14

Applications of the advisory system

The applications described here confirm usefulness of the developed theory and algorithms. At the same time, the applications proved to be significant for developing both the theory and algorithms. They helped us to discover errors, inconsistencies, drawbacks and bugs. They stimulated solutions of many subproblems that arose in attempts to implement the theoretical results. The process is still unfinished but the discussed interaction can be unanimously claimed to be useful.

Three rather different applications are outlined here. Advising at cold rolling mill is described in Section 14.1. Nuclear medicine application related to treatment of thyroid gland cancer is reflected in Section 14.2. Prediction problems related to urban traffic control are in Section 14.3. Practical conclusions are made in Section 14.4; see also [177].

The presented experience with *applications* we are working with reflects the research stage that corresponds with the last arrow in the verification chain $\boxed{\text{theory} \rightarrow \text{software} \rightarrow \text{offline experiments} \rightarrow \text{implementation.}}$

14.1 Operation of a rolling mill

Support of rolling mill operators in their decisions how to adjust key operating parameters of the machine was the main application output of the project that stimulated the results presented in this text. In this application, the fixed advisory system with periodically refreshed advisory mixture is used.

14.1.1 Problem description

Nowadays rolling mills are usually equipped with a control system enabling high quality production for properly adjusted manual settings. It is, however, difficult to find optimal settings for all possible working conditions and every type of the material being processed.

The rolling mill

A cold rolling mill is a complex machine used for reduction of metal strip thickness. The thickness is reduced between working rolls of the mill. It is affected by the rolling force in conjunction with input and output strip tensions being applied. For reversing mills, the strip is moved forward and backward in several passes until the desired thickness is achieved.

Two types of cold rolling mills were selected for full-scale experiments. They differ in the arrangement of rolls. The 4-high rolls arrangement consists of two working rolls each being supported by a single back-up roll. Data from two mills of this type were used for offline experiments.

A fine reversing cold rolling mill with 20-high arrangement of rolls, Fig. 14.1, was employed for both offline test and final online advisory system implementation. About 12 material types — alloys of iron, copper, nickel, zinc, etc. — are processed on the machine. Contact meters on both sides of the rolling mill provide measurements of strip thickness with $\pm 1\,\mu$m accuracy. For illustration, rolling forces are of order of 10^6 N, electric currents of drives about 10^2 A, strip tensions about 10^4 N and rolling speed in orders of 0.1–$1\,\mathrm{ms}^{-1}$.

The machine has been already equipped with the two-level control and information system including the adaptive thickness controller [178]. For properly adjusted settings, the controller is capable to keep deviations of the output strip thickness on the level of the measurement accuracy.

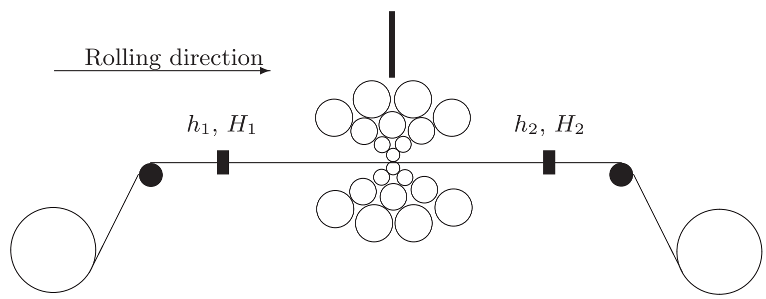


Fig. 14.1. Schematic diagram of the 20-high reversing rolling mill. For the selected rolling direction, H_1 and h_1 denote the input strip thickness and deviation from its nominal value, respectively; similarly, H_2 and h_2 for the output thickness.

Data collection

The modern distributed control system of rolling mills enables to archive every sample of process data. The collection is triggered by the strip movement.

Several tens of signals are collected on each particular mill. Thus, for each pass, $(\mathring{d}, \mathring{t})$ data matrix is produced where $\mathring{d}$ is the number of channels specific for a particular mill and the number of samples $\mathring{t}$ varies from 3 000 to 30 000 depending on a particular pass. Data samples are recorded each 4 cm of the strip, i.e., sampling period varies according to the strip speed. For instance, for the speed 1 m/s, the sampling period is 40 ms. The rolling mill experts selected $\mathring{d} = 10$ most important data channels as adequate for advising purposes. The specific selection depends on a particular rolling mill as described below.

14.1.2 Problem and its solution

The advisory system is to help adjust the main process quantities in order to maintain the best possible product quality. Moreover, for the 20-high mill, the goal is a better utilization of the rolling speed range in order to increase production potential without a loss of the high product quality. The practical implementation aspects described here relate to this problem.

Offline phase

The offline phase of the solution consists of data preprocessing, Section 6.2, estimation of the prior mixture information, Section 6.4, structure and parameter estimation, Sections 6.6 and 6.5, and the design of the advisory system; see Chapters 7 and 9.

Data

Data from two 4-high rolling mills were utilized for offline experiments. 58 quantities are recorded. The number of considered quantities was reduced to 10.

Most important selected data channels for mill 1 (rolling copper and its alloys) and mill 2 (rolling iron and its alloys) differ. The difference is caused by differences of control systems at respective mills. In contrast with the 20-high rolling mill, they contain explicitly rolling forces and hydraulic pressures on rolling cylinders.

Additional offline tests and a final online implementation concern the 20-high rolling mill. The available 44 data channels were also reduced to 10 relevant ones shown in Table 14.1.

For reversing rolling mill input/output pairs of signals refer to right/left sides of the mill, or vice versa according to the rolling direction. The front/rear pairs of signals refer to operator/drive sides of a rolling mill.

Particular sets of important data channels for the given rolling mills were selected by combining experience, correlation analysis, availability and reliability of particular data channels. The quantities measured with insufficient precision, heavily correlated or irrelevant for the addressed problem were omitted. Formally, they were treated as surplus data d_{o+} of the o-system.

Table 14.1. Offline and online processed quantities for the 20-high rolling mill

#	Symbol	Description	Typical range	Unit	Type	Filter
1	T_1	Input strip tension	$\langle 0; 50\rangle$	kN	u_o	Outliers 3
2	T_2	Output strip tension	$\langle 0; 50\rangle$	kN	u_o	Outliers 3
3	v_R	Ratio of input and output strip speeds	$\langle 0.5; 0.99\rangle$		u_o	Outliers 5
4	v_1	Input strip speed	1^{st} pass: $\langle 0.1; 0.3\rangle$	m/s	u_o	Outliers 5
			another: $\langle 0.5; 0.7\rangle$	m/s		
5	v_2	output strip speed	1^{st} pass: $\langle 0.1; 0.3\rangle$	m/s	u_o	Outliers 5
			another: $\langle 0.5; 0.7\rangle$	m/s		
6	I_1	Electric current of the input coiler	$\langle 50; 200\rangle$	A	Δ_{p+}	Smooth
7	I_2	Electric current of the output coiler	$\langle 50; 200\rangle$	A	Δ_{p+}	Smooth
8	I	Electric current of the mill main drive	$\langle 20; 180\rangle$	A	Δ_{p+}	Smooth
9	h_1	Deviation of input thickness from nominal value	$\langle -50; 50\rangle$	μm	Δ_{p+}	—
10	h_2	Deviation of output thickness from nominal value	$\langle -10; 10\rangle$	μm	Δ_o	Limits $\langle -10; 10\rangle$

Quality markers

The concept of quality markers is introduced in Section 5.1.4. The deviation of the output strip thickness from its nominal value is the key quality marker, which must be kept as low as possible, practically in units of μm. To evaluate the output quality, three statistical performance indicators (capability coefficients), usual in statistical process control [179], are taken as quality markers:

1. Statistical coefficient C_p defined

$$C_p = \frac{tol^+_{h_2} + |tol^-_{h_2}|}{6\,\sigma_{H_2}} , \tag{14.1}$$

   where H_2 denotes output thickness, h_2 is its deviation from the nominal value H_{2nom}, $tol^+_{h_2}$, $tol^-_{h_2}$ are boundaries of tolerance range of h_2 and $\bar{H}_2$, σ_{H_2} are mean and standard deviation of the output thickness H_2, respectively. The coefficient C_p describes variability of output thickness H_2 despite its magnitude, e.g., bias.
2. Statistical coefficient C_{pk} defined

$$C_{pk} = \frac{\min(\bar{h}_2 - tol^-_{h_2},\ tol^+_{h_2} - \bar{h}_2)}{3\,\sigma_{H_2}} , \tag{14.2}$$

where $\bar{h}_2$ denotes the mean of h_2. The coefficient C_{pk} describes the relative difference of the output thickness deviation h_2 from the mean of the tolerance range related to the variability of H_2.

3. The coefficient C_{per} representing the percentage of h_2 being within the tolerance range $\langle tol_{h_2}^{-},\ tol_{h_2}^{+} \rangle$.

The aim of the quality control is to keep values of the coefficients C_p, C_{pk} and C_{per} as high as possible.

These markers were used according to the customer's wish to compare the product quality before and after implementation of the advisory system.

Preprocessing of data files

Data files corresponding to particular passes were grouped according to the material type. Within each group, three subgroups were created for the first, and next even and odd passes through the mill. The corresponding data files within a subgroup were merged into a file having 5×10^5 samples on average. The subsequent mixture estimation provided acceptable results but it was time consuming. Therefore only representative shorter parts from particular data files were merged, preserving mutual ratios of data lengths.

Consequently, a typical data file within each subgroup contains roughly 30,000 samples of 10 selected data channels. As a result, n files have been available for each rolling mill, where $n = 3 \times n_{\mathrm{mat}}$ and n_{mat} is the number of material types processed on the given rolling mill. The selection and merging procedures were automated to allow repetition for new sets of process data.

The selected quantities and the way of their filtering for the 20-high rolling mill are in Table 14.1. There, the number given for outliers removal means a multiple of standard deviation as a half-width of an acceptance interval around the estimated mean; see 6.2.2 and 6.2.3.

Before further analysis, data in each channel were scaled to zero mean and unit variance.

Static mixture estimation

The static mixtures can be considered as a smooth approximation of multidimensional histograms representing the number of occurrences of data points in the 10-dimensional (10-D) space. As an input, the data gained during the operation of an experienced operator, yielding a high-quality product, are used. The adjustment of the actual working point into a location where data points occurred most often during the good rolling should lead to a high-quality product.

Preprocessed data files were used for an iterative estimation in the offline mode using the Mixtools function mixinit [180] that provides initialization by hierarchical factor splitting and implements the normal version of Algorithm 6.8. Typically, ten iterations per data file were used. The runtime of estimation for a single file with 30,000 records on a 1 GHz machine was about

10 minutes for default options, 160 minutes when batch quasi-Bayes estimation was applied. On average, nine components were estimated.

Comparison of the estimated mixture with the empirical pdf constructed from multidimensional histograms served as the main indicator of the estimation success. The estimated mixture seems to be quite fair for some selected 2-D projections, while for another projection with crumbled data clusters it is problematic. Improvements have been achieved by utilizing nondefault options for the identification procedure. Essentially, the changes have to reflect high differences of noise-level in processed data-record entries. At the end, an acceptable compromise was achieved considering different importance of data channels from the application point of view. A two-dimensional projection of a typical static mixture is shown in the Fig. 14.2. For the 20-high rolling mill, estimated static mixtures are used as inputs for online operation according to the type of material and pass number.

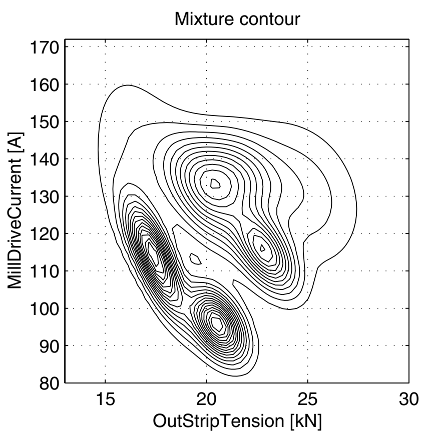


Fig. 14.2. Left plot: 2-D marginal projection of a static mixture for two selected data channels. Right plot: composition of the mixture from particular components.

Dynamic mixture estimation

Despite its clear interpretation, the static approach does not respect dependence of a current state of the system on its recent history, i.e., does not model the system evolution. The first-order auto-regression was found sufficient to describe estimated dynamic components. The order of the model was chosen with respect to sampling rate, human reaction times and closed-loop behavior. The estimates were periodically corrected (after each pass) and used in the fixed advisory system. The fully adaptive advisory system was not used yet.

Zero- and second-order models were also tried The quality of these alternatives was compared according to values of logarithmic v-likelihood and by inspecting prediction errors. Based on these tests, the first-order model was chosen as optimal. The same comparison was used for a finer tuning of optional parameters of the estimation procedure.

Typical processing time of the dynamic estimation on the same machine as above is approximately 5 minutes for default options and 120 minutes for the batch quasi-Bayes option with three detected components in average.

As an alternative to a dual static/dynamic estimation, a pseudo-dynamic approach has been investigated; cf. Section 6.4.9. The static mixture describing data vector, containing also one delayed data record was estimated. The processing combines intuitively appealing histogram interpretation while respecting the dynamic nature of dependencies among data records. Dynamic models are obtained by appropriate conditioning; see Proposition 7.2.

Simultaneous design

The *simultaneous design*, Section 5.4.6, was applied for advising at the 20-high rolling mill.

The user's ideal pdf ${}^{\lfloor U}f$ for the 20-high rolling mill was constructed as a static normal pdf that respects the treated regulated problem and aims of the control.

Specifically, the following correspondence to general notions applies

$$\begin{aligned} u_o &= (T_1, T_2, v_R, v_1, v_2) \ \text{ recognizable actions} \\ \Delta_o &= (h_2) \ \text{ o-innovations} \\ \Delta_{p+} &= (I_1, I_2, I, h_1) \ \text{ surplus p-data.} \end{aligned}$$

The user's ideal pdf is expressed as (see Section 5.4.6)

$$\begin{aligned} {}^{\lfloor U}f(d(\mathring{t})) &= \prod_{t=1}^{\mathring{t}} {}^{\lfloor U}f(d_t|d(t-1)) \\ &= \prod_{t=1}^{\mathring{t}} {}^{\lfloor U}f(\Delta_{o;t})\, {}^{\lfloor U}f(u_{o;t})\, {}^{\lfloor U}f(c_t)\, {}^{\lfloor I}f(\Delta_{p+;t}|d(t-1)). \quad (14.3) \end{aligned}$$

The user's ideal pf ${}^{\lfloor U}f(c_t)$ served only for exclusion of dangerous components. It was set as uniform one on nondangerous components as the user cannot make any statement about them.

The term ${}^{\lfloor I}f(\Delta_{p+;t}|d(t-1))$ concerns quantities that cannot be influenced by the operator and therefore they are "left to their fate", i.e., no requirements are put on them.

The term ${}^{\lfloor U}f(\Delta_{o;t})$ is the key factor in the true user's ideal. It expresses the main management aim. As the innovation h_2 is continuous, ${}^{\lfloor U}f(\Delta_{o;t})$ has a form of normal pdf $\mathcal{N}_{h_2}(0, \sigma_{h_2})$. The zero expectation expresses the wish to get zero deviations of the output thickness from the technologically prescribed value. The value of σ_{h_2} expresses acceptable spread of these deviations. It is related to the desired coefficient C_p (14.1). For instance, if C_p is required to be 4/3, then $\sigma_{h_2} \equiv \sigma_{H_2}$ must be 8-times contained in the interval $\langle tol_{h_2}^-, tol_{h_2}^+ \rangle$. Then $\sigma_{h_2} = (tol_{h_2}^+ + |tol_{h_2}^-|)/8$.

The pdf ${}^{\lfloor U}f(u_{o;t})$ concerning recognizable actions was chosen as a product of one-dimensional normal pdfs. Their means and variances express the technologically desirable ranges of these quantities; see Table 14.1. Their choice was made similarly as for the output thickness.

The parameters of time-invariant pdfs ${}^{\lfloor U}f(c_t)$, ${}^{\lfloor U}f(\Delta_{o;t})$ and ${}^{\lfloor U}f(u_{o;t})$ together with a list of corresponding data channels are passed to the design procedures that generate the optimized ideal mixture ${}^{\lfloor I}f(d_t|d(t-1))$.

The advisory mixture is obtained by projection of ${}^{\lfloor I}f$ to the space (u_o, Δ_o), i.e., the advisory mixture is

$$ {}^{\lfloor I}f(u_{o;t}, \Delta_{o;t}|\Delta_{p+;t}, d(t-1)). \tag{14.4} $$

The overall procedure of simultaneous design is described in Algorithm 7.9.

Online phase

The online phase was fully implemented for the 20-high rolling mill.

Generating advices

Based on *offline* results of mixture estimations, the advisory part of the system uses the newest available data for generating recommendations to operators. The shapes and weights of advisory mixture components are recalculated for each new data record and assign the probability of the best possible location of the working point. Marginal pdfs of the advisory mixture (14.4) are evaluated for all o-data channels $d_{i;t}$, $i = 1, \ldots, \mathring{d}_o$. The recommended values correspond with the modes of these one-dimensional projections.

The limited extent of the o-data allowed us to avoid a presentation problem.

A heuristic solution of signaling problem is temporarily implemented. A simple distance between the actual working point and its best recommended location is evaluated permanently and set up as an “alarm” when it is found that the setting should be changed.

Mixture updating and adaptivity

The adaptive advisory system updates online the mixture estimate, repeats the design with each new data record and uses the updated advisory mixture. Historical (slower machines) and practical (separation of advising and updating) reasons make us use a fixed advisory system. To allow learning, the estimated mixtures are updated and redesigned online by a parallel process during each pass, for each material and each pass subgroup (first, odd and even; see paragraph concerning data preprocessing). For the new pass of the given subgroup, the updated advisory mixtures are used.

14.1.3 Implementation

The advisory system has been implemented into the distributed control system by Compureg Plzeň, s.r.o. The implementation was realized in the form of a dedicated server and a visualization node equipped with a special graphical user interface (*GUI*). A specific effort was made to enable the bulk of computation to be executed either under MS Windows or Linux operating systems.

The server

The server executes the following two main applications.

Adviser loads a proper mixture file at the beginning of the pass and then, when appropriate, uses new data record for evaluating the advisory mixture and its marginal projections to be displayed for operators;

Updater uses data samples corresponding to a good rolling for updating the mixture. Thus, the system can refine the estimates and learn new possible locations of "good" working points.

Both applications use a multithreading technique that allows us to optimize the system performance. Programs are coded in ANSI C, and the Tcl language/interpreter is used for interface to a human supervisor. This together with the appropriate version of the Mixtools library [180] makes porting to another computer platform easy.

Graphical user interface

A visualization node was introduced for the control system on which the dedicated GUI for operators is executed. The application can be used in a one-dimensional (1-D; Fig. 14.3) and two-dimensional (2-D; Fig. 14.4) modes, which project selected quantities of (14.4), and compares them with their actual values. Values of the projected pdfs are transformed into a color scale and actual working points are marked by a black line. Its basic properties and appearance can be easily modified through a configuration file. Whatever mode is selected, the GUI displays the overall status of the system in the form of traffic lights where "green" status means "no recommendations". The quantity to be changed primarily is emphasized, together with its recommended value. For 2-D mode, the operator can select data channels whose marginal mixture projection should be plotted. OpenGL functions are used for smooth online animations.

14.1.4 Results

Offline processing of data from two 4-high rolling mills was made mainly to confirm functionality of developed algorithms and to verify generality of the

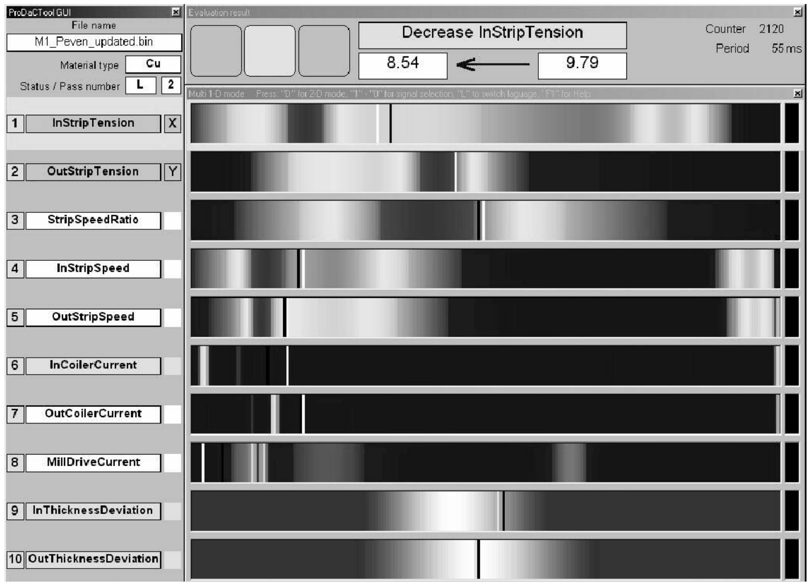


Fig. 14.3. Screenshot of the Graphical User Interface for 1-D mode.

approach. Qualitatively, the results were satisfactory. As the advisory loop was not closed, no overall quantitative evaluation was made.

The case of the 20-high rolling mill was brought to online use enabling us to compare statistically the production before and after the implementation.

The tolerance range was set to $tol^{+}_{h_2} = 10\ \mu$m, $tol^{-}_{h_2} = -10\ \mu$m for all cases. Data collected within 20 months prior to implementation were taken as the basis for comparisons. More than two months of the new system operation were taken into account for the final evaluation. Comparisons were made for several sorts of materials, which were processed in both the periods.

Numbers for comparisons were obtained by SQL queries to production databases. The following values were evaluated:

- Averages of the rolling speed,
- Averages of coefficients C_p, C_{pk} and C_{per}.

Quality markers for results evaluation are represented by relative differences of these values (in percent) between the production before and after installation of the advisory system. For example, $\Delta\bar{C}_p$ is defined as

$$\Delta\bar{C}_p = \frac{(\bar{C}_p)_{\text{after}} - (\bar{C}_p)_{\text{before}}}{(\bar{C}_p)_{\text{before}}} \cdot 100\%.$$

The values $\Delta\bar{C}_{pk}$, $\Delta\bar{C}_{per}$ and $\Delta\bar{v}_2$ are defined similarly.

Six common material types, marked by capital letters, were tested in both periods. Percentage improvements of monitored values are summarized in Table 14.2. First pass and further passes are distinguished in comparisons since

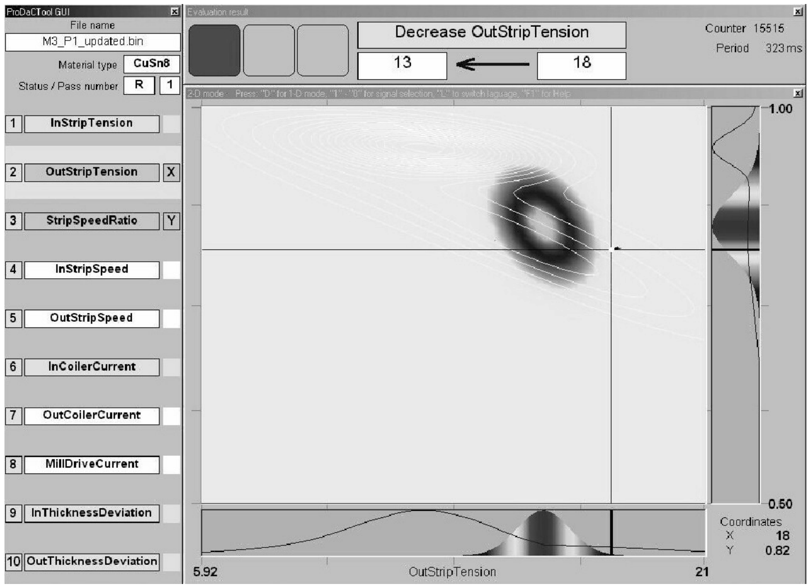


Fig. 14.4. Screenshot of the Graphical User Interface for 2-D mode.

conditions for both cases differ. The C_{per} coefficient was evaluated for last passes only.

Table 14.2. Percentage improvements of quality markers for the 20-high mill.

Material / Quality Marker		A	B	C	D	E	F	Weighted mean
Pass 1	$\Delta\bar{C}_p$	0.98	0.50	-0.69	-0.14	0.66	0.72	0.45
	$\Delta\bar{C}_{pk}$	43.36	15.45	28.57	16.13	39.39	0.00	4.84
	$\Delta\bar{v}_2$	13.33	29.41	58.33	17.65	47.37	10.53	17.08
Pass 2+	$\Delta\bar{C}_p$	31.87	24.68	83.33	-2.94	89.29	-1.08	9.51
	$\Delta\bar{C}_{pk}$	25.30	6.25	105.41	5.00	85.45	1.43	12.36
	$\Delta\bar{v}_2$	16.67	42.86	62.96	16.67	26.83	29.41	33.41
Last pass	$\Delta\bar{C}_{per}$	42.76	16.22	33.33	25.81	45.95	0.68	5.50

14.1.5 Conclusions

The system turned out to stabilize the highest product quality and delimit secure machine settings for the whole variety of working conditions. Obtained results, Table 14.2, validate prior expectations that the rolling speed can be increased by up to 40% (depending on the pass number and material type) while preserving the highest product quality.

The results are encouraging in spite of the fact that the improvements are limited by the amount of information contained in the available data — relatively small and fine mill and low rolling speed. At the same time, we are aware that utilization for high-speed steel mills will undoubtedly need more computing power and probably a further development as well.

14.2 Treatment of thyroid gland cancer

Treatment of thyroid gland tumor using ^{131}I is another important application area tried.

The task is to help physicians to decide on administration of an individual amount of radioactive ^{131}I for a specific patient using information hidden in retrospective data. The decisions in this application were supported by a fixed advisory system. Its advisory mixture can be modified periodically after collecting enough new patient data records.

14.2.1 Problem description

Treatment of thyroid gland carcinoma is a complex process involving endocrinology, histology, surgery, radiology and other branches of medicine. The current section pertains to support of the radiological treatment [181].

Thyroid gland and iodine

The thyroid gland accumulates anorganic iodine from blood and incorporates it into thyroid hormones. A stable isotope of iodine is $^{127}_{53}$I, simply denoted as ^{127}I. Radioactive isotope is chemically identical to the stable one but it has a different number of neutrons in the nucleus. Such a nucleus is unstable, i.e., it can spontaneously change and emit one or more ionizing particles. The nucleus of ^{131}I produces ionizing particles of β- and γ-radiation due to nuclear decays. Thyroid accumulating ^{131}I becomes a source of ionizing radiation. The thyroid tissue is almost unaffected by γ-particles (high-energy photons). The majority of them are not absorbed in the tissue and they can be detected outside the body. Almost all β-particles (electrons) are absorbed in the tissue and their energy is passed to the tissue. The half-life ${}^{\lfloor \mathrm{p}}T$ of ^{131}I — the time in which one half of the radioactive nuclei takes a change — is approximately 8 days.

Radiotherapy of thyroid cancer

A thyroid tumor is usually removed by surgery. What remains or reappears is destroyed by radiotherapy. The radiotherapy has a diagnostic and a therapeutic stage.

In the diagnostic stage, ^{131}I in the form of sodium or potassium iodide of a low diagnostic (tracer) activity is administered orally to a patient. This activity is about $70\,\mathrm{MBq} = 70{\times}10^6$ changes per second in mean. Then, several measurements are performed, detecting γ-radiation from the accumulating tissues, and some quantities describing the kinetics of ^{131}I are evaluated. This information is used for the decision about the next therapeutic stage.

In the therapeutic stage, the patient is administered the solution with ^{131}I of a high activity in the range 3–10 GBq. The therapeutic dose, representing the absorbed energy of β-radiation caused by ^{131}I accumulated in the thyroid, destroys the thyroid tissue.

About 3–6 months later, the diagnostic stage is repeated to examine the therapy result.

Notations, terms and quantities

The notation adopted in this section follows the conventions introduced in this book. Hence, it differs from a notation usual in other thyroid-related papers.

Continuous time will be denoted as ρ, discrete time index as t. Discrete (indexed) time is denoted as ρ_t. A value of a quantity X in a time instant t is denoted as $X_t \equiv X_{\rho_t}$. A quantity X as a function of time ρ is denoted as X_ρ.

After the administration, ^{131}I is distributed over the body and accumulated in thyroid gland and, possibly, other tissues, called *lesions*, and partially over a patient's whole body. Lesions are usually metastases of the primary tumor. Here, we include thyroid gland among them.

The thyroid activity rapidly increases, reaches its maximum within hours and then slowly decreases within days or weeks. The activity of lesions and the whole body can be evaluated and sequences $\{A_t, \rho_t\}$ of activity A_t sampled in time ρ_t. However, activity is not directly measurable and just a few, rather noisy, indirect measurements are available. A nontrivial, theoretical and Bayesian algorithmic solution of its estimation is summarized in [182].

Administered activity A_0 is activity of ^{131}I salt solution drunk by the patient. The diagnostic administration is denoted as ${}^{\lfloor \mathrm{d}}A_0$ and the therapeutic one as ${}^{\lfloor \mathrm{t}}A_0$.

Dosimetric data are sequences of activity in time $\{(A_t, \rho_t)\}$ concerning the given lesion or the whole body.

The *effective half-life* ${}^{\lfloor \mathrm{ef}}T$ is a parameter of a mono-exponential model describing decrease of activity of a lesion in time. The mono-exponential model uses the maximum A_1 of the activity course in the time ρ_1 and predicts the activity A_ρ in time ρ as follows

$$A_\rho = A_1 \ \exp\left(-\frac{\rho - \rho_1}{{}^{\lfloor \mathrm{ef}}T}\ \ln 2\right). \tag{14.5}$$

The model describes the activity course only for $\rho > \rho_1$. Effective half-life ${}^{\lfloor \mathrm{ef}}T$ combines physical and biological mechanisms of activity decrease and quantifies its rate.

Residence time τ is a time for which all the administered activity hypothetically “resided” in a selected tissue to cause the same irradiation effect, i.e.,

$$\tau = \frac{1}{A_0} \int_0^{+\infty} A_\rho \, \mathrm{d}\rho. \tag{14.6}$$

The residence time is proportional to radiation dose — energy per mass unit — absorbed in the tissue. Values concerning the diagnostic or therapeutic administration are denoted ${}^{\lfloor \mathrm{d}}\tau$ or ${}^{\lfloor \mathrm{t}}\tau$, respectively.

Relative activity (uptake) ${}^{\lfloor \mathrm{r}}A$ is a percentage of the absolute lesion activity A from the administered activity corrected to the physical decay, i.e.,

$${}^{\lfloor \mathrm{r}}A = \frac{A}{A_0 \; \exp\left(-\frac{\rho}{{}^{\lfloor \mathrm{p}}T} \ln 2\right)} \; 100\,\%. \tag{14.7}$$

Time ρ is set to zero at the administration moment. Uptake informs about the ability of the lesion to accumulate ^{131}I. Usual values of the maximum thyroid uptake are up to 5 %.

Excretion of the activity ${}^{\lfloor \mathrm{r}}E$ is a relative activity diluted from the body within some time interval. Typically, the intervals used for its description are 0–24 hours, ${}^{\lfloor \mathrm{r}}E_2$, and 24–48 hours, ${}^{\lfloor \mathrm{r}}E_3$, after the administration. These values are estimated from measurements of the whole-body activity and carry additional information about ^{131}I kinetics in the organism. They are evaluated after the diagnostic administration only.

14.2.2 Problem and its solution

The aim is to recommend such a value of administered activity ${}^{\lfloor \mathrm{t}}A_0$ for therapy so that 3–6 months later the maximum uptake ${}^{\lfloor \mathrm{rmn}}A$ is less than 0.18 %. The value ${}^{\lfloor \mathrm{t}}A_0$ should be chosen as low as possible.

The solution consists of offline estimation of a static mixture on the historical patient data and online generation of a recommendation on ${}^{\lfloor \mathrm{t}}A_0$.

Therapeutic activity ${}^{\lfloor \mathrm{t}}A_0$ is a key quantity influencing the treatment success, because the dose absorbed in the thyroid tissue is proportional to it. A decision on the optimal value of ${}^{\lfloor \mathrm{t}}A_0$ is the responsibility of the physician. The value of ${}^{\lfloor \mathrm{t}}A_0$ must be high enough so that the affected tissue is destroyed. This is quantified by the negligible accumulation ability after the therapy. However, the administered activity must be low enough to meet prescribed irradiation limits and to minimize secondary radiation risks.

Offline phase

Data

The data used for the mixture analysis were collected from 1992 on more than 6500 patients. After reorganizing and preprocessing, the final data file contains about 1200 complete records of 12 biophysical quantities directly measured or estimated using Bayesian methodology [183, 184].

The quantities chosen for each record are described in the Table 14.3.

Table 14.3. Quantities used in the advisory system for nuclear medicine

#	Symbol	Description	Typical range	Unit	Type
1	g	Patient's gender: 0=female, 1=male	$\{0,1\}$		Δ_{p+}
2	a	Patient's age	$\langle 7;85\rangle$	year	Δ_{p+}
3	${}^{\lfloor \mathrm{d}}A_0$	Diagnostic administered activity before therapy	$\langle 10;130\rangle$	MBq	Δ_{p+}
4	${}^{\lfloor \mathrm{l}}n$	Number of lesions	$\langle 1;5\rangle$		Δ_{p+}
5	${}^{\lfloor \mathrm{efd}}T$	Diagnostic thyroidal effective half-life	$\langle 0.1;8\rangle$	day	Δ_{p+}
6	${}^{\lfloor \mathrm{d}}\tau$	Diagnostic thyroidal residence time	$\langle 0.000\,4;1\rangle$	day	Δ_{p+}
7	${}^{\lfloor \mathrm{rmd}}A$	Maximum diagnostic thyroid uptake before therapy	$\langle 0.01;25\rangle$	%	Δ_{p+}
8	${}^{\lfloor \mathrm{r}}E_2$	Excretions 0–24 hours	$\langle 15;88\rangle$	%	Δ_{p+}
9	${}^{\lfloor \mathrm{r}}E_3$	Excretions 24–48 hours	$\langle 0.8;34\rangle$	%	Δ_{p+}
10	${}^{\lfloor \mathrm{t}}n$	Number of previous therapies	$\langle 0;5\rangle$		Δ_{p+}
11	${}^{\lfloor \mathrm{t}}A_0$	Therapeutic administered activity	$\langle 1\,700;8\,300\rangle$	MBq	u_o
12	${}^{\lfloor \mathrm{rmn}}A$	Maximum diagnostic thyroid uptake after therapy	$\langle 0.01;1.22\rangle$	%	Δ_o

One data record contains information about one patient concerning his single therapeutic and two diagnostic administrations: 2–3 days before and 3–6 months after the therapy. Data records of different patients are undoubtedly mutually independent. One patient can have more records if more therapies have been absolved. Although an individual patient's history is causal, time interval between two therapies is in the order of months to years. This allows us to assume weak mutual dependence of records even for one patient. Therefore a static model of the whole patient population provides "natural" data description. With the "delayed" diagnostic uptake after therapy in the record, the model is pseudo-dynamic.

Quality marker

The less thyroid tissue remained after therapy, the less activity it accumulates. The therapy is taken as successful if the accumulation activity is negligible.

The marker of the therapy success is a maximum relative activity ${}^{\lfloor \mathrm{rmn}}A$ (uptake) reached by the thyroid gland in the diagnostic administration of ^{131}I after the therapy. Ideally, after a successful therapy, ${}^{\lfloor \mathrm{rmn}}A = 0$. Practically, ${}^{\lfloor \mathrm{rmn}}A < 0.18\,\%$ of the administered activity represents a successful therapy.

Preprocessing of the data files

Records with missing data were excluded. For the complete preserved records, the data were tested for physically and medically meaningful ranges to avoid cases with mistyped values or outliers of another origin. Due to this operation, other records of the original number were lost.

Because of numerical reasons, each quantity was scaled to zero mean and unit variance. All the estimations and computations were performed with the scaled data. The results are presented in their original scales.

Static mixture estimation

The addressed problem and its technical conditions lead to the following correspondence of data entries to the general notions presented in Chapter 5

$$\begin{aligned} u_o &= {}^{\lfloor \mathrm{t}}A_0 \ \text{ recognizable action} \\ \Delta_o &= {}^{\lfloor \mathrm{rmn}}A \text{ o-innovation} \\ \Delta_{p+} &= (g, a,\ {}^{\lfloor \mathrm{d}}A_0,\ {}^{\lfloor \mathrm{l}}n,\ {}^{\lfloor \mathrm{efd}}T,\ {}^{\lfloor \mathrm{d}}\tau,\ {}^{\lfloor \mathrm{rmd}}A,\ {}^{\lfloor \mathrm{r}}E_2,\ {}^{\lfloor \mathrm{r}}E_3,\ {}^{\lfloor \mathrm{t}}n) \ \text{ surplus p-data.} \end{aligned} \tag{14.8}$$

The meaning of respective quantities is summarized in Table 14.3.

The 12-dimensional data space $({}^{\lfloor \mathrm{t}}A_0,\ {}^{\lfloor \mathrm{rmn}}A, \Delta_{p+})$ was examined for the occurrence of clusters. The parameters and structure of the normal static mixture $f({}^{\lfloor \mathrm{t}}A_0,\ {}^{\lfloor \mathrm{rmn}}A, \Delta_{p+}|\Theta)$ (see 5.9) were estimated. Three processing ways were tried.

BMTB algorithm: An older, less elaborated variant of the sandwich algorithm providing Bayesian extension of the mean-tracking algorithm [154] was tried; see Chapter 12. It is computationally very cheap, but, unfortunately, it failed in the sparsely populated data space available.

AutoClass software: This freely available software [42] detects clusters using the EM algorithm (see Chapter 8) with random starts. Repetitive searches increase the probability of finding a better description of clusters but the computational cost increases, too. Over 700 000 random starts have been tried to get stable results. Low dimensions of the data set have allowed it. The best result was used as partial prior information for processing corresponding fully to the theory and algorithm described in Chapter 8.

Hierarchical splitting, quasi-Bayes algorithm and AutoClass: An older version of initialization [44] and quasi-Bayes algorithm as implemented in Mixtools [180] were used. The results were enhanced by exploiting partially results obtained by AutoClass.

Simultaneous design

The simultaneous design, described in Section 5.4.6 and elaborated in Proposition 9.15, was applied. It means that the component weights were interpreted more as elements of the mixture approximating the objective distribution of data records than objective values reflecting characteristics of the patient population.

As the parameters Θ are assumed to be known in the design, they will be omitted in the notation below. Using the notation introduced in (14.8), the user ideal pdf ${}^{\lfloor U}f(d(\mathring{t}))$ was defined in the same way like in (14.3). Uniform pf ${}^{\lfloor U}f(c_t)$ was chosen. The true user's ideal for the recognizable action ${}^{\lfloor U}f(u_{o;t})$ was set to $\mathcal{N}_{{}^{\lfloor \mathrm{t}}A_0}(3\,500, 800)$. This pdf has practical support within the usual range of administered activities. The true user ideal pdf of $\Delta_{o;t}$ was chosen conditioned by the value of $u_{o;t}$ in the following way:

$$ {}^{\lfloor U}f(\Delta_{o;t}|u_{o;t}) = \mathcal{N}_{{}^{\lfloor \mathrm{rmn}}A}(g({}^{\lfloor \mathrm{t}}A_0), 0.03), $$

where $g({}^{\lfloor \mathrm{t}}A_0) = a\,{}^{\lfloor \mathrm{t}}A_0 + b$, $a < 0$. The values of a and b were chosen so that the range of ${}^{\lfloor \mathrm{t}}A_0$ shown in the Table 14.3 is linearly mapped to the required range of ${}^{\lfloor \mathrm{rmn}}A \in (0; 0.18)$ with a negative slope. Using this formulation, the intention to minimize therapeutic administered activity ${}^{\lfloor \mathrm{t}}A_0$ was expressed.

After the design made by Algorithm 7.9, the constructed ideal pdf was projected into $d_o^* \equiv {}^{\lfloor \mathrm{rmn}}A,\ {}^{\lfloor \mathrm{t}}A_0{}^*$ giving the advisory mixture ${}^{\lfloor I}f(u_o, \Delta_o|\Delta_{p+})$.

Online phase

The statistics of the ideal pdf ${}^{\lfloor I}f({}^{\lfloor \mathrm{rmn}}A,\ {}^{\lfloor \mathrm{t}}A_0|\Delta_{p+})$ conditioned by the actual patient's data Δ_{p+} were evaluated. The maximum of the pdf indicates the recommended value of ${}^{\lfloor \mathrm{t}}A_0$ with predicted value of corresponding ${}^{\lfloor \mathrm{rmn}}A$. The example with a two-dimensional pdf map and one-dimensional marginal pdfs of both quantities is shown in the Figure 14.5.

14.2.3 Implementation

The initial versions of advisory modules were programmed in C++. An original API capable of transferring MATLAB mex-functions into stand-alone applications was used to take advantage of ready-made Mixtools algorithms. The application accepting data Δ_{p+} and computing a grid of the corresponding ${}^{\lfloor I}f({}^{\lfloor \mathrm{rmn}}A,\ {}^{\lfloor \mathrm{t}}A_0|\Delta_{p+})$ was integrated into the system *Iodine III* [185]. This system, written in MS Visual FoxPro, cares about data management and evaluation of a range of Bayesian estimates of various biophysical quantities. The presented GUI is a part of this system.

Graphical presentations of advices are shown in Figure 14.5. The interactive color pdf map allows physicians to try, for the given patient, various planned administrations by clicking on the map and examining the consequences, i.e., to predict values of $^{\lfloor \mathrm{rmn}}A$. The GUI allows to classify the advices for testing purposes, save the data, restore data of previously processed patients and adjust GUI.

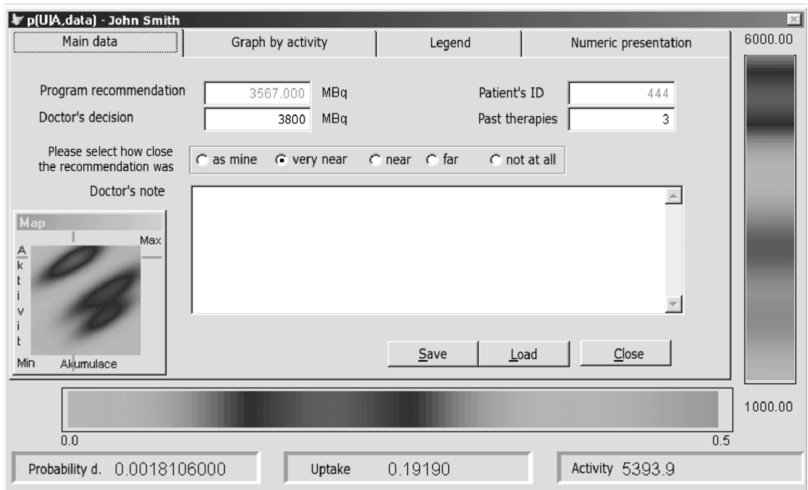


Fig. 14.5. Screenshot of *Iodine III* with interactive probability density map on the left and bars with marginal pdfs at the bottom and on the right

14.2.4 Results

With the cooperation of physicians from the Clinic of Nuclear Medicine and Endocrinology, Motol Hospital, Prague, the advices were tested for 101 patients. The recommended values of $^{\lfloor \mathrm{t}}A_0$ were compared to those decided by the physicians.

It was observed that the best coincidence was reached in the cases when a patient first visited the clinic for a radio-destruction of thyroid remnants after thyroid removal during surgery. These patients usually suffer from a primary tumor, i.e., only the thyroid is affected and no metastases are usually developed.

In these cases, the treatment procedure usually does not depend on other quantities like histology of the tumor, biochemical analyzes, etc. The differences between the recommended and decided value of $^{\lfloor \mathrm{t}}A_0$ in these cases are below 15 %. These cases represent 46 % of the tested patients. A difference below 10 % was observed in 31 % of the tested patients.

In cases of therapy repetition due to the tumor reappearance, the coincidence was much lower. A difference above 50 % was observed in 15 % of patients. Generally, the advisory system recommends lower values than the physicians. Decision in this category depends also on other examinations than the used dosimetric ones. Unfortunately, these vital data have not been available in the clinic in a form suitable for automatic processing.

14.2.5 Conclusions

The described advisory system is an original application of the mixture theory in this field. However, the task is specific to a small amount of the data available; therefore the results of the learning stage depend heavily on prior information. The AutoClass initialization has improved performance of the tested version of the Mixtools. Even with the prior obtained on the basis of 700 000 random starts, the model designed for the advising has acceptable physical interpretation only in about one-half of the cases. The advices are sensitive to mixture estimation. The key assumption, that the requested information is contained in the available data set, seems to be violated by using only a subset of data that is only accessible in a systematic way. The *GIGO principle* (garbage in, garbage out) is manifested here.

Anyway, the subset of cases relevant to the data set and the way of its processing, 46 % in this application, exhibits satisfactory results. Moreover, the quantitatively demonstrated conclusion that dosimetric data are insufficient for controlling the therapy has initiated the reorganization of data acquisition. Consequently, the advising based also on additional informative data will be possible in the near future. The nature of the considered data implies that the use of mixed data mixtures, Chapter 13, will be inevitable.

14.3 Prediction of traffic quantities

With an increasing number of cars, various significant transportation problems emerge, especially those concerning urban traffic, [186], [187]. Feedback control via existing traffic lights is a viable way to decrease them.

14.3.1 Problem description

The control design for complex traffic systems has to be build by solving partial, mutually harmonized, subtasks. Among them, the prediction of the traffic state is of primary importance. Here, we demonstrate appropriateness of predicting traffic quantities based on a mixture model.

Prediction of urban traffic state

Solution of urban transportation problems via reconstruction of the street network is expensive and very limited as it has to respect the existing urban conditions. Thus, the capacity of the network has to be efficiently exploited. This makes feedback traffic control utilizing the available traffic lights extremely important. Its design has to be able to predict a future traffic state for a given regime of traffic lights. Only with such a model, the regime can be varied to increase the permeability of city crossroads network. For it, mixture modelling seems to be suitable. Changes of daily and seasonal traffic indicate it clearly.

Notations, terms and quantities

Controlled networks are split into microregions. They are logically self-contained transportation areas of several cross-roads with their adjoining roads. Their modelling and feedback control exploit data measured by detectors based on inductive electric coils placed under the road surface. The presence of a huge metallic object above the coil changes its magnetic properties and thus individual cars are detected. Each detector signals a presence or absence of a car above it. From this signal, basic transportation quantities are evaluated:

- *Occupancy* o, which is defined as the portion (in %) of the time when the inspected place is occupied by cars.
- *Intensity* q expressing the number of cars per hour.
- *Density* ρ, which is defined as the number of cars per kilometer of the traffic flow. Specifying an average length of cars, it is computed as a ratio of occupancy and average car length.

Intensity and density describe the traffic state at the detector position.

14.3.2 Problem and its solution

Crossroads in cities are controlled by setting a proper green proportion and cycle length of the signal lights. Mostly, these signals are set according to a working plan whose changes during the day are given by a fixed schedule. Automatic feedback in selection of these plans can improve quality of the traffic significantly.

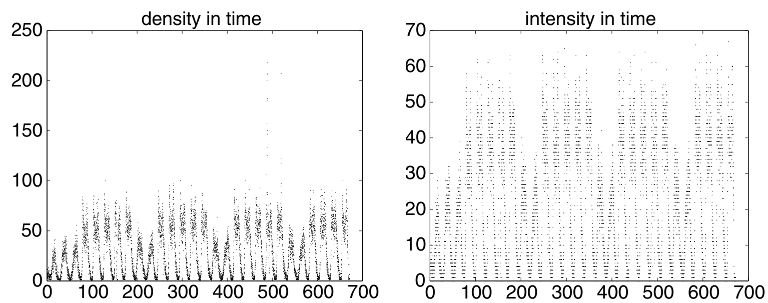
For the control, knowledge of traffic flow state incoming into the microregion in the near future is necessary. Reliable and possibly multistep prediction of the traffic flow can decide about practical success of such a feedback control.

Offline phase

Data

For experiments, 20-dimensional data records measured along two traffic lanes of the Strahov tunnel in Prague were used. The length of the tunnel is 2 km with two lanes in each direction. The detectors are placed after 500 m under each lane. Traffic in the tunnel is relatively fluent and almost no traffic congestions are observed.

Time course of the measured quantities on detectors reflects natural transportation periodicity. It is clearly seen on the Fig. 14.6, showing a typical time course of intensity and density of traffic flow for about 4 weeks.



Time course of density ρ Time course of intensity q

Fig. 14.6. Traffic quantities from the middle of the tunnel

The daily periodicity is visible there. At night, the traffic intensity is very low. In the morning, the demands rapidly rise due to the cars of commuters and increased commercial transport. The intensity of the traffic reaches soon its maximum and a slight fall is observed around noon. Then, it rises again and the saturation lasts until the evening when the intensity starts to fall gradually to zero. The weekly periodicity connected with alternating work days and weekends is also strongly reflected in the data.

The available detectors provide 5-minutes samples of the traffic intensity and density. Data from 10 detectors are recorded giving 20-dimensional data record each 5 minutes, 288 vectors per day.

About 8 500 data records reflecting approximately 4 weeks period were used for learning and 300 data records corresponding to 1 day served for testing.

Quality marker

Quality of models used as predictors is judged using the relative standard deviation R_i of the prediction error $\hat{e}_{i;t} \equiv d_{i;t} - \hat{d}_{i;t}$ of the ith data entry $d_{i;t}$,

ith channel,

$$R_i \equiv \sqrt{\frac{\sum_{t\in t^*} \hat{e}_{i;t}^2}{\sum_{t\in t^*}(d_{i;t}-\bar{d}_i)^2}},\ \ i = 1,\ldots,\mathring{d} = 20,\ \bar{d}_i \equiv \text{sample mean of } d_{i;t}\ t \in t^*. \tag{14.9}$$

In (14.9), the point prediction $\hat{d}_{i;t}$ is an approximate conditional expectation of $d_{i;t}$.

Preprocessing of data

Experience shows that detectors are fault-prone and their reparation is far from being easy. Therefore the filtration of their data is necessary. Here, outlier filtration and normalization to zero mean and unit standard deviation were applied on the raw data.

Model for prediction

The traffic system switches among several very different states. A mixture model has been chosen as a proper tool for modelling of this effect.

The model is built in the learning offline phase and is used for prediction with another part of the data. During the learning, the joint pdf of the modelled data record is estimated. Then, the pdf predicting selected channel(s) conditioned on other measured data is created. It is used for point or interval predictions of the channels of interest.

Estimation of mixture models

Both static and dynamic normal mixture models were used.

Experiments with the static case inspected the contribution of multivariate modelling and of the switching among various components.

Auto-regressions of the first-order used in the dynamic case were found adequate for modelling and predicting 20-dimensional data records. Mixtures with components having higher order brought no improvements.

The initialization based on hierarchical factor splitting and implementing the normal version of Algorithm 6.8, [44], was used. It determined a finer structure of components and their number. Up to six components were found. In all cases, at least three of these components had non-negligible weights. Some components had almost zero weights but it is interesting that their omission decreased the prediction quality visibly. The rarely populated components described outlying data and consequently improved parameter estimates of the remaining components.

Online phase

The traffic application has not reached the implementation phase yet mainly because of economical reasons. Thus, testing of predictive capabilities of the learnt models on validation data was used as a substitution of their online use. It imitates well application of the models as fixed predictors.

The test were performed in the experimental and development environment provided by the Mixtools toolbox [180].

14.3.3 Experiments

The experiments verified the ability of the estimated mixture model to predict well the state of the traffic flow at the northern exit of the tunnel using the data along the whole tunnel.

Prediction quality of the inspected mixture model (MIX) was compared with the one-component mixture, i.e., with the auto-regression (AR) model. The comparison of results is used both for the static and dynamic case.

The choice of the best model variant was based on the corresponding v-likelihood. The function of the obtained predictors was quantified by the performance index R_i (14.9). Its dependence on the number of modelled channels $\mathring{d}$ and the number of prediction steps, are presented in Tables 14.4 and 14.5.

Static mixture estimation

Static normal components predict the future data by their offsets only. The prediction made with them shows the need for dynamic modelling and superiority of the mixture model in comparison to the corresponding AR model. This superiority stems from the data-dependent switching between components, i.e., between different offsets.

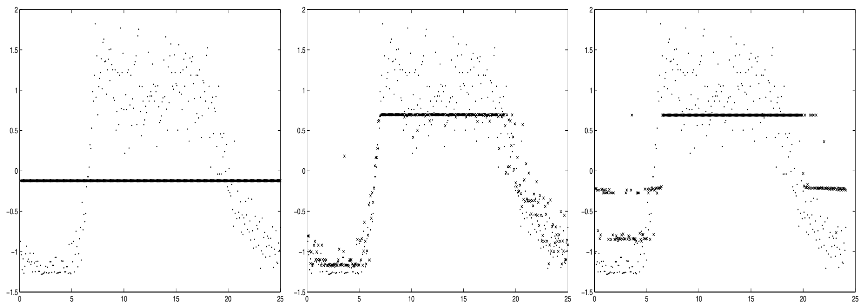
The two-dimensional model of intensity and density at the tunnel exit, 1^{st} and 2^{nd} channel, $\mathring{d} = 2$, was estimated only.

Table 14.4 shows prediction quality expressed by the marker $R_i, i = 1, 2$ (14.9) for two modelled channels and the prediction horizon ranging from 1 to 18 steps, which corresponds to the range from 5 to 90 minutes.

Table 14.4. Multistep ahead predictions with static two-dimensional model. R_i is the marker (14.9) for channels $i = 1, 2$ ($\mathring{d} = 2$). Results related to the mixture and auto-regression models are marked “MIX” and “AR”, respectively.

	R_1		R_2	
Steps ahead	MIX	AR	MIX	AR
1 (5 min)	0.418	1.021	0.422	1.020
6 (30 min)	0.512	1.025	0.516	1.024
12 (60 min)	0.634	1.031	0.638	1.030
18 (90 min)	0.756	1.028	0.760	1.037

Figure 14.7 shows typical behavior of the compared autoregression “AR” and static mixture “MIX” models on one-step and 18-step ahead predictions.



1-step-ahead "AR" 1-step-ahead "MIX" 18-step-ahead "MIX"

Fig. 14.7. Prediction with auto-regression "AR" and static mixture "MIX" models. Dots denote measured data, $\times$ predictions. Data are normalized to zero mean and unit variance.

Dynamic mixture estimation

First-order normal components have been used. Mixtures with $\mathring{d} = 2$ and $\mathring{d} = 20$ were estimated, the latter was projected into the space of the channels 1 and 2.

Table 14.5 shows a comparison of the marker R_i, (14.9), for different number of modelled channels used for multiple-step predictor. The length of predictions covers the range from 5 to 90 minutes (one step means 5 minutes).

Table 14.5. Multistep predictions with dynamic models. $\mathring{d}$ — the number of modelled channels, R_i the prediction marker (14.9) for channels $i = 1, 2$. The results are related to the automatically initialized mixture model with four components found.

	R_1				R_2			
Steps ahead	1	6	12	18	1	6	12	18
$\mathring{d} = 2$	0.34	0.45	0.56	0.68	0.48	1.05	1.09	1.12
$\mathring{d} = 20$	0.32	0.41	0.54	0.67	0.86	0.91	0.93	0.96

14.3.4 Results

Comparison of auto-regression and mixture models

It can be seen, that the mixture models predict better, even though the signal lights are preset into the mode preventing congestions inside the Strahov tunnel and thus limiting other traffic modes. The mixture model also utilize better the information from the additional entries of data records.

Static models

The good property of mixtures is clearly seen on Fig. 14.7. The static AR model is able to predict only the mean value of the whole data sample. The projected mixture, due to the data-dependent weights of individual components, is able to follow even the ascending and descending trends of the data; see Section 7.1.2. With the increasing length of prediction the ability to follow data declines but the switching among components is clearly visible.

Dynamic models

They describe and thus predict data of this kind much better. For short-term predictions, projected high dimensional model does not dominate over the low-dimensional one. For longer-term predictions, advantage of the high-dimensional model is clearly visible as it includes more information about the traffic state.

14.3.5 Conclusions

In summary, mixtures are worth being considered for modelling and prediction of high-dimensional transportation data. Their ability to make reasonable multistep ahead predictions indicates their ability to grasp a significant portion of physical properties of the modelled systems. This is a necessary precondition for applicability of a model in a control design.

14.4 Conclusions on the applications

14.4.1 Lessons learned

The applicability of the advisory system based on mixture models has been tested on three real processes: rolling of metal strips, treatment of thyroid gland tumor and prediction of transportation data. All of them have a multimodal nature that makes the use of ordinary linear models inefficient.

The rolling mill application confirmed that the adopted methodological basis is sound and that the resulting algorithms can be applied with a measurable increase of efficiency of the managed system. At the same time, it has provided invaluable feedback to the reported research and clarified priorities for continuing research.

A specific trait seen on a nuclear medicine application is insufficiency of data given by the technical conditions of the treatment and data measurement. It makes the most visible the key application problem: the amount of information on the system carried by the data has to be sufficient. According to the *GIGO principle* (garbage in, garbage out), the quality of advices depends on how much information in the data represents the system behavior in its entirety. It is clear that lack of information in the data is crucial for successful estimation and advising.

This application has also confirmed the real need for a full coverage of mixed-mixtures modelling.

All cases, and especially the experiments with traffic data, confirmed that mixture models, in spite of necessary approximations in their estimation, are an excellent tool for description of multimodal systems.

14.4.2 Other application areas

The described applications are just samples from an extreme application domain of the described theory and algorithms. During the development of the advisory system we met, for instance, the following tasks that can use efficiently the research outcomes

- prediction of nonlinear phenomena like load in energy networks (electricity, gas, heat),
- modelling, prediction and evaluation of life-saving signals of prematurely born children,
- use of mixtures for speculative short-term monetary operations,
- support of operators of large furnaces for producing glass or metal,
- call for creating operation model to be used for training of operators of complex processes,
- attempt to use mixture models for fault detection and isolation [188],
- effort to use mixture models in complex social phenomena studied, for instance, in connection with research in electronic democracy [129],

Optimized Bayesian Dynamic Advising
Theory and Algorithms
Karny, M. (Ed.)
2006, XVII, 529 p., Hardcover
ISBN: 978-1-85233-928-9