

Bayesian Estimation of Analyte Concentrations using Sensor Responses and Design Optimization of a Sensor System

David Han*

Abstract

Using an array of sensors with well calibrated but different tuning curves, it is possible to appreciate a wide range of stimuli. In this work, we first revisit the statistical estimation of the stimuli concentrations given the responses of a sensor array. Since it is not a typical regression problem, the Bayesian concept is adopted to develop an estimation method by elucidating the dynamic and uncertain nature of the environment-dependent stimuli with a proper choice of the probability distribution. Other studies confirm that the proposed method can demonstrate a superior performance in terms of accuracy and precision when compared to the popular frequentist methods in addition to the theoretical soundness it enjoys as a statistical estimation problem. Under the proposed framework, the design optimization of an artificial sensory system is also formulated using the expected Bayes risk as an objective function to minimize. The same approach may be equally applied to any sensory system in order to optimize its performance within a population of sensors. Finally, an illustrative example is provided to describe how the proposed method can be applied for the optimal configuration of a sensory system for a given sensing task.

Key Words: Bayesian analysis, chemical sensors, noninformative prior, optimal design of sensory arrays, sensitivity, statistical inference

1. Introduction

A chemical olfactory system is a sensor array consisting of hundreds of olfactory receptor neurons. Generally, these receptor neurons do not exhibit specificity to any single chemical compound but rather provide varying levels of response to multiple compounds. The precise mechanism for olfactory perception remains still unclear but the aggregate responses from a sensor array can provide the fundamental chemical information to detect tens of thousands of unique chemical vapors. By combining non-specific, general-purpose sensors with well calibrated but different tuning curves, it is possible to distinguish a wide range of stimuli or achieve a given sensing task without striving to develop fully selective or specific sensors to each chemical analyte [4]. Hence, sensor arrays are often proposed as potentially powerful and relatively inexpensive methods to characterize complex chemical mixtures.

Moreover, using information theoretic approaches, neural receptor systems have been explored in order to understand how these systems are structured, how the structure informs chemical recognition capability, and what the resulting implications are when designing a sensor-based chemical detection system [1, 6]. Despite these efforts, the literature and application of non-specific sensor arrays for general-purpose chemical detection present general disappointment upon implementation. Even though such arrays have been reported frequently and have been the subject of numerous reviews over the past decades, relatively few instances of successfully commercialized devices exist to date. One reason is because the research community has mainly focused on the development of individual sensors or on the application for which the sensor array is to be used without considering the design and evaluation issues of sensor arrays. Consequently, design, optimization, and implementation

*Department of Management Science and Statistics, University of Texas at San Antonio, TX 78249

of sensor arrays still remain time-intensive and costly, leading to sensory systems which tend to underperform when fielded, compared to their laboratory performance.

In this work, we revisit the statistical estimation problem of chemical stimuli given the responses of a sensor array, discussed in [18]. The concept of Bayesian analysis is adopted to develop an estimation method by explaining the dynamic and uncertain nature of the environment-dependent stimuli through a choice of the prior distribution. Under the proposed framework, the optimal configuration of an artificial sensory system is then discussed using the expected Bayes risk as a suitable objective function to characterize the performance of the sensory system. The proposed approach is generalizable and could be applied to other sensory systems for the stimuli estimation and/or the optimal sensory system designs. The rest of the paper is organized as follows. The formal model based on the Bayesian framework is developed in Section 2, and the estimation of analyte concentrations is discussed in Section 3 while Section 4 addresses the optimal design of a sensory system using the expected Bayes risk. An illustrative example is provided in Section 5 to describe how the proposed method can be applied for the optimal configuration of a sensory system. Finally, Section 6 concludes the paper.

2. Bayesian-based Model

Since the stimulus population and their respective concentrations can vary greatly from environment to environment the sensor array is exposed to, the covariate vector \mathbf{x} cannot be treated as static parameters to be estimated like in [18]. Rather, its dynamic and uncertain nature has to be elucidated in the model via a proper choice of its probabilistic distribution (*i.e.*, prior). For the efficient estimation of the unknown stimuli given an observed set of the sensory responses, here we adopt the concept of the Bayesian framework by treating the unknown stimuli vector as a random vector \mathbf{X} . We also express the prior historical information, the experts' opinions or beliefs about the stimuli through a distribution function of choice, denoted by $f_{\mathbf{x}}(\mathbf{x}; \boldsymbol{\psi})$ with predetermined hyperparameters $\boldsymbol{\psi}$. Then, by the Bayes' theorem, the (posterior) probability distribution function of $\langle \mathbf{X} | \mathbf{y} \rangle$ is expressed as

$$f_{\mathbf{x}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}; \boldsymbol{\theta}, \boldsymbol{\psi}) = \frac{f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}; \boldsymbol{\theta})f_{\mathbf{x}}(\mathbf{x}; \boldsymbol{\psi})}{\int_{\mathbf{x}} f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}; \boldsymbol{\theta})f_{\mathbf{x}}(\mathbf{x}; \boldsymbol{\psi})d\mathbf{x}}, \quad (1)$$

where $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}; \boldsymbol{\theta})$ is the joint distribution function of $\langle \mathbf{Y} | \mathbf{x} \rangle$ or the likelihood function of $\boldsymbol{\theta}$ while the denominator is called the marginal likelihood. Based on our previous assumption, $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}; \boldsymbol{\theta})$ is the joint density function of the multivariate normal distribution with the mean vector specified by $h^*(\boldsymbol{\theta}^*, \mathbf{x})$ and the variance-covariance matrix by $\boldsymbol{\Sigma}$.

As shown in Eq. (1), the posterior distribution is a combination of the prior (determined by the researcher) and the likelihood (determined by the data). The contribution of these two quantities to the posterior is not equal though. With more sensory responses (*i.e.*, $m \gg n$), the likelihood is given much more relative weight in calculating the posterior for the unknown stimuli [21]. This may seem trivial but it can have enormous implications when the number of sensors is small as the posterior distribution is highly reliant on how the prior was specified. When the information contained in the likelihood is relatively small due to a limited number of the sensory responses, the prior will play a key role in the posterior for estimating the stimuli and each additional piece of information will have a pronounced impact. Thus, it is necessary to include external information in the form of informative (or subjective) priors. For instance, one might need to consult application-specific experts, meta-analyses, or review studies in the area of interest to obtain informative, accurate priors that can meaningfully contribute to the posterior distribution of \mathbf{X} . Previous knowledge or information about the environment the sensor array is exposed to can also be incorporated

in the prior. Specifying the priors of \mathbf{X} based on expert opinions or previous studies can potentially improve the inferential performance since it allows to base results on more information than what is strictly provided in the sensory responses, which is especially helpful with small data sizes.

If one has very little or no prior information about the potential stimuli, noninformative priors such as Jeffreys prior, which is proportional to the square root of the Fisher information, or a reference prior, which maximizes the expected Kullback-Leibler divergence (*i.e.*, mutual information), can be employed to conduct the objective Bayesian inference. Gelman [8] also recommended using half-Cauchy priors for variance components while O'Malley & Zaslavsky [16] outlined a multivariate extension to the half-Cauchy prior. If the prior is set in the vague vicinity of the population value of \mathbf{X} , even with a fairly large variance, the advantages of the Bayesian framework can be realized. Previous studies in the literature reported that in terms of bias, power, appropriateness of coverage intervals, and efficiency, the weakly informative prior performs nearly as well as a strongly informative prior or the frequentist maximum likelihood estimators with small sample corrections. This finding is particularly helpful because weakly informative priors are widely applicable to various model types, are not overly difficult to implement, and reduce the chance that the posterior can be misleading due to an overly precise or extensive prior distribution.

3. Estimation of Analytes

Once the posterior is specified, the inference for $\langle \mathbf{X} | \mathbf{y} \rangle$ can be performed using the decision theoretic approach with a loss function of choice. The usual quantity of interest is the measurement of central tendency or location of the posterior distribution. Then, Bayes estimator (or action) is an estimator or decision rule that minimizes the posterior expected loss (*viz.*, Bayes risk). Equivalently, it maximizes the posterior expected utility function. Depending on the property of the posterior, one could utilize the posterior mean, which minimizes the expected loss with respect to the quadratic error loss function while the posterior median is the robust estimator and minimizes the expected loss under the absolute difference loss function in a univariate case. If a value with the greatest posterior probability is desired, the maximum a posteriori (MAP) estimate (*i.e.*, posterior mode(s)) can be used as the measurement of center, and it also minimizes the expected loss with respect to the 0-1 loss function. If one wants to quantify the uncertainty about the posterior center in addition to a point estimate, the HPD (highest posterior density) credible interval or set can be derived, which is analogous to the frequentist confidence interval but has a more intuitive interpretation [10].

As an illustration, let us consider n analyte concentrations to sense with a sensor array composed of n (not necessarily identical) sensors. It is also assumed that a special linear regression model holds for \mathbf{Y} given \mathbf{x} such that $h^*(\boldsymbol{\theta}^*, \mathbf{x}) = \mathbf{B}\mathbf{x}$ where \mathbf{B} is the $n \times n$ orthogonal matrix of the parameters $\boldsymbol{\theta}^*$, and $\langle \mathbf{Y} | \mathbf{x} \rangle$ consists of independent and homoscedastic random responses with the common variance σ^2 , which has no dependence upon the stimulus. Moreover, to represent the lack of information on the potential analytes to sense, equal likelihood is assigned on all possible analyte concentrations by defining a flat, uniform prior over the domain of \mathbf{X} (*i.e.*, noninformative), expressed as

$$f_{\mathbf{X}}(\mathbf{x}; \psi) = \begin{cases} 1/\psi & \text{if } x_i \geq 0, \forall i = 1, 2, \dots, n \\ 0 & \text{otherwise} \end{cases},$$

where x_i is the i -th element of \mathbf{x} and ψ is the normalizing constant of the joint density function of \mathbf{X} . Since $f_{\mathbf{X}}(\mathbf{x}; \psi)$ does not integrate to 1 or a finite value, it is an improper prior while the posterior is still proper. Then, it can be shown that the posterior distribution

of \mathbf{X} given \mathbf{y} is a product of left truncated normal probability densities at zero. Using the decision theoretic approach, the expected quadratic loss is minimized by the posterior mean vector given as $\boldsymbol{\mu}^{\text{EQL}} = (e_1^{\text{EQL}}, e_2^{\text{EQL}}, \dots, e_n^{\text{EQL}})^\top$ where

$$e_i^{\text{EQL}} = \mu_i + \sigma \frac{\phi(-\mu_i/\sigma)}{1 - \Phi(-\mu_i/\sigma)}, \quad i = 1, 2, \dots, n \quad (2)$$

with $\mu_i = \mathbf{b}_i^\top \mathbf{y}$ and \mathbf{b}_i is the i -th column vector of \mathbf{B} . Here, $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal density and cumulative distribution functions, respectively. If one utilizes the expected absolute difference loss to base the estimation decision, the mean loss could be minimized by the posterior median vector given as $\boldsymbol{\mu}^{\text{EAL}} = (e_1^{\text{EAL}}, e_2^{\text{EAL}}, \dots, e_n^{\text{EAL}})^\top$ where

$$e_i^{\text{EAL}} = \mu_i + \sigma \Phi^{-1} \left(\frac{1 + \Phi(-\mu_i/\sigma)}{2} \right), \quad i = 1, 2, \dots, n \quad (3)$$

with μ_i as defined above. The posterior joint density of $\langle \mathbf{X} | \mathbf{y} \rangle$ is also maximized by the MAP vector given as $\boldsymbol{\mu}^{\text{MAP}} = (e_1^{\text{MAP}}, e_2^{\text{MAP}}, \dots, e_n^{\text{MAP}})^\top$ where

$$e_i^{\text{MAP}} = \begin{cases} \mu_i & \text{if } \mu_i \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

It is not surprising to see that all the estimators of the stimuli concentrations presented above heavily depend on the accuracy and precision of the estimates of the parameters $\boldsymbol{\theta}$ in the likelihood portion of the posterior. In the case these estimators cannot be obtained analytically, one could utilize a popular stochastic simulation based approach such as the Markov chain Monte Carlo (MCMC) sampling method implementing the Metropolis-Hastings algorithm with the Gibbs sampler in order to elicit the posterior distribution of the stimuli concentrations. Through appropriate selection of a prior for \mathbf{X} that is restricted to have non-negative support, the proposed method can also prevent inadmissible estimates (*e.g.*, negative variances). Models that are difficult or inestimable with frequentist methods (*i.e.*, $m < n$) can be fit straightforwardly under the Bayesian framework [14, 15]. Since Bayesian methods do not rely on large sample asymptotics, they are also better equipped to handle small sample situations but the estimates can be sensitive to the specification of the prior. Hence, if one decides to utilize informative priors, it is important to make sure that the information on which such priors are based is accurate with small samples. Otherwise, the resulting estimates and variance can be biased due to misleading informative priors [5].

4. Optimal Design of Sensory System

In the process of selecting which sensors to be incorporated into a custom, sensing task-driven chemical olfactory system, a typical objective is to maximize the accuracy and precision with which the sensory system can estimate the stimulus or optimally discriminate between neighboring stimuli [17]. Sanchez-Montanes & Pearce [18] discussed how Fisher information matrix could be used to achieve this objective. However, the Cramer-Rao lower bound obtained by inverting the Fisher information is useful as a performance metric only if the variance of the chosen estimator for a given sensory system can actually attain the lower bound. In the statistical inferential theory, even the best unbiased estimator for a given system may not achieve the lower bound as its variance. Therefore, if the Cramer-Rao lower bound or the Fisher information is used to characterize or compare the performance of sensor arrays, it could dangerously overestimate the true merit of the systems in an unequal manner and mislead the system design to a non-optimal, inferior one. Also, if one desires to achieve the lower bound, it usually requires an unrealistically large

number of s -independent and identical (*iid*) sensors per sensor array, which is practically impossible due to the cost and technology constraints. In addition, the Fisher information as a function depends on the (true) stimulus types and/or concentrations, which are unknown. Thus, calculation of the Fisher information requires the estimation of these unknown quantities first, hence carries a certain but unknown amount of propagated error in an uncontrolled manner. Moreover, treating the stimulus types and/or concentrations as the parameters of a statistical model for estimation in the first place is problematic as a theoretically sound approach since the stimulus types and their respective concentrations can change dynamically depending on the environment to which the sensory system is exposed in the field studies.

Addressing all these fundamental issues, a new way of estimating the stimulus using the sensory receptor responses was proposed in the previous sections by adopting the Bayesian framework. For estimating the individual stimulus within a complex chemical vapor mixture, the optimal estimator is obtained by minimizing the expected loss, also known as the Bayes risk. We now steer our attention to the optimal configuration of a sensory system under the same framework. Here we propose the expected Bayes risk as a metric to determine the merit of a sensory system and discuss its connection to the chemometric sensitivity. This can be used to characterize the analytical capability of a sensory system and to quantify its performance. It can be also used as part of an optimization procedure for selecting noisy olfactory sensors within a population in order to make as accurate and precise estimate of the real chemical stimulus exposed to the overall sensory system as possible. The expected Bayes risk serves as a suitable objective function to achieve the optimization purpose, and a sensory system which minimizes the expected Bayes risk can be considered the optimally configured system for a given task.

As a simple motivating example, let us consider estimating the concentration of a single analyte of interest X with a sensor array composed of m *iid* sensors. It is also assumed that the receptor response of a sensor was linearly calibrated to changes in the stimulus concentration (in the dynamic range of interest) so that $\langle Y_j|x \rangle$ is normally distributed with the conditional mean response of $(\theta_0 + \theta_1 x)$ and the common variance σ_Y^2 for $j = 1, 2, \dots, m$. These assumptions are valid for olfactory sensors based on electrochemical cell technology as well as for measurements from metal-oxide semiconductor and conducting polymer chemosensors [2]. If one wishes to model fluorescence-based optical chemosensors used in the artificial olfactory systems, Laplace or double-exponential distribution could be utilized while Poisson (discrete) distribution is appropriate for approximating the spiking counts of olfactory receptor neurons [3, 18]. Here, θ_0 is a constant contribution to the measured response (background or constant interfering constituents) while the slope of the calibration curve, θ_1 is precisely the chemometric *figure of merit*, known as sensitivity to the presence of the analyte. In addition, a normal conjugate prior is assumed for X with mean ψ_X and variance σ_X^2 . With this prior, one could express the lack of information on the analyte by assigning an arbitrarily large value to σ_X with a suitable choice of ψ_X to favor the objective inference for X (*i.e.*, noninformative).

Then, it can be shown that the posterior distribution of X given the observed responses y_1, y_2, \dots, y_m is identical to that of $\langle X|\bar{y} \rangle$, which is again normal. Here, \bar{y} is the arithmetic sample average of y_1, y_2, \dots, y_m . Using the decision theoretic approach, the estimator of the analyte concentration, which minimizes the Bayes risk with respect to the squared error loss, is the posterior mean of X expressed by

$$\mu_{x|\bar{y}} = \left(\frac{\bar{y} - \theta_0}{\theta_1} \right) \frac{\theta_1^2 \sigma_X^2}{\sigma_Y^2/m + \theta_1^2 \sigma_X^2} + \psi_X \frac{\sigma_Y^2/m}{\sigma_Y^2/m + \theta_1^2 \sigma_X^2}, \quad (5)$$

and the corresponding minimum Bayes risk, R_B^{\min} is the posterior mean squared error

(MSE) or the posterior variance of X , expressed as the inverse of precision ($m\theta_1^2\sigma_Y^{-2} + \sigma_X^{-2}$). Since it is a constant in this case, the expected Bayes risk is still

$$E[R_B^{\min}] = \frac{\sigma_X^2 \sigma_Y^2}{m\theta_1^2 \sigma_X^2 + \sigma_Y^2}, \quad (6)$$

and it depends on m , the number of *iid* sensor responses, and the sensitivity θ_1 along with the measurement error parameter σ_Y . As σ_X tends to ∞ for the objective inference, the estimator in Eq. (5) becomes the classical MLE from the likelihood based approach with the observed responses y_1, y_2, \dots, y_m , and the expected Bayes risk in Eq. (6) reduces to $\sigma_Y^2/(m\theta_1^2)$. Hence, if one needs to choose a sensor to construct a sensor array from an arbitrary population of sensors, each with well calibrated response curves, the sensor which minimizes this quantity should be selected in order to maximize the accuracy and precision in estimating the analyte of interest. Clearly, there is no unique solution but the idea is that for a given task, the best sensor is the one with the highest sensitivity and the smallest measurement error, and the estimate precision can be further improved by increasing the sensor density of a sensor array (*viz.*, increasing m).

Let us now revisit the example considered in Section 3. In this situation, the estimator to minimize the Bayes risk under the squared error loss is the posterior mean vector e_i^{EOL} specified in Eq. (2) and the corresponding minimum Bayes risk is $R_B^{\min}(\mathbf{y}) = \sum_{i=1}^n \sigma_{X_i|\mathbf{y}}^2$, where $\sigma_{X_i|\mathbf{y}}^2$ is the posterior MSE or the posterior variance of X_i given by

$$\sigma_{X_i|\mathbf{y}}^2 = \sigma^2 \left\{ 1 - \frac{\mathbf{b}_i^\top \mathbf{y}}{\sigma} \frac{\phi(-\mathbf{b}_i^\top \mathbf{y}/\sigma)}{[1 - \Phi(-\mathbf{b}_i^\top \mathbf{y}/\sigma)]} - \left[\frac{\phi(-\mathbf{b}_i^\top \mathbf{y}/\sigma)}{[1 - \Phi(-\mathbf{b}_i^\top \mathbf{y}/\sigma)]} \right]^2 \right\}, \quad i = 1, 2, \dots, n. \quad (7)$$

Here, the minimum Bayes risk is equivalent to the trace of the variance-covariance matrix of \mathbf{X} given \mathbf{y} , and it can be understood as the A -optimality criterion of the statistical design of experiment. It provides an overall measure of the average posterior variance and gives the sum of the eigenvalues of the variance-covariance matrix. The expected Bayes risk is then defined by

$$\begin{aligned} E[R_B^{\min}(\mathbf{Y})] &= \int_{\mathbf{y}} R_B^{\min}(\mathbf{y}) f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y} \\ &= \int_{\mathbf{y}} \int_{\mathbf{x}} R_B^{\min}(\mathbf{y}) f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}; \boldsymbol{\theta}^*, \sigma^2) f_{\mathbf{X}}(\mathbf{x}; \psi) d\mathbf{x} d\mathbf{y}, \end{aligned} \quad (8)$$

where $f_{\mathbf{Y}}(\mathbf{y})$ is the marginal distribution of \mathbf{Y} as defined in the denominator of Eq. (1). The resulting expected risk in Eq. (8) is highly nonlinear but depends only on the *known* calibration parameters, thus a similar observation can be made as in the previous example. It can be understood how different tuning curves within a sensor array can affect the accuracy and precision of stimulus estimation. If one desires to construct a sensor array composed of n sensors selected from a population, then the set of sensors which minimizes the expected Bayes risk should be utilized (*i.e.*, sensors with high sensitivities and small measurement errors). It is noted that the same techniques described here can be applied to analyze any sensory system, including biological and artificial chemical olfactory systems, that exploits a population coding of the stimulus to optimize its performance. The expected Bayes risk provides a quantitative measure regarding the fundamental capability of a given measurement system to estimate the chemical mixture composition. This measure is also independent of the *post hoc* data analysis techniques used to estimate the analyte concentrations. Hence, it can be used to measure the *potential* analytical capability of a sensory

Table 1: Sensitivity values of 24 simulated linear sensors in [11] for estimating three potential analytes

Sensor Number	Analyte 1	Analyte 2	Analyte 3
1	0.0476	0.5864	0.7360
2	0.3488	0.6751	0.7947
3	0.4513	0.3610	0.5449
4	0.2409	0.6203	0.6862
5	0.7150	0.8112	0.8936
6	0.8562	0.0193	0.0548
7	0.2815	0.0839	0.3037
8	0.7311	0.9748	0.0462
9	0.1378	0.6513	0.1955
10	0.8367	0.2312	0.7202
11	0.1386	0.4035	0.7218
12	0.5882	0.1220	0.8778
13	0.3662	0.2684	0.5824
14	0.8068	0.2578	0.0707
15	0.5038	0.3317	0.9227
16	0.4896	0.1522	0.8004
17	0.8770	0.3480	0.2859
18	0.3531	0.1217	0.5437
19	0.4494	0.8842	0.9848
20	0.9635	0.0943	0.7157
21	0.0423	0.9300	0.8390
22	0.9730	0.3990	0.4333
23	0.1892	0.0474	0.4706
24	0.6671	0.3424	0.5607

system under consideration, rather than a direct prediction of what may be observed in practice.

5. Illustrative Examples

Let us consider the design optimization of a simulated sensor array described in [17]. It is desired to select three sensors from a pool of 125 possible sensor types encompassing every possible permutation of five different sensitivities $\{0, 0.25, 0.50, 0.75, 1\}$ for each of three analytes. This creates a total of $\binom{125}{3} = 317,750$ sensor array configurations. It is assumed that the sensor response is additive in the presence of multiple analytes with the concentration-independent measurement error distributed as standard normal (*viz.*, zero mean and unit variance). Further, a proper conjugate multivariate normal prior is assumed with the generalized variance tending to ∞ for the objective inference. Then, it can be shown that the posterior variance-covariance matrix of the analyte concentrations is obtained as $(\mathbf{B}^\top \mathbf{B})^{-1}$, where \mathbf{B} is an array configuration of the chosen sensors. The expected Bayes risk is then calculated as the expected sum of the diagonal entries of $(\mathbf{B}^\top \mathbf{B})^{-1}$. In such a system, the array configuration that minimizes the expected Bayes risk consists of three different sensors, each of which exhibits the minimum sensitivity of zero for one of the three analytes and the maximum sensitivity of one for the other two.

This configuration can be represented by a matrix of sensitivity values, given by

$$\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

where each column corresponds to a different analyte and each row corresponds to a different sensor. The expected Bayes risk of this array configuration is 2.250 with the marginal variance for each analyte being 0.750. Minimizing the expected Bayes risk to optimize the sensor array design assumes that one wishes to quantify all three analytes. If one wishes to quantify only a particular analyte, however, it is possible to find other array configurations that are optimal for that sensing task. Consider a scenario where quantification of analyte 1 only is of interest with the other two regarded as an interfering background. The array configuration which minimizes the marginal variance for analyte 1 is obtained as

$$\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0.25 \\ 1 & 0.25 & 0 \end{pmatrix}$$

with the marginal variance of 0.516. Compared to the array configuration with the minimum expected Bayes risk, this configuration improves the performance capability when the sensing task is to determine analyte 1. This 31.2% reduction in the marginal variance for analyte 1 comes at a substantial cost, however, since the marginal variances for the other two analytes drastically increase to 8.250 or 1100%. The expected Bayes risk of this array configuration is 17.016, which is considerably worse than the previous configuration. Nevertheless, this may not be an issue if precise quantification of the other analytes is not desired. For instance, a sensor array configured with three identical sensors, each specific to analyte 1 with the sensitivity of one, has a decreased variance of 0.333 via signal averaging but with no collateral capabilities to sense analytes 2 or 3. This clearly demonstrates that the effective sensor array optimization is highly contingent upon proper consideration of the analytical task to be addressed.

As pointed out in [11], the design of a sensory system is often significantly more constrained than the previous example. Sensors exhibiting specificity for a target analyte or analyte set are often not available, leaving a sensor array designer to select a subset of arrays from a pool of candidate sensors with considerable cross-sensitivities, especially as the set of potential analytes gets larger. Let us revisit the synthetic example considered in [11], where a sensor array is drawn from a pool of candidate sensors with linear response functions. The sensitivity values for each sensor-analyte pair were generated from a standard uniform distribution. Table 1 reproduces the sensitivity values of these 24 simulated linear sensors for each of three potential analytes.

For a sensor array with m sensors, there are $\binom{24}{m}$ possible array configurations, producing a total of nearly 17 million configurations with three or more sensors. In order to independently estimate three analytes from a mixture, a sensor array should consist of at least three sensors, and a total of 2,024 configurations is possible for this minimal-sized sensor array. The expected Bayes risk spans a wider range of orders of magnitude over different configurations. Table 2 details the three-sensor array configurations selected from the sensor library in Table 1 under different design criteria. The optimal array configuration which minimizes the expected Bayes risk consists of the sensors 8, 20, and 21, each exhibiting relatively low sensitivity to one analyte compared to the other two. Examining the marginal variance of each analyte, one can see that the optimal configuration for estimating analyte 1 only reduces the marginal variance for analyte 1 to less than three quarters of that

Table 2: Various design criteria for a three-sensor system and the corresponding expected Bayes risks, analyte-specific marginal variances along with the optimal array configurations drawn from the sensor library in Table 1

Optimal Sensor Array Design Criterion	Sensor Numbers	Expected Bayes Risk	Marginal Variances		
			Analyte 1	Analyte 2	Analyte 3
minimum expected Bayes risk	(8,20,21)	3.498	1.085	0.920	1.493
minimum variance for analyte 1	(6,17,21)	890.0	0.760	393.7	495.5
minimum variance for analyte 2	(8,20,21)	3.498	1.085	0.920	1.493
minimum variance for analyte 3	(1,8,15)	8.967	4.926	2.994	1.047

observed under the minimum expected Bayes risk. However, this produces much greater Bayes risk due to significant increases in error for analytes 2 and 3. Similarly, the optimal configuration for estimating analyte 3 only presents an improved marginal variance for analyte 3 at the expense of greater error for analytes 1 and 2. Since the optimal configuration for estimating analyte 2 only happens to minimize the expected Bayes risk, it exhibits the best compromise, should all analytes need to be estimated using the sensor array.

It is demonstrated that the expected Bayes risk of the optimal array configuration monotonically decreases with the array size, but by selecting larger array subsets, the expected Bayes risk associated with the worst performing array decreases much more rapidly than for the optimal configurations. Hence, if one chooses sensor array subsets at random, selecting larger arrays would present a significant risk mitigation against selecting a poor performing sensor array. When designing the optimal sensor arrays, this assessment would enable the quantitative cost versus performance trade off analysis regarding the array size. Other analytical tasks such as single-analyte estimation could be similarly assessed by first associating that analytical task with an orientation in the sample space and then calculating the marginal variance along that axis [11].

6. Conclusion

In this work, the estimation problem of the stimuli concentrations was revisited. Since it is actually not a typical regression problem, we developed a statistical method to analyze the chemical stimuli given the responses of cross-sensitive arrays, using the concept of Bayesian analysis. The dynamic and uncertain nature of the environment-dependent stimuli was elucidated via a choice of the prior distribution. Under this framework, the expected Bayes risk was proposed as a performance metric of a sensor array for a given sensing task, and it can be used to select an optimal combination of sensors when designing a sensor array. The proposed approach is generalizable and could be applied to other sensory systems for the stimuli estimation and/or the optimal sensory system designs. Future work will explore the general variance-covariance matrix of \mathbf{X} given \mathbf{y} as a metric to determine the merit of a sensory system and its connection to the chemometric concepts of net analyte signal, sensitivity and selectivity.

References

- [1] Alkasab, T.K., White, J., Kauer, J.S. (2002). A computational system for simulating and analyzing arrays of biological and artificial chemical sensors. *Chemical Senses*, **27**: 261–275.
- [2] Barsan, N., Stetter, J.R., Findlay, M., Gopel, W. (1999). High performance gas sensing of CO: comparative tests for semiconducting (SnO₂-based) and for amperometric gas sensors. *Analytical Chemistry*, **71**: 2512–2517.

- [3] Boeckh, J., Ernst, K.D. (1987). Contribution of single unit analysis in insects to an understanding of olfactory function. *Journal of Comparative Physiology A*, **161**: 549–565.
- [4] Boeker, P. (2014). On ‘electronic nose’ methodology. *Sensors and Actuators B: Chemical*, **204**: 2–17.
- [5] Depaoli, S. (2014). The impact of inaccurate informative priors for growth parameters in Bayesian growth mixture modeling. *Structural Equation Modeling*, **21**: 239–252.
- [6] Fonollosa, J., Fernandez, L., Huerta, R., Gutierrez-Galvez, A., Marco, S. (2013). Temperature optimization of metal oxide sensor arrays using mutual information. *Sensors and Actuators B: Chemical*, **187**: 331–339.
- [7] Gardner, J.W., Bartlett, P.N. (1996). Performance definition and standardization of electronic noses. *Sensors and Actuators B: Chemical*, **33**: 60–67.
- [8] Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, **1**: 515–534.
- [9] Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, **72**: 320–338.
- [10] Jackman, S. (2009). *Bayesian Analysis for the Social Sciences*. Wiley & Sons, New York, NY.
- [11] Johnson, K.J., Knapp, A. (2017). Selectivity measure for arrays of non-specific sensors. *Sensors and Actuators B: Chemical*, **251**: 1076–1088.
- [12] Kenward, M.G., Roger, J.H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, **53**: 983–997.
- [13] Kenward, M.G., Roger, J.H. (2009). An improved approximation to the precision of fixed effects from restricted maximum likelihood. *Computational Statistics and Data Analysis*, **53**: 2583–2595.
- [14] Levy, R. (2009). The rise of Markov chain Monte Carlo estimation for psychometric modeling. *Journal of Probability and Statistics*, 1–18.
- [15] Muthen, B., Asparouhov, T. (2012). Bayesian structural equation modeling: a more flexible representation of substantive theory. *Psychological Methods*, **17**: 313–335.
- [16] O’Malley, A.J., Zaslavsky, A.M. (2005). Variance-covariance functions for domain means of ordinal survey items. *Survey Methodology*, **31**: 169–182.
- [17] Pearce, T.C., Sanchez-Montanes, M.A. (2004). Chemical sensor array optimization: geometric and information theoretic approaches, in: Pearce, T.C., Schiffman, S.S., Nagle, H.T., Gardner, J.W. (eds), *Handbook of Machine Olfaction: Electronic Nose Technology*. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, FRG, pp.347–375.
- [18] Sanchez-Montanes, M.A., Pearce, T.C. (2001). Fisher information and optimal odor sensors. *Neurocomputing*, **38-40**: 335–341.
- [19] Skene, S.S., Kenward, M.G. (2010a). The analysis of very small samples of repeated measurements I: an adjusted sandwich estimator. *Statistics in Medicine*, **29**: 2825–2837.
- [20] Skene, S.S., Kenward, M.G. (2010b). The analysis of very small samples of repeated measurements II: a modified Box correction. *Statistics in Medicine*, **29**: 2838–2856.
- [21] van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J.B., Neyer, F.J., Aken, M.A. (2014). A gentle introduction to Bayesian analysis: applications to developmental research. *Child Development*, **85**: 842–860.