section 5, the framework is used to generate on-line advice for operators of a metal rolling mill, so as to increase output while maintaining quality. At all stages, the link between effective dynamic prediction of the controlled output, and effective design of the advice itself, is stressed.

2. THE ELEMENTS OF DYNAMIC MIXTURE MODELLING AND ADVISORY SYSTEM DESIGN

The Process-Operator Loop (POL) is assumed to migrate between $\mathring{c}$ modes, each described by a probability model called a *component.* The notation and assumptions of the paper which accompanies this one [5] are adopted. The parameters of each component are assumed to be *fixed* unknowns, as are the mixing parameters: $\Theta_{c;t} = \Theta_c$, $\alpha_{c;t} = \alpha_c$. Hence, a time-invariant multiple model is assumed. Furthermore, the switching process, c_t, is assumed to be non-Markov (white). Each of the $\mathring{c}$ components is decomposed into $\mathring{d}$ *factors* via application of the chain rule in one of $\mathring{d}!$ ways. Each factor expresses a regression model of flexible kind for each scalar channel. Hence, there are $\mathring{c}\mathring{d}$ factors (time-variant, structure-invariant scalar regressions) associated with the model.

2.1. Off-Line Analysis

An archive of off-line data gathered from the POL is analyzed sequentially via the Quasi-Bayes (QB) algorithm [5], to yield an estimated mixture of dynamic models, ${}^{[E]}$. The implied mathematical objects, and their interrelationship, are summarized in the following structure:

`identify` $\mathring{c}$
`for` $t = 1, \ldots, \mathring{t}$
  `for` $c = 1, \ldots, \mathring{c}$
    $d_t = [y_t, u_t]$ `(off-line archive)`
    `for` $i = 1, \ldots, \mathring{d}$
      `extract` $\psi_{ic;t}$ `(regressors of channel` i `in component` c`, time` t`)`
      `identify` Θ_{ic} `(parameters of factor` i `in component` c`)`
      ${}^{[E]}f(d_{i;t}|\psi_{ic;t}, \Theta_{ic})$ `(factor` i `of component` c`)`
    `end`
  `end`
  ${}^{[E]}f(d_t|\phi_{c;t-1}, \Theta_c, c) = \prod_{i=1}^{\mathring{d}} {}^{[E]}f(d_{i;t}|\psi_{ic;t}, \Theta_{ic})$ `(component` c`)`
`end`
${}^{[E]}f(d_t|d(t-1), \Theta) = \sum_{c=1}^{\mathring{c}} \alpha_c \, {}^{[E]}f(d_t|\phi_{c;t-1}, \Theta_c, c)$ `(mixture of dynamic models)`

2.1.1. Advisory Design Stage of the Off-Line Analysis The identified dynamic mixture, ${}^{[E]}f$, above (in factor form typically), forms one key input to the design of optimized advice. The other is the *user-ideal probability (density) function (p.f.)*, ${}^{[U]}f$. In simultaneous design [5], there is a full definition, as follows, at $t \in t^*$:

$\texttt{for}\ c = 1, \ldots, \mathring{c}$
$\qquad \texttt{specify} \quad {}^{[U]}f(c_t|d(t-1)), \quad {}^{[U]}f(u_t|d(t-1)), \quad {}^{[U]}f(y_t|u_t, d(t-1))$
$\texttt{end}$

$$ {}^{[U]}f(d_t, c_t|d(t-1)) \;=\; {}^{[U]}f(y_t|u_t, d(t-1))\, {}^{[U]}f(u_t|d(t-1))\, {}^{[U]}f(c_t|d(t-1)) $$

Typically, the user-ideal kernels are defined for a chosen control horizon (also called $\mathring{t}$ in [5]). The distributions above express the favoured component(s), the favoured process inputs, and favoured process response, respectively, at time t, and are subject to natural conditions of control [5]. In this paper, these distributions are elicited using one of the following approaches:

1. quantifiable management aims concerning desirable ranges of certain variables are expressed as distributions on these variables. Typically, the chosen p.f.s are normals concentrated on these ranges. In cases where there are to be several constraints, Maximum Entropy elicitation can be used [7];
2. data corresponding to desirable modes of operation of the POL can be marked in the off-line archive, and a mixture identified for this subset of the data. This mixture—or its dominant component—then constitutes ${}^{[U]}f$.

The optimized design, ${}^{[I]}f$, is found from ${}^{[E]}f$ and ${}^{[U]}f$, via a Kullback-Leibler (KL) minimization with respect to those parameters of ${}^{[E]}f$ deemed controllable in the user-ideal [5].

2.2. On-Line Analysis

${}^{[I]}f$ alone is passed to the on-line stage. It is evaluated in real-time at the current operating point. Marginals of ${}^{[I]}f$ on prioritized controllable variables are presented to the operator as advice. On-line data may be used to update (in recursive or batch mode) ${}^{[E]}f$, in which case ${}^{[I]}f$ is recalculated. Hence, the generated advice tracks the changing behaviour of the POL.

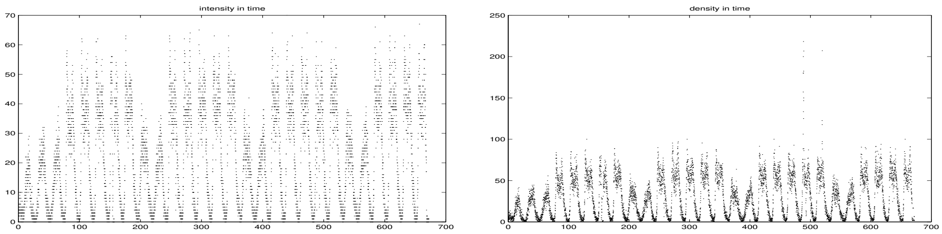
3. URBAN VEHICULAR TRAFFIC PREDICTION

As urbanization and car-ownership increase apace, traffic congestion is impinging ever more seriously on the lives of urban dwellers, and on the commerical and economic welfare of society [11]. Any attempt to address the traffic control problem must respect the difficult constraints imposed by the urban environment. Mixture models seem particularly appropriate in traffic modelling, because (i) epochs can be clearly identified, corresponding to rush-hours, seasonal factors (school holidays, climate influences), etc.; and (ii) current practice in traffic control is to switch between a finite number of pre-programmed traffic light sequences in an attempt to increase the permeability of city crossroads.

In this study, mixtures of static and dynamic components are examined in an attempt to validate their use for traffic data. Specifically, the multistep predictive capability of the identified models is reported. Advisory design (Section 2.1.1) will not be considered for this application.

3.1. Traffic Measurements in the Strahov Tunnel, Prague

Traffic flow can be quantifed at a particular time, t, and place via a magnetic detector under the road. Essentially, a detector records the amount of time, in a short (5 minute) interval around t, for which that point in the road was occupied. Using knowledge of the average length of a vehicle, this record can be converted into two estimated traffic measures: (i) the traffic intensity, q, being the number of cars which pass that point per minute, measured around time t; (ii) the density, ρ, being the number of cars per km at that point and at time t. In Fig. 1, the time course of these variables is shown over a four-week period, for a specific detector placed near the centre of the busy Strahov road tunnel in Prague. It clearly shows the daily and weekly periodicities which are to be expected.



Time course of intensity q Time course of density ρ

Figure 1. Traffic Intensity and Density records for the Strahov Tunnel, Prague.

3.1.1. The Off-line Data Set In this study, a snapshot, d_t, of the traffic state for the entire tunnel is provided via an array of synchronized sensors placed at 5 regularly-spaced points along each of two northbound lanes of the tunnel. Hence, $\mathring{d} = 2 \times 2 \times 5 = 20$. Measurements are recorded every 5 minutes for 4 weeks, generating an archive of $\mathring{t} = 8064$ data items. An extra one day's data (288 items) are reserved for validation of predictions. The task examined here is to predict future traffic state (system output), $y_t = [d_{1;t}, d_{2;t}]'$, at the northbound exit of the Strahov Tunnel, using this off-line archive. The system input is then considered to be the traffic states recorded by the intra-tunnel sensors: $u_t = [d_{3;t}, d_{20;t}]'$. Clearly, there will be strong correlation (i) between sensors, encouraging a regression model between the channels, $d_{i;t}$, and (ii) from snapshot to snapshot, encouraging a dynamic model. The prospect of capturing multimodal behaviour (Section 3) via mixtures of these model types (multiple models) is examined.

3.2. Model Identification for the Off-Line Archive

Two modelling strategies are briefly outlined: (i) *static* models, ignoring correlations between d_t, $t \in t^*$; and (ii) *dynamic* models, seeking to capture these correlations.

3.2.1. Static Modelling The d_t are modelled as (conditionally) mutually independent, *i.e.* $f(d(\mathring{t})|\Theta) = \prod_{t \in t^*} f(d_t|\Theta)$. It follows (Section 2 and [5]) that $\phi_{c;t-1} = \{\}$, $\forall c \in c^*$ and $\forall t \in t^*$. Assuming factorization of components in natural order, then $\psi_{ic;t} = \psi_{i;t} = [d_{i+1;t}, \ldots, d_{\mathring{d};t}]'$. All factors are normal:

$$f(d_{i;t}|\psi_{i;t}, \Theta_{ic_t}, c_t) = \mathcal{N}(\Theta_{ic_t}^T \psi_{i;t}, \sigma_{ic_t}^2),$$

where $\sigma_{ic_t}^2$ is the conditional variance of the ith variable in the c_tth component. The identified mixture therefore seeks to confine each datum, d_t, to a manifold formed as the union of $\mathring{c}$ lines in the 20-dimensional data-space, d^*. The mixture components, $f(d_t|\Theta_{c_t}, c_t)$ are hyper-ellipsoids each orientated with its principal axis along one of these lines, and centred on the identified component mean, μ_{c_t}. Furthermore, (i) a single-component regression, where $\mathring{c} = 1$ *a priori*, is compared to (ii) a multiple model mixture, with $\mathring{c}$ chosen between 3 and 7 (depending on the mixture initialization used).

3.2.2. Dynamic Modelling First-order temporal regression is assumed for all components, and so the factor regression vectors are again component-independent: $\psi_{ic_t;t} = \psi_{i;t} = [d_{i+1;t}, \ldots, d_{\mathring{d};t}, d'_{t-1}]'$. A single dynamic component is again compared to a mixture of dynamic components.

3.3. Prediction of Traffic State

The identified factors (Section 2.1), are used to construct a $\tau > 0$-step-ahead prediction of the 20-dimensional Tunnel traffic state: ${}^{[E]}f(d_{t+\tau}|d(t), \Theta)$. Marginalization over $u_{t+\tau}$, generates the two-variable prediction on the Tunnel output, ${}^{[E]}f(y_{t+\tau}|d(t), \Theta)$. These p.f.s are evaluated on the 288 on-line data items.

3.3.1. Static Model Predictions: ${}^{[E]}f(y_{t+\tau}|d(t), \Theta)$ is used to generate the predictions in Fig. 2. The one-component regression cannot react to changing modes in the data, and simply

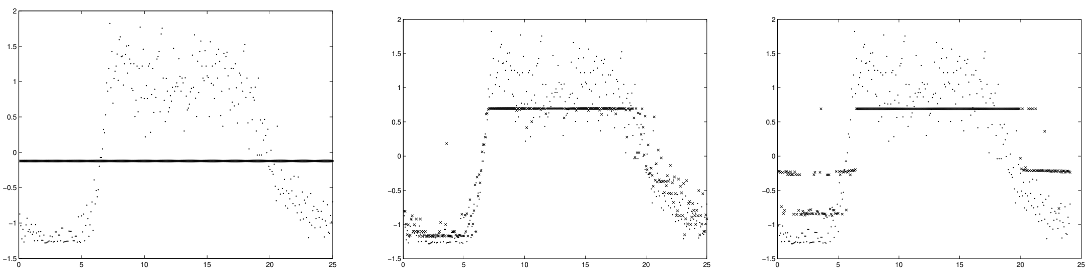


Figure 2. STATIC Predictions of Traffic Intensity at Tunnel Output. **Left:** one-component regression, $\tau = 1$; **centre:** mixture, $\tau = 1$; **right:** mixture, $\tau = 12$. (On-line data (•); predictions (×)).

predicts the conditional mean of the traffic intensity. The mixture model identifies variations in the component pointer, $c_t \in c^*$, permitting far superior prediction of the multimodal data.

For *static* one- and multicomponent models, Table I shows the Prediction Error coefficient (PE), expressed as the ratio of the standard deviation of the prediction error to that of the respective output channel. It confirms the enhanced predictive capability of the mixture. PE increases, as expected, with τ.

τ	$\mathring{c}=1$		$\mathring{c}>1$	
	Intensity	Density	Intensity	Density
1	1.021	1.020	0.418	0.422
6	1.025	1.024	0.512	0.516
12	1.031	1.030	0.634	0.638
18	1.028	1.037	0.756	0.760

Table I. PE for τ-step-ahead STATIC Predictions of Traffic Tunnel Output.

τ	marginal prediction of output				joint prediction of full Tunnel state			
	$\mathring{c}=1$		$\mathring{c}>1$		$\mathring{c}=1$		$\mathring{c}>1$	
	Intensity	Density	Intensity	Density	Intensity	Density	Intensity	Density
1	0.376	0.378	0.391	0.392	0.330	0.331	0.328	0.329
6	0.526	0.521	0.487	0.483	0.444	0.436	0.421	0.416
12	0.711	0.706	0.614	0.615	0.576	0.566	0.524	0.526
18	0.840	0.836	0.758	0.766	0.697	0.688	0.654	0.644

Table II. PE for τ-step-ahead DYNAMIC Predictions of Traffic Tunnel Output (based on marginal prediction, $y_{t+\tau}$, and full state prediction, $d_{t+\tau}$, respectively).

3.3.2. Dynamic Model Predictions: In Table II, *output channel* PEs are recorded for both the 2-D predictive marginal and for the full 20-D prediction. Once again, the improvement offered by the mixture model over the single-component regression is evident. In general, PE increases with τ, as before. Comparing with Table I, the following conclusions are reached:

- PEs for the dynamic models are significantly lower than for the static ones. The improvement is more marked for the single-component regressions than for the mixtures, and tends to diminish with increasing τ;
- Mixtures lead to better prediction than single regressions under both the static and dynamic modelling assumptions, though the improvement is less marked in the dynamic case (being between about 1 to 5 % in terms of PE). In off-line identification of the model, some components had small weights, but their omission caused a clearly observable reduction in prediction quality.

The current study clearly indicates the benefits of dynamic mixture modelling for high-dimensional traffic data analysis. Dynamic mixtures are capable of accurate prediction many steps (e.g. 90 minutes) ahead. Hence, it is reasonable to expect that they can serve for effective advisory design, since components arise, not only because of intrinsic changes in traffic state, but also as a result of traffic sequence choices made by the operator [1].

4. AN ADVISORY SYSTEM FOR RADIOTHERAPY OF THYROID CARCINOMA

A probabilistic advisory system has been installed at the Clinic of Nuclear Medicine in Motol Hospital, Prague, Czech Republic. The purpose of the system is to recommend to the physician a patient-specific administration of radioactive ^{131}I in the treatment of thyroid cancer.

4.1. Radiotherapy of Thyroid Carcinoma

Treatment of thyroid gland carcinoma is a complex process involving endocrinology, histology, surgery, radiology and other branches of medicine. The current work pertains to support of the radiological treatment [8]. The thyroid gland accumulates iodine from the blood and creates thyroid hormones. Injected radioactive ^{131}I emits ionizing β- and γ-particles. The majority of γ-particles (high energy photons) is not absorbed in the tissue and is detectable outside the body. Almost all β-particles (electrons) pass their energy to the tissue by absorption, and in sufficient concentrations will destroy the thyroid tissue.

The usual medical practice is to remove a thyroid tumour by surgery. What remains or reappears is destroyed by radiotherapy, which has two distinct stages, *diagnostic* and *therapeutic*. In the diagnostic stage, a low-activity ^{131}I tracer (about 70 MBq) is administered orally to the patient. The detected γ-radiation emanating from the accumulating tissues is measured, and some quantities describing the kinetics of ^{131}I are evaluated. The information is used by the physician to decide on an appropriate high-activity therapeutic administration of ^{131}I, usually in the range of 3–10 GBq. The β-radiation absorbed by the thyroid tissues (the dose) destroys them. Finally, 3 to 6 months after therapy, the diagnostic stage is repeated in order to determine how much, if any, of the thyroid tissue remains.

4.1.1. The Patient Archive Data from more than 6000 patients treated in the manner above have been archived since 1992 at Motol. Following various pre-processing and scaling steps, and elimination of spurious records [10], about $\mathring{t} = 1100$ patient patients were actually used for off-line mixture identification. The record, d_t, for the t-th patient, contained $\mathring{d} = 12$ bio-physical variables. The variables summarize (i) general patient information (gender, age, number of therapies previously undergone), (ii) biophysical information obtained from the diagnostic stage (administered activity of ^{131}I, number of tumour sites accumulating ^{131}I, effective half-life and residence time [8], etc), (iii) the activity, A_0, of the therapeutic administration, and (iv) the maximum thyroid activity, A_{rm}, expressed as a percentage of A_0, detected during the further diagnostic stage after therapy. A_0—as determined by the physician—is, of course, the key quantity to influence the absorbed dose and, therefore, the treatment success. A_{rm} is the agreed quantifier of success. Ideally, $A_{rm} = 0$ % after a successful therapy. However, in practice, $A_{rm} < 0.18$ % is defined as successful. The aim of the current work is to recommend a value of A_0 to the physician which is designed to meet this criterion of success.

4.2. Off-Line Processing and Static Modelling of the Data Archive

Clusters are evident in the 12-dimensional data space, d^*, caused by such factors as (i) the existence of distinct population sub-groups susceptible to this cancer, and (ii) distinct approaches to therapy guided by different physicians at Motol. The static modelling of Section 3.2.1 is appropriate, since each d_t is an independently treated patient. Smooth non-linear relationships between variables can be accommodated by inferring $\mathring{c}$ sufficiently large, but the total number of parameters in the mixture model, $\mathring{\Theta}$, is monotonically rapidly increasing with $\mathring{c}$. With identification depending on a strictly limited set of data, estimation accuracy can therefore be expected to diminish as a consequence of these non-linearities. To counteract this effect, several of the variables, $d_{i;t}$, are chosen to be non-linear functions of actual patient measurements, the choice being guided by physiologically-based prior knowledge of these non-linearities [10]. The Mixtools [6] suite, appropriately initialized, was used to identify the mixture. The marginal for key variables A_0 and A_{rm} is shown in Fig. 3.

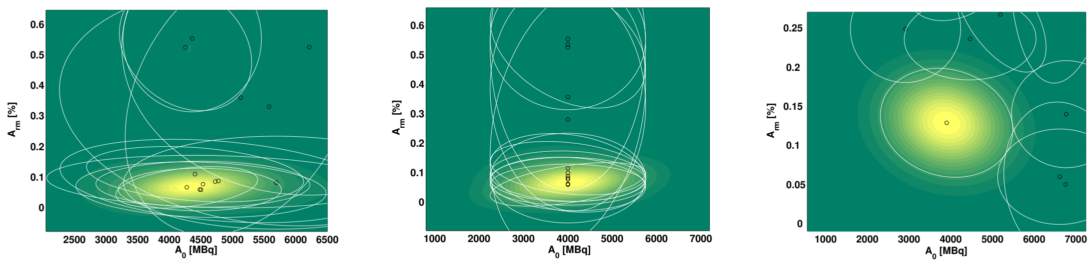


Figure 3. Static Mixtures associated with Advice Generation in Nuclear Medicine Dosimetry. **Left:** marginal mixture, ${}^{[E]}f(A_{0,t}, A_{rm,t}|\Theta)$, estimated from patient archive data; **centre** optimized marginal p.f., ${}^{[I]}f(A_{0,t}, A_{rm,t}|\Theta)$; **right:** patient-specific optimized conditional p.f., ${}^{[I]}f(A_{0,t}, A_{rm,t}|x_t, \Theta)$.

4.2.1. Specification of User-Ideal, ${}^{[U]}f$ The first approach outlined in Section 2.1.1 is adopted:

- A_0 must not result in doses which exceed limits imposed by radiological protection standards. If A_0 is set too low, it is likely that any residual thyroid tissue after therapy will have been damaged by β-radiation, and be incapable of any further uptake of ^{131}I. This limits the treatment options available to the physician [8]. Current medical practice at Motol confines A_0 to the range 3–7.5 GBq. The non-normal distribution of A_0, with mean 4 GBq, was approximated by a normal, yielding the following user-ideal for A_0:

$$ {}^{[U]}f(A_0|\mu_{A_0}, \sigma_{A_0}) = \mathcal{N}(\mu_{A_0}, \sigma^2_{A_0}). $$

  Here, $\mu_{A_0} = 4000$ and $\sigma_{A_0} = 1200$, giving 3 standard deviations to the upper limit.
- As stated in Section 4.1.1, therapy is regarded as successful if $A_{rm} < 0.18$. This is expressed also via a normal user-ideal density, with $\mu_{A_{rm}} = 0.09$, and $\sigma_{A_{rm}} = 0.015$.

Finally:

$$ {}^{[U]}f(A_0, A_{rm}|\mu_{A_0}, \sigma_{A_0}, \mu_{A_{rm}}, \sigma_{A_{rm}}) = \mathcal{N}(\mu_{A_0}, \sigma^2_{A_0})\mathcal{N}(\mu_{A_{rm}}, \sigma^2_{A_{rm}}). \tag{1} $$

This user-ideal models A_{rm} as independent of A_0, a clearly unrealistic assumption. Introducing correlations, as well as modifications of $\mu_{A_{rm}}, \sigma_{A_{rm}}$, did not, however, affect the quality of the resulting advisory system, and so the independence assumption was kept.

4.3. Recommendation of Patient-Specific A_0

Two distinct approaches have been attempted. These are briefly described now.

4.3.1. Deteministic Recommendation of A_0 Consider a patient, labelled by $t \in t^*$ (the set of 'on-line' patients), for whom an advised therapeutic administration, A_0, is to be provided. $d_t \sim {}^{[E]}f(d_t|\Theta)$, whose structure and parameters, $\Theta = [\Theta_1', \ldots, \Theta_{\mathring{c}}', \alpha']'$, have been identified from the patient archive (Section 4.1.1). Let $d_t = [A_{0,t}, A_{rm,t}, x_t']'$, where x_t are the diagnostic and general variables for patient t. Then, it may be shown [10] that

$$\hat{A}_{rm,t} = \mathcal{E}(A_{rm,t}|A_{0,t}, x_t, \Theta) = \sum_{i \in c*} a_{i,t}(A_{0,t}|x_t, \Theta_i, \alpha)\ \ \mathcal{E}_i(A_{rm,t}|A_{0,t}, x_t, \Theta_i) = g(A_{0,t}|x_t, \Theta), \tag{2}$$

where $\mathcal{E}$ denotes expectation with respect to the (conditional) full mixture, ${}^{[E]}f$, and $\mathcal{E}_i$ denotes expectation with respect to the ith component, whose (estimated) parameters are Θ_i. $a_{i,t}$ are weights determined by their arguments in (2). Hence, $\hat{A}_{rm,t}$ is a deterministic function, g, of the administered activity, $A_{0,t}$. The monotonically decreasing nature of g has been confirmed for many patients, as explained in [10]. The deterministic recommendation for the t-th patient, $\hat{A}_{0,t}$, is then given as the minimum $A_{0,t}$ for which the expected maximum thyroid activity in diagnosis after therapy is less than 0.18%; i.e. $\hat{A}_{0,t} = g^{-1}(0.18|x_t, \Theta)$.

4.3.2. Probabilistic Recommendation of A_0 The simultaneous control design procedure (Section 2.1.1 and [5]) is employed. It uses the archive data mixture, ${}^{[E]}f$, and the user-ideal, ${}^{[U]}f$ (1), to produce the joint optimized p.f., ${}^{[I]}f(A_{0,t}, A_{rm,t}, x_t|\Theta)$. Marginalizing with respect to x_t produces the bivariate optimized p.f. in Fig. 3. Simple conditioning ('slicing') of the joint p.f. with respect to the patient data, x_t, for the t-th new patient, provides the patient-specific bivariate optimized p.f., ${}^{[I]}f(A_{0,t}, A_{rm,t}|x_t, \Theta)$. An example is given in Fig. 3. This, or a certainty equivalent (typically the Maximum *a Posteriori* (MAP) estimate from this patient-specific p.f.), are available as advisory support for the physician in determining the administration appropriate for the t-th new patient.

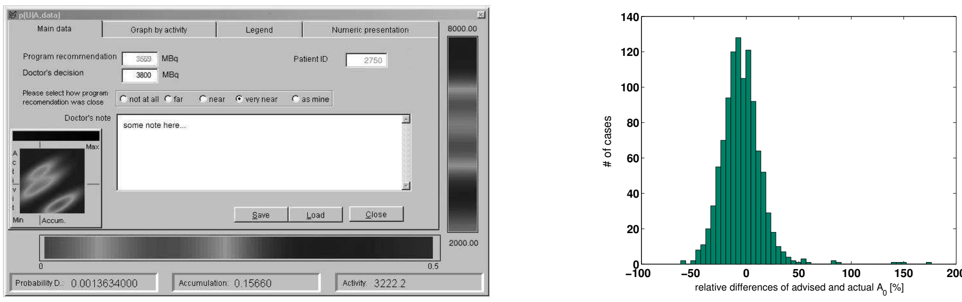


Figure 4. **Left:** screen of *Iodine III* with interactive probability density map; **right:** Relative differences of advised and actual A_0, calculated from archive data.

4.4. Current Clinical Experience

The advisory module was programmed in C++ and integrated into the system *Iodine III*, which handles data management and computations at Motol. Graphical presentation of the generated advice is shown in Fig. 4 (left). The interactive colour probability density map allows doctors to try, for the current patient, various hypothetical administrations and to examine the consequences — in terms of A_{rm} — predicted by the advisory modules.

Correlation between the *Iodine III* recommendations and the physician's actual choice has been investigated for historical cases where the therapeutic administration was recorded. In Fig. 4 (right), the distribution of relative differences is plotted in histogram form. The number of outliers and the negative bias ($-3.8\,\%$ with sample deviation $21.2\,\%$, indicative of a systematic recommendation *below* the physician's choice) are small. Recently, *Iodine III* has been upgraded to record systematically the recommended and actual therapeutic administrations for each new patient case, thereby greatly enhancing the quality of the emerging archive. This will allow not only further exploration of the correlations above, but also quantitative assessment of (i) any shifts in physician practice in consequence of their use of the system, and (ii) the impact of the advisory system, which will come via comparison of the actual realized 'on-line' $A_{rm,t}$ values with those observed before the system was used.

5. ADVISORY SYSTEM DESIGN FOR A METAL ROLLING MILL

This process has already been equipped with a control system enabling high quality production of metal strips when the adjustable settings are well chosen by the operator. The need for a mixture-based advisory system stems from the difficulty in finding optimal settings for all possible working conditions and processed metal type. Success is reflected in a quantified improvement of process performance when the operator is supported by the advisory system.

5.1. The Rolling Mill

A fine reversing cold rolling mill with a 20-high arrangement of rolls (Fig. 5) is considered. Strips of metal of 12 types (alloys of iron, copper, nickel, zinc, etc.) are processed on it, by being passed alternately forwards and backwards through the roller assembly. Contact meters on both sides of the rollers provide measurements of strip thickness with $\pm 1\,\mu$m accuracy. The machine is equipped with a two-level control and information system, including an adaptive thickness controller which uses Automatic Gauge Control (AGC) [13]. With appropriate settings, it is possible to keep deviations of the output strip thickness to the scale of the measurement accuracy. This is difficult to achieve for all possible working conditions encountered.

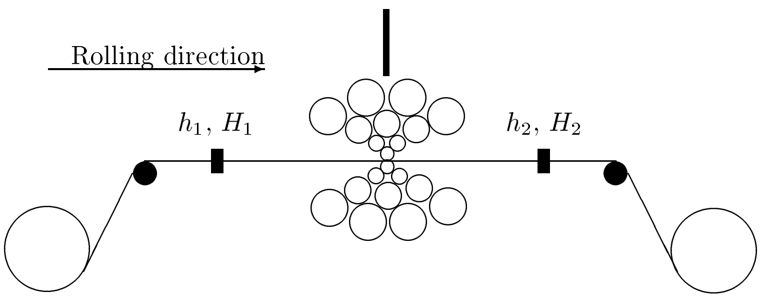


Figure 5. 20-high reversing rolling mill. For the rolling direction indicated, H_2 and h_2 denote output strip thickness and deviation from its desired value, $H_{2\,nom}$, respectively. H_1 and h_1 are input values.

The quality of the output is quantified via the following measures:

$$C_p = \frac{tol_{h_2}^+ + |tol_{h_2}^-|}{6\,\sigma_{H_2}}, \tag{3}$$

$$C_{pk} = \frac{\min(\bar{h}_2 - tol_{h_2}^-,\ tol_{h_2}^+ - \bar{h}_2)}{3\,\sigma_{H_2}}. \tag{4}$$

Here, $tol_{h_2}^+$ and $tol_{h_2}^-$ are boundaries of the tolerance range of h_2 (Fig. 5), σ_{H_2} is the standard deviation of the output thickness, H_2, and $\bar{h}_2$ denotes the mean of h_2. Another quantifier of output quality used in the advisory design is C_{per}, which is the percentage of the output thickness deviation, h_2, within the tolerance range $[tol_{h_2}^-,\ tol_{h_2}^+]$. The aim of the quality control is to maximize the coefficients C_p, C_{pk} and C_{per}.

5.1.1. The POL Data Archive The distributed control system is triggered to record a snapshot of all process variables, d_t, for every 4 cm of strip movement. In the physical implementation, 44 channels are recorded, but this is reduced to $\mathring{d} = 10$ (the most significant channels) in the final archive. Let $M = m \in \{1, \ldots, 12\}$ be a pointer to the metal type, and let $R = r \in \{1, 2, 3\}$ label the *first* pass, any *forward-moving* pass, and any *reverse-moving* pass, respectively [1]. The archive is assembled by forming 36 aggregate records, each obtained by grouping together all data associated with each of the possible 12×3 realizations of (M, R). Each record, initially of typical size 5×10^5, is reduced to 3×10^4 samples, following systematic pre-processing steps which improve their amenability to mixture modelling. Mixture estimation, advisory design and advice generation—as described in the following sections—are then undertaken in an identical manner for *each* of the 36 (M, T) cases above. Hence, all identified p.f.s are conditionals with respect to (M, T) (though this fact is suppressed in the notation), with $\mathring{d} = 10$, and $\mathring{t} = 3 \times 10^4$ approximately.

5.2. Off-Line Processing

For each record, two mixture modelling cases were considered, as described next.

5.2.1. Static Modelling Observations, d_t, are modelled as conditionally independent, i.e. $f(d_t|d(t-1), \Theta_{c_t}, c_t) = f(d_t|\Theta_{c_t}, c_t)$, so that $\phi_{c_t;t-1} = \{\}$, $\forall c_t \in c^*$ (Section 2.1). Typically, 10 iterations per data file were used [6] in the estimation of the associated static mixture, with average processing times of 80 minutes on a 1 GHz machine.

In Fig. 6, a projection of a typical estimated mixture onto two signal channels (*i.e.* the marginal) is illustrated. Many show excellent correspondence to data histograms, but the current challenge is to capture structure in marginals where many small, low-weight components are evident. The static model cannot be used in prediction, which is a key element in effective advisory design for POLs with delayed control responses, such as here. The problem is overcome by capturing the intrinsic temporal correlations in the data via a dynamic model.

5.2.2. Dynamic Modelling A first-order (in time) temporal regression is employed for all components:

$$f(d_t|d(t-1), \Theta_{c_t}, c_t) = f(d_t|d_{t-1}, \Theta_{c_t}, c_t).$$

Equivalently, $\phi_{c_t;t-1} = d_{t-1}$, $\forall c_t \in c^*$. It is appropriate to capture the dynamic relationships on a time-scale appropriate to the human operator, since this is the time-scale for which the

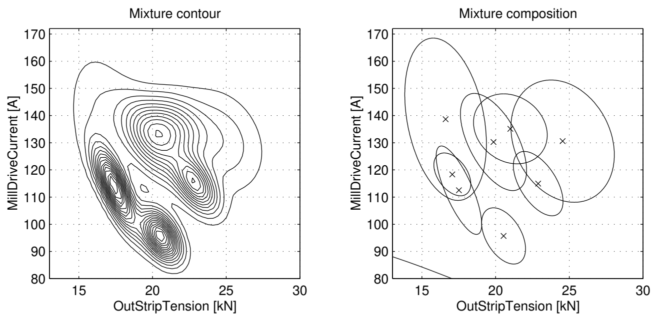


Figure 6. Typical STATIC archive mixture, ${}^{[E]}f$, for the rolling mill. **Left:** 2-D marginal (projection) for two selected channels. **Right:** composition of mixture from its components.

advisory design is required. Hence, the data set is decimated: $q_t = d_{Dt}$, $t = 1, \ldots, \lfloor \mathring{t}/D \rfloor$. With a typical POL control system sampling period of 40 ms (for rolling speed 1 m/s), and an operator perception time of about 250 ms, a typical value for D is 6. Typical processing time for estimation of the dynamic mixture for one decimated record is 7 minutes. The quality of the estimation is assessed via comparison of predictions of the dynamic mixture model with actually realized POL states. The results are encouraging, and indicate that a first order temporal regression on decimated data is sufficient for effective advice generation.

5.2.3. Specification of the User-Ideal Probability Function Formal probabilistic advisory design [5] requires quantification—via the user-ideal p.f, ${}^{[U]}f$—of the management ambitions for high-quality operation of the POL. Both of the policies outlined in Section 2.1.1 have been tested [1]:

1. 1-D normal p.f.s are specified for channels of interest. For example, the p.f. for H_2 is defined with mean $H_{2\ nom}$ (Fig. 5) and standard deviation derived from the tolerances in (3). These p.f.s are multiplied together to yield the necessary multidimensional p.f. on the specified channels.
2. Periods of high quality operation of the controlled system are marked, and the static mixture identified for these data is used directly as the user-ideal.

The second approach, while requiring extra off-line processing effort, immediately overcomes the two principal limitations of the first. Specifically:

- natural correlations between variables are preserved;
- the user-ideal has support in the full data space, d^*. Its marginals in variables not specified by the manager under the first policy above therefore reflect *actual* physical states of the controlled system observed in the past. The failure of the first policy to constrain these variables to physically-meaningful intervals risks an advisory design where these are steered towards unacceptable values.

5.3. On-Line Advice Generation

${}^{[I]}f$, deduced off-line from ${}^{[E]}f$ and ${}^{[U]}f$ [5], is passed to the on-line stage of the algorithm, where tailored advice is made available to the operator in real time. Let the on-line data be denoted by $r_t, t \in t^*$ (Section 2.1.1). Let $r_t = [s_t', z_t']'$ be a partition into a subset, s_t, of channels for which advised values for the operator are to be calculated, and let their complement be z_t. The following marginal predictive p.f. is readily available:

$$^{[I]}f(s_{t+1}|r(t),\Theta) = \int_{z_{t+1}\in z^*} {}^{[I]}f(r_{t+1}|r(t),\Theta)\ dz_{t+1}. \tag{5}$$

These predict optimal ranges for the chosen variables, D steps (about 250 ms, as explained in Section 5.2.2) into the future, constituting on-line advice. The numerically *un*intensive task of evaluating the off-line marginals at the operating point, r_t, is achieved using the MixTools suite [6]. The prioritization of advised variables is based on comparison of r_t with a certainty equivalent (MAP estimate, typically) extracted from $^{[I]}f(r_{t+1}|r(t),\Theta)$. If the separation between these exceeds specified limits, priority-ranked warnings are generated along with recommended adjustments [3]. The estimated mixture, $^{[E]}f$, and therefore $^{[I]}f$, are updated periodically (current practice is to do this at the end of each pass) to benefit from enriched information available from the advisory-system-supported POL.

5.4. Implementation of Advisory System

The advisory system was integrated into the distributed control system of the rolling mill in the form of a dedicated ProDaCTool server and a visualization node equipped with a special Graphical User Interface (GUI). Both MS Windows and Linux implementations were provided. The server executes two main applications: (i) an *Adviser*, which loads $^{[I]}f$ from the off-line

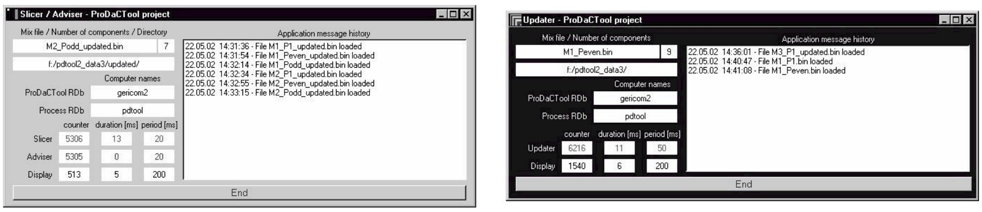


Figure 7. Screenshots of Adviser and Updater applications.

phase, evaluates marginals, and generates recommendations as appropriate (Section 5.3); and (ii) an *Updater*, which re-estimates $^{[E]}f$ and $^{[I]}f$ using on-line data from epochs of good rolling mill performance, potentially uncovering new high-quality modes in the POL data space, q^*. Both applications use a multi-threading technique to optimize system performance. The programs are coded in ANSI C, and the Tcl language/interpreter is used for interfacing to a human supervisor (Fig. 7). This ensures ease of migration to another software platform, when the appropriate version of the Mixtools library [6] is employed.

5.4.1. The Graphical User Interface (GUI) A new visualization node was provided for the control system on which a dedicated GUI for operators was executed [1] (Fig. 8). Its display styles and options can be easily modified via a configuration file. In all cases, the GUI displays the overall status of the system in the form of traffic lights, where *Green* indicates 'no recommendations'. The variable which is recommended to be changed first is highlighted, together with the recommended value. In 2-D mode, the operator can select channels for

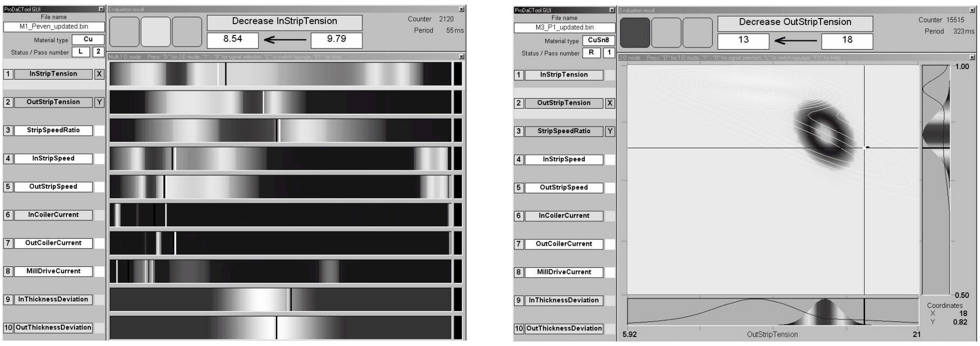


Figure 8. Screenshots of the Graphical User Interface in 1-D and 2-D modes respectively.

Increase of the average rolling speed in % for rolling between coilers	
1st month (average for all materials)	6.7
1st month (best result for a single material)	25.9
2nd month (average for all materials)	24.0
2nd month (best result for a single material)	48.6

Table III. The increased average rolling speed achieved with installation of advisory systems: academic design (1st month), simultaneous design (2nd month). All quality measures were improved.

which the projection, ${}^{[I]}f(s_{t+1}|r(t), \Theta)$, is displayed in real time. `OpenGL` functions are used for smooth on-line plot animations.

5.5. Current Experience of the Rolling Mill Advisory System

Both academic and simultaneous designs [5] of ${}^{[I]}f$ for the advisory system have been tested *in situ* for one month apiece. In all cases, it was possible to *increase* the rolling speed while actually *increasing* all three quality markers outlined in Section 5.1. The speed increases are summarized in Table III. The quoted increases have been calculated with respect to 20-month averages measured before installation of the advisory system. The implementation site is a relatively small and fine mill with low rolling speed. Implementation of the advisory system in a high-speed steel mill will undoubtedly require more computing power, and further consideration of the advisory design strategy. Nevertheless, the performance improvements are impressive, and point to the significant economic benefits that can accrue from installation of dynamic-mixture-based advisory systems in complex industrial plants.

6. CONCLUSIONS

The variety of application contexts studied in this paper underlines the prevalence of high-dimensional multimodal data problems. All of the processes studied have been characterized as POLs: the measured system is influenced by the actions of an operator. The processes themselves have an intrinsic tendency to operate in a countable number of distinct ways (modes), and the operator compounds the multimodality of the measurements by the tendency to operate on the system also in a countable number of ways. This paper presents powerful evidence for the efficacy of mixtures of dynamic regression models in estimating and predicting such POLs, and the quantifiable improvement they offer over the use of single regressions. Long-range temporal predictions using dynamic models (*e.g.* the traffic study in Section 3) *and* channel predictions in static models (the nuclear medicine study in Section 4) have been achieved. A guiding principle is *that effective prediction in controlled systems is the basis for good control and advisory design*.

The full probabilistic control design presented in the accompanying paper [5] has been used for design of an advisory system in two of the three application studies presented here. In nuclear medicine, recommended therapeutic applications of a radiopharmaceutical closely follow those chosen independently by the physician, and work is ongoing to demonstrate that expected loss will be reduced when the physician follows the advice of Iodine III. In the rolling mill application, results following from 'closing the loop' between advisory system and POL have already led to quantifiable improvements in performance, as reported in Section 5.

The ability to perform the mixture estimation and advisory design off-line is a key advantage. On-line advice generation constitutes a small number of inexpensive steps (essentially, evaluation of kernels with respect to current data). Fast, real-time advisory systems for high-dimensional problems (such as the installed rolling-mill advisory system) are therefore feasible.

The optimized p.f., ${}^{[I]}f$, reflects a realistic aim for the POL, since it takes into account actually observed dynamics, via ${}^{[E]}f$, *and* informed management aspirations, ${}^{[U]}f$. For success, the former must capture well the effect of control actions (inputs) on outputs, and the latter must be realistically formulated with respect to what can be achieved. If either does not hold, the advisory design will be undermined. For example, no user-ideal has yet been defined for the traffic-light POL.

With regard to effective identification of ${}^{[E]}f$, the application studies have underlined two key problems commonly arising in practice:

- poor-quality data (such as the electric current and strip tension channels in the rolling mill); the pre-processing suite [12] in MixTools is therefore a vital part of the overall system, in yielding data amenable to mixture modelling;
- inadequate off-line populating of significant modes (*e.g.* rare inter-sequence switching in the traffic application; expensive and incomplete historical data records in the nuclear medicine application); the need for a sufficient and well-recorded archive is vital.

All applications are currently being extended, but the generic nature of the advisory design, underlined by the *range* of successful applications presented in this paper, recommends it for many applications.

REFERENCES

1. Project IST-1999-12058, *Decision Support Tool for Complex Industrial Processes Based on Probabilistic Data Clustering (ProDaCTool).* 1999-2002.
2. J. Aguilarmartin, *Qualitative control, diagnostic and supervision of complex processes.* Mathematics and Computers In Simulation, 36(2):115–127, 1994.

3. M. Kárný, J. Böhm, T. Guy, L. Jirsa, I. Nagy, P. Nedoma, A. Quinn, L. Tesař, M. Maisey, M. Tichý, *ProDaCTool background.* Tech. rep., ÚTIA AV ČR, Praha, 2001. Deliverable of the ProDaCTool project.
4. A. Polymenis, D.M. Titterington, *On the determination of the number of components in a mixture.* Stat. Probabil. Lett., 4:295–298, 1998.
5. M. Kárný, J. Bohm, T.V. Guy, P. Nedoma, *Mixture-based adaptive probabilistic control.* This issue.
6. P. Nedoma, M. Kárný, I. Nagy, *MixTools, MATLAB toolbox for mixtures: user's guide.* Tech. rep., ÚTIA AV ČR, Praha, 2001. Deliverable of the ProDaCTool project.
7. R. D. Rosenkrantz, Editor, *E.T. Jaynes: Papers on Probability, Statistics and Statistical Physics.* D. Reidel, Dordrecht-Holland, 1983.
8. J. Harbert, W. Eckelman, R. Neumann, *Nuclear Medicine. Diagnosis and Therapy.* New York: Thieme Medical Publishers Inc., 1996.
9. G. A. Siegel, S. R. Thomas, J. B. Stubbs, M. G. Stabin, M. T. Hays, *MIRD pamphlet No.16: Techniques for quantitative radiopharmaceutical bio-distribution, data acquisition and analysis.* Journal of Nuclear Medicine, 40(2), 1999.
10. L. Jirsa, A. Quinn, *Mixture analysis of nuclear medicine data: medical decision support.* Proc. Irish Signals and Systems Conference, 393–398, NUI Maynooth, 2001.
11. D. Helbing. *Traffic data and their implications for consistent traffic flow modelling.* Transportation Systems, Chania, 809–814, Greece, 1997.
12. L. Tesař, A. Quinn, *Method for artefact detection and suppression using α-stable distributions.* Proc. Int. Conf. on Artifical Neural Networks and Genetic Algorithms (ICANNGA), Prague, 2001.
13. P. Ettler, F. Jirkovský, *Digital controllers for ŠKODA rolling mills.* In *Advanced Methods in Adaptive Control for Industrial Applications*: Lecture Notes in Control and Information Sciences, No. 158, Springer-Verlag, Berlin/Heidelberg, 31–35, 1991.

