ECEA 2016

sciforum

Conference Proceedings Paper

Comparison Between Bayesian and Maximum Entropy Analysis of Flow Networks

Steven H. Waldrip [1,†] and Robert K. Niven [1,†,*]

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

* Correspondence: r.niven@adfa.edu.au; Tel.: +61-262-688-330

† These authors contributed equally to this work.

Abstract: Both the maximum entropy (MaxEnt) and Bayesian methods update a prior to a posterior probability density function (pdf) by the inclusion of new information, respectively in the form of constraints or data. To find the posterior, the MaxEnt method maximizes an entropy function subject to constraints, using the method of Lagrange multipliers, whereas the Bayesian method finds its posterior by multiplying the prior with likelihood functions, in which the measured data are substituted into the appropriate terms. The purpose of this work is to develop a Bayesian method to analyze flow networks and compare it to the MaxEnt method. Flow networks include, among others, water and electrical distribution networks and transport networks. The purpose of using probabilistic methods to model these networks is to predict the flow rates (and other variables on the network) when there is not enough information to model them deterministically, and also to incorporate the effects of uncertainty. After developing the Bayesian method, we show that the Bayesian and Maxent methods obtain the same posterior means but, when the prior is a normal distribution, their covariances are different. The Bayesian method incorporates interactions between variables through the likelihood function. It achieves this through second-order or higher cross-terms within the posterior pdf. The MaxEnt method however, incorporates interactions between variables using Lagrange multipliers, avoiding second order correlation terms in the posterior covariance. Therefore, the mean value inferences made by the MaxEnt and Bayesian methods are similar, but the MaxEnt method has a numerical advantage in its integrations, since the correlation terms can be avoided.

Keywords: maximum entropy method; Bayesian inference; probability; flow networks

1. Introduction

The analysis of flow rates on networks is required for the design and monitoring of electrical distribution, water distribution systems, sewer systems, irrigation systems, fire suppression systems, drainage systems and any other system through which conserved quantities are transported. Their analysis is an important engineering problem. Traditionally, these systems have been analyzed using deterministic methods. These methods incorporate physical laws such as the conservation of mass and path-independence of the potential function (e.g. electrical potential or pressure), Kirchhoff's first and second laws respectively. A closed set of equations is required for a unique solution. These methods yield precise parameter values but do not consider uncertainty due to a lack of knowledge of the state of the system or arising from flow variability. To account for uncertainty, a probabilistic framework is required.

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 analyzed. Therefore, "*...a single application of Bayes' theorem gives us only a probability; not a probability distribution*" [1]. To analyze the data, a set of data values are evaluated in the likelihood function, then normalized 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 each datum is analyzed does not impact the final result.

In this paper we develop a Bayesian method to analyze flow networks, presented in Section 2. This theory builds upon the maximum entropy (MaxEnt) method of [2]. In Section 3, we compare the distribution obtained with the Bayesian method with the MaxEnt method. Finally in Section 4 we discuss our results.

2. 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, intern assembled into the vector

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

The Bayesian method can contain a subset of the flow rates as parameters of the probability distribution, to avoid potential inconsistencies from different network representations [2]. The reduced parameter set $\boldsymbol{X}$, consists of a fixed set of n flow rates selected from $\boldsymbol{\Psi}$. The indices of $\boldsymbol{X}$ in $\boldsymbol{\Psi}$ are given by the set $\mathcal{E}$. The parameters not in $\boldsymbol{X}$ can be assembled into $\bar{\boldsymbol{X}}$ and the indices of $\bar{\boldsymbol{X}}$ in $\boldsymbol{\Psi}$ are given by the set $\mathcal{D}$. At least $N-1$ flow rates must be chosen to be able to form a deterministic equation for $\bar{\boldsymbol{X}}$ from $\boldsymbol{X}$ using Kirchoff's laws but up to $N+M$ can be chosen, where every flow rate will be included in the probability distribution function (pdf). The derivation of the Bayesian method requires a prior belief of the system, represented as a prior pdf, to be updated using observed data to a posterior pdf according to Bayes' theorem:

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

where $p(\boldsymbol{X}|\boldsymbol{y})$ is the posterior pdf, $p(\boldsymbol{y}|\boldsymbol{X})$ is the likelihood function, $q(\boldsymbol{X})$ is the prior pdf, $\boldsymbol{X}$ is the vector of selected model parameters and $\boldsymbol{y}$ is the vector of observed data. The flow rates $\bar{\boldsymbol{X}}$ which are not included as part of the model parameters are taken as functions of the model parameters $\boldsymbol{X}$, given by:

$$\bar{\boldsymbol{X}} = \boldsymbol{V}\boldsymbol{X} \tag{3}$$

where

$$\boldsymbol{V} = -\boldsymbol{A}^{-1}_{i\in\mathcal{V},j\in\mathcal{D}}\boldsymbol{A}_{i\in\mathcal{V},j\in\mathcal{E}} \tag{4}$$

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

where diag() places the elements of a vector on the diagonal of a square matrix of zeros and the set $\mathcal{V}$ contains the $N+M-n$ indices of the equations required to uniquely define $\bar{\boldsymbol{X}}$ from $\boldsymbol{X}$. The square matrix $\boldsymbol{A}_{i\in\mathcal{V},j\in\mathcal{D}}$ and normally non-square matrix $\boldsymbol{A}_{i\in\mathcal{V},j\in\mathcal{E}}$ contain a subset of the rows of $\boldsymbol{A}$ corosponding to the chosen equations but have the columns cosponsoring to those not included and included in the pdf respectively. The matrix $\boldsymbol{C}$ is an $N\times(N+M)$ connectivity matrix containing elements $\{-1,0,1\}$. The entries $C_{i,r}, \forall i \in \{1,..N\}$ indicate the connectivity 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 $\boldsymbol{K}$ is an $(M+N)\times 1$ vector of flow resistances. The matrix $\boldsymbol{W}$ 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. The 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 $\boldsymbol{F}$ 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 number of flow rate observation locations for flows entering/exiting or within the network respectively. The matrix $\boldsymbol{T}$ 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 $\langle \boldsymbol{Y_T} \rangle$ 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. Further details of these matrix constructions are given in [2,3], with reference to the MaxEnt method.

The prior can be chosen to represent a 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, a multidimensional normal distribution is selected, here defined over the real domain.

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

where $\boldsymbol{m}$ is the $n \times 1$ vector of mean flow rates and $\boldsymbol{\Sigma}$ is the $n \times n$ matrix of co-variances. To incorporate the physics of the system and observations of the system likelihood functions are defined. The likelihood function to incorporate the remaining conservation of mass (or flow rate for incompressible systems) is given by a delta function

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

where

$$\boldsymbol{C_X} = \boldsymbol{C}_{i \notin \mathcal{V}, j \in \mathcal{E}} \qquad \boldsymbol{C_{\bar{X}}} = \boldsymbol{C}_{i \notin \mathcal{V}, j \in \mathcal{D}} \boldsymbol{V} \tag{8}$$

This delta function is proportional to the normal distribution

$$-2\ln(p(\boldsymbol{0}|\boldsymbol{X})) \propto \lim_{\boldsymbol{\Sigma_C} \to \boldsymbol{0}} \left(\boldsymbol{0} - \left(\boldsymbol{C_X} + \boldsymbol{C_{\bar{X}}}\right)\boldsymbol{X}\right)^{\top} \boldsymbol{\Sigma}_{\boldsymbol{C}}^{-1} \left(\boldsymbol{0} - \left(\boldsymbol{C_X} + \boldsymbol{C_{\bar{X}}}\right)\boldsymbol{X}\right) \tag{9}$$

The likelihood function to incorporate the remaining loop laws is given by a delta function

$$p(\boldsymbol{0}|\boldsymbol{X}) = \delta\left(\boldsymbol{0} - \left(\boldsymbol{W_X} + \boldsymbol{W_{\bar{X}}}\right)\boldsymbol{X}\right) \tag{10}$$

where

$$\boldsymbol{W_X} = \boldsymbol{W}_{i \notin \mathcal{V}, j \in \mathcal{E}} \mathrm{diag}\left(\boldsymbol{K}_{i \in \mathcal{E}}\right) \qquad \boldsymbol{W_{\bar{X}}} = \boldsymbol{W}_{i \notin \mathcal{V}, j \in \mathcal{D}} \mathrm{diag}\left(\boldsymbol{K}_{i \in \mathcal{D}}\right) \boldsymbol{V} \tag{11}$$

This delta function is proportional to the normal distribution

$$-2\ln(p(\boldsymbol{0}|\boldsymbol{X})) \propto \lim_{\boldsymbol{\Sigma_W} \to \boldsymbol{0}} \left(\boldsymbol{0} - \left(\boldsymbol{W_X} + \boldsymbol{W_{\bar{X}}}\right)\boldsymbol{X}\right)^{\top} \boldsymbol{\Sigma}_{\boldsymbol{W}}^{-1} \left(\boldsymbol{0} - \left(\boldsymbol{W_X} + \boldsymbol{W_{\bar{X}}}\right)\boldsymbol{X}\right) \tag{12}$$

Observed flow rates can be constrained with the likelihood function

$$-2\ln(p(\boldsymbol{Y_F}|\boldsymbol{X})) \propto \left(\langle \boldsymbol{Y_F} \rangle - \left(\boldsymbol{F_X} + \boldsymbol{F_{\bar{X}}}\right)\boldsymbol{X}\right)^{\top} \boldsymbol{N_F} \boldsymbol{\Sigma}_{\boldsymbol{F}}^{-1} \left(\langle \boldsymbol{Y_F} \rangle - \left(\boldsymbol{F_X} + \boldsymbol{F_{\bar{X}}}\right)\boldsymbol{X}\right) \tag{13}$$

where $\langle \boldsymbol{Y_F} \rangle$ is a $(N_{\hat{\Theta}} + N_{\hat{Q}}) \times 1$ vector that has the mean flow rate of all observations for a link in each element, $\boldsymbol{N_F}$ is a $(N_{\hat{\Theta}} + N_{\hat{Q}}) \times (N_{\hat{\Theta}} + N_{\hat{Q}})$ diagonal matrix with the number of observations of $\langle Y_{Fi} \rangle$

placed on the element $N_{Yi,i}$, $\boldsymbol{\Sigma_F}$ is the $(N_{\hat{\Theta}} + N_{\hat{Q}}) \times (N_{\hat{\Theta}} + N_{\hat{Q}})$ covariance matrix of the observations and

$$\boldsymbol{F_X} = \boldsymbol{F}_{i \notin \mathcal{V}, j \in \mathcal{E}} \qquad \boldsymbol{F_{\bar{X}}} = \boldsymbol{F}_{i \notin \mathcal{V}, j \in \mathcal{D}} \boldsymbol{V} \tag{14}$$

Observed potential differences can be constrained with the likelihood function

$$-2\ln(p(\boldsymbol{Y_T}|\boldsymbol{X})) \propto (\langle \boldsymbol{Y_T} \rangle - (\boldsymbol{T_X} + \boldsymbol{T_{\bar{X}}})\,\boldsymbol{X})^\top \boldsymbol{N_T} \boldsymbol{\Sigma_T^{-1}} (\langle \boldsymbol{Y_T} \rangle - (\boldsymbol{T_X} + \boldsymbol{T_{\bar{X}}})\,\boldsymbol{X}) \tag{15}$$

where $\langle \boldsymbol{Y_T} \rangle$ is a $h_c \times 1$ vector that has the mean potential difference of all observations between two points in each element, $\boldsymbol{N_T}$ is a $h_c \times h_c$ diagonal matrix with the number of observations of $\langle Y_{Ti} \rangle$ placed on the element $N_{Ti,i}$, $\boldsymbol{\Sigma_T}$ is the $h_c \times h_c$ covariance matrix of the observations and

$$\boldsymbol{T_X} = \boldsymbol{T}_{i \notin \mathcal{V}, j \in \mathcal{E}} \mathrm{diag}\,(\boldsymbol{K}_{i \in \mathcal{E}}) \qquad \boldsymbol{T_{\bar{X}}} = \boldsymbol{T}_{i \notin \mathcal{V}, j \in \mathcal{D}} \mathrm{diag}\,(\boldsymbol{K}_{i \in \mathcal{D}})\,\boldsymbol{V} \tag{16}$$

These likelihood function assumes that every observation is taken with the same uncertainty and is in a form which allows multiple observations on different links to be incorporated into the one function. They can be obtained by applying Bayes rule multiple times for each observation on each link. Applying Bayes rule with each of the likelihood functions, expanding and dropping all terms which are not functions of $\boldsymbol{X}$ gives

$$-2\ln(p(\boldsymbol{X}|\boldsymbol{y})) \propto \boldsymbol{X}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{X} - \boldsymbol{X}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{m} - \boldsymbol{m}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{X} + \boldsymbol{X}^\top \boldsymbol{O}^\top \boldsymbol{S}^{-1} \boldsymbol{O} \boldsymbol{X} - \boldsymbol{y}^\top \boldsymbol{S}^{-1} \boldsymbol{O} \boldsymbol{X} - \boldsymbol{X}^\top \boldsymbol{O}^\top \boldsymbol{S}^{-1} \boldsymbol{y} \tag{17}$$

where

$$\boldsymbol{O} = \begin{bmatrix} \boldsymbol{C_X} + \boldsymbol{C_{\bar{X}}} \\ \boldsymbol{W_X} + \boldsymbol{W_{\bar{X}}} \\ \boldsymbol{F_X} + \boldsymbol{F_{\bar{X}}} \\ \boldsymbol{T_X} + \boldsymbol{T_{\bar{X}}} \end{bmatrix} \tag{18}$$

$$\boldsymbol{S}^{-1} = \begin{bmatrix} \boldsymbol{\Sigma_C^{-1}} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma_W^{-1}} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{N_F}\boldsymbol{\Sigma_F^{-1}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{N_T}\boldsymbol{\Sigma_T^{-1}} \end{bmatrix} \tag{19}$$

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ \langle \boldsymbol{Y_F} \rangle \\ \langle \boldsymbol{Y_T} \rangle \end{bmatrix} \tag{20}$$

combining like factors

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

Completing the square gives

$$-2\ln(p(\boldsymbol{X}|\boldsymbol{y})) \propto (\boldsymbol{X} - \langle \boldsymbol{X} \rangle)^\top \boldsymbol{\Sigma_p^{-1}} (\boldsymbol{X} - \langle \boldsymbol{X} \rangle) \tag{22}$$

where the variance and mean flow rates are given by

$$\boldsymbol{\Sigma_p} = \left(\boldsymbol{\Sigma}^{-1} + \boldsymbol{O}^\top \boldsymbol{S}^{-1} \boldsymbol{O} \right)^{-1} \tag{23}$$

$$\langle \boldsymbol{X} \rangle = \boldsymbol{\Sigma_p} \left(\boldsymbol{\Sigma}^{-1} \boldsymbol{m} + \boldsymbol{O}^\top \boldsymbol{S}^{-1} \boldsymbol{y} \right) \tag{24}$$

Using the Woodbury matrix identity to find the posterior covariance gives

$$\boldsymbol{\Sigma}_p = \boldsymbol{\Sigma} - \boldsymbol{\Sigma}\boldsymbol{O}^\top \left(\boldsymbol{S} + \boldsymbol{O}\boldsymbol{\Sigma}\boldsymbol{O}^\top\right)^{-1} \boldsymbol{O}\boldsymbol{\Sigma} \tag{25}$$

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

$$\boldsymbol{\Sigma}_p^{-1}\boldsymbol{\Sigma}\boldsymbol{O}^\top = \boldsymbol{O}^\top + \boldsymbol{O}^\top\boldsymbol{S}^{-1}\boldsymbol{O}\boldsymbol{\Sigma}\boldsymbol{O}^\top = \boldsymbol{O}^\top\boldsymbol{S}^{-1}\left(\boldsymbol{S} + \boldsymbol{O}\boldsymbol{\Sigma}\boldsymbol{O}^\top\right) \tag{26}$$

left multiply with the posterior covariance

$$\boldsymbol{\Sigma}\boldsymbol{O}^\top = \boldsymbol{\Sigma}_p\boldsymbol{O}^\top\boldsymbol{S}^{-1}\left(\boldsymbol{S} + \boldsymbol{O}\boldsymbol{\Sigma}\boldsymbol{O}^\top\right) \tag{27}$$

obtain $\boldsymbol{\Sigma}_p\boldsymbol{O}^\top\boldsymbol{S}^{-1}$ by right multiply by $(\boldsymbol{S} + \boldsymbol{O}\boldsymbol{\Sigma}\boldsymbol{O}^\top)^{-1}$

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

The mean flow rate can now be found from (24) by substituting (25) and (28) to give

$$\langle \boldsymbol{X} \rangle = \left(\boldsymbol{\Sigma} - \boldsymbol{\Sigma}\boldsymbol{O}^\top \left(\boldsymbol{S} + \boldsymbol{O}\boldsymbol{\Sigma}\boldsymbol{O}^\top\right)^{-1} \boldsymbol{O}\boldsymbol{\Sigma}\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{m} + \boldsymbol{\Sigma}\boldsymbol{O}^\top \left(\boldsymbol{S} + \boldsymbol{O}\boldsymbol{\Sigma}\boldsymbol{O}^\top\right)^{-1} \boldsymbol{y} \tag{29}$$

$$= \boldsymbol{m} + \boldsymbol{\Sigma}\boldsymbol{O}^\top \left(\boldsymbol{S} + \boldsymbol{O}\boldsymbol{\Sigma}\boldsymbol{O}^\top\right)^{-1} (\boldsymbol{y} - \boldsymbol{O}\boldsymbol{m}) \tag{30}$$

3. Comparison

The MaxEnt probability distribution with normalization, Kirchhoff's first and second law, potential difference and flow rate constraints with the prior (6) is proportional to [2–6]

$$\ln(p^*(\boldsymbol{X})) \propto -\frac{1}{2}\left(\boldsymbol{X}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{X} - \boldsymbol{X}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{m} - \boldsymbol{m}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right) - \gamma^\top\boldsymbol{O}\boldsymbol{X} \tag{31}$$

where $\gamma = \begin{bmatrix}\boldsymbol{\alpha} & \boldsymbol{\beta} & \boldsymbol{\lambda} & \boldsymbol{\eta}\end{bmatrix}^\top$ and $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\lambda}$ and $\boldsymbol{\eta}$ (row vectors) are the Lagrange multipliers for Kirchhoff's first and second law, flow rate and potential difference constraints respectively. Combining terms of the same order and assuming $\boldsymbol{\Sigma}$ is symmetric and positive definite

$$\ln(p^*(\boldsymbol{X})) \propto -\frac{1}{2}\boldsymbol{X}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{X} - \left(\boldsymbol{m}^\top\boldsymbol{\Sigma}^{-1} + \gamma^\top\boldsymbol{O}\right)\boldsymbol{X} \tag{32}$$

completing the square

$$\ln(p^*(\boldsymbol{X})) \propto -\frac{1}{2}\left(\boldsymbol{X} - \boldsymbol{m} + \boldsymbol{\Sigma}\boldsymbol{O}^\top\gamma\right)^\top \times \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{X} - \boldsymbol{m} + \boldsymbol{\Sigma}\boldsymbol{O}^\top\gamma\right) \tag{33}$$

The above form allows the mean to be obtained as

$$\langle \boldsymbol{X} \rangle = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \boldsymbol{X}p(\boldsymbol{X})d\boldsymbol{X} = \boldsymbol{m} - \boldsymbol{\Sigma}\boldsymbol{O}^\top\gamma \tag{34}$$

using the constraint equations the Lagrange multipliers can be found from

$$\gamma = \left(\boldsymbol{O}\boldsymbol{\Sigma}\boldsymbol{O}^\top\right)^{-1}(\boldsymbol{O}\boldsymbol{m} - \boldsymbol{y}) \tag{35}$$

Substituting (35) into (34) gives

$$\langle \boldsymbol{X} \rangle = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \boldsymbol{X} p(\boldsymbol{X}) d\boldsymbol{X} = \boldsymbol{m} - \boldsymbol{\Sigma} \boldsymbol{O}^{\top} \left(\boldsymbol{O} \boldsymbol{\Sigma} \boldsymbol{O}^{\top} \right)^{-1} (\boldsymbol{O} \boldsymbol{m} - \boldsymbol{y}) \tag{36}$$

therefore the MaxEnt and Bayesian methods give the same mean flow predictions when all data are known precisely, i.e. incorporated with delta functions.

4. Discussion and Conclusions

The MaxEnt and Bayesian methods rest on different theoretical foundations although they are both able to predict flows on networks by updating a prior belief to a posterior with the inclusion of new information in the form of constraints or uncertain data. Through deriving a Bayesian method to analyze flow networks and comparing it with the equivalent MaxEnt method, it has been shown that both the MaxEnt and Bayesian methods give the same mean flow rate prediction when normal distribution priors are used but different covariance predictions.

The reason for the differences in the covariance predictions is that the interactions between variables in the Bayesian method are incorporated in the posterior covariance matrix whereas these interactions can be incorporated into the Lagrange multipliers in the MaxEnt method. The inclusion of the interactions between variables within the Lagrange multipliers allows the MaxEnt method to have a computational advantage over the Bayesian method. This is because in some large dimensional problems the MaxEnt method can be integrated by multiplying one dimensional integrals together rather than integrating the system as a whole.

Acknowledgments: This project acknowledges funding support from the Australian Research Council Discovery Projects grant DP140104402, Go8/DAAD Australia-Germany Joint Research Cooperation Scheme RG123832 and the French Agence Nationale de la Recherche Chair of Excellence (TUCOROM) and Institute Pprime, Poitiers, France. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jaynes, E.T. The Relation of Bayesian and Maximum Entropy Methods. In *Maximum-Entropy and Bayesian Methods in Science and Engineering*; Kluwer Academic Publishers: Amsterdam, The Netherlands, 1988; Vol. 1, pp. 25–29.
2. 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.
3. Waldrip, S.H.; Niven, R.K.; Abel, M.; Schlegel, M. Maximum Entropy Analysis of Hydraulic Pipe Flow Networks. *J. Hydraul. Eng.* **2016**, *142*, 04016028.
4. Waldrip, S.H.; Niven, R.K.; Abel, M.; Schlegel, M. Maximum entropy analysis of hydraulic pipe networks. Bayesian Inference and Maximum Entropy Methods In Science And Engineering: Proceedings of the 33rd International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MaxEnt 2013); Niven, R.K.; Brewer, B.; Paull, D.; Shafi, K.; Stokes, B., Eds.; AIP: Canberra, ACT, Australia, 2014; Vol. 180, pp. 180–186.
5. Waldrip, S.H.; Niven, R.K.; Abel, M.; Schlegel, M.; Noack, B.R. MaxEnt analysis of a water distribution network in Canberra, ACT, Australia. Bayesian Inference and Maximum Entropy Methods In Science And Engineering (MaxEnt 2014); Mohammad-Djafari, A.; Barbaresco, F., Eds.; AIP: Clos Lucé, Amboise, France, 2015; Vol. 479, pp. 479–486.
6. Waldrip, S.H. Probilistic Analysis of Flow Networks. PhD thesis, The University of New South Wales, Canberra, Australia, 2017.