Adjustment of Nonuniform Sampling Locations in Spatial Data Sets

Managing rail geometry variations to keep trains running safely

Masako Kamiyama and Tomoyuki Higuchi

This article describes a procedure for adjusting sampling locations in one spatially discretized data set to those in another when the value differences between these sets are mainly caused by the sampling intervals that locally lengthen and shorten. This adjustment is formulated into an optimization form that can be solved by dynamic programming. Unknown parameters involved in the form can be identified using the maximum likelihood procedure that employs nonlinear filtering for a generalized state-space model. This procedure is based on the fact that the optimal solution in dynamic programming is equivalent to the "maximum a posteriori estimate" in a Bayesian framework.

Introduction

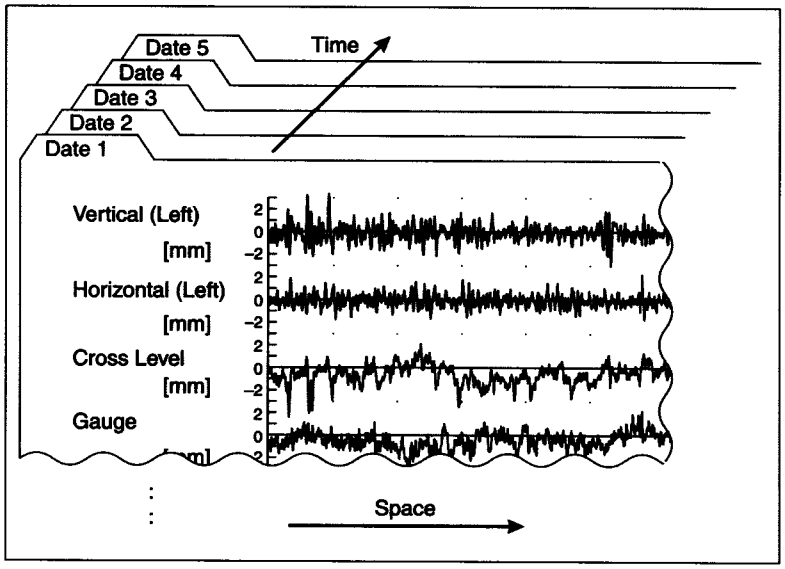
To manage the rail geometry of a railway track, a special rolling stock called the "track inspection car" periodically measures the rail geometry because it varies slightly under the load of passing trains. These geometry variations must be managed to keep trains running safely. While running on the rails, the track inspection car continuously measures various aspects of rail geometry, for example, vertical irregularity and the distance between the two rails (called "gauge"). These geometry measurements are simultaneously discretized at fixed spatial intervals and are recorded in digital data sets as illustrated in Figure 1.

Although it is desirable that these locations can be fixed to observe variations in the rail geometry, the set of the discretized locations on the rail changes slightly with each measurement, as illustrated in Figure 2. These changes are caused by the pulse used for selecting the sampling locations (called a "wheel-rotation pulse") being linked to the rotation of the car wheel as illustrated in Figure 3. Thus, identical spatial discretization cannot be reproduced.

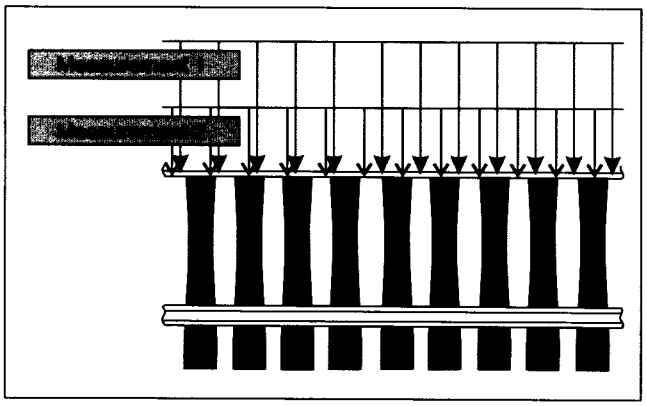
Moreover, it is difficult to adjust these location gaps after the discretization. If the spatial intervals between the discretized locations (called a "sampling interval") keep constant in two measuring runs, these gaps could be easily adjusted by calculating the correlation coefficient distance between the two data sets, even if the locations themselves change. However, in reality, some sampling intervals shorten or lengthen locally due to slipping or sliding of the car wheel (illustrated in Figure 4). Unfortunately, the length and location of these locally irregular intervals cannot be detected. This causes difficulty in adjusting the location gaps.

An example of these irregular sampling intervals is shown in Figure 5. The upper two curves show the values of the track gauges measured in the same railway section and demonstrate similar behavior since track gauge does not usually vary under the load of passing trains. The difference sequences between them are also shown below. Note that a difference operation is

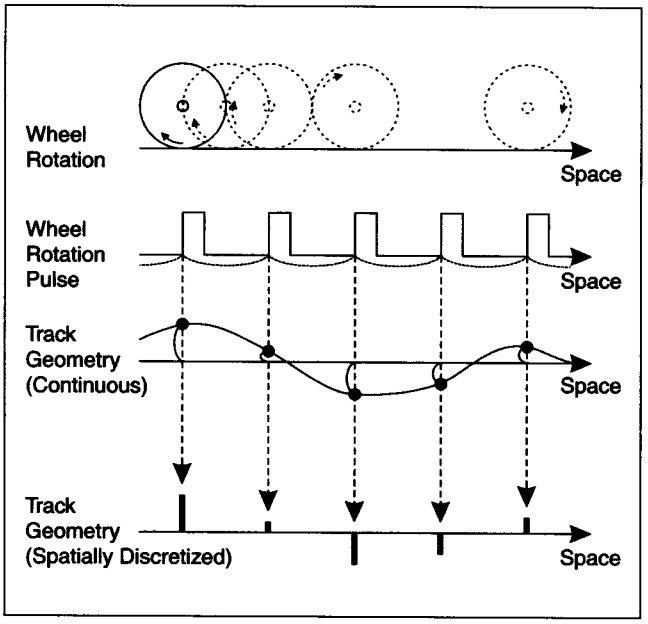
1053-5888/04/$20.00©2004IEEE



▲ *1. Scheme for measured railway track geometry and its database.*



▲ *2. Scheme for discretized-location gaps in a railway track with two measurements.*



▲ *3. Scheme for spatially discretized geometry of a railway track with wheel-rotation pulses.*

performed on the different intervals of (A). The amplitude of both difference sets changes abnormally in the second half although the two gauge waveforms appear similar. Hence, the wheel appears to slip or slide during one or the other of the measuring runs.

Until now, the variation in track geometry could not be computed from the differences between two data sets obtained with a track inspection car because of this location problem. The variation has been evaluated only by comparing representative statistics such as maxima. If the variation can be computed for every sampling location, it would be helpful in understanding the various mechanisms that influence rail geometry.

We have developed a procedure for adjusting the difference between sampling locations in two data sets when these sets tend toward similar values and nominally have the same sampling intervals. Furthermore, our procedure is applicable to noisy data sets. Figure 5 shows an example of adjustment of the locations with our proposed procedure. The abnormal amplitude, which is seen in the difference data sets mentioned above, is suppressed. Our procedure reveals that the wheel slid about 1 m when the training data set was measured.

Our procedure is based on the fact that the optimal solution of dynamic programming is equivalent to a maximum a posteriori (MAP) estimate in a Bayesian framework [1].

The procedure is outlined as follows.

▲ 1) Select a supervised data set and a training data set that satisfy the several criteria for adjustment purposes.

▲ 2) Model a mechanism to yield the nonuniform sampling, i.e., the wheel rotation including slip and slide.

▲ 3) Formulate this model in an optimization problem that can be solved by dynamic programming, which is a general method for solving nonlinear discrete optimization problems. However, this form contains unknown parameters.

▲ 4) Represent this nonlinear optimization problem with a generalized state-space representation and identify the unknown parameters (called hyperparameters in a Bayesian framework) using a nonlinear filtering algorithm based on the maximum likelihood method.

▲ 5) Adjust the sampling location differences by dynamic programming with the identified parameters.

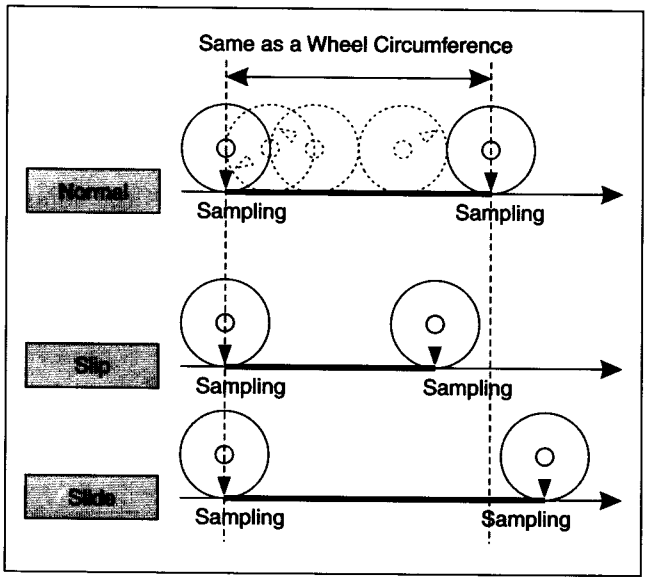
Data Set Properties

Properties of a Gauge Data Set

Since the inspection car simultaneously measures several aspects of rail geometry, it is possible to select the most suitable measurement from the various measurements to adjust the sampling locations. For convenience in this adjustment, the selected measurement should have the smallest variation over time. If these

values vary radically with time, the supervised data set values are completely different from those of the training set. In the case of rail geometry, the "gauge" (shown in Figure 6) data set is best since it has the smallest variation under trainloads.

Examples of the measured gauge sets are shown as the top two curves in Figure 5. These sets are already discretized at about 30 cm and have the following properties: steady-state random signals that contain noise arising from the measuring device. The average of the data set is about 0 mm (zero means the gauge has the regular value) with a standard deviation of about 0.7 mm. The noise is thought to be Gaussian white noise with a 0 mm nominal mean value and independent of the gauge value. Moreover, an analog low-pass filter smooths the measured signal before the spatial discretization.

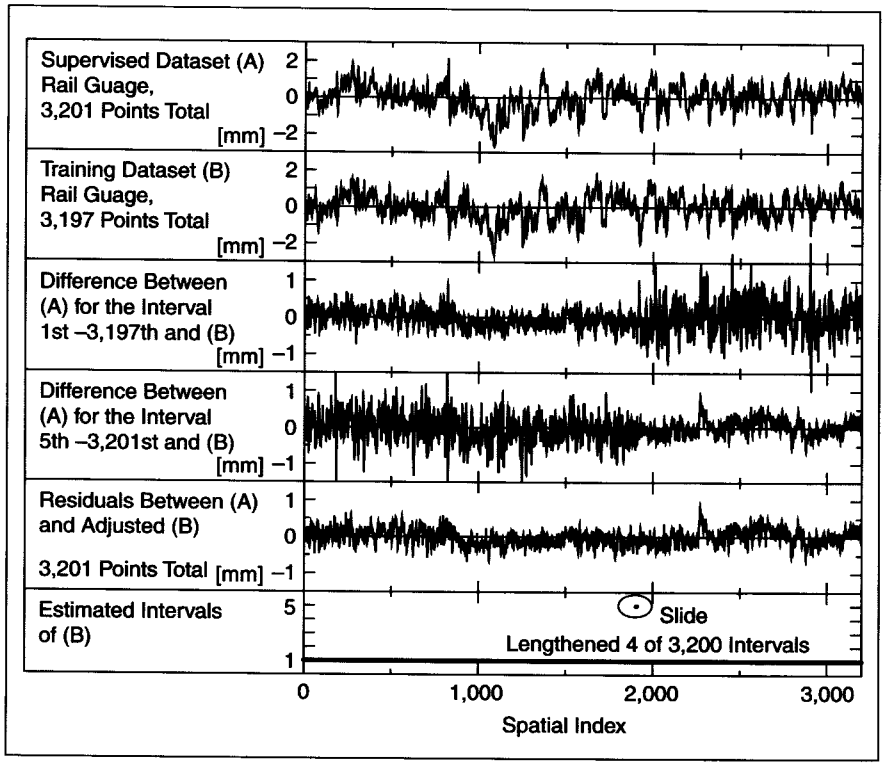


▲ 4. Scheme for variations in spatial-sampling intervals with wheel-rotation states.

Selection of a Data Interval for Analysis

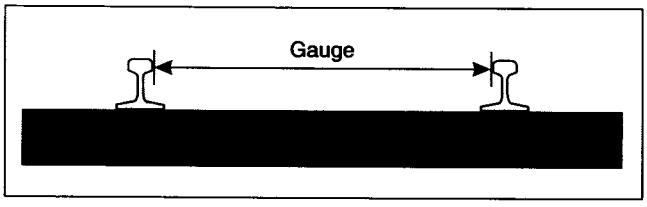
Since inspection cars run hundreds of kilometers in one measuring run and produce large data sets, adjusting all sampling locations simultaneously is impractical. Thus, it is necessary to extract two partial data sets from the two original long sets obtained on different measuring runs. To accomplish this, these extracted sets must be measured over approximately the same track section.

Incidentally, some data points in the measured data sets can roughly match the fix points in the railway track. These points are identified by peaks in the pulse signals generated by the ground transmitters embedded in the railway track and detected by the detector in the car as illustrated in Figure 7. We call this a "positioning pulse." Note that this pulse is different from a "wheel-rotation pulse." This positioning pulse is recorded and discretized at the same time as the other measured signals. After the discretization, the peaks in the discrete positioning pulse indicate that these data points are equivalent to the positions above the ground transmitters.



▲ 5. Example of measured data sets from two measurements in the same railway section, their differences, and residuals after location adjustment.

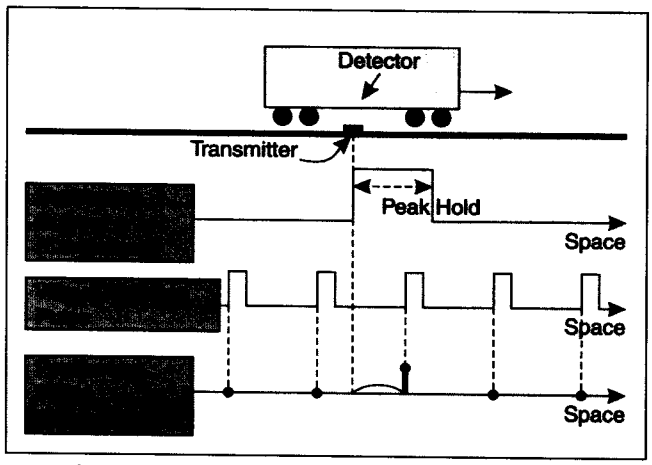
However, these peaks do not indicate the positions right above the transmitters since the peaks of the signal are held before the spatial discretization, as illustrated in Figure 7. This is done so as not to lose these peaks in the discretization. Hence, the peak corresponds to the first location at which the wheel-rotation pulse was generated after the detector in the car passed the transmitter. Figure 8 demonstrates the possible-existence range of pulse-generated location P estimated from discretized positioning pulse S. After the spatial discretization, the position of P is estimated to



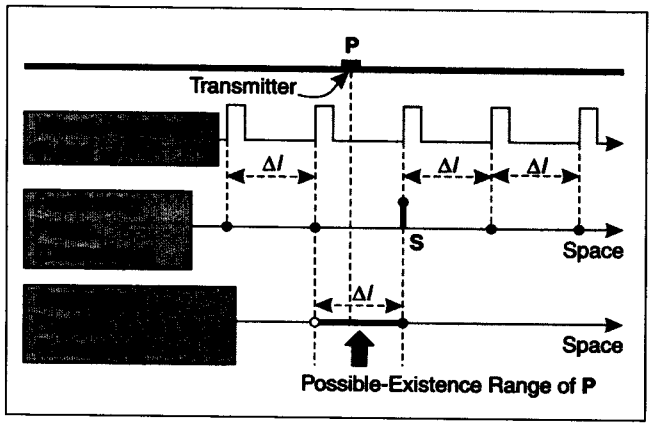
▲ 6. Illustration of rail gauge.

be somewhere between S and position Δl back from S; here, Δl represents the sampling interval. Thus, when S and S' are obtained near the same transmitter as the supervised set and the training set, the relative position of S (seen from S') is possible only in the range illustrated in Figure 9.

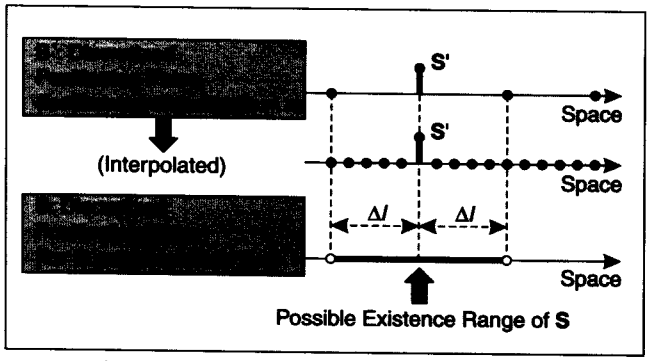
The remainder of this article considers the adjustment issue as adjusting the sampling locations between the extracted data sets. We select the supervised data set so that both ends of the set (y_1 and y_T, where $y_{1:T}$ is the selected supervised set) are equivalent to S. The training set is correspondingly selected. The gauge sets in Figure 5 are selected according to this procedure.



▲ 7. Scheme for a positioning-pulse signal generated in a railway track and discretized with wheel-rotation pulses.



▲ 8. Possible-existence range of pulse-generated location P estimated from discretized positioning-pulse S.



▲ 9. Possible-existence range of discretized location S in a supervised data set as seen from S' in a training data set when Δl is the sampling interval.

Selection of a Supervised Data Set

Selecting a suitable supervised data set for a certain training data set is important. We decided that the supervised set should satisfy the following two conditions.

▲ The first condition is that track maintenance work should not be conducted between the two measurement dates for the supervised and the training data sets. Track maintenance work noticeably changes rail geometry, including the gauge, whereas trainloads produce only slight changes.

▲ The second condition is that the sampling interval of the supervised set should be uniform because we aim to estimate the frequency characteristics of the geometrical variation after this adjustment. Hence, uniformity of the spatial sampling intervals is desirable for our supervised set; however, strict uniform sampling is impossible.

For these reasons, a specific criterion is required for estimating the uniformity of the sampling intervals, and we decided to estimate the uniformity from T: the number of data included. This number varies slightly with the measuring run even though the car measures between the same two ground transmitters. A series of measurements yields a distribution of these numbers. An example is shown in Figure 10. Thus, we decided that the supervised data set should be with 3,201, the most frequently arising number in this distribution. In other words, this mode number is a representation of a measuring run in which the inspection car did not slip or slide.

If several data sets satisfy these conditions, we select the supervised data set that was measured closest in time to the training set since daily trainloads also change the geometry of the rails.

Modeling Adjustment of Nonuniform Sampling Locations

Transformation of an Adjustment Problem

We adjust the nonuniform sampling locations according to the following procedure:

▲ 1) Select a supervised data set suitable for a training data set.

▲ 2) Divide the original sampling interval of the training set by positive integer α, and approximate the values for the newly interpolated data (interpolation).

▲ 3) Select data points corresponding to each supervised data from the interpolated training set.

Steps 2 and 3 are illustrated in Figure 11.

In this procedure, the adjustment is transformed into the selection of data points from the interpolated data set. Let n_t be the data point index of the training set, and select it corresponding to the tth supervised data.

This article expresses $\{n_1, n_2, \cdots, n_{T-1}, n_T\}$ as $\mathbf{n}_{1:T}$. The other variables are similarly expressed. Thus, the adjustment is transformed into construction of optimal sequence $\mathbf{n}_{1:T}$, where T is the number of the supervised data. Consequently, the number of the data contained in the interpolated training data set is approximately αT.

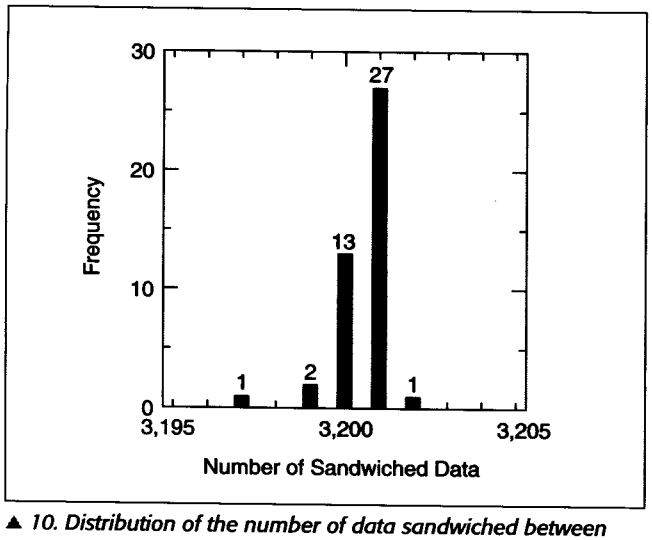
Hence, if $n_t - n_{t-1}$ is larger than α, the wheel is regarded as having slipped since the sampling interval of the training set lengthens with adjustment. Similarly, if $n_t - n_{t-1}$ is smaller than α, the wheel is regarded as having slid. Note that the uniformity of the intervals of the supervised data set cannot be assured. Hence, "slip" means $n_t - n_{t-1} > \alpha$, and "slide" means $n_t - n_{t-1} < \alpha$, although the occurrence of slip or slide is not assured.

Step 2 in the described procedure approximates the values of the data newly generated by the interpolation, so as not to change the frequency characteristic of the original data set [2]. This is because a low-pass filter has already smoothed the data sets before sampling. That is to say, this filter assures that the data set before sampling does not contain higher frequency components.

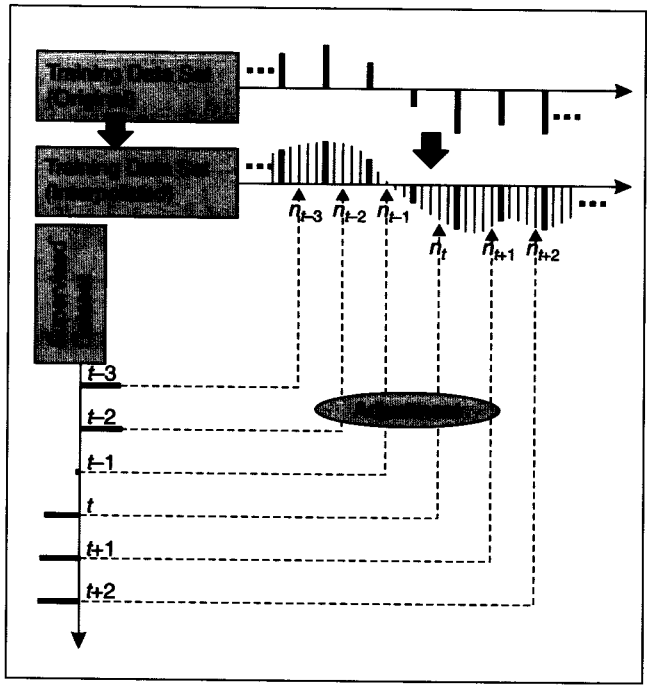
The interpolated training data set is called the "training data set" in the rest of this article.

Selection of a Training Data Set

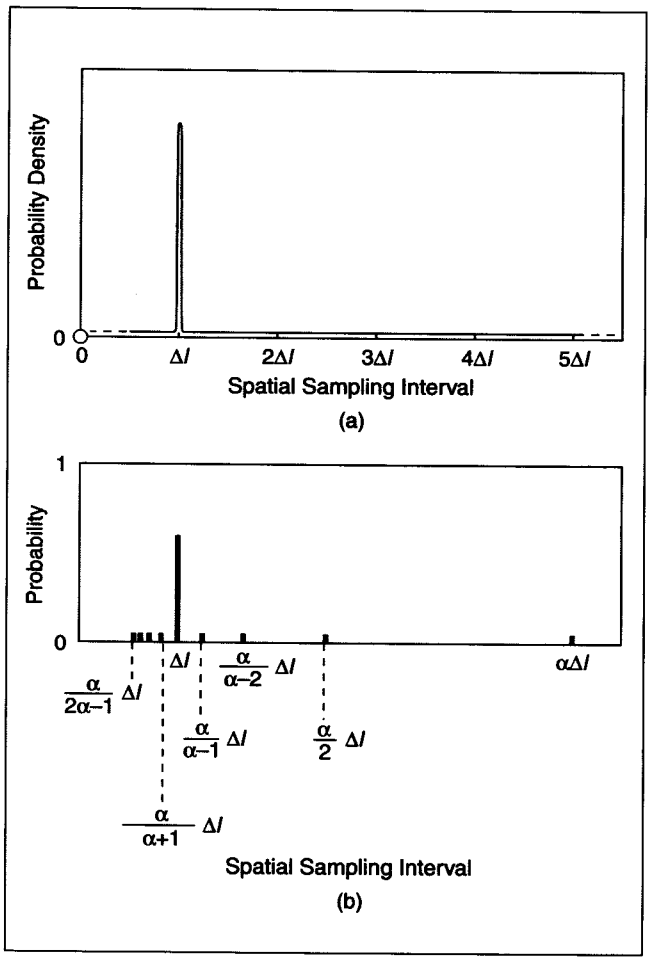
Before computation for the adjustment, consider the choice of training data set from the large set after interpolation. This selected training set must include all data that can be measured right above the locations of all supervised data points. In the discussion about the interval of the supervised data set, both ends of supervised set y_1 and y_T are equivalent to S in Figure 8. Hence, the training set is sufficient to include all data that are in the two possible-existence ranges of S in Figure 9 for y_1 and y_T and that are sandwiched by these two ranges. Note that it is not right above the



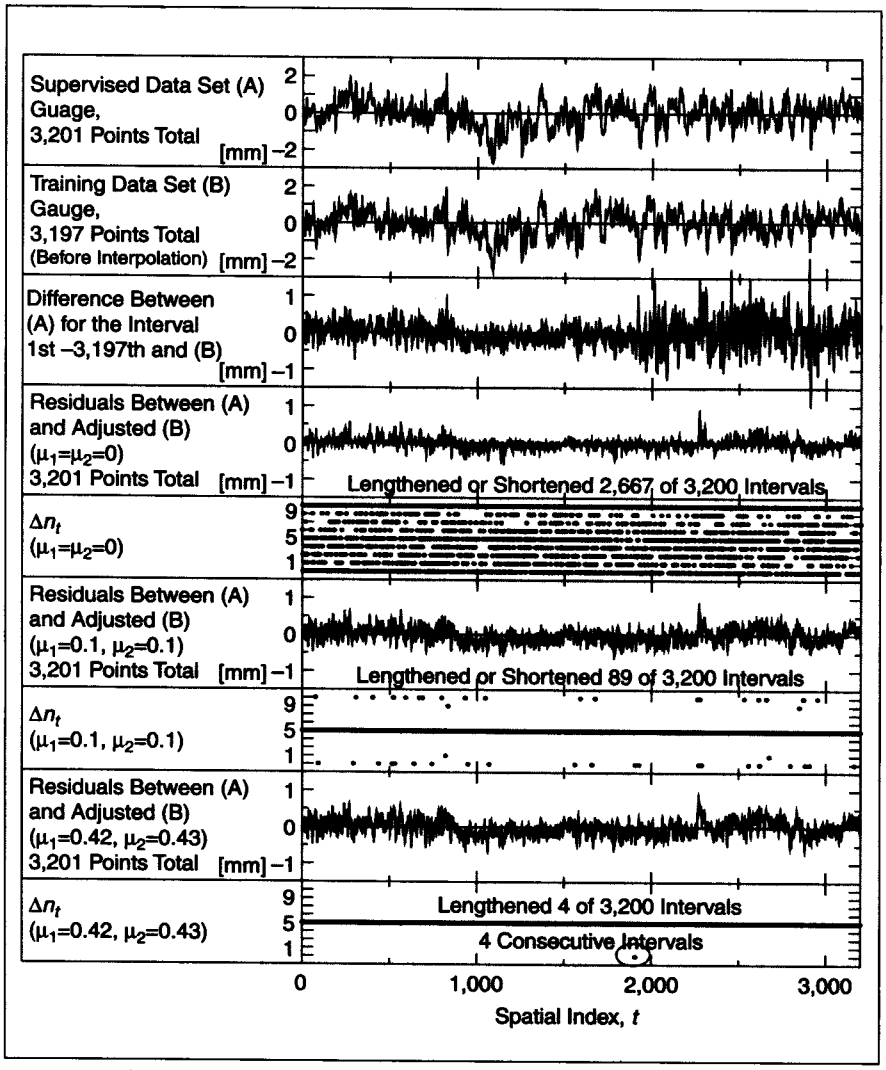
▲ *10. Distribution of the number of data sandwiched between two positioning pulses generated by two specific transmitters (including both boundary pulses).*



▲ *11. Scheme for the basic idea for adjusting the discretized-location gaps.*



▲ *12. Modeled probability distribution of spatial sampling intervals: (a) continuous and (b) discretized when $\alpha = 5$ and Δl is the spatial interval for a supervised data set.*



▲ *13. Typical results of adjusting sampling locations against three sets of μ_1 and μ_2 when $\alpha = 5$.*

ground transmitter. As illustrated in Figure 9, the possible-existence range of S (seen from S') is from $-\Delta l$ to Δl. Thus, all data that can be obtained above point S are the following: S' itself and forward and backward $\alpha - 1$ data from S'. These are $2\alpha - 1$ in all. Here, $\mathbf{z}_{1:N}$ is the selected training-data set and N is the number of the included data ($N \approx \alpha T$), z_1 is configured at $(\alpha - 1)$th forward data from $S'$and z_N at $(\alpha - 1)$-th backward data from S'.

Properties of the Optimal Data-Point Sequence

Let n'_t be a data-point index of the training set, selected to match the tth supervised data point y_t, and let n_t be the optimal index. $\mathbf{n}_{1:T}$ is the sequence constructed by all the optimal data-points. Hence, this location adjustment is equivalent to searching for optimal solution $\mathbf{n}_{1:T}$ from all feasible solutions $\mathbf{n}'_{1:T}$. This section considers the desirable properties of $\mathbf{n}_{1:T}$ for the optimization.

Firstly, $\mathbf{n}'_{1:T}$ can be evaluated by $\sum(y_t - z_{n'_t})^2$, where $z_{n'_t}$ is the value of the n'_tth training data. This is because the supervised set and the training set (before interpolation) are thought to be individually contaminated by Gaussian white noise sequences when the track inspection car obtains these data sets.

Second, consider the properties of $\Delta n_t \stackrel{\text{def}}{=} n_t - n_{t-1}$, which represents the sampling interval. If the sampling intervals in the training set are the same as those in the supervised set, $\Delta n_t = \alpha$. On the other hand, if the training intervals are locally different from the supervised intervals, Δn_t is a different integer from α. The training data interval is estimated at $(\alpha/\Delta n_t)\Delta l$, where Δl is length of the supervised data interval.

Obviously, the true sampling intervals have continuous values. The precise distribution of these is unknown. However, the probability of the slipping and sliding is assumed to be very low. The total running-distance with slipping and sliding is empirically assumed to be less than 0.1% of all.

For the range of values of Δn_t, this article employs $1 \le \Delta n_t \le 2\alpha - 1$. When $\Delta n_t = 1$, the wheel has slid, and the interval has locally lengthened to $\alpha\Delta l$. Similarly, when $\Delta n_t = 2\alpha - 1$, the wheel has slipped, and the interval has locally shortened to $(\alpha/2\alpha - 1)\Delta l$.

In addition, we assume that the probability distribution of Δn_t is the same except for $\Delta n_t = \alpha$. The modeled distribution is illustrated in Figure 12. The assumed interval that is continuously distributed around Δl is modeled as the discrete probability.

Finally, consider the properties of $\Delta^2 n_t \stackrel{\text{def}}{=} n_t - 2n_{t-1} + n_{t-2}$. We assume $\Delta^2 n_t = 0$ for almost all t ($3 \le t \le T$). This is because $\Delta^2 n_t$ is naturally zero when the wheel rotates regularly. Similarly, even when the wheel slips or slides, $\Delta^2 n_t$ is also assumed to be zero. In other words, the wheel in the course of slipping ($\Delta n_t > \alpha$) seems to recover normal rotation ($\Delta n_{t+1} = \alpha$) or to continue slipping at the same rate ($\Delta n_{t+1} = \Delta n_t$). This also occurs when the wheel slides. Thus, $-2\alpha + 2 \le \Delta^2 n_t \le 2\alpha - 2$ is obtained.

Formulating Adjustment into an Optimization Problem and Its Interpretation

Formulation into an Optimization Problem

Sequence $\mathbf{n}_{1:T}$ is the optimal solution that minimizes the following constrained nonlinear target function:

$$\mathbf{n}_{1:T} = \arg\min \sum_{t=1}^{T} F(n'_t)$$
$$= \arg\min_{n'_{1:T}} \left[\sum_{t=1}^{T} (y_t - z_{n'_t})^2 + \mu_1 \sum_{t=2}^{T} \xi(\Delta n'_t - \alpha, \alpha - 1) + \mu_2 \sum_{t=3}^{T} \xi(\Delta^2 n'_t, 2\alpha - 2) \right], \quad (1)$$

subject to

$$1 \le n'_1 \le 2\alpha - 1,$$
$$N - (2\alpha - 1) + 1 \le n'_T \le N,$$

where

- y_t is the value of the tth supervised data set
- T is the number of the supervised data
- $z_{n'_t}$ is the value of the n'_tth training data
- $\mu_1 \ge 0$ and $\mu_2 \ge 0$ are the weighting coefficients
- ξ is the function defined by

$$\xi(n, \gamma) = \begin{cases} 0, & n = 0, \\ 1, & 0 < |n| \le \gamma, \\ \infty, & \gamma < |n|. \end{cases}$$

To simplify this notation, we replace n'_t and n'_{t-1} with vector $\mathbf{x}'_t \stackrel{\text{def}}{=} [n'_t n'_{t-1}]^T$ (where a "^{T}" represents transpose) with $\mathbf{x}'_t \in \mathbf{S}_t \stackrel{\text{def}}{=} \{[n'_t n'_{t-1}]^T | n'_t \in S_t, n'_{t-1} \in S_{t-1}\}$, where S_t is the search area given for n'_t. By substituting these equations for $F(n'_t)$, function $g(\mathbf{x}'_t) \stackrel{\text{def}}{=} \min \sum_{j=1}^{t} F(n'_j)$ can be obtained as the following recursive formulation:

$$g(\mathbf{x}'_t) = \min \left[g(\mathbf{x}'_{t-1}) + f(\mathbf{x}'_t, \mathbf{x}'_{t-1}) | \mathbf{x}'_{t-1} \in \mathbf{S}_{t-1} \right], \quad \text{for} \quad 3 \le t \le T,$$

where

$$f(\mathbf{x}'_t, \mathbf{x}'_{t-1}) = (y_t - z_{n'_t})^2 + \mu_1 \xi(\Delta n'_t - \alpha, \alpha - 1) + \mu_2 \xi(\Delta^2 n'_t, 2\alpha - 2),$$

and "min[·]" represents the minimum element among "[·]."

The value of $g(\mathbf{x}'_t)$ depends on $g(\mathbf{x}'_{t-1})$, $\mathbf{x}'_t$, and $\mathbf{x}'_{t-1}$ alone, if the values of μ_1, μ_2, and α are all given. Therefore, this problem observes the principle of optimality, which underlies the following "dynamic programming" technique.

Optimization by Dynamic Programming

The following "dynamic programming" technique is effective for optimizing this type of problem.

1) Initialization

For each $n'_1 \in S_1 = \{1, 2, \ldots, 2\alpha - 1\}$,
$g(\mathbf{x}'_1) \stackrel{\text{def}}{=} (y_1 - z_{n'_1})^2$.

For each $\mathbf{x}'_2 \in \mathbf{S}_2 = \{[n'_2 n'_1]^T | n'_2 \in S_2, n'_1 \in S_1\}$,
$g(\mathbf{x}'_2) \stackrel{\text{def}}{=} \min \left[g(\mathbf{x}'_1) + (y_2 - z_{n'_2})^2 + \mu_1 \xi(n'_2 - n'_1 - \alpha, \alpha - 1) | n'_1 \in S_1 \right]$.

2) Recursion

For each $3 \le t \le T$ and each $\mathbf{x}'_t \in \mathbf{S}_t$,
$g(\mathbf{x}'_t) = \min \left[g(\mathbf{x}'_{t-1}) + f(\mathbf{x}'_t, \mathbf{x}'_{t-1}) | \mathbf{x}'_{t-1} \in \mathbf{S}_{t-1} \right]$,
$d(\mathbf{x}'_t) = \arg\min_{\mathbf{x}'_{t-1}(2)} \left[g(\mathbf{x}'_{t-1}) + f(\mathbf{x}'_t, \mathbf{x}'_{t-1}) | \mathbf{x}'_{t-1} \in \mathbf{S}_{t-1} \right]$
where $\mathbf{x}'_{t-1}(2) = n'_{t-2}$.

3) Termination

$$\min F(n'_t) = \min \left[g(\mathbf{x}'_T) | \mathbf{x}'_T \in \mathbf{S}_T \right] \left(\text{the minimized value of } \sum_{t=1}^{T} F(n'_t) \right),$$
$$\begin{bmatrix} n_T \\ n_{T-1} \end{bmatrix} = \arg\min_{\mathbf{x}'_T} \left[g(\mathbf{x}'_T | \mathbf{x}'_T \in \mathbf{S}_T) \right].$$

4) Backtrace

For $t = T, T-1, \ldots, 4, 3$,
$n_{t-2} = d(\mathbf{x}'_t | \mathbf{x}'_t(2) = n_{t-1})$.

Figure 13 shows examples of the optimal solution, which depends on the value of weighting coefficients μ_1 and μ_2. We should carefully select these parameters.

For this selection, consider an interpretation of solutions $\mathbf{n}_{1:T}$ and parameters μ_1 and μ_2.

Dynamic Programming and MAP Estimation

Assume that $(y_t - z_{n_t})$ is a normal random variable with a zero mean and variance σ^2. By multiplying $\sum F(n'_t)$ in (1) by $-1/(2\sigma^2)$ and exponentiating it, we obtain

$$\mathbf{n}_{1:T} = \arg\min_{n'_t} \left\{ \prod_{t=1}^{T} \exp\left[-\frac{1}{2\sigma^2} (y_t - z_{n'_t})^2 \right] \right\} \times \exp\left\{ -\frac{1}{2\sigma^2} \left[\mu_1 \sum_{t=2}^{T} \xi(\Delta n'_t - \alpha, \alpha - 1) + \mu_2 \sum_{t=3}^{T} \xi(\Delta^2 n'_t, 2\alpha - 2) \right] \right\}. \quad (2)$$

Next, by normalizing the first term on the right side of (2), we find that

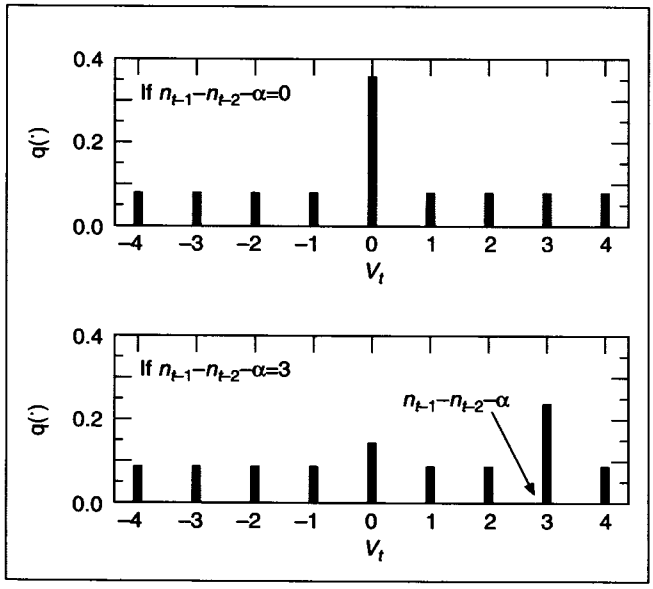
$$\prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2} (y_t - z_{n'_t})^2 \right]. \quad (3)$$

This function can be interpreted as the conditional density $p(\mathbf{y}_{1:T} | \mathbf{n}_{1:T})$ assuming normality. Similarly, by using normalization factor C, the second term of (2) is represented by

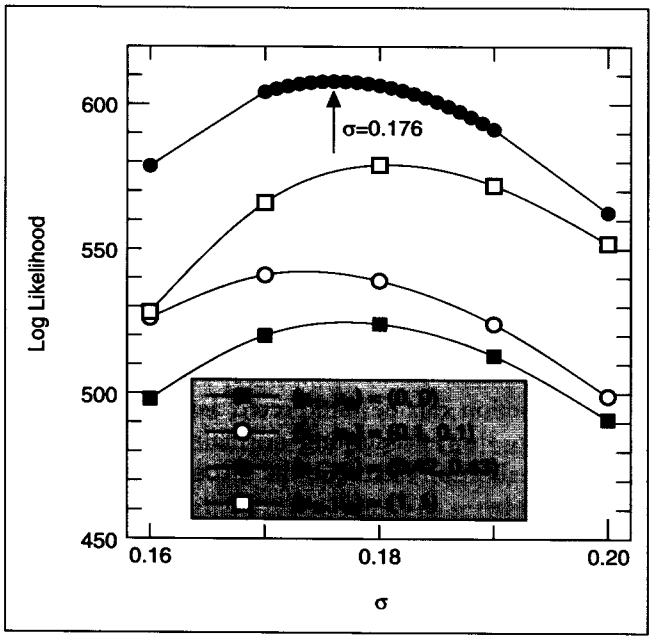
$$C \exp\left\{ -\frac{1}{2\sigma^2} \left[\mu_1 \sum_{t=2}^{T} \xi(\Delta n'_t - \alpha, \alpha - 1) + \mu_2 \sum_{t=3}^{T} \xi(\Delta^2 n'_t, 2\alpha - 2) \right] \right\}. \quad (4)$$

This function can be interpreted as probability distribution $p(\mathbf{n}_{1:T})$. In this way, (2) yields the following interpretation within a Bayesian framework: $p(\mathbf{n}_{1:T}|\mathbf{y}_{1:T}) \propto p(\mathbf{y}_{1:T}|\mathbf{n}_{1:T})p(\mathbf{n}_{1:T})$, where $p(\mathbf{n}_{1:T}|\mathbf{y}_{1:T})$ is the posterior distribution of $\mathbf{n}_{1:T}$, $p(\mathbf{y}_{1:T}|\mathbf{n}_{1:T})$ is the data distribution conditional on $\mathbf{n}_{1:T}$, and $p(\mathbf{n}_{1:T})$ is the prior distribution.

Hence, minimizing $\sum F(n'_t)$ by dynamic programming is equivalent to constructing a sequence that maximizes the posterior distribution of $\mathbf{n}_{1:T}$ with a given $\mathbf{y}_{1:T}$; $\mathbf{n}_{1:T}$ is then called the MAP estimate.



▲ *14. Typical distribution $q(\cdot)$for system noise v_t when $\alpha = 5$, $\mu_1 = 1$, $\mu_2 = 2$, and $\sigma^2 = 1$.*



▲ *15. Log-likelihood LL values change with μ_1, μ_2, and σ computed from data sets (A) and (B) in Figure 13.*

Identifying the Hyperparameters Included in an Optimization Form

Method of Computing Log-Likelihood

According to the above interpretation, the likelihood of a model specified by μ_1, μ_2, and σ is obtained by

$$L(\mu_1, \mu_2, \sigma) \stackrel{\text{def}}{=} p(\mathbf{y}_{1:T}|\mu_1, \mu_2, \sigma) = \prod_{t=1}^{T} p(y_t|\mathbf{y}_{1:t-1}, \mu_1, \mu_2, \sigma).$$

Hence, the values of the hyperparameters (μ_1, μ_2, σ) can be evaluated using the value of L [3]. Note that the concept of "hyperparameter" in a Bayesian framework is the same as that of "parameter" in a state-space model [4]. The most suitable hyperparameters $(\tilde{\mu}_1, \tilde{\mu}_2, \tilde{\sigma})$ are identified by maximizing L and LL where $LL \stackrel{\text{def}}{=} \log L$.

Log-likelihood LL is given by

$$LL \stackrel{\text{def}}{=} \log L(\mu_1, \mu_2, \sigma) = \sum_{t=1}^{T} \log p(y_t|\mathbf{y}_{1:t-1}, \mu_1, \mu_2, \sigma), \quad (5)$$

and obtained as the by-product of the computation for $p(\mathbf{n}_{1:T}|\mathbf{y}_{1:T})$.

Transformation into a Generalized State-Space Representation

Bayesian interpretations (3) and (4) can be transformed into the following generalized state-space representation:

(System model)

$$\mathbf{x}_t = f(\mathbf{x}_{t-1}) = \begin{bmatrix} n_{t-1} + \alpha + v_t(n_{t-1}, n_{t-2}) \\ n_{t-1} \end{bmatrix},$$

(Observation model)

$$y_t = h(\mathbf{x}_t) + w_t = z_{n_t} + w_t,$$

where $\mathbf{x}_t \stackrel{\text{def}}{=} [n_t \; n_{t-1}]^T$, $v_t \sim q(\cdot|\mathbf{x}_{t-1}, \mu_1, \mu_2, \sigma)$, $h(\mathbf{x}_t) \stackrel{\text{def}}{=} z_{n_t}$, and $w_t \sim N(0, \sigma^2)$. Distribution $q(\cdot)$ is defined as follows:

▲ If $n_{t-1} - n_{t-2} = \alpha$, then

$$q(v_t|\mathbf{x}_{t-1}, \mu_1, \mu_2, \sigma) \stackrel{\text{def}}{=} \begin{cases} 1/\beta_1, & v_t = 0, \\ \exp\left(-\frac{\mu_1+\mu_2}{2\sigma^2}\right)/\beta_1, & 1 \le |v_t| \le \alpha - 1, \end{cases}$$

where $\beta_1 = 1 + 2(\alpha - 1)\exp(-(\mu_1 + \mu_2/2\sigma^2))$ for $\sum q(\cdot) = 1$.

▲ If $n_{t-1} - n_{t-2} \neq \alpha$, then

$$q(v_t|\mathbf{x}_{t-1}, \mu_1, \mu_2, \sigma) \stackrel{\text{def}}{=} \begin{cases} \exp\left(-\frac{\mu_2}{2\sigma^2}\right)/\beta_2, & v_t = 0, \\ \exp\left(-\frac{\mu_1}{2\sigma^2}\right)/\beta_2, & v_t = n_{t-1} - n_{t-2} - \alpha, \\ \exp\left(-\frac{\mu_1+\mu_2}{2\sigma^2}\right)/\beta_2, & \text{other than the above} v_t, \text{ and} \\ & 1 \le |v_t| \le \alpha - 1, \end{cases}$$

where $\beta_2 = \exp(-(\mu_1/2\sigma^2)) + \exp(-(\mu_2/2\sigma^2)) + [2(\alpha-1)-1]\exp(-(\mu_1+\mu_2/2\sigma^2))$ for $\sum q(\cdot) = 1$. These distributions are illustrated in Figure 14.

This generalized state-space model is characterized as follows:

▲ Former state vector $\mathbf{x}_{t-1}$ controls the distribution of current system noise v_t

▲ y_t and w_t are real continuous variables, whereas $\mathbf{x}_t$ takes on integer values.

Identification of Hyperparameters with a Maximum Likelihood Function

The generalized state-space representation has the advantage of computation of aforementioned likelihood $p(\mathbf{y}_{1:T}|\mu_1, \mu_2, \sigma)$ since the following recursive relations between state distributions are available [4].

(Prediction)

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})\,d\mathbf{x}_{t-1} \quad (6)$$

(Filtering)

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \frac{p(y_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})}{\int p(y_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})d\mathbf{x}_t} = \frac{p(y_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})}{p(y_t|\mathbf{y}_{1:t-1})}. \quad (7)$$

Note that $p(y_t|\mathbf{y}_{1:t-1})$ appears as the denominator in (7). Thus, log-likelihood LL in (5) is obtained after these recursive computations. In addition, $p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$ and $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ can be computed arithmetically without complicated numerical integration such as [4]. This is because the values of $\mathbf{x}_t$ are discrete. We maximize LL using a grid search over a hyperparameter space since the number of hyperparameters is only three.

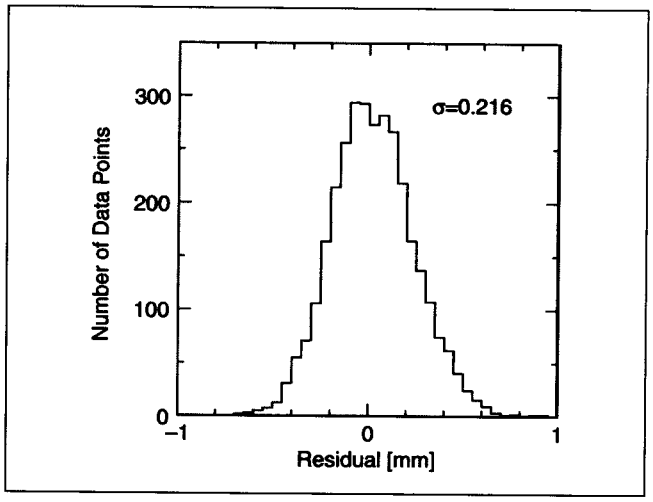
Incidentally, the recursive computation for LL requires initial state distribution $p(\mathbf{x}_1|\mathbf{y}_{0:0})$. A uniform distribution is assumed here. The reasons for this treatment are given in "Selection of a Data Interval for Analysis."

Discussion

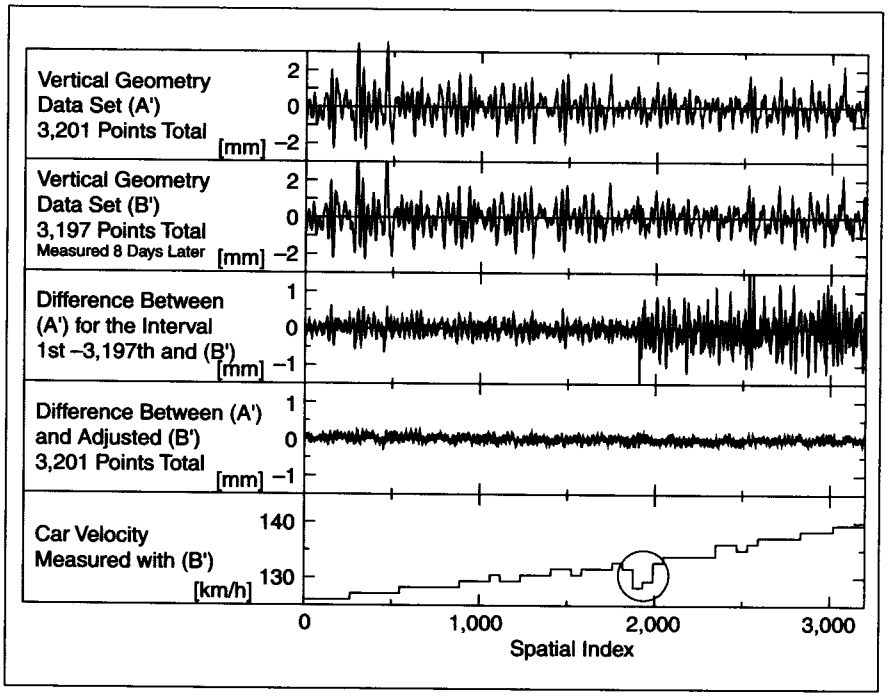
Application

We adjusted the sampling locations of (B) to those of (A) in Figure 5 where $\alpha = 5$. Data sets (A) and (B) are actually measured by a track inspection car in the same railway section. Figure 15 displays the values of log-likelihood $LL(\mu_1, \mu_2, \sigma)$ computed to estimate the optimal parameters $(\tilde{\mu}_1, \tilde{\mu}_2, \tilde{\sigma})$. Table 1 is a summary of the estimation results [$LL_{\max}$ in Table 1 means the maximum $LL(\mu_1, \mu_2, \sigma)$]. These results indicate that the most likely adjustment is shown in the lowest case in Figure 13 (the same as the adjustment results in Figure 5). In addition, these results also predict that the wheel has slid over a distance equivalent to four sampling intervals on measuring run (B).

$LL_{\max}$	$\tilde{\mu}_1$	$\tilde{\mu}_2$	$\tilde{\sigma}$	σ_{DP}
608	0.42	0.43	0.176	0.216



▲ *16. The value distribution of residuals between (A) and most likely adjusted (B) in Figure 5.*



▲ *17. Vertical geometry data sets (A') and (B') obtained at the same time as (A) and (B) in Figure 5, respectively, adjusted (B'), and the velocity of the inspection car: the sampling locations of (B') are adjusted according to those of the most likely adjusted (B) in Figure 5.*

Next, Figure 16 displays the distribution of the estimated residuals shown in Figure 5. Let σ_{DP} be the standard deviation of this distribution to distinguish it from $\tilde{\sigma}$. As summarized in Table 1, σ_{DP} is 0.216 and is larger than $\tilde{\sigma} = 0.176$. This difference is probably based on the properties of the residuals. In other words, the residual sequence differs from the ideal Gaussian white noise in the following respect; nonstationary behaviors are seen approximately $t = 900$ and $t = 2,300$.

In addition, we also moved the sampling locations of the vertical-irregularity data set (B′), simultaneously measured with (B), according to the above adjustment. The comparison between adjusted (B′) and the vertical-irregularity data set (A′), simultaneously measured with (A), is shown in Figure 17. Although the data-point sequence for adjusting (B′) was constructed independently of the values of (A′) and (B′), the difference sequence between (A′) and adjusted (B′) seems reasonable.

Figure 17 also displays the train speed obtained with (B) and (B′); this speed data usually cannot be referred to. The fact that the speed decreased with no other apparent reason to around $t = 1,900$ indicates that the wheel probably slid.

Applicability of the Present Procedure from a Methodological Viewpoint

Our procedure is based on the fact that a maximum a posteriori (MAP) estimate (in a Bayesian framework) is equivalent to the optimal solution obtained by dynamic programming. Dynamic programming is a general method for solving nonlinear discrete optimization problems that observe the "principle of optimality."

Here, the equivalence mentioned above is summarized. Consider a generalized state-space model for observed time series $\mathbf{y}_{1:T}$,

$$\begin{aligned} \text{System distribution:} \quad & \mathbf{x}_t \sim p_s(\mathbf{x}_t|\mathbf{x}_{t-1}, \boldsymbol{\theta}_s), \\ \text{Observation distribution:} \quad & y_t \sim p_o(y_t|\mathbf{x}_t, \boldsymbol{\theta}_o), \end{aligned} \tag{8}$$

where $\mathbf{x}_t$ is an unknown state vector, $\boldsymbol{\theta}_s$ and $\boldsymbol{\theta}_o$ are hyper-parameter vectors in a Bayesian framework, and y_t is the supervised data. Joint distribution $p(\mathbf{x}_{1:T}|\mathbf{y}_{1:T})$ has the following form:

$$p(\mathbf{x}_{1:T}|\mathbf{y}_{1:T}) = \prod_{t=1}^{T} p_s(\mathbf{x}_t|\mathbf{x}_{t-1}, \boldsymbol{\theta}_s) p_o(y_t|\mathbf{x}_t, \boldsymbol{\theta}_o).$$

MAP estimation maximizes this joint distribution under the given $\mathbf{y}_{1:T}$ and can be obtained by maximizing the logarithm of this distribution

$$\sum_{t=1}^{T} [\log p_s(\mathbf{x}_t|\mathbf{x}_{t-1}, \boldsymbol{\theta}_s) + \log p_o(y_t|\mathbf{x}_t, \boldsymbol{\theta}_o)]. \tag{9}$$

Fortunately, (9) can be maximized by dynamic programming [1].

The most important contribution of this article is that when a certain optimization problem can be transformed into form (9), it is possible to derive generalized state-space model (8). In this case, $\mathbf{x}_t$ consists of n_t and n_{t-1}, and $\boldsymbol{\theta} \stackrel{\text{def}}{=} [\boldsymbol{\theta}_s^T \boldsymbol{\theta}_o^T]^T = [\mu_1 \ \mu_2 \ \sigma]^T$.

When the most suitable values of $\boldsymbol{\theta}$ are necessary, they can be obtained by maximizing log-likelihood $LL(\boldsymbol{\theta})$ defined as follows [3]:

$$LL(\boldsymbol{\theta}) = \log p(\mathbf{y}_{1:T}, \boldsymbol{\theta}) = \sum_{t=1}^{T} \log p(y_t|\mathbf{y}_{1:t-1}, \boldsymbol{\theta}),$$

since interpretation using the maximum likelihood principle allows us to determine unknown parameter vector $\boldsymbol{\theta}$ objectively. Conditional distribution $p(y_t|\mathbf{y}_{1:t-1}, \boldsymbol{\theta})$ can be obtained as the by-product of nonlinear filtering.

Conclusion

Gaps in sampling locations for rail geometry data sets obtained with a track inspection car can be adjusted by optimization with dynamic programming after modeling the wheel rotation and detection of the ground transmitter. The unknown parameters introduced by modeling can be identified using the maximum likelihood procedure that employs nonlinear filtering for a generalized state-space model. This is because this optimization function can be interpreted within a framework of generalized state-space representation. This procedure is based on the fact that the optimal solution of dynamic programming is equivalent to the maximum a posteriori estimate in a Bayesian framework.

This adjustment procedure will be useful in observing variations in rail geometry.

Masako Kamiyama received the B.S. and M.S. degrees in resource engineering from Waseda University, Tokyo, Japan, in 1990 and 1992, respectively. She is currently pursuing the Ph.D. degree at Department of Statistical Science, the Graduate University for Advanced Studies, Japan. Since 1992, she has been a researcher with Railway Technical Research Institute (RTRI), Tokyo, Japan.

Tomoyuki Higuchi is currently director of the Prediction and Knowledge Discovery Research Center (PKDRC) at the Institute of Statistical Mathematics (ISM). He is also a professor of ISM and of the Graduate University for Advanced Studies. He obtained the B.S. degree in 1984, the M.S. degree in 1986, and the Ph.D. degree in 1989 from the University of Tokyo.

References

[1] S. Godsill, A. Doucet, and M. West, "Maximum a posteriori sequence estimation using Monte Carlo particle filters," *Ann. Inst. Statist. Math.*, vol. 53, pp. 82–96, Mar. 2001.

[2] S.L. Marple, Jr., *Digital Spectral Analysis with Applications.* Englewood Cliffs, NJ: Prentice Hall, 1987, pp. 43–44.

[3] H. Akaike, "Likelihood and the Bayes procedure," in *Bayesian Statistics*, J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M. Smith, Eds. Valencia, Spain: University Press, 1980, pp. 143–166.

[4] G. Kitagawa, "Non-Gaussian state-space modeling of nonstationary time series," *J. Amer. Stat. Assoc.*, vol. 82, pp. 1032–1063, Dec. 1987.