Article

Comparison Between Bayesian and Maximum Entropy Analyses of Flow Networks [†]

Steven H. Waldrip and Robert K. Niven *

School of Engineering and Information Technology, The University of New South Wales, Canberra, ACT 2600, Australia; Steven.Waldrip@student.adfa.edu.au

*   Correspondence: r.niven@adfa.edu.au; Tel.: +61-2-6268-8330

†   This paper is an extended version of our paper published in the 3rd International Electronic and Flipped Conference on Entropy and Its Applications, 1–10 November 2016.

Abstract: We compare the application of Bayesian inference and the maximum entropy (MaxEnt) method for the analysis of flow networks, such as water, electrical and transport networks. The two methods have the advantage of allowing a probabilistic prediction of flow rates and other variables, when there is insufficient information to obtain a deterministic solution, and also allow the effects of uncertainty to be included. Both methods of inference update a prior to a posterior probability density function (pdf) by the inclusion of new information, in the form of data or constraints. The MaxEnt method maximises an entropy function subject to constraints, using the method of Lagrange multipliers, to give the posterior, while the Bayesian method finds its posterior by multiplying the prior with likelihood functions incorporating the measured data. In this study, we examine MaxEnt using soft constraints, either included in the prior or as probabilistic constraints, in addition to standard moment constraints. We show that when the prior is Gaussian, both Bayesian inference and the MaxEnt method with soft prior constraints give the same posterior means, but their covariances are different. In the Bayesian method, the interactions between variables are applied through the likelihood function, using second or higher-order cross-terms within the posterior pdf. In contrast, the MaxEnt method incorporates interactions between variables using Lagrange multipliers, avoiding second-order correlation terms in the posterior covariance. The MaxEnt method with soft prior constraints, therefore, has a numerical advantage over Bayesian inference, in that the covariance terms are avoided in its integrations. The second MaxEnt method with soft probabilistic constraints is shown to give posterior means of similar, but not identical, structure to the other two methods, due to its different formulation.

Keywords: maximum entropy analysis; Bayesian inference; probability; flows; networks

1. Introduction

The analysis of flow rates on networks is required for the design and monitoring of electrical, water, sewer, irrigation, fire suppression, drainage, oil, gas and any other networks through which fluids or energy are transported. Their analysis is an important engineering problem. Traditionally, these systems have been analysed using deterministic methods. These methods incorporate physical laws such as Kirchhoff's first and second laws (conservation of mass and equivalence of potentials at nodes) and sufficient known parameter values, giving a closed-form set of equations which is solved for the (deterministic) solution. Deterministic methods yield precise parameter values but do not consider uncertainty, either due to a lack of knowledge of the state of the system or flow variability. To account for uncertainty, a probabilistic framework is required. There are two primary

methods for probabilistic inference: Bayesian inference using Bayes' rule, and maximum entropy (MaxEnt) analysis.

Bayes' theorem comes from the product rule of probabilities. To use Bayes' theorem, the prior and likelihood functions need to be chosen before the data are analysed. To analyse the data, a set of data values are incorporated in the likelihood function, which is then updated by Bayes' rule to obtain the posterior. This process can be repeated for each data set by using the posterior as the prior for the next data set. The order in which the data sets are analysed should not impact the final result.

The analysis presented here follows the well-known Bayesian parameter estimation or regression procedure as described in [1]. Although the applications of this procedure are extremely vast, including least-squares regression, the authors are not aware of it being applied in its general form to estimate flows on a network. An example of using Bayes' theorem with transient pipe flows is presented in Rougier and Goldstein [2] who solve the water hammer partial differential equations, incorporating uncertainty from the pipeline characteristics and the boundary conditions. Bayes' theorem is used to estimate the flows, pressures and pipeline characteristics as time progresses, using data obtained through real-time monitoring of the pipeline in a few locations. As this method requires the solution of a partial differential equation which incorporates time and uncertainty, its computational cost is high and therefore is restricted to small networks; in the example, a single pipe is analysed. The Bayesian method can also be used to calibrate model parameters, often using least-squares regression. Savic et al. [3] provide a comprehensive review of calibration techniques used with water networks. As an alternative to predicting model parameters, Hutton et al. [4] use Bayes' theorem to update the coefficients in an autoregressive, data-driven model to predict future flow rates (at two locations in their example) using current and previous flow rate observations. In their case study, they were able to provide accurate one-hour forecasts for the monitored locations. Hutton and Kapelan [5] extended their previous analysis to predict pipe bursts by considering the difference between their predicted and observed flow rates. They were able to detect abnormal flow conditions representing pipe bursts greater than 5% of normal flow conditions.

Entropy is a measure of uncertainty [6–10]. The MaxEnt method for inference can be derived from an axiomatic approach based on the axioms of locality, coordinate invariance and subsystem independence [11–13]. Alternatively the MaxEnt method can be derived from a combinatorial approach [14–17], which shows that the MaxEnt method infers the most probable distribution, subject to the constraints and prior [6,14–20]. The maximum relative entropy method (MaxEnt), equivalent to the minimum Kullback–Leibler divergence [21], is a method of inference used to infer or update a probability distribution describing an under-determined system, which respects all constraints imposed on the system and is closest to the prior distribution [8]. However, the MaxEnt method is a method of inference, with no guarantee that the inferred solution will be realised [6,8,9]. The validity of the distribution will depend on the assumptions used to construct the MaxEnt model. The MaxEnt method has been applied to predict the flows on water distribution networks [22–25], transportation networks [26], electrical networks [26] and generic flow networks [27–29].

There have been many studies on the connection between Bayes' theorem and the MaxEnt method, with some authors suggesting that one can be obtained from the other (in either direction) [30–33]. Giffin and Caticha's method [31–33] to obtain Bayes' rule using MaxEnt requires the relative entropy function to be defined over the model parameters and the data. The normalisation constraint is applied, and the variables representing the data are constrained with a pdf (probability density function) representing the Bayesian likelihood function. The pdf constraint is applied to the parameters defined to be the prior in the Bayesian method. Bayes' rule is obtained by dividing the inferred distribution by the pdf defined over the data parameters, i.e., dividing by the pdf constraint. Although this equivalence is mathematically correct, the second constraint appears somewhat contrived as it is applied over the parameters of the Bayesian prior.

The current authors have compared the probability distributions of quasi-Newton rules obtained when inferring the Jacobian or Hessian using Bayesian inference [34,35] and the MaxEnt method [36].

In both methods, the same Gaussian prior was used. In the MaxEnt method, secant equations were used as the constraints. In Bayes' method, delta likelihood functions incorporating the secant equation were used to represent the data. It was found that both methods obtained the same posterior means, but the covariance matrices were found to be different.

In this study, we develop a Bayesian method to analyse flow networks (Section 2). This theory contains many features in common with the MaxEnt method of [25]. In Section 3.1, we present a MaxEnt theory using soft constraints that are implemented in the prior pdf. In Section 3.2, we compare the distributions obtained by the two methods. In Section 4.1, we also present a MaxEnt theory with soft constraints implemented using pdfs as constraints, and, in Section 4.2, we compare this to the Bayesian method. Finally, in Section 5, we discuss our findings.

2. Bayesian Analysis

Consider a flow network with N external flow rates and M internal flow rates, which can be assembled into the vectors $\boldsymbol{\Theta}$ and $\boldsymbol{Q}$, respectively, which, in turn, can be assembled into the vector

$$\boldsymbol{\Psi} = \begin{bmatrix} \boldsymbol{\Theta} \\ \boldsymbol{Q} \end{bmatrix}, \tag{1}$$

In the Bayesian method, to avoid inconsistencies due to different network representations, we consider a basis set X of n flow rates selected from $\boldsymbol{\Psi}$ as parameters of the pdf used to represent the uncertainty. The indices of the basis set X in $\boldsymbol{\Psi}$ are given by the set $\mathcal{B}$, while the indices of the complementary non-basis set of flow rates in $\boldsymbol{\Psi}$ are given by $\mathcal{N}$. For closure, at least $N-1$ basis flow rates must be chosen but up to $N+M$ can be chosen. The derivation of the Bayesian method requires a prior belief of the state of the system, represented as a prior pdf $q(X)$, which is updated using observed data to a posterior pdf according to Bayes' rule:

$$p(X|y) = \frac{p(y|X)q(X)}{\underset{\Omega}{\int \cdots \int} p(y|X)q(X)dX}, \tag{2}$$

where $p(y|X)$ is the likelihood function, the denominator allows for normalisation, X is the basis set of flow rates, y is the vector of observed data and Ω is the domain of X. The flow rates $\bar{X}$ not included in the basis set are taken as functions of the model parameters X, using:

$$\bar{X} = VX, \tag{3}$$

where

$$V = -A_{i\in\mathcal{V},j\in\mathcal{N}}^{-1}A_{i\in\mathcal{V},j\in\mathcal{B}}, \tag{4}$$

$$A = \begin{bmatrix} C \\ W\text{diag}(K) \\ F \\ T\text{diag}(K) \end{bmatrix}, \tag{5}$$

in which

- diag() places the elements of a vector on the diagonal of a square matrix of zeros;
- the set $\mathcal{V}$ contains the $N+M-n$ indices of the equations required to uniquely define $\bar{X}$ from X;
- the matrix C is an $N \times (N+M)$ connectivity matrix containing elements $\{-1,0,1\}$. Its entries $C_{i,r}, \forall i \in \{1,...N\}$ indicate membership of edge r to the node i, given by 0 if the edge is not connected to the node, 1 if the assumed direction of Q_m or Θ_i is entering the node and -1 otherwise;
- the vector K is an $(M+N) \times 1$ vector of flow resistances;

- the matrix W is a $w \times (N + M)$ loop matrix containing elements $\{-1, 0, 1\}$, where w is the number of independent cycles (loops) within the network. Its entries $W_{i,r}, \forall i \in \{1, ..., w\}$ indicate membership of edge r within loop i, given by 0 if the edge is not in the loop, 1 if the assumed direction of Q_m is in a clockwise direction around the loop and -1 otherwise;
- the matrix F is a $N_{\hat{\Theta}} + N_{\hat{Q}} \times N + M$ matrix containing either 0 or 1 in each of its elements. Each row will have a single 1 on the index corresponding to the dimension of the observed link, with the remaining elements set to 0;
- $N_{\hat{\Theta}}$ and $N_{\hat{Q}}$ are the numbers of flow rate observation locations for flows entering/exiting or within the network respectively; and
- the matrix T is a $h_c \times (M + N)$ pseudo-loop matrix containing $\{-1, 0, 1\}$, where h_c is the number of potential difference constraints applied. The pseudo-loop matrix contains paths between nodes of known pressure or potential values. For convenience, these are referenced to the potential at a single reference node H_0; this gives $\langle Y_T \rangle$ as the $h_c \times 1$ vector of mean potential differences between H_0 and H_j, for all nodes with potential observations. The entries in $T_{i,r}, \forall i \in \{1, ..., h_c\}$ indicate membership of edge r within the potential difference constraint index i, given by 0 if the edge is not in the constraint, 1 if the assumed direction of Q_m is defined as in the direction from node 0 to node j, and -1 otherwise.

The prior is chosen to represent one's belief of the system state before incorporating any measured data. Although any distribution which represents what is believed about the system state could be chosen, in this study, a multidimensional Gaussian distribution is selected, defined over the real domain:

$$q(X) = \frac{\exp\left(-\frac{1}{2} (X - m)^\top \Sigma^{-1} (X - m)\right)}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}}, \tag{6}$$

where m is the $n \times 1$ vector of prior mean flow rates and Σ is the $n \times n$ matrix of prior covariances.

In Bayes' method, likelihood functions are used to incorporate the physics of the system as well as any observed data, as follows:

- The likelihood function to incorporate conservation of mass at each node or Kirchhoff's first law (or the flow rate for incompressible systems) is given by a delta function

$$p(0|X) = \delta \left(0 - (C_X + C_{\bar{X}}) X\right), \tag{7}$$

where

$$C_X = C_{i \notin \mathcal{V}, j \in \mathcal{B}}, \tag{8}$$

$$C_{\bar{X}} = C_{i \notin \mathcal{V}, j \in \mathcal{N}} V. \tag{9}$$

This delta function is defined by the limit of a Gaussian distribution

$$-2 \ln(p(0|X)) \propto \lim_{\Sigma_C \to 0} \left(0 - (C_X + C_{\bar{X}}) X\right)^\top \Sigma_C^{-1} \left(0 - (C_X + C_{\bar{X}}) X\right). \tag{10}$$

- The likelihood function to incorporate the loop laws for each loop, Kirchhoff's second law, is given by a delta function

$$p(0|X) = \delta \left(0 - (W_X + W_{\bar{X}}) X\right), \tag{11}$$

where

$$W_X = W_{i \notin \mathcal{V}, j \in \mathcal{B}} \operatorname{diag}(K_{i \in \mathcal{B}}), \tag{12}$$

$$W_{\bar{X}} = W_{i \notin \mathcal{V}, j \in \mathcal{N}} \operatorname{diag}(K_{i \in \mathcal{N}}) V. \tag{13}$$

This delta function is defined by the limit of the Gaussian distribution

$$-2 \ln(p(0|X)) \propto \lim_{\Sigma_W \to 0} \left(0 - (W_X + W_{\bar{X}}) X\right)^\top \Sigma_W^{-1} \left(0 - (W_X + W_{\bar{X}}) X\right). \tag{14}$$

- Observed flow rates can be constrained using the likelihood function

$$-2\ln(p(Y_F|X)) \propto (Y_F - (F_X + F_{\bar{X}})X)^\top \Sigma_F^{-1} (Y_F - (F_X + F_{\bar{X}})X),\tag{15}$$

where Y_F is a $N_{\hat{\Theta}} + N_{\hat{Q}} \times 1$ vector that has the flow rate of each observation for a link in its elements, Σ_F is the $N_{\hat{\Theta}} + N_{\hat{Q}} \times N_{\hat{\Theta}} + N_{\hat{Q}}$ covariance matrix of the observations and

$$F_X = F_{i\notin\mathcal{V},j\in\mathcal{B}},\tag{16}$$

$$F_{\bar{X}} = F_{i\notin\mathcal{V},j\in\mathcal{N}}V.\tag{17}$$

- Observed potential differences can be constrained with the likelihood function

$$-2\ln(p(Y_T|X)) \propto (Y_T - (T_X + T_{\bar{X}})X)^\top \Sigma_T^{-1} (Y_T - (T_X + T_{\bar{X}})X),\tag{18}$$

where Y_T is a $h_c \times 1$ vector that has the potential difference of an observation between two points in each element, Σ_T is the $h_c \times h_c$ covariance matrix of the observations and

$$T_X = T_{i\notin\mathcal{V},j\in\mathcal{B}}\mathrm{diag}\,(K_{i\in\mathcal{B}}),\tag{19}$$

$$T_{\bar{X}} = T_{i\notin\mathcal{V},j\in\mathcal{N}}\mathrm{diag}\,(K_{i\in\mathcal{N}})\,V.\tag{20}$$

Applying Bayes' rule with each of the likelihood functions, and expanding and dropping all terms which are not functions of X gives the posterior in the form

$$-2\ln(p(X|y)) \propto X^\top\Sigma^{-1}X - X^\top\Sigma^{-1}m - m^\top\Sigma^{-1}X + X^\top O^\top S^{-1}OX - y^\top S^{-1}OX - X^\top O^\top S^{-1}y,\tag{21}$$

where

$$O = \begin{bmatrix} C_X + C_{\bar{X}} \\ W_X + W_{\bar{X}} \\ F_X + F_{\bar{X}} \\ T_X + T_{\bar{X}} \end{bmatrix},\tag{22}$$

$$S^{-1} = \begin{bmatrix} \Sigma_C^{-1} & 0 & 0 & 0 \\ 0 & \Sigma_W^{-1} & 0 & 0 \\ 0 & 0 & \Sigma_F^{-1} & 0 \\ 0 & 0 & 0 & \Sigma_T^{-1} \end{bmatrix},\tag{23}$$

$$y = \begin{bmatrix} 0 \\ 0 \\ Y_F \\ Y_T \end{bmatrix}.\tag{24}$$

Combining like factors gives

$$-2\ln(p(X|y)) \propto X^\top\left(\Sigma^{-1} + O^\top S^{-1}O\right)X - X^\top\left(\Sigma^{-1}m + O^\top S^{-1}y\right) - \left(m^\top\Sigma^{-1} + y^\top S^{-1}O\right)X.\tag{25}$$

Completing the square gives

$$-2\ln(p(X|y)) \propto (X - \langle X\rangle)^\top \Sigma_p^{-1} (X - \langle X\rangle),\tag{26}$$

where the mean flow rates and variance matrix are given by

$$\langle X\rangle = \Sigma_p\left(\Sigma^{-1}m + O^\top S^{-1}y\right),\tag{27}$$

$$\Sigma_p = \left(\Sigma^{-1} + O^\top S^{-1} O \right)^{-1}. \tag{28}$$

Using the Woodbury matrix identity [37] to find the posterior covariance gives

$$\Sigma_p = \Sigma - \Sigma O^\top \left(S + O\Sigma O^\top \right)^{-1} O\Sigma. \tag{29}$$

The following algebra is needed to find a form which does not require an inversion of a zero matrix arising from the delta functions. Right multiplying the inverse posterior covariance by $\Sigma O^\top$ gives

$$\Sigma_p^{-1} \Sigma O^\top = O^\top + O^\top S^{-1} O\Sigma O^\top = O^\top S^{-1} \left(S + O\Sigma O^\top \right), \tag{30}$$

then left multiplying with the posterior covariance

$$\Sigma O^\top = \Sigma_p O^\top S^{-1} \left(S + O\Sigma O^\top \right). \tag{31}$$

Extracting $\Sigma_p O^\top S^{-1}$ by right multiplying by $\left(S + O\Sigma O^\top \right)^{-1}$ gives

$$\Sigma O^\top \left(S + O\Sigma O^\top \right)^{-1} = \Sigma_p O^\top S^{-1}. \tag{32}$$

The posterior mean flow rates can then be found from Equation (27) by substituting Equation (29) and Equation (32) to give

$$\langle X \rangle = m + \Sigma O^\top \left(S + O\Sigma O^\top \right)^{-1} \left(y - Om \right). \tag{33}$$

3. MaxEnt Analysis with Soft Constraints Implemented in the Prior

3.1. Formulation

The maximum entropy method is defined by the following algorithm [6,9]: (i) define a probability measure over the uncertainties of interest; (ii) construct a relative entropy function; (iii) define a prior probability function and constraints; (iv) maximise the entropy subject to the constraints and prior, to infer the probability distribution which describes the system; and, if desired, (v) extract statistical moments of quantities of interest. Soft MaxEnt constraints have previously been suggested by the authors [24,25] but have not been formally derived. To implement soft constraints, we define a pdf which expresses the uncertainty in the system defined over a reduced parameter set X, consisting of a basis set of n flow rates selected from Ψ and also the parameter observations Y_F and Y_T. The indices of X in Ψ are again given by the set $\mathcal{B}$. Again for closure, at least $N - 1$ basis flow rates must be chosen but up to $N + M$ can be chosen. The joint probability is defined to be:

$$p(X)dX = \text{Prob}(X \leq Y_X \leq X + dX, Y_F \leq Y_{Y_F} \leq Y_F + dY_F, Y_T \leq Y_{Y_T} \leq Y_T + dY_T), \tag{34}$$

where Y_X, Y_{Y_F} and Y_{Y_T} are the vectors of the random variables for X, Y_F and Y_T, respectively. We also assume that each of the flow rate and potential difference constraints are applied as soft constraints, but this does not restrict this method to only use soft constraints, and strict constraints defined by expectations can still be applied. This choice of pdf gives the following relative entropy or negative Kullback–Leibler function [21], over the space of uncertainties used in this formulation:

$$\mathfrak{H} = - \int_{l_1}^{u_1} \cdots \int_{l_{n+n_o}}^{u_{n+n_o}} p(X, Y_F, Y_T) \ln \frac{p(X, Y_F, Y_T)}{q(X, Y_F, Y_T)} dX dY_F dY_T, \tag{35}$$

where $n_o = N_{\hat{\Theta}} + N_{\hat{Q}} + h_c$, the number of data observations, $q(X, Y_F, Y_T)$ is the prior pdf, and l_i and u_i are the lower and upper bounds of the *i*th flow rate. The relative entropy is then maximised subject to the constraints on the system. The following constraints are always required:

- Normalisation of probability:

$$1 = \int_{l_1}^{u_1} \cdots \int_{l_{n+n_o}}^{u_{n+n_o}} p(X, Y_F, Y_T) dX dY_F dY_T. \tag{36}$$

- Kirchhoff's first law, for the conservation of flow rates at each internal node, here imposed in the mean:

$$0 = (C_X + C_{\bar{X}}) \left(\int_{l_1}^{u_1} \cdots \int_{l_{n+n_o}}^{u_{n+n_o}} X p(X, Y_F, Y_T) dX dY_F dY_T \right). \tag{37}$$

- Kirchhoff's second law, which requires the potential difference to vanish around each enclosed loop, again imposed in the mean:

$$0 = (W_X + W_{\bar{X}}) \left(\int_{l_1}^{u_1} \cdots \int_{l_{n+n_o}}^{u_{n+n_o}} X p(X, Y_F, Y_T) dX dY_F dY_T \right). \tag{38}$$

We also allow for any of the following constraints:

- A set of specified inflow/outflow and internal flow rate constraints:

$$0 = \int_{l_1}^{u_1} \cdots \int_{l_{n+n_o}}^{u_{n+n_o}} \left((F_X + F_{\bar{X}}) X - Y_F \right) p(X, Y_F, Y_T) dX dY_F dY_T. \tag{39}$$

- Potential difference constraints between pairs of nodes:

$$0 = \int_{l_1}^{u_1} \cdots \int_{l_{n+n_o}}^{u_{n+n_o}} \left((T_X + T_{\bar{X}}) X - Y_T \right) p(X, Y_F, Y_T) dX dY_F dY_T. \tag{40}$$

After identifying the constraints, the entropy (35) is then maximised subject to Equations (36)–(38) and whichever of Equations (39) and (40) apply. Applying the calculus of variations, we form the Lagrangian:

$$
\begin{aligned}
\mathfrak{L} = & -\int_{l_1}^{u_1} \cdots \int_{l_{n+n_o}}^{u_{n+n_o}} p(X, Y_F, Y_T) \ln \frac{p(X, Y_F, Y_T)}{q(X, Y_F, Y_T)} dX dY_F dY_T - \hat{\kappa} \left(\int_{l_1}^{u_1} \cdots \int_{l_{n+n_o}}^{u_{n+n_o}} p(X, Y_F, Y_T) dX dY_F dY_T - 1 \right) \\
& - \alpha \left((C_X + C_{\bar{X}}) \left(\int_{l_1}^{u_1} \cdots \int_{l_{n+n_o}}^{u_{n+n_o}} X p(X, Y_F, Y_T) dX dY_F dY_T \right) \right) \\
& - \beta \left((W_X + W_{\bar{X}}) \left(\int_{l_1}^{u_1} \cdots \int_{l_{n+n_o}}^{u_{n+n_o}} X p(X, Y_F, Y_T) dX dY_F dY_T \right) \right) \\
& - \lambda \left(\int_{l_1}^{u_1} \cdots \int_{l_{n+n_o}}^{u_{n+n_o}} \left((F_X + F_{\bar{X}}) X - Y_F \right) p(X, Y_F, Y_T) dX dY_F dY_T \right) \\
& - \eta \left(\int_{l_1}^{u_1} \cdots \int_{l_{n+n_o}}^{u_{n+n_o}} \left((T_X + T_{\bar{X}}) X - Y_T \right) p(X, Y_F, Y_T) dX dY_F dY_T \right),
\end{aligned} \tag{41}
$$

where $\hat{\kappa}$, (scalar) α, β, λ and η (row vectors) are the Lagrange multipliers for the normalisation, Kirchhoff's first and second laws, flow rates and the head loss constraints, respectively. The variation of $\mathfrak{L}$ is given by $\delta \mathfrak{L} = 0$. Extremizing Equation (41) by taking the functional derivative with respect to $p(X)$ and combining integrals gives:

$$
\begin{aligned}
\delta \mathfrak{L} = 0 = \int_{l_1}^{u_1} \cdots \int_{l_{n+n_o}}^{u_{n+n_o}} \Big[& -\ln \frac{p(X, Y_F, Y_T)}{q(X, Y_F, Y_T)} - \kappa - \alpha (C_X + C_{\bar{X}}) X - \beta (W_X + W_{\bar{X}}) X \\
& -\lambda \left((F_X + F_{\bar{X}}) X - Y_F \right) - \eta \left((T_X + T_{\bar{X}}) X - Y_T \right) \Big] dX,
\end{aligned} \tag{42}
$$

where $\kappa = \hat{\kappa} + 1$. Rearrangement gives the following solution for $p(X)$ (the Boltzmann distribution):

$$p^*(X, Y_F, Y_T) = q(X, Y_F, Y_T)e^{-\kappa - \alpha(C_X + C_{\bar{X}})X - \beta(W_X + W_{\bar{X}})X - \lambda((F_X + F_{\bar{X}})X - Y_F) - \eta((T_X + T_{\bar{X}})X - Y_T)}. \tag{43}$$

This can be solved, in conjunction with the constraints (36)–(40), to give $p^*(X, Y_F, Y_T)$ and the Lagrange multipliers κ, α, β, λ and η.

3.2. Solution and Comparison to Bayesian Solution

In the MaxEnt method, we choose the prior pdf as the multidimensional Gaussian

$$q(X, Y_F, Y_T) \propto \exp\left(-\frac{1}{2}\left(\begin{bmatrix} X \\ Y_F \\ Y_T \end{bmatrix} - \begin{bmatrix} m \\ m_F \\ m_T \end{bmatrix}\right)^\top \begin{bmatrix} \Sigma^{-1} & 0 & 0 \\ 0 & \Sigma_F^{-1} & 0 \\ 0 & 0 & \Sigma_T^{-1} \end{bmatrix} \left(\begin{bmatrix} X \\ Y_F \\ Y_T \end{bmatrix} - \begin{bmatrix} m \\ m_F \\ m_T \end{bmatrix}\right)\right), \tag{44}$$

where m ($n \times 1$ vector) and Σ ($n \times n$ matrix) are the mean and covariance of the prior flow rates within the entropy function, m_F ($N_{\hat{\Theta}} + N_{\hat{Q}} \times 1$ vector) and m_T ($h_c \times 1$ vector) are the values of the observations of the flow rate and potential differences, respectively, and Σ_F ($N_{\hat{\Theta}} + N_{\hat{Q}} \times N_{\hat{\Theta}} + N_{\hat{Q}}$ matrix) and Σ_F ($h_c \times h_c$ matrix) are their respective covariances. The resulting MaxEnt pdf with normalisation, Kirchhoff's first and second law, potential difference and flow rate constraints is proportional to

$$\begin{aligned}
\ln(p^*(X, Y_F, Y_T)) \propto -\frac{1}{2}&\left(\begin{bmatrix} X \\ Y_F \\ Y_T \end{bmatrix}^\top \begin{bmatrix} \Sigma^{-1} & 0 & 0 \\ 0 & \Sigma_F^{-1} & 0 \\ 0 & 0 & \Sigma_T^{-1} \end{bmatrix} \begin{bmatrix} X \\ Y_F \\ Y_T \end{bmatrix} - \begin{bmatrix} X \\ Y_F \\ Y_T \end{bmatrix}^\top \begin{bmatrix} \Sigma^{-1} & 0 & 0 \\ 0 & \Sigma_F^{-1} & 0 \\ 0 & 0 & \Sigma_T^{-1} \end{bmatrix} \begin{bmatrix} m \\ m_F \\ m_T \end{bmatrix}\right. \\
&\left. - \begin{bmatrix} m \\ m_F \\ m_T \end{bmatrix}^\top \begin{bmatrix} \Sigma^{-1} & 0 & 0 \\ 0 & \Sigma_F^{-1} & 0 \\ 0 & 0 & \Sigma_T^{-1} \end{bmatrix} \begin{bmatrix} X \\ Y_F \\ Y_T \end{bmatrix}\right) - \gamma^\top\left(\begin{bmatrix} C_X + C_{\bar{X}} & 0 & 0 \\ W_X + W_{\bar{X}} & 0 & 0 \\ F_X + F_{\bar{X}} & -I & 0 \\ T_X + T_{\bar{X}} & 0 & -I \end{bmatrix} \begin{bmatrix} X \\ Y_F \\ Y_T \end{bmatrix}\right),
\end{aligned} \tag{45}$$

where $\gamma = \begin{bmatrix} \alpha & \beta & \lambda & \eta \end{bmatrix}^\top$. Combining terms of the same order and assuming the covariance is symmetric and positive definite

$$\begin{aligned}
\ln(p^*(X, Y_F, Y_T)) \propto -\frac{1}{2}&\begin{bmatrix} X \\ Y_F \\ Y_T \end{bmatrix}^\top \begin{bmatrix} \Sigma^{-1} & 0 & 0 \\ 0 & \Sigma_F^{-1} & 0 \\ 0 & 0 & \Sigma_T^{-1} \end{bmatrix} \begin{bmatrix} X \\ Y_F \\ Y_T \end{bmatrix} - \left(\begin{bmatrix} m \\ m_F \\ m_T \end{bmatrix}^\top \begin{bmatrix} \Sigma^{-1} & 0 & 0 \\ 0 & \Sigma_F^{-1} & 0 \\ 0 & 0 & \Sigma_T^{-1} \end{bmatrix}\right. \\
&\left. + \gamma^\top \begin{bmatrix} C_X + C_{\bar{X}} & 0 & 0 \\ W_X + W_{\bar{X}} & 0 & 0 \\ F_X + F_{\bar{X}} & -I & 0 \\ T_X + T_{\bar{X}} & 0 & -I \end{bmatrix}\right) \begin{bmatrix} X \\ Y_F \\ Y_T \end{bmatrix}
\end{aligned} \tag{46}$$

and completing the square

$$
\begin{aligned}
\ln(p^*(X, Y_F, Y_T)) \propto -\frac{1}{2} & \left(\begin{bmatrix} X \\ Y_F \\ Y_T \end{bmatrix} - \begin{bmatrix} m \\ m_F \\ m_T \end{bmatrix} + \begin{bmatrix} \Sigma & 0 & 0 \\ 0 & \Sigma_F & 0 \\ 0 & 0 & \Sigma_T \end{bmatrix} \begin{bmatrix} C_X + C_{\bar{X}} & 0 & 0 \\ W_X + W_{\bar{X}} & 0 & 0 \\ F_X + F_{\bar{X}} & -I & 0 \\ T_X + T_{\bar{X}} & 0 & -I \end{bmatrix}^\top \gamma \right)^\top \\
\times \Sigma^{-1} & \left(\begin{bmatrix} X \\ Y_F \\ Y_T \end{bmatrix} - \begin{bmatrix} m \\ m_F \\ m_T \end{bmatrix} + \begin{bmatrix} \Sigma & 0 & 0 \\ 0 & \Sigma_F & 0 \\ 0 & 0 & \Sigma_T \end{bmatrix} \begin{bmatrix} C_X + C_{\bar{X}} & 0 & 0 \\ W_X + W_{\bar{X}} & 0 & 0 \\ F_X + F_{\bar{X}} & -I & 0 \\ T_X + T_{\bar{X}} & 0 & -I \end{bmatrix}^\top \gamma \right).
\end{aligned}
\tag{47}
$$

The above form allows the mean to be obtained as

$$
\begin{bmatrix} \langle X \rangle \\ \langle Y_F \rangle \\ \langle Y_T \rangle \end{bmatrix} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \begin{bmatrix} X \\ Y_F \\ Y_T \end{bmatrix} p(X, Y_F, Y_T) dX dY_F dY_T = \begin{bmatrix} m \\ m_F \\ m_T \end{bmatrix} - \begin{bmatrix} \Sigma & 0 & 0 \\ 0 & \Sigma_F & 0 \\ 0 & 0 & \Sigma_T \end{bmatrix} \begin{bmatrix} C_X + C_{\bar{X}} & 0 & 0 \\ W_X + W_{\bar{X}} & 0 & 0 \\ F_X + F_{\bar{X}} & -I & 0 \\ T_X + T_{\bar{X}} & 0 & -I \end{bmatrix}^\top \gamma.
\tag{48}
$$

Using the constraint equations, the Lagrange multipliers can be found from

$$
\gamma = \left(\begin{bmatrix} C_X + C_{\bar{X}} & 0 & 0 \\ W_X + W_{\bar{X}} & 0 & 0 \\ F_X + F_{\bar{X}} & -I & 0 \\ T_X + T_{\bar{X}} & 0 & -I \end{bmatrix} \begin{bmatrix} \Sigma & 0 & 0 \\ 0 & \Sigma_F & 0 \\ 0 & 0 & \Sigma_T \end{bmatrix} \begin{bmatrix} C_X + C_{\bar{X}} & 0 & 0 \\ W_X + W_{\bar{X}} & 0 & 0 \\ F_X + F_{\bar{X}} & -I & 0 \\ T_X + T_{\bar{X}} & 0 & -I \end{bmatrix}^\top \right)^{-1} \left(\begin{bmatrix} C_X + C_{\bar{X}} & 0 & 0 \\ W_X + W_{\bar{X}} & 0 & 0 \\ F_X + F_{\bar{X}} & -I & 0 \\ T_X + T_{\bar{X}} & 0 & -I \end{bmatrix} \begin{bmatrix} m \\ m_F \\ m_T \end{bmatrix} \right).
\tag{49}
$$

Substituting Equation (49) into Equation (48) gives

$$
\begin{aligned}
\begin{bmatrix} \langle X \rangle \\ \langle Y_F \rangle \\ \langle Y_T \rangle \end{bmatrix} &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \begin{bmatrix} X \\ Y_F \\ Y_T \end{bmatrix} p(X, Y_F, Y_T) dX dY_F dY_T = \begin{bmatrix} m \\ m_F \\ m_T \end{bmatrix} - \begin{bmatrix} \Sigma & 0 & 0 \\ 0 & \Sigma_F & 0 \\ 0 & 0 & \Sigma_T \end{bmatrix} \begin{bmatrix} C_X + C_{\bar{X}} & 0 & 0 \\ W_X + W_{\bar{X}} & 0 & 0 \\ F_X + F_{\bar{X}} & -I & 0 \\ T_X + T_{\bar{X}} & 0 & -I \end{bmatrix}^\top \\
\times & \left(\begin{bmatrix} C_X + C_{\bar{X}} & 0 & 0 \\ W_X + W_{\bar{X}} & 0 & 0 \\ F_X + F_{\bar{X}} & -I & 0 \\ T_X + T_{\bar{X}} & 0 & -I \end{bmatrix} \begin{bmatrix} \Sigma & 0 & 0 \\ 0 & \Sigma_F & 0 \\ 0 & 0 & \Sigma_T \end{bmatrix} \begin{bmatrix} C_X + C_{\bar{X}} & 0 & 0 \\ W_X + W_{\bar{X}} & 0 & 0 \\ F_X + F_{\bar{X}} & -I & 0 \\ T_X + T_{\bar{X}} & 0 & -I \end{bmatrix}^\top \right)^{-1} \left(\begin{bmatrix} C_X + C_{\bar{X}} & 0 & 0 \\ W_X + W_{\bar{X}} & 0 & 0 \\ F_X + F_{\bar{X}} & -I & 0 \\ T_X + T_{\bar{X}} & 0 & -I \end{bmatrix} \begin{bmatrix} m \\ m_F \\ m_T \end{bmatrix} \right).
\end{aligned}
\tag{50}
$$

Extracting the posterior means then gives

$$
\langle X \rangle = m + \Sigma O^\top \left(\begin{bmatrix} 0 & 0 & 0 \\ 0 & \Sigma_F & 0 \\ 0 & 0 & \Sigma_T \end{bmatrix} + O\Sigma O^\top \right)^{-1} \left(\begin{bmatrix} 0 \\ m_F \\ m_T \end{bmatrix} - Om \right).
\tag{51}
$$

Applying the limit to Σ_C and Σ_W gives Equation (33). In consequence, the MaxEnt formulation with soft prior constraints Section 3.1 and the Bayesian formulation Section 2 give the same mean flow rate prediction (33), but with different covariance matrices.

4. MaxEnt Analysis with Soft Probabilistic Constraints

4.1. Formulation

The MaxEnt method can also incorporate soft constraints using a probabilistic representation of the observed data. To implement this, we define a pdf that expresses the uncertainty in the system defined over a reduced parameter set X, again consisting of a basis set of n flow rates selected from

$\boldsymbol{\Psi}$. The indices of $\boldsymbol{X}$ in $\boldsymbol{\Psi}$ are again given by the set $\mathcal{B}$, while their complement is given by the set $\mathcal{N}$. Again, at least $N-1$ basis flow rates must be chosen but up to $N+M$ can be chosen. The joint probability is defined to be:

$$p(\boldsymbol{X})d\boldsymbol{X} = \mathrm{Prob}(\boldsymbol{X} \leq \mathrm{Y}_{\boldsymbol{X}} \leq \boldsymbol{X} + d\boldsymbol{X}), \tag{52}$$

where $\mathrm{Y}_{\boldsymbol{X}}$ is the vectors of the random variables for $\boldsymbol{X}$. We again assume that each of the flow rate and potential difference constraints are applied as soft constraints, but this does not restrict this method to only use soft constraints, and strict constraints can also be applied. This choice of pdf gives the following relative entropy or negative Kullback–Leibler function [21], over the space of uncertainties used in this formulation:

$$\mathfrak{H} = -\int_{l_1}^{u_1} \cdots \int_{l_n}^{u_n} p(\boldsymbol{X}) \ln \frac{p(\boldsymbol{X})}{q(\boldsymbol{X})} d\boldsymbol{X}, \tag{53}$$

where $q(\boldsymbol{X})$ is the prior pdf, and l_i and u_i are the lower and upper bounds of the ith flow rate. The relative entropy is then maximised subject to the constraints on the system. The following constraints are always required:

- Normalisation of probability:

$$1 = \int_{l_1}^{u_1} \cdots \int_{l_n}^{u_n} p(\boldsymbol{X}) d\boldsymbol{X}. \tag{54}$$

- Kirchhoff's first law, for the conservation of flow rates at each internal node, here imposed in the mean:

$$\boldsymbol{0} = (\boldsymbol{C}_{\boldsymbol{X}} + \boldsymbol{C}_{\bar{\boldsymbol{X}}}) \left(\int_{l_1}^{u_1} \cdots \int_{l_n}^{u_n} \boldsymbol{X} p(\boldsymbol{X}) d\boldsymbol{X} \right). \tag{55}$$

- Kirchhoff's second law, which requires the potential difference to vanish around each enclosed loop, again imposed in the mean:

$$\boldsymbol{0} = (\boldsymbol{W}_{\boldsymbol{X}} + \boldsymbol{W}_{\bar{\boldsymbol{X}}}) \left(\int_{l_1}^{u_1} \cdots \int_{l_n}^{u_n} \boldsymbol{X} p(\boldsymbol{X}) d\boldsymbol{X} \right). \tag{56}$$

We also allow for any of the following constraints:

- A set of specified inflow/outflow and internal flow rate constraints assuming the uncertainty is described by a Gaussian distribution:

$$\langle \boldsymbol{F_p} \rangle = \int_{l_1}^{u_1} \cdots \int_{l_n}^{u_n} \ln \left(\frac{\exp\left(-\frac{1}{2}((\boldsymbol{F_X}+\boldsymbol{F_{\bar{X}}})\boldsymbol{X}-\boldsymbol{Y_F})^\top \boldsymbol{\Sigma}_{\boldsymbol{F}}^{-1}((\boldsymbol{F_X}+\boldsymbol{F_{\bar{X}}})\boldsymbol{X})-\boldsymbol{Y_F}\right)}{(2\pi)^{\frac{N_{\hat{\Theta}}+N_{\hat{Q}}}{2}} |\boldsymbol{\Sigma}_{\boldsymbol{F}}|} \right) p(\boldsymbol{X}) d\boldsymbol{X}. \tag{57}$$

- Potential difference constraints between pairs of nodes again assuming the uncertainty is described by a Gaussian distribution:

$$\langle \boldsymbol{T_p} \rangle = \int_{l_1}^{u_1} \cdots \int_{l_n}^{u_n} \ln \left(\frac{\exp\left(-\frac{1}{2}((\boldsymbol{T_X}+\boldsymbol{T_{\bar{X}}})\boldsymbol{X}-\boldsymbol{Y_T})^\top \boldsymbol{\Sigma}_{\boldsymbol{T}}^{-1}((\boldsymbol{T_X}+\boldsymbol{T_{\bar{X}}})\boldsymbol{X}-\boldsymbol{Y_T}\right)}{(2\pi)^{\frac{h_c}{2}} |\boldsymbol{\Sigma}_{\boldsymbol{T}}|} \right) p(\boldsymbol{X}) d\boldsymbol{X}. \tag{58}$$

After identifying the constraints, the entropy (53) is then maximised subject to Equations (54)–(56) and whichever of Equations (57) and (58) apply. Applying the calculus of variations, we form the Lagrangian:

$$
\begin{aligned}
\mathfrak{L} = & -\int_{l_1}^{u_1} \cdots \int_{l_n}^{u_n} p(X) \ln \frac{p(X)}{q(X)} dX - \hat{\kappa} \left(\int_{l_1}^{u_1} \cdots \int_{l_n}^{u_n} p(X) dX - 1 \right) \\
& - \boldsymbol{\alpha} \left((C_X + C_{\bar{X}}) \left(\int_{l_1}^{u_1} \cdots \int_{l_n}^{u_n} X p(X) dX \right) \right) \\
& - \boldsymbol{\beta} \left((W_X + W_{\bar{X}}) \left(\int_{l_1}^{u_1} \cdots \int_{l_n}^{u_n} X p(X) dX \right) \right) \\
& - \boldsymbol{\lambda} \left(\int_{l_1}^{u_1} \cdots \int_{l_n}^{u_n} \ln \left(\frac{\exp\left(-\frac{1}{2} \left((F_X + F_{\bar{X}}) X - Y_F \right)^\top \Sigma_F^{-1} \left((F_X + F_{\bar{X}}) X \right) - Y_F \right)}{(2\pi)^{\frac{N_{\hat{\Theta}} + N_{\hat{Q}}}{2}} |\Sigma_F|} \right) p(X) dX - \langle F_p \rangle \right) \\
& - \boldsymbol{\eta} \left(\int_{l_1}^{u_1} \cdots \int_{l_n}^{u_n} \ln \left(\frac{\exp\left(-\frac{1}{2} \left((T_X + T_{\bar{X}}) X - Y_T \right)^\top \Sigma_T^{-1} \left((T_X + T_{\bar{X}}) X \right) - Y_T \right)}{(2\pi)^{\frac{h_C}{2}} |\Sigma_T|} \right) p(X) dX - \langle T_p \rangle \right),
\end{aligned}
\tag{59}
$$

where $\hat{\kappa}$, (scalar) $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\lambda}$ and $\boldsymbol{\eta}$ (row vectors) are the Lagrange multipliers for the normalisation, Kirchhoff's first and second laws, flow rates and the head loss constraints, respectively. The variation of $\mathfrak{L}$ is given by $\delta\mathfrak{L} = 0$. Extremizing Equation (59) by taking the functional derivative with respect to $p(X)$ and combining integrals gives:

$$
\begin{aligned}
\delta\mathfrak{L} = 0 = \int_{l_1}^{u_1} \cdots \int_{l_n}^{u_n} & \Bigg[-\ln \frac{p(X)}{q(X)} - \kappa - \boldsymbol{\alpha}(C_X + C_{\bar{X}})X - \boldsymbol{\beta}(W_X + W_{\bar{X}})X \\
& - \boldsymbol{\lambda} \left(\ln \left(\frac{\exp\left(-\frac{1}{2} \left((F_X + F_{\bar{X}}) X - Y_F \right)^\top \Sigma_F^{-1} \left((F_X + F_{\bar{X}}) X \right) - Y_F \right)}{(2\pi)^{\frac{N_{\hat{\Theta}} + N_{\hat{Q}}}{2}} |\Sigma_F|} \right) \right) \\
& - \boldsymbol{\eta} \left(\ln \left(\frac{\exp\left(-\frac{1}{2} \left((T_X + T_{\bar{X}}) X - Y_T \right)^\top \Sigma_T^{-1} \left((T_X + T_{\bar{X}}) X \right) - Y_T \right)}{(2\pi)^{\frac{h_C}{2}} |\Sigma_T|} \right) \right) \Bigg] dX,
\end{aligned}
\tag{60}
$$

where $\kappa = \hat{\kappa} + 1$. Rearrangement gives the following solution for $p(X)$ (the Boltzmann distribution):

$$
\begin{aligned}
p^*(X) = q(X) \exp \Bigg(& -\kappa - \boldsymbol{\alpha}(C_X + C_{\bar{X}})X - \boldsymbol{\beta}(W_X + W_{\bar{X}})X \\
& - \boldsymbol{\lambda} \left(\ln \left(\frac{\exp\left(-\frac{1}{2} \left((F_X + F_{\bar{X}}) X - Y_F \right)^\top \Sigma_F^{-1} \left((F_X + F_{\bar{X}}) X \right) - Y_F \right)}{(2\pi)^{\frac{N_{\hat{\Theta}} + N_{\hat{Q}}}{2}} |\Sigma_F|} \right) \right) \\
& - \boldsymbol{\eta} \left(\ln \left(\frac{\exp\left(-\frac{1}{2} \left((T_X + T_{\bar{X}}) X - Y_T \right)^\top \Sigma_T^{-1} \left((T_X + T_{\bar{X}}) X \right) - Y_T \right)}{(2\pi)^{\frac{h_C}{2}} |\Sigma_T|} \right) \right) \Bigg).
\end{aligned}
\tag{61}
$$

This can be solved, in conjunction with the constraints (36)–(40), to give $p^*(X)$ and the Lagrange multipliers κ, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\lambda}$ and $\boldsymbol{\eta}$. As the purpose of the soft constraints is to incorporate a distribution as a constraint, we take $\boldsymbol{\lambda} = -\mathbf{1}$ and $\boldsymbol{\eta} = -\mathbf{1}$.

4.2. Solution and Comparison to Bayesian Solution

If the prior is chosen to be proportional to Equation (6), the MaxEnt probability distribution with normalisation, Kirchhoff's first and second law, potential difference and flow rate constraints is proportional to

$$\ln(p^*(X)) \propto -\frac{1}{2}\left(X^\top \Sigma^{-1} X - X^\top \Sigma^{-1} m - m^\top \Sigma^{-1} X\right) - \alpha(C_X + C_{\bar{X}})X - \beta(W_X + W_{\bar{X}})X$$

$$+ \left(-\frac{1}{2}\left(X^\top (F_X + F_{\bar{X}})^\top \Sigma_F^{-1} (F_X + F_{\bar{X}}) X - X^\top (F_X + F_{\bar{X}})^\top \Sigma_F^{-1} Y_F - Y_F^\top \Sigma_F^{-1} (F_X + F_{\bar{X}}) X\right)\right) \quad (62)$$

$$+ \left(-\frac{1}{2}\left(X^\top (T_X + T_{\bar{X}})^\top \Sigma_T^{-1} (T_X + T_{\bar{X}}) X - X^\top (T_X + T_{\bar{X}})^\top \Sigma_T^{-1} Y_T - Y_T^\top \Sigma_T^{-1} (T_X + T_{\bar{X}}) X\right)\right).$$

Combining terms of the same order and assuming the covariance is symmetric and positive definite

$$\ln(p^*(X)) \propto -\frac{1}{2}X^\top \left(\Sigma^{-1} + (F_X + F_{\bar{X}})^\top \Sigma_F^{-1} (F_X + F_{\bar{X}}) + (T_X + T_{\bar{X}})^\top \Sigma_T^{-1} (T_X + T_{\bar{X}})\right) X$$

$$+ \left(m^\top \Sigma^{-1} - \alpha(C_X + C_{\bar{X}}) - \beta(W_X + W_{\bar{X}}) + Y_F^\top \Sigma_F^{-1} (F_X + F_{\bar{X}}) + Y_T^\top \Sigma_T^{-1} (T_X + T_{\bar{X}})\right) X, \quad (63)$$

let

$$\hat{O} = \begin{bmatrix} C_X + C_{\bar{X}} \\ W_X + W_{\bar{X}} \end{bmatrix}, \quad (64)$$

$$\tilde{O} = \begin{bmatrix} F_X + F_{\bar{X}} \\ T_X + T_{\bar{X}} \end{bmatrix}, \quad (65)$$

$$\tilde{S}^{-1} = \begin{bmatrix} \Sigma_F^{-1} & 0 \\ 0 & \Sigma_T^{-1} \end{bmatrix}, \quad (66)$$

$$\tilde{y} = \begin{bmatrix} Y_F \\ Y_T \end{bmatrix}, \quad (67)$$

so

$$\ln(p^*(X)) \propto -\frac{1}{2}X^\top \left(\Sigma^{-1} + \tilde{O}^\top \tilde{S}^{-1} \tilde{O}\right) X + \left(m^\top \Sigma^{-1} - \begin{bmatrix} \alpha & \beta \end{bmatrix} \hat{O} + \tilde{y}^\top \tilde{S}^{-1} \tilde{O}\right) X, \quad (68)$$

and completing the square

$$\ln(p^*(X)) \propto -\frac{1}{2}\left(X - \langle X \rangle\right)^\top \tilde{\Sigma}_P^{-1} \left(X - \langle X \rangle\right), \quad (69)$$

where

$$\langle X \rangle = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} X p(X) dX = \tilde{\Sigma}_P \left(\Sigma^{-1} m - \hat{O}^\top \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \tilde{O}^\top \tilde{S}^{-1} \tilde{y}\right) \quad (70)$$

and

$$\tilde{\Sigma}_P = \langle XX^\top \rangle - \langle X \rangle\langle X \rangle^\top = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} XX^\top p(X) dX - \langle X \rangle\langle X \rangle^\top = \left(\Sigma^{-1} + \tilde{O}^\top \tilde{S}^{-1} \tilde{O}\right)^{-1}. \quad (71)$$

Using the Woodbury matrix identity [37] gives the posterior covariance matrix

$$\tilde{\Sigma}_P = \Sigma - \Sigma\tilde{O}^\top \left(\tilde{S} + \tilde{O}\Sigma\tilde{O}^\top\right)^{-1} \tilde{O}\Sigma. \quad (72)$$

The following algebra follows that of the previous derivations. Right multiplying the inverse posterior covariance by $\Sigma\tilde{O}^\top$ gives

$$\tilde{\Sigma}_P^{-1}\Sigma\tilde{O}^\top = \tilde{O}^\top + \tilde{O}^\top \tilde{S}^{-1} \tilde{O}\Sigma\tilde{O}^\top = \tilde{O}^\top \tilde{S}^{-1} \left(\tilde{S} + \tilde{O}\Sigma\tilde{O}^\top\right). \quad (73)$$

Left multiplying with the posterior covariance then gives

$$\boldsymbol{\Sigma}\tilde{\boldsymbol{O}}^{\top} = \tilde{\boldsymbol{\Sigma}}_P \tilde{\boldsymbol{O}}^{\top} \tilde{\boldsymbol{S}}^{-1} \left(\tilde{\boldsymbol{S}} + \tilde{\boldsymbol{O}} \boldsymbol{\Sigma} \tilde{\boldsymbol{O}}^{\top} \right) \tag{74}$$

and obtaining $\tilde{\boldsymbol{\Sigma}}_P \tilde{\boldsymbol{O}}^{\top} \tilde{\boldsymbol{S}}^{-1}$ by right multiplying by $\left(\tilde{\boldsymbol{S}} + \tilde{\boldsymbol{O}} \boldsymbol{\Sigma} \tilde{\boldsymbol{O}}^{\top} \right)^{-1}$ gives

$$\boldsymbol{\Sigma}\tilde{\boldsymbol{O}}^{\top} \left(\tilde{\boldsymbol{S}} + \tilde{\boldsymbol{O}} \boldsymbol{\Sigma} \tilde{\boldsymbol{O}}^{\top} \right)^{-1} = \tilde{\boldsymbol{\Sigma}}_P \tilde{\boldsymbol{O}}^{\top} \tilde{\boldsymbol{S}}^{-1}. \tag{75}$$

The posterior mean flow rates can now be expressed using Equation (70) by substituting Equations (72) and (75) to give

$$\langle X \rangle = \left(\boldsymbol{\Sigma} - \boldsymbol{\Sigma}\tilde{\boldsymbol{O}}^{\top} \left(\tilde{\boldsymbol{S}} + \tilde{\boldsymbol{O}} \boldsymbol{\Sigma} \tilde{\boldsymbol{O}}^{\top} \right)^{-1} \tilde{\boldsymbol{O}} \boldsymbol{\Sigma} \right) \boldsymbol{\Sigma}^{-1} \boldsymbol{m} + \boldsymbol{\Sigma}\tilde{\boldsymbol{O}}^{\top} \left(\tilde{\boldsymbol{S}} + \tilde{\boldsymbol{O}} \boldsymbol{\Sigma} \tilde{\boldsymbol{O}}^{\top} \right)^{-1} \tilde{\boldsymbol{y}} - \tilde{\boldsymbol{\Sigma}}_P \hat{\boldsymbol{O}}^{\top} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix}, \tag{76}$$

$$\langle X \rangle = \boldsymbol{m} + \boldsymbol{\Sigma}\tilde{\boldsymbol{O}}^{\top} \left(\tilde{\boldsymbol{S}} + \tilde{\boldsymbol{O}} \boldsymbol{\Sigma} \tilde{\boldsymbol{O}}^{\top} \right)^{-1} \left(\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{O}} \boldsymbol{m} \right) - \tilde{\boldsymbol{\Sigma}}_P \hat{\boldsymbol{O}}^{\top} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix}. \tag{77}$$

Using the constraint equations, the Lagrange multipliers can be found from

$$\begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} = \left(\hat{\boldsymbol{O}} \tilde{\boldsymbol{\Sigma}}_P \hat{\boldsymbol{O}}^{\top} \right)^{-1} \hat{\boldsymbol{O}} \left(\boldsymbol{m} + \boldsymbol{\Sigma}\tilde{\boldsymbol{O}}^{\top} \left(\tilde{\boldsymbol{S}} + \tilde{\boldsymbol{O}} \boldsymbol{\Sigma} \tilde{\boldsymbol{O}}^{\top} \right)^{-1} \left(\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{O}} \boldsymbol{m} \right) \right). \tag{78}$$

Substituting Equation (78) into Equation (77) gives the posterior means

$$\langle X \rangle = \boldsymbol{m} + \boldsymbol{\Sigma}\tilde{\boldsymbol{O}}^{\top} \left(\tilde{\boldsymbol{S}} + \tilde{\boldsymbol{O}} \boldsymbol{\Sigma} \tilde{\boldsymbol{O}}^{\top} \right)^{-1} \left(\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{O}} \boldsymbol{m} \right) - \tilde{\boldsymbol{\Sigma}}_P \hat{\boldsymbol{O}}^{\top} \left(\hat{\boldsymbol{O}} \tilde{\boldsymbol{\Sigma}}_P \hat{\boldsymbol{O}}^{\top} \right)^{-1} \hat{\boldsymbol{O}} \left(\boldsymbol{m} + \boldsymbol{\Sigma}\tilde{\boldsymbol{O}}^{\top} \left(\tilde{\boldsymbol{S}} + \tilde{\boldsymbol{O}} \boldsymbol{\Sigma} \tilde{\boldsymbol{O}}^{\top} \right)^{-1} \left(\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{O}} \boldsymbol{m} \right) \right) \tag{79}$$

or

$$\langle X \rangle = \left(\boldsymbol{m} + \boldsymbol{\Sigma}\tilde{\boldsymbol{O}}^{\top} \left(\tilde{\boldsymbol{S}} + \tilde{\boldsymbol{O}} \boldsymbol{\Sigma} \tilde{\boldsymbol{O}}^{\top} \right)^{-1} \left(\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{O}} \boldsymbol{m} \right) \right) \left(\boldsymbol{I} - \tilde{\boldsymbol{\Sigma}}_P \hat{\boldsymbol{O}}^{\top} \left(\hat{\boldsymbol{O}} \tilde{\boldsymbol{\Sigma}}_P \hat{\boldsymbol{O}}^{\top} \right)^{-1} \hat{\boldsymbol{O}} \right). \tag{80}$$

As evident, the first bracketed term is of similar structure to Equations (33) and (51), although it contains parameters related to the soft probabilistic constraints. The second term gives a second-order expansion of that solution relating to the interaction between the hard moment constraints and soft probabilistic constraints. If all constraints were applied as soft probabilistic constraints, Equation (33) would be obtained. Numerical experiments suggest that the means obtained in Equation (80) are equal to the Bayesian posterior means Equation (33), in several examples considered, but the covariances are different.

5. Discussion

The MaxEnt and Bayesian methods rest on different theoretical foundations but are both able to predict flows on networks by updating the prior belief to the posterior with the inclusion of new information in the form of constraints or uncertain data. This study compares the application of Bayesian inference and the MaxEnt method for the analysis of flow networks, for the latter using soft constraints—included in the prior or imposed as probabilistic constraints—in addition to standard moment constraints. It is shown that both the Bayesian method and MaxEnt method with soft prior constraints, implemented using a multidimensional Gaussian prior pdf, infer the same mean flow rates but different covariance matrices. In the Bayesian method, the interactions between variables are applied through the likelihood function, using second or higher-order cross-terms within the posterior pdf. In contrast, the MaxEnt method incorporates interactions between variables using Lagrange multipliers, avoiding second-order correlation terms in the posterior covariance. The MaxEnt method with soft prior constraints therefore has a numerical advantage in its integrations, in that the covariance terms are avoided.

In contrast, the second MaxEnt method with probabilistic and moment constraints is shown to give a posterior mean of similar, but not identical, structure to the other two methods. Due to the mixture of constraint types, some of the interactions between variables are incorporated in the Lagrange multipliers and some are incorporated in the covariance matrix, leading to a more complicated formulation.

For both MaxEnt formulations given herein, the equivalence between the posterior means inferred by the Bayesian and MaxEnt methods is dependent on the choice of a multidimensional Gaussian prior and its parameterisation. Further research is required to classify the effect of other prior distributions on the MaxEnt and Bayesian formulations, and whether these lead to equivalences between the means or higher-order moments of the inferred posterior pdf.

Author Contributions: Steven H. Waldrip conducted the analysis and prepared this manuscript for this work, under the guidance and supervision of Robert K. Niven. Both authors have read and approved the final manuscript.

References

1. Sivia, D.; Skilling, J. *Data Analysis: A Bayesian Tutorial*, 2nd ed.; Oxford University Press: Oxford, UK, 2006.
2. Rougier, J.; Goldstein, M. A Bayesian analysis of fluid flow in pipe-lines. *J. R. Stat. Soc.* **2001**, *50*, 77–93.
3. Savic, D.A.; Kapelan, Z.S.; Jonkergouw, P.M. Quo vadis water distribution model calibration? *Urban Water J.* **2009**, *6*, 3–22.
4. Hutton, C.J.; Kapelan, Z.; Vamvakeridou-Lyroudia, L.; Savic, D. Real-time demand estimation in water distrubtion systems under uncertainty. In Proceedings of the 14th Water Distribution Systems Analysis Conference (WDSA 2012), Adelaide, Australia, 24–27 September 2012; pp. 1374–1385.
5. Hutton, C.; Kapelan, Z. Real-time Burst Detection in Water Distribution Systems Using a Bayesian Demand Forecasting Methodology. *Procedia Eng.* **2015**, *119*, 13–18.
6. Jaynes, E.T. Information Theory and Statistical Mechanics. *Phys. Rev.* **1957**, *106*, 620–630.
7. Jaynes, E.T. Information Theory and Statistical Mechanics. II. *Phys. Rev.* **1957**, *108*, 171–190.
8. Kapur, J.N.; Kesavan, H.K. *Entropy Optimization Principles with Applications*; Academic Press: Boston, MA, USA, 1992.
9. Jaynes, E.T. *Probability Theory: The Logic of Science*; Cambridge University Press: Cambridge, UK, 2003.
10. Pressé, S.; Ghosh, K.; Lee, J.; Dill, K.A. Principles of maximum entropy and maximum caliber in statistical physics. *Rev. Mod. Phys.* **2013**, *85*, 1115–1141.
11. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 623–656.
12. Shore, J.; Johnson, R. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Trans. Inf. Theory* **1980**, *26*, 26–37.
13. Caticha, A. Relative entropy and inductive inference. In Proceedings of the 23rd International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Jackson Hole, Wyoming, 3–8 August 2003.
14. Boltzmann, L. Über die Beziehung zwischen dem zweiten Hauptsatz der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung respektive den Sätzen über das Wärmegleichgewicht (On the Relationship between the Second Fundamental Theorem of the Mechanical Theory of H. *Wiener Berichte* **1877**, *2*, 373–435. (In German)
15. Sharp, K.; Matschinsky, F. Translation of Ludwig Boltzmann's Paper "On the Relationship between the Second Fundamental Theorem of the Mechanical Theory of Heat and Probability Calculations Regarding the Conditions for Thermal Equilibrium" Sitzungberichte der Kaiserlichen Akademie der Wissenschaften. *Entropy* **2015**, *17*, 1971–2009.

16. Jaynes, E.T. Prior Probabilities. *IEEE Trans. Syst. Sci. Cybern.* **1968**, *4*, 227–241.

17. Niven, R.K. Combinatorial entropies and statistics. *Eur. Phys. J. B* **2009**, *70*, 49–63.

18. Planck, M. Ueber das Gesetz der Energieverteilung im Normalspectrum. *Annalen der Physik* **1901**, *309*, 553–563. (In German)

19. Tribus, M. *Thermostatics and Thermodynamics: An Introduction to Energy, Information and States of Matter, with Engineering Applications*; Van Nostrand: New York, NY, USA, 1961.

20. Ellis, R.S. *Entropy, Large Deviations, and Statistical Mechanics*; Springer: Berlin/Heidelberg, Germany, 1985.

21. Kullback, S.; Leibler, R.A. On Information and Sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86.

22. Waldrip, S.H.; Niven, R.K.; Abel, M.; Schlegel, M. Maximum entropy analysis of hydraulic pipe networks. In Proceedings of the 33rd International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MaxEnt 2013), Canberra, Australia, 15–20 December 2013; pp. 180–186.

23. Waldrip, S.H.; Niven, R.K.; Abel, M.; Schlegel, M.; Noack, B.R. MaxEnt analysis of a water distribution network in Canberra, ACT, Australia. In Proceedings of the 2014 Bayesian Inference and Maximum Entropy Methods In Science and Engineering (MaxEnt 2014), Amboise, France, 21–26 September 2014; pp. 479–486.

24. Waldrip, S.H.; Niven, R.K.; Abel, M.; Schlegel, M. Maximum Entropy Analysis of Hydraulic Pipe Flow Networks. *J. Hydraul. Eng.* **2016**, *142*, 04016028.

25. Waldrip, S.; Niven, R.K.; Abel, M.; Schlegel, M. Reduced-Parameter Method for Maximum Entropy Analysis of Hydraulic Pipe Flow Networks. *J. Hydraul. Eng.* **2016**, submitted.

26. Waldrip, S.H. The Probabilistic Analysis of Flow Networks. PhD Thesis, The University of New South Wales, Canberra, Australia, 2017.

27. Niven, R.K.; Abel, M.; Schlegel, M.; Waldrip, S.H. Maximum entropy analysis of flow networks. In Proceedings of the 33rd International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MaxEnt 2013), Canberra, Australia, 15–20 December 2013; pp. 159–164.

28. Niven, R.K.; Abel, M.; Schlegel, M.; Waldrip, S.H. Maximum entropy analysis of flow and reaction networks. In Proceedings of the 2014 Bayesian Inference and Maximum Entropy Methods In Science and Engineering (MaxEnt 2014), Amboise, France, 21–26 September 2014; pp. 271–278.

29. Niven, R.; Waldrip, S.; Abel, M.; Schlegel, M.; Noack, B. Maximum Entropy Analysis of Flow Networks with Nonlinear Constraints. In Proceedings of the 2nd International Electronic Conference on Entropy and Its Applications, Basel, Switzerland, 15–30 November 2015; p. A012.

30. Williams, P.M. Bayesian Conditionalisation and the Principle of Minimum Information. *Br. J. Philos. Sci.* **1980**, *31*, 131–144.

31. Caticha, A.; Giffin, A. Updating Probabilities. In Proceedings of the 26th International Workshop on Bayesian Inference and Maximum Entropy Methods, Paris, France, 8–13 July 2006; pp. 31–42.

32. Giffin, A.; Caticha, A. Updating probabilities with data and moments. In Proceedings of the 27th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MaxEnt 2007), Saratoga Springs, NY, USA, 8–13 July 2007; pp. 74–84.

33. Giffin, A. Maximum Entropy: The Universal Method for Inference. PhD Thesis, University at Albany, State University of New York, Albany, NY, USA, 2008.

34. Hennig, P.; Kiefel, M. Quasi-Newton Methods: A New Direction. In Proceedings of the 29th International Conference on Machine Learning (ICML-12), Edinburgh, Scotland, 26 June–1 July 2012; pp. 25–32.

35. Hennig, P.; Kiefel, M. Quasi-Newton Methods: A New Direction. *J. Mach. Learn. Res.* **2013**, *14*, 843–865.

36. Waldrip, S.H.; Niven, R.K. Maximum Entropy Derivation of Quasi-Newton Methods. *SIAM J. Optim.* **2016**, *26*, 2495–2511.

37. Woodbury, M.A. *Inverting Modified Matrices*; Memorandum Report 42; Princeton University: Princeton, NJ, USA, 1950.