Uncertainty Propagation in Long-term Structured Regression on Evolving Networks

Djordje Gligorijevic, Jelena Stojanovic and Zoran Obradovic
Center for Data Analytics and Biomedical Informatics, Temple University, Philadelphia, PA 19122 USA
{gligorijevic, jelena.stojanovic, zoran.obradovic}@temple.edu

Abstract

In long-term forecasting it is important to estimate the confidence of predictions, as they are often affected by errors that are accumulated over the prediction horizon. To address this problem, an effective novel iterative method is developed for Gaussian structured learning models in this study for propagating uncertainty in temporal graphs by modeling noisy inputs. The proposed method is applied for three long-term (up to 8 years ahead) structured regression problems on real-world evolving networks from the health and climate domains. The obtained empirical results and use case analysis provide evidence that the new approach allows better uncertainty propagation as compared to published alternatives.

Introduction

Conditional probabilistic graphical models provide a powerful framework for structured regression in spatio-temporal datasets with complex correlation patterns. It has been shown that models utilizing underlying correlation patterns (structured models) can significantly improve predictive accuracy as compared to models not utilizing such information (Radosavljevic, Vucetic, and Obradovic 2010; 2014; Ristovski et al. 2013; Wytock and Kolter 2013; Stojanovic et al. 2015) .

This study is aimed to support structured regression for long-term decision making, which has been of interest in many high impact applications. For example, in order to prepare beds, personnel and medications, hospitals are interested in estimating future number of patients admitted in different departments. Providing long-term predictions of healthcare trends could immensely increase quality of decisions, which could lead to better hospital care (Dey et al. 2014; Stiglic et al. 2014; Ghalwash, Radosavljevic, and Obradovic 2014; Gligorijevic, Stojanovic, and Obradovic 2015), drug coverage and health insurances for clients in need. On the other hand, long-term predictions of geophysical variables, such are predictions of precipitation and lightning strikes, are very important in agriculture, telecommunications, power systems and elsewhere (Sivakumar and Berndtsson 2010; Dokic et al. 2016), thus enabling more efficient fund management and providing more stable services.

However, good predictive accuracy is not always sufficient for the long–term decision making. Uncertainty estimation, a tally of reliability for model predictions, is an important quality indicator used for decision making (Smith 2013). It is also reasonable to assume that reliability of a model decreases when predicting further in the future. This is due to *noisiness in input data*, caused by a change in distribution or accumulated error of iterative predictions, and it should be reflected in the increase of estimated uncertainty of the model predictions. This effect is often referred to as *uncertainty propagation*. Therefore, to decide the time point on the prediction horizon, or level of certainty with which the predictive model could be considered as useful and reliable, it is important to have a proper uncertainty propagation estimate for reasoning under uncertainty.

Thus, a particular interest of this paper is long-term forecasting on non-static networks with continuous target variables (structured regression) and proper uncertainty propagation estimate in such evolving networks. This is motivated by climate modeling of long–term precipitation prediction in spatio–temporal weather station networks, as well as prediction of different disease trends in temporal disease-disease networks.

Methods that address uncertainty propagation in multiple steps-ahead prediction can be viewed as direct and iterative (Smith 2013). In both types, error estimates made by the models should be taken into account, to ensure uncertainty propagation. Direct methods are the ones where uncertainties can be explicitly computed while predicting in the future. They assume that input variables will be available in the entire prediction horizon. However, this is a strong assumption when predicting far ahead, thus limiting many of the models. Moreover, these methods usually need more training data than iterative methods to produce useful predictions. Therefore, we will not focus on direct methods in this paper. Iterative methods, on the other hand, can provide prediction any number of steps ahead, up to the desired horizon. These methods are iteratively predicting one step ahead and use lagged predictions as model inputs, as shown in Figure 1. In this study we assume that input variables in each of the steps are normally distributed as $\mathcal{N}(\mu_{X_{T+k}}, \Sigma_{X_{T+k}})$, and that the new point estimate $\hat{y}_{T+k} = \mu_{T+k}$ is obtained

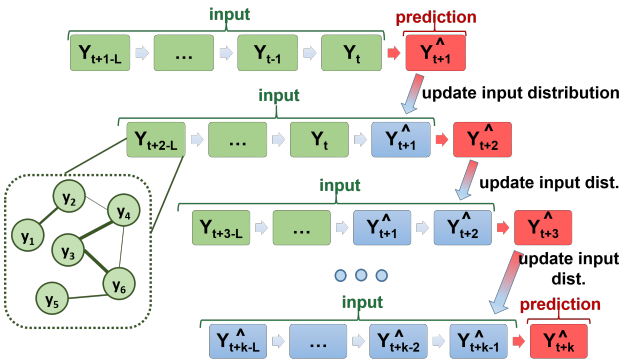


Figure 1: Iterative multiple–steps–ahead prediction in a temporal network represented by a vector of input attributes (X_t) and target variables y_t for each time step t. Initial $\mathcal{L}$ time steps (green boxes) are observed and remaining steps (blue boxes) are predicted iteratively in a one time step ahead process (red box).

using a predictive model.

To address uncertainty propagation in multiple steps ahead forecasting on evolving networks we propose a novel iterative uncertainty propagation model for structured regression that extends Continuous (Gaussian) Conditional Random Fields (GCRF) (Radosavljevic, Vucetic, and Obradovic 2010; 2014). In the proposed model, uncertainty that naturally comes from the data itself is taken into account when estimating uncertainty of the model predictions. Such setup enables iterative multiple–steps–ahead prediction with the GCRF as an iterative uncertainty propagation method.

In the past, iterative methods were developed for the Gaussian models (Girard et al. 2003; Candela et al. 2003; Kocijan et al. 2004; 2005), however, without strong empirical analysis. Moreover, to the best of our knowledge this is the first study addressing uncertainty propagation through iterative propagation methods for regression *on graphs*. To evaluate the quality of the proposed method, we compare to the several benchmarks from the group of unstructured iterative models: iterative Linear Regression (ILR) (Smith 2013) and iterative Gaussian Processes (IGP) (Girard et al. 2003; Candela et al. 2003). Results show evidence of benefits of using structure information for prediction and uncertainty propagation estimation in multiple–steps–ahead setup.

Key contributions of this paper are summarized bellow:

- A novel extension is provided for the GCRF model where we enable it to perform structure learning, rather than using sub-optimal predefined structure;
- A statistically sound practical approach is proposed to correct GCRF's uncertainty estimation that improves applicability via modeling uncertain inputs;
- A novel approach is developed for iterative multiple-steps-ahead prediction with propagation of errors expressed in terms of uncertainty estimation;
- Robustness of the proposed approach is demonstrated by applications to high impact real-world problems in climate and healthcare.

The supplement materials with additional experiments and theoretical derivations, as well as the Matlab implementation are available for download at the authors' websites.

Structured Regression Models

In this section a graph based structured regression model is described first, followed by a description of the proposed long–term predictive model as well as mathematical background for forecasting from noisy inputs. For readability purposes we provide a table of notation symbols:

Symbol	Notation meaning
X_t/Y_t	Input/Output variables for entire network in time step t
x_i/y_i	Input/Output variable for particular node in network
$x_i^{(d)}$	d'th dimension of input variable for node i
μ_{x^*}	Means of inputs distributions
$\{\Sigma_{X^*}\}$	Covariances of inputs distributions
μ_*	Predictive mean
Σ_{**}	Predictive variance

Gaussian Conditional Random Fields

Gaussian Conditional Random Fields (GCRF) (Radosavljevic, Vucetic, and Obradovic 2010; 2014) is a structured regression model that captures both the network structure and the mapping from attribute values of the nodes (X) to variables of interest (y). It is a model over a general graph structure, and can represent the structure as a function of time, space, or other user-defined structures. It models the structured regression problem as estimation of a joint continuous distribution over all nodes in a graph. For our purposes the model takes the following log-linear form

$$p(y|x) = \frac{1}{Z} exp\Big(\sum_{k=1}^{K}\sum_{i=1}^{N} \alpha_k (y_i - R_k(X,\theta_k))^2 + \sum_{l=1}^{L}\sum_{i\sim j} \beta_l S^{(l)}(x_i, x_j, \psi_l)(y_i - y_j)^2\Big) \quad (1)$$

where α and β are parameters of the feature functions, which model the association of each y_i and X, and the interaction between different y_i and y_j in the graph, respectively. Here, $R_k(X,\theta_k)$ functions are any specified unstructured predictors that map $X \rightarrow y_i$ independently, and might also be used to incorporate domain specific models. θ_k are parameters of the k-th unstructured model. Similarity metric $S^{(l)}(X,\psi_l)$ is used to define the weighted undirected graph structure between labels, for which parameters ψ_l might be specified.

Learning the structure via a predefined similarity metric, rather than simply using a given graph structure, is a novelty for the GCRF model considered in this study. We have, thus, enabled this model to perform structure learning of node labels correlations.

The final form of this GCRF model is defined by its mean μ and covariance Σ^{-1} matrices which we specify as

$$\Sigma^{-1} = \begin{cases} 2\alpha + 2\sum_l \sum_g \beta_l S^{(l)}(x_i, x_g, \psi_l), i = j \\ -2\sum_l \beta_l S^{(l)}(x_i, x_j, \psi_l), i \neq j \end{cases} \quad (2)$$

and:

$$\mu = 2\Sigma \left(\sum_k \alpha_k R(x_i, \theta_k) \right). \quad (3)$$

Quadratic form in Eq. 1 can be represented as a multivariate Gaussian. This specific way of modeling allows efficient inference and learning. Additionally, the GCRF model can, due to its Gaussian form, intrinsically highlight areas of the input space where prediction quality is poor by indicating the higher variance around the predicted mean.

Learning the GCRF model: The learning task is to optimize parameters $\hat{\alpha}, \hat{\beta}, \hat{\theta}, \hat{\psi}$ by maximizing the conditional log–likelihood,

$$(\hat{\alpha}, \hat{\beta}, \hat{\theta}, \hat{\psi}) = \underbrace{argmax}_{\alpha,\beta,\theta,\psi} logP(y|X; \alpha, \beta, \theta, \psi). \quad (4)$$

All the parameters are learned by a gradient-based optimization. Note that both the unstructured predictors and similarity metrics should be differentiable such that all parameters can be optimized using gradient methods.

Gradients of the conditional log-likelihood are

$$\begin{aligned}
&\frac{\partial \mathcal{L}}{\partial \alpha_k} = -\frac{1}{2}(y-\mu)^T \frac{\partial \Sigma^{-1}}{\partial \alpha_k}(y-\mu) + \\
&+ (\frac{\partial b^T}{\partial \alpha_k} - \mu^T \frac{\partial \Sigma^{-1}}{\partial \alpha_k})(y-\mu) + \frac{1}{2} Tr(\Sigma \frac{\partial \Sigma^{-1}}{\partial \alpha_k}) \\
&\frac{\partial \mathcal{L}}{\partial \beta_l} = -\frac{1}{2}(y+\mu)^T \frac{\partial \Sigma^{-1}}{\partial \beta_l}(y-\mu) + \frac{1}{2} Tr(\Sigma \frac{\partial \Sigma^{-1}}{\partial \beta_l}) \\
&\frac{\partial \mathcal{L}}{\partial \theta_k} = 2\alpha_k (y-\mu) \frac{\partial R_k}{\partial \theta_k} \\
&\frac{\partial \mathcal{L}}{\partial \psi_l} = -\frac{1}{2}(y-\mu) \frac{\partial \Sigma^{-1}}{\partial S_l} \frac{\partial S_l}{\partial \psi_l}(y-\mu)^T + \frac{1}{2} Tr(\Sigma \frac{\partial \Sigma^{-1}}{\partial S_l} \frac{\partial S_l}{\partial \psi_l}).
\end{aligned} \quad (5)$$

Maximizing the conditional log–likelihood is a non-convex, however smooth objective, and can be optimized using standard Quasi-Newton optimization techniques. Partial derivatives in Eq. 5 are specific for the choice of unstructured predictor and similarity metric. Note that $R_k(X, \theta_k)$ and $S^{(l)}(x_i, x_j, \psi_l)$ functions can be any differentiable predictor and similarity function. The GCRF model is Gaussian and, therefore, the maximum a posteriori estimate of y is obtained at the expected value μ of the GCRF distribution.

In this paper, for the simplicity, the choice of the unstuctured predictor is a *linear function* $R_k(x, \theta_k) = x\theta_k$, and choice of parametrized similarity metric is the *Gaussian kernel similarity*:

$$S(x_i, x_j, \psi) = \psi_0 exp\left(-\frac{1}{2} \sum_{d=1}^{D} \frac{(x_i^{(d)} - x_j^{(d)})^2}{\psi_d^2}. \right) \quad (6)$$

The optimization problem of this model is non–convex. Therefore, there are no guarantees that the solution will be optimal, as the model can be potentially optimized to a local minimum (Radosavljevic, Vucetic, and Obradovic 2014). However, a good initialization of parameters based approach should lead to a close to optimal solution for such deep architectures as the one proposed in this paper (Bengio 2012).

Uncertainty propagation by modeling noisy inputs

Uncertainty estimation should always take into account uncertainty that naturally comes from the data itself. Our approach gravitates around inclusion of uncertainty coming from input variables, when previously obtained predictions are used as inputs. Such setup enables multiple–steps–ahead prediction with the GCRF as an iterative uncertainty propagation method, which could be applied in practice for the problems addressed in this study.

In order to model the distribution of input variables, a reasonable Gaussian error assumption is made about generating process u of input variables x. Thus, the distribution of input variables can be presented as $p(x) = \mathcal{N}(u, \Sigma_x)$. The new data point for prediction will be annotated as x_*. In the general case, we predict on the entire set of points representing a single snapshot of a network, so we annotate testing points with X_*.

The distribution of the target variable can then be expressed by the marginalization of input variables distribution:

$$p(y_*|\mathcal{D}) = \int p(y_*|X_*, \mathcal{D}) p(X_*) d_{X_*}. \quad (7)$$

As the distribution of $p(y_*|X_*, \mathcal{D})$ is Gaussian in the GCRF model, and the distribution of X_* is conjugate to the target variable distribution, marginal distribution $p(y_*|\mathcal{D})$ is a Gaussian as well. Since this integral is intractable for estimation in most of the cases, potential ways of solving it include sampling methods, variational Bayes or direct approximation of the moments of distribution as shown in (Girard 2004). For large or complex non–linear parametrized models, sampling–based uncertainty propagation is often computationally infeasible. This work is focused on approximating moments of the resulting distribution in Eq. 7, similarly to (Girard et al. 2003), however extended for evolving networks.

It is useful to first formalize the conditional Gaussian prediction form of the GCRF at point X_*

$$P(y_*|X_*) = \mathcal{N}\left(\begin{bmatrix} \mu \\ \mu_* \end{bmatrix}, \begin{bmatrix} \Sigma & \Sigma_* \\ \Sigma_*^T & \Sigma_{**} \end{bmatrix} \right), \quad (8)$$

where predictive mean μ_* and variance Σ_{**} of the network are defined in Eq. 3 and as inverse of Eq. 2 respectively.

In order to approximate the resulting distribution in Eq. 7, we approximate its first two moments (Girard et al. 2003). Moments can be expressed using the law of iterated expectation and conditional variance and solved using Laplace's method. Such methods of uncertainty propagation that are done by truncating multi–dimensional Taylor expansions of quantities of interest in order to approximate uncertainty criteria are called perturbation methods in the literature. Accuracy of such approach is governed by the order of Taylor expansion (Smith 2013).

The first moment of the distribution specified in Eq. 7 does not provide any correction over the zero'th order, within the first order Taylor expansion.

The second moment ($v(X_*)$) is approximated by the sec-

ond order Taylor expansion and its approximation yields:

$$v(X_*) = \Sigma_{**}\Big|_{X=\mu_{X_*}} + \frac{1}{2}Tr\left[H_{\Sigma_{**}}\{\Sigma_{X_*}\}\right] + J_{\mu_*}^T\{\Sigma_{X_*}\}J_{\mu_*}, \tag{9}$$

where we find several new terms. Σ_{X_*} is introduced as variance from distribution of X_*. The notation $\{\Sigma_{X_*}\}$ serves to signify that rather than maintaining a single covariance matrix for all nodes in the graph, we can opt for maintaining a covariance matrix for each node in the graph. Jacobian $J_{\Sigma_{**}}$ simplifies $\nabla_d \frac{\partial \Sigma_{**}}{\partial X_*^{(d)}}\Big|_{X=\mu_{X_*}}$, and Hessian $H_{\Sigma_{**}}$ simplifies $\nabla_{d,e} \frac{\partial^2 \Sigma_{**}}{\partial X_*^{(d)} \partial X_*^{(e)T}}\Big|_{X=\mu_{X_*}}$.

This is a point where information from distribution of input variables X provides a correction over predictive uncertainty of the GCRF. We see from Eq. (9) that there is a correction of the predictive variance: $\frac{1}{2}Tr\left[H_{\Sigma_{**}}\{\Sigma_{X_*}\}\right] + J_{\mu_*}^T\{\Sigma_{X_*}\}J_{\mu_*}$, influenced by the distribution of input variables via $\{\Sigma_{X_*}\}$. By solving partial derivatives J_{μ_*}, $J_{\Sigma_{**}}$ and $H_{\Sigma_{**}}$, we obtain corrected predictive variance that includes uncertainty coming from input variables. As we cannot analytically determine Σ_{**} we use the derivative of an inverse rule to solve $J_{\Sigma_{**}}$:

$$J_{\Sigma_{**}} = -\nabla_d \Sigma_{**} \frac{\partial \Sigma_{**}^{-1}}{\partial x_*^{(d)}} \Sigma_{**}, \tag{10}$$

and for the Hessian $H_{\Sigma_{**}}$:

$$H_{\Sigma_{**}} = \nabla_{d,e}\Sigma_{**}\left(2\frac{\partial \Sigma_{**}^{-1}}{\partial X_*^{(d)}}\Sigma_{**}\frac{\partial \Sigma_{**}^{-1}}{\partial X_*^{(e)}} - \frac{\partial^2 \Sigma_{**}^{-1}}{\partial X_*^{(d)} X_*^{(e)T}}\right)\Sigma_{**}. \tag{11}$$

$$J_{\mu_*} = \nabla_d - \Sigma_{**}\frac{\partial \Sigma_{**}^{-1}}{\partial x_*^{(d)}} 2\alpha X_* \theta + \Sigma_{**} 2\alpha \theta^{(d)T}, \tag{12}$$

where Jacobian in Eq. 12 is solved for the case when only one linear predictor is used. First and second derivatives of Σ_{**} can be calculated from Eq. 2.

Using derivations from Eq. (10), (11), (12), which are specific to the GCRF model, in the equation of approximated variance (9), we obtain corrected variance for the GCRF model. The model's predictive variance is dependent on variance of input data, assuming input data has a Gaussian error. This allows the GCRF model to be sensitive to significant changes on input data distribution, which results in higher predictive variance when predicting in the unknown.

To ensure propagation of uncertainty we apply the iterative approach to multiple-steps-ahead prediction, since we include uncertainty that is accumulating from the input variables (Candela et al. 2003; Girard et al. 2003).

Uncertainty propagation In order to properly model previous outputs as inputs as we predict ahead in time, lagged outputs are observed as random variables. The input vectors, will also be random variables, as they incorporate predictions recursively, $X_T \sim \mathcal{N}(\mu_{X_{T+k}}, \Sigma_{X_{T+k}})$. Note that for each node in a network we will maintain a $\mathcal{N}(\mu_{X_{T+k}}, \Sigma_{X_{T+k}})$ distribution. After each successive prediction, as new predicted values become inputs for the next prediction, Σ_{X_*} needs to be updated accordingly. In order to update $\Sigma_{X_{T+k}}$ for the new input $\hat{y}_{T+k}$, we need to calculate cross-covariance terms $\Sigma_{X_{T+k}}$ using

$$cov(y_{T+k}, X_{T+k}) = J_{\mu_{T+k}}^T\{\Sigma_{X_{T+k}}\}. \tag{13}$$

Now that we have all components needed, an inference procedure that handles noisy inputs defined as lagged predictions is described as Algorithm 1.

Algorithm 1 Multiple–steps–ahead GCRF regression

Input: Test data **X**, **model** $(\alpha_k, \beta_l, \theta_k, \psi_l)$
1. Initialize Σ_{X_*} for each node in a graph with all zeroes
2. Make a one–step–ahead prediction of $\hat{y}_{T+1}$
for $k = 2...K$ **do**
3. Update inputs according to the previous predictions $\hat{y}_{T+k-1}$
4. Update $\{\Sigma_{X_*}\}$ for the previously introduced noisy input using Eq. (13)
5. Predict following time step $\hat{y}_{T+k}$ using Eq. 3 and Eq. 9
end for

Data

Inpatient discharge data: We used the State Inpatient Databases (SID)[1] California database provided by the Agency for Healthcare Research and Quality and is included in the Healthcare Cost and Utilization Project (HCUP) . This dataset contains 35,844,800 inpatient discharge records over 9 years (from January 2003 to December 2011) collected from 474 hospitals. For each patient there are up to 25 diagnosis codes together with additional inpatient information. Problems considered in this study are long-term prediction of admission and mortality rate for each diagnoses out of 260 CCS coded diagnoses groups in a comorbidity graph that is constructed in monthly resolution for these 9 years of data.

Climate precipitation data: A dataset of precipitation records from meteorological stations across the USA has been acquired from NOAA's National Climate Data Center (NCDC) (Menne, Williams, and Vose 2009). A temporal graph is constructed on monthly resolution such that nodes at each time slice represent one hundred stations from the North West region of the USA, where we observed less than 5% of missing data in the months used for evaluations. Predictive problem from this domain is long–term monthly precipitation amount prediction in different weather stations.

As mentioned in the methodology section, graph structure for the described datasets will be learned within the proposed structured model, such that predictive power of the model is maximized.

[1]HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2005-2009. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp

Experimental results

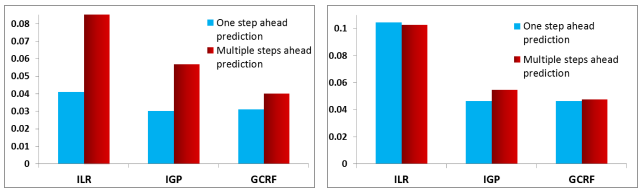
Set-up of the experiments conducted on two real-world datasets from the medical and climate domains, and results in terms of predictive error (Mean Squared Error - MSE) and plots of predictions with propagating uncertainty are reported in this section. The results of the three iterative models clearly demonstrate benefits of the structured GCRF model, which, in addition to learning a linear mapping $x \rightarrow y$, improved accuracy by including information from the graph structure.

The obtained propagation of uncertainty as the model incrementally predicts further in the future was significantly better than alternatives and it follows the change of data distribution for the GCRF model, while previously developed non structured model IGP often fails to do so. Specific findings on predicting admission and mortality rate based on inpatient discharge data and on predicting precipitation over the North West region of the US are reported in the following subsections.

Experiments on disease networks

For disease admission and for mortality rate prediction we have trained our models on 36 monthly snapshots and iteratively predicted for the following 48 months. For each disease we have used 18 months of previous target variable values as inputs for training and prediction. Admission count for each disease has been normalized. Mortality rate is defined as the number of patients that have died with a disease as the primary cause divided by the number of patients who were admitted to hospitals with the same disease as the primary one.

Experimental results on admission rate prediction Mean Squared Error for three algorithms for one step and multiple steps ahead prediction are shown at Figure 2(a). As

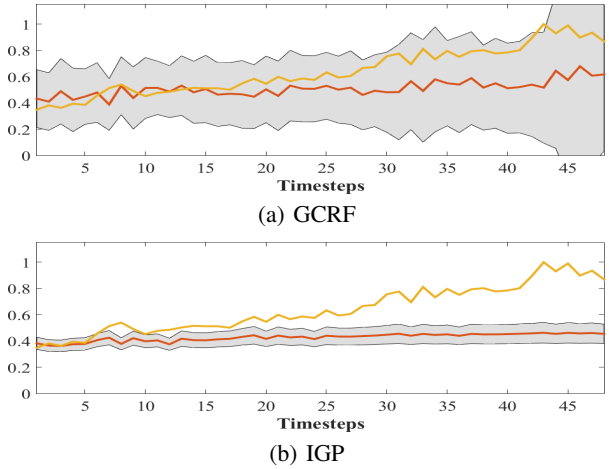


(a) admission rate (b) mortality rate

Figure 2: MSE of one (blue) and multiple (red) 48 months ahead predictions of admission rate (a) and mortality (b) for all 260 disease categories obtained by 3 methods.

expected, multiple–steps–ahead prediction has larger MSE when compared to the prediction of the first time step only. While accuracy of the proposed GCRF model with linear unstructured predictor is comparable to nonlinear IGP for a single step prediction of admissions and mortality rate, in both applications the extended GCRF was more accurate for the long–term predictions, which clearly demonstrates benefits of using the information from the structure.

In Figure 3, we show prediction and uncertainty propagation of GCRF and IGP for Septicemia disease (we do not show ILR since the accuracy was bad for multiple steps ahead prediction and the model does not provide satisfactory uncertainty propagation). We observe that there was a huge change in the test data distribution of admission rate of Septicemia vs. training distribution and so models failed to predict a huge increase in the number of sepsis related admissions that occurred after some point in future. As predic-



(a) GCRF

(b) IGP

Figure 3: [y–axis]: Predictions (red lines) and uncertainty estimates (gray area) of GCRF and IGP models for Septicemia disease **admission rate** (orange line) for up to 48 months (4 years) ahead [x–axis].

tion error was accumulating, the uncertainty propagation for the extended GCRF model properly followed the errors the model was making, which is a desirable feature for a predictive model. This is due to the change of distribution of input variables, which are moving away from the distribution of the input features on the training data, causing the correction term from Eq. 9 to become larger and larger. However, such a behavior is not observed when using the IGP model, which is due to the stale predictions where the model's inputs do not change as IGP's predictions vary only slightly.

Depending on the purpose and the precision needed for decision making, we may propose that when predicting the number of admissions up to 24 months ahead, results obtained by the extended GCRF model were acceptably reliable, however after that, we should consider waiting for more input values as indicated by increased uncertainty. On the other hand, if we were to trust the IGP based self-estimate of confidence, we would make a huge error in prediction estimate, as early as after 7-th month of prediction and uncertainly bounds would not provide any evidence that these predictions are poor.

Experimental results on mortality rate prediction Results for the mortality rate prediction are shown in the Figure 2(b), and follow similar pattern as results for disease admission rate prediction. In the following several plots presented at Figure 4, we show predictions using the extended GCRF with their confidence intervals for the top six killing diseases in California for the period of 2003–2011. In all

experiments the uncertainty of the model rightly reflected the noisiness of the real time signal, which is illustrating the power of the extended GCRF model to properly model uncertainty. Additionally, in all experiments we found that there was no uncertainty propagation unless the model starts making errors in the prediction and the distribution of the inputs changes, in which case uncertainty of the model increases properly.

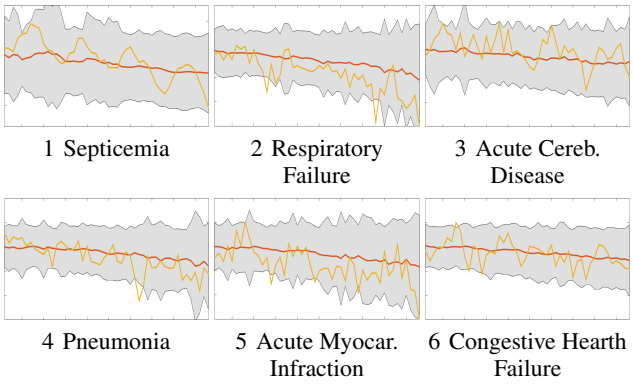


Figure 4: [y–axis]: **Mortality rate** predictions (red lines) and uncertainty estimates (gray area) by the extended GCRF for six of the most frequent diseases causing death in California for up to 48 months (4 years) ahead [x–axis]. True mortality is shown as orange lines.

Experiments on precipitation network

Precipitation models were trained on 48 months of data and tested on the following 96 months. As inputs, models were learned using 12 previous values of precipitation variables for each station in the region. Prediction for 96 months ahead is done such that lagged predictions of the previous 12 months are used as input attributes of the models and prediction is done iteratively for those 96 snapshots of the network. Accuracy results obtained by three models for immediate and for long-term (96 steps ahead) predictions are shown in Figure 5. Again the extended GCRF model provided the best accuracy of the used uncertainty propagation models in both one–step–ahead and multiple–steps–ahead prediction. The predictive accuracy results clearly demonstrate that the structured GCRF model with linear predictor outperforms both linear ILR and nonlinear IGP models, which were previously successfully applied on the climate domain (Drignei 2009; Sivakumar and Berndtsson 2010), on both single and multiple–steps–ahead prediction.

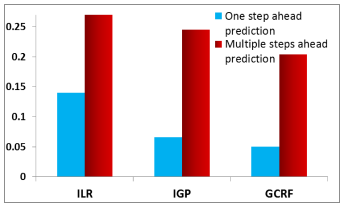
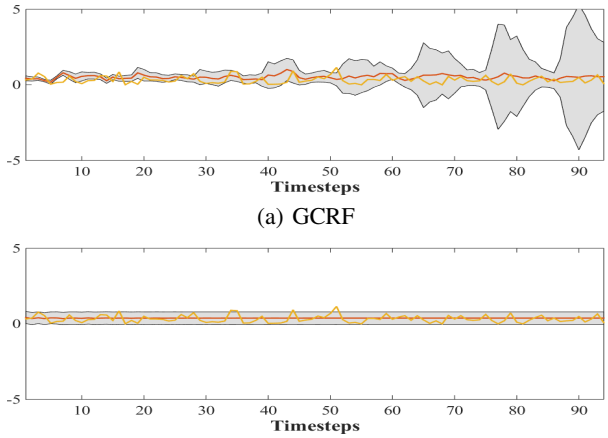


Figure 5: MSE of one (blue) and multiple (red) 96 months ahead predictions of **precipitation** on all stations by 3 methods.

The uncertainty propagation when predicting multiple steps ahead is demonstrated in Figure 6, on the previously described setup. Plots show 96 months of prediction for the GCRF model in Figure 6(a) and results are compared to those of the Gaussian Processes regression in Figure 6(b). The plot reflects that the GCRF can offer certain predictions



(a) GCRF

(b) IGP

Figure 6: [y-axis]: Precipitation predictions (red lines) and uncertainty estimates (gray area) by GCRF and IGP iterative models for up to 96 months (8 years) ahead [x–axis]. True precipitation is shown as orange lines.

up to about 30 months ahead, after which uncertainty is accumulated on the displayed station of the network. GCRF also accurately highlights areas with larger error across the entire predictive horizon. On the other hand, IGP in Figure 6(b) shows virtually no uncertainty propagation due to the smoothness of the prediction, even though the model is making errors. After a relatively small number of steps, predictions by the IGP model have converged and as lagged inputs did not change, there was no change in input distribution. Thus, there was no uncertainty propagation as the IGP model has fully propagated after only a few steps.

Conclusion

In this study we developed the iterative uncertainty propagation method for the GCRF model that allows controlled multiple-steps-ahead prediction with modeling uncertain inputs, as well as graph structure learning. The experiments were conducted on real-world networks from medical and climate domains for the long-term predictions, where the proposed model outperformed baselines in terms of both accuracy and uncertainty propagation. The proposed method is also readily applicable to other long-term structured regression models with Gaussian canonical form and for applications where correlation among outputs carries information.

Acknowledgments

The authors gratefully acknowledge the support of DARPA grant FA9550-12-1-0406 negotiated by AFOSR, NSF BIGDATA grant 14476570 and ONR grant N00014-15-1-2729.

References

Bengio, Y. 2012. Practical recommendations for gradient-based training of deep architectures. In *Neural Networks: Tricks of the Trade*. Springer. 437–478.

Candela, J. Q.; Girard, A.; Larsen, J.; and Rasmussen, C. E. 2003. Propagation of uncertainty in bayesian kernel models-application to multiple-step ahead forecasting. In *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*, volume 2, II–701. IEEE.

Dey, S.; Gyorgy, J.; B.L., W.; Steinbach, M.; and Kumar, V. 2014. Mining interpretable and predictive diagnosis codes from multi-source electronic health records. In *SIAM International Conference on Data Mining*.

Dokic, T.; Dehghanian, P.; Chen, P.-C.; Kezunovic, M.; Medina-Cetina, Z.; Stojanovic, J.; and Obradovic, Z. 2016. Risk assesment of a transmission line insulation breakdown due to lightning and sever weather. In *Hawaii Intermational Conference on System Sciences, HICSS-49*.

Drignei, D. 2009. A kriging approach to the analysis of climate model experiments. *Journal of agricultural, biological, and environmental statistics* 14(1):99–114.

Ghalwash, M. F.; Radosavljevic, V.; and Obradovic, Z. 2014. Utilizing temporal patterns for estimating uncertainty in interpretable early decision making. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 402–411. ACM.

Girard, A.; Rasmussen, C. E.; Quinonero-Candela, J.; and Murray-Smith, R. 2003. Gaussian process priors with uncertain inputs – application to multiple-step ahead time series forecasting. In *Neural Information Processing Systems*.

Girard, A. 2004. *Approximate methods for propagation of uncertainty with Gaussian process models*. Ph.D. Dissertation.

Gligorijevic, D.; Stojanovic, J.; and Obradovic, Z. 2015. Improving confidence while predicting trends in temporal disease networks. In *4th Workshop on Data Mining for Medicine and Healthcare, 2015 SIAM International Conference on Data Mining*.

Kocijan, J.; Murray-Smith, R.; Rasmussen, C. E.; and Girard, A. 2004. Gaussian process model based predictive control. In *American Control Conference*. IEEE.

Kocijan, J.; Girard, A.; Banko, B.; and Murray-Smith, R. 2005. Dynamic systems identification with gaussian processes. *Mathematical and Computer Modelling of Dynamical Systems* 11(4):411–424.

Menne, M.; Williams, C.; and Vose, R. 2009. The us historical climatology network monthly temperature data.

Radosavljevic, V.; Vucetic, S.; and Obradovic, Z. 2010. Continuous conditional random fields for regression in remote sensing. In *19th European Conf. on Artificial Intelligence*.

Radosavljevic, V.; Vucetic, S.; and Obradovic, Z. 2014. Neural gaussian conditional random fields. In *Proc. European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*.

Ristovski, K.; Radosavljevic, V.; Vucetic, S.; and Obradovic, Z. 2013. Continuous conditional random fields for efficient regression in large fully connected graphs. In *Association for the Advancement of Artificial Intelligence Conference*.

Sivakumar, B., and Berndtsson, R. 2010. *Advances in data-based approaches for hydrologic modeling and forecasting*. World Scientific.

Smith, R. C. 2013. *Uncertainty Quantification: Theory, Implementation, and Applications*, volume 12. SIAM.

Stiglic, G.; Wang, F.; Davey, A.; and Obradovic, Z. 2014. Readmission classification using stacked regularized logistic regression models. In *AMIA 2014 Annual Symposium*.

Stojanovic, J.; Jovanovic, M.; Gligorijevic, D.; and Obradovic, Z. 2015. Semi-supervised learning for structured regression on partially observed attributed graphs. In *SIAM International Conference on Data Mining*.

Wytock, M., and Kolter, Z. 2013. Sparse gaussian conditional random fields: Algorithms, theory, and application to energy forecasting. In *Proc. of the 30th International Conference on Machine Learning (ICML-13)*, 1265–1273.