Supplementary Material: Uncertainty Propagation in Long-term Structured Regression on Evolving Networks

Djordje Gligorijevic, Jelena Stojanovic and Zoran Obradovic
{gligorijevic, jelena.stojanovic, zoran.obradovic}@temple.edu

Appendix A: Modelling noisy inputs derivations

In order to model the distribution of input variables, a reasonable assumption is that input variables x are generated by some process u, and that process has a Gaussian error. 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 these testing points with X_*.

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

$$p(y_*|\mathcal{D}) = \int p(y_*|X_*, \mathcal{D}) p(X_*) d_{X_*}. \tag{1}$$

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 [1]. For large or complex non–linear parametrized models, sampling–based uncertainty propagation is often computationally infeasible, thus this work is focused on approximating moments of the resulting distribution in Eq. 1, similarly to [2], however applied to evolving networks.

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

$$P(y_*|X_*) = \mathcal{N}\left(\begin{bmatrix}\mu \\ \mu_*\end{bmatrix}, \begin{bmatrix}\Sigma & \Sigma_* \\ \Sigma_*^T & \Sigma_{**}\end{bmatrix}\right), \tag{2}$$

with predictive mean μ_* and variance Σ_{**}.

In order to approximate the resulting distribution in Eq. 1, we approximate its first two moments. They 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 an approach is governed by the order of Taylor expansion [3].

Approximating first moment - the mean

The first moment of the distribution specified in Eq. 1 can be estimated by the Law of iterated expectations:

$$m(X_*) = E_{X_*}[E[y_*|X_*]] = E_{X_*}[\mu_*]. \tag{3}$$

The predictive mean $m(X_*)$ can be estimated by approximating μ_* by its first order Taylor expansion around μ_{X_*} (by μ_{X_*}, we annotate the mode of the distribution of input variables of all nodes in the graph X_*)

$$m(X_*) = \mu_*\Big|_{X=\mu_{X_*}} + J_{\mu_*}^T(X_* - \mu_{X_*}) + O(\|X_* - \mu_{X_*}\|^2), \tag{4}$$

where Jacobian

$$J_{\mu_*} = \nabla_d \frac{\partial \mu_*}{\partial X_*^{(d)}}\Big|_{X=\mu_{X_*}}. \tag{5}$$

The expected value of this Taylor expansion yields

$$E_{X_*}[\mu_*] = E_{X_*}\left[\mu_*\Big|_{X=\mu_{X_*}} + J_{\mu_*}^T(X_* - \mu_{X_*})\right] = \mu_* \tag{6}$$

We can see that, within the first order Taylor expansion, the prediction mean at any y_* does not provide any correction over the zero'th order.

Approximating second moment - the variance

The second moment is estimated by the Law of conditional variance:

$$v(X_*) = E_{X_*}[var_{y_*}(y_*|X_*)] + var_{X_*}(E_{y_*}[y_*|X_*]) = E_{X_*}[\Sigma_{**}] + var_{X_*}(\mu_*)$$

In order to obtain predictive variance, $v(X_*)$, on the other hand, we need to approximate $E_{X_*}[\Sigma_{**}]$ and $var_{X_*}(\mu_*)$. The natural choice for approximating $E_{X_*}[\Sigma_{**}]$ is second order Taylor expansion:

$$\Sigma_{**} = \Sigma_{**}\Big|_{X=\mu_{X_*}} + J_{\Sigma_{**}}^T(X_* - \mu_{X_*}) + \frac{1}{2}(X_* - \mu_{X_*})^T H_{\Sigma_{**}}(X_* - \mu_{X_*}) + O(\|X_* - \mu_{X_*}\|^3), \tag{7}$$

where Jacobian and Hessian are:

$$J_{\Sigma_{**}} = \nabla_d \frac{\partial \Sigma_{**}}{\partial X_*^{(d)}}\Big|_{X=\mu_{X_*}}, \tag{8}$$

$$H_{\Sigma_{**}} = \nabla_{d,e} \frac{\partial^2 \Sigma_{**}}{\partial X_*^{(d)} \partial X_*^{(e)T}}\Big|_{X=\mu_{X_*}}. \tag{9}$$

The part of the Eq. 7, $\frac{1}{2}(X_* - \mu_{X_*})^T H_{\Sigma_{**}}(X_* - \mu_{X_*}))$ is solved using the expression of quadratic form under Gaussian:

$$\int (y-\mu)^T \Sigma^{-1}(y-\mu)\mathcal{N}(\mu_X, \Sigma_X)d_x = (y-\mu)^T\Sigma^{-1}(y-\mu) + Tr[\Sigma^{-1}\Sigma_X]. \tag{10}$$

The expected value of then becomes:

$$E_{X*}[\Sigma_{**}] = \Sigma_{**}\Big|_{X=\mu_{X_*}} + \frac{1}{2}Tr\left[H_{\Sigma_{**}}\{\Sigma_{X_*}\}\right], \tag{11}$$

where we find a new term Σ_{X_*} 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. This is a point where information from distribution of input variables X provides a correction over predictive uncertainty of the GCRF.

Now, we can calculate $var_{X_*}(\mu_*)$ using previously obtained $m(X_*)$

$$var_{X_*}(\mu_*) \approx var_{X_*}\left(\mu_*\Big|_{X=\mu_{X_*}} + J_{\mu_*}^T(X_* - \mu_{X_*})\right) = J_{\mu_*}^T\{\Sigma_{X_*}\}J_{\mu_*} \tag{12}$$

After combining the previous two results for $E_{X*}[\Sigma_{**}]$ and $var_{X_*}(\mu_*)$, we obtain the expression for predictive variance:

$$v(X_*) = \Sigma_{**}\Big|_{X=\mu_{X_*}} + \frac{1}{2}Tr\left[H_{\Sigma_{**}}\{\Sigma_{X_*}\}\right] + J_{\mu_*}^T\{\Sigma_{X_*}\}J_{\mu_*}. \tag{13}$$

We see from Eq. (13) that there is a correction of the predictive variance influenced by the distribution of input variables via $\{\Sigma_{X_*}\}$. By solving partial derivatives, we can obtain corrected predictive variance that includes uncertainty coming from input variables. Finally solutions of the three partial derivatives are needed to complete the correction term expression: J_{μ_*}, $J_{\Sigma_{**}}$ and $H_{\Sigma_{**}}$. As we cannot analytically determine Σ_{**} we use the derivative of an inverse rule to solve $J_{\Sigma_{**}}$:

$$J_{\Sigma_{**}} = -\nabla_d \Sigma_{**}\frac{\partial \Sigma_{**}^{-1}}{\partial x_*^{(d)}}\Sigma_{**}, \tag{14}$$

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

$$H_{\Sigma_{**}} = \nabla_{d,e}\Sigma_{**}\left(2\frac{\partial \Sigma_{**}^{-1}}{\partial X_*^{(d)}}\Sigma_{**}\frac{\partial \Sigma_{**}^{-1}}{\partial X_*^{(e)}} - \frac{\partial^2 \Sigma_{**}^{-1}}{\partial X_*^{(d)}X_*^{(e)T}}\right)\Sigma_{**}. \tag{15}$$

$$J_{\mu_*} = \nabla_d - \Sigma_{**}\frac{\partial \Sigma_{**}^{-1}}{\partial x_*^{(d)}}2\alpha\theta^T X_* + \Sigma_{**}2\alpha\theta^{(d)T}, \tag{16}$$

where Jacobian in Eq. 16 is solved for case when only one linear predictor is used. First and second derivatives of Σ_{**} can be calculated from the Precision matrix of the GCRF model.

$$\frac{\partial \Sigma_{**}}{\partial X_*^d} = \begin{cases} 2\sum_g \beta_l S(x_i, x_g, \psi_l)\frac{\partial S(x_i,x_g,\psi_l)}{\partial x_i^{(d)}}, i = j \\ -2\sum_g \beta_l S(x_i, x_g, \psi_l)\frac{\partial S(x_i,x_g,\psi_l)}{\partial x_i^{(d)}}, i \neq j \end{cases} \tag{17}$$

$$\frac{\partial \Sigma_{**}^{-1}}{\partial x_*^d x_*^e} = \begin{cases} 2\sum_g \beta_l (S(x_i, x_g, \psi_l)\frac{\partial S(x_i,x_g,\psi_l)}{\partial x_i^{(d)}}\frac{\partial S(x_i,x_g,\psi_l)}{\partial x_i^{(e)}} + S(x_i, x_g, \psi_l)\frac{\partial S(x_i,x_g,\psi_l)}{\partial^2 x_i^{(d)}x_i^{(e)}}), i = j \\ -2\sum_g \beta_l S(x_i, x_g, \psi_l)\frac{\partial S(x_i,x_g,\psi_l)}{\partial x_i^{(d)}}\frac{\partial S(x_i,x_g,\psi_l)}{\partial x_i^{(e)}} + S(x_i, x_g, \psi_l)\frac{\partial S(x_i,x_g,\psi_l)}{\partial^2 x_i^{(d)}x_i^{(e)}}), i \neq j \end{cases} \tag{18}$$

Using derivations obtained in the Eq. (14), (15), (16), which are specific to the GCRF model, in the equation of approximated variance (13), we obtain corrected variance for the GCRF model. Now 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 should then apply the iterative approach to multiple-steps-ahead prediction, since we now include uncertainty that is accumulating from the input variables [2, 4].

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 $\hat{y}_{T+k}$, all we need to do is to calculate

$$cov(\hat{y}_{T+k}, X_{T+k}) = E_x[E_y[\hat{y}_{T+k} \times X_{T+k}]] - E[\hat{y}_{T+k}]E[X_{T+k}], \tag{19}$$

with $E[\hat{y}_{T+k}]$ given as the prediction of the model and $E[X_{T+k}] = \mu_{X_{T+k}}$. We only have to estimate expected value of product of the two random variables which can be expressed as:

$$E_X[E_y[y_{T+k} \times X_{T+k}]] = \int X_{T+k} \left[\mu_{T+k}\Big|_{X=\mu_{X_{T+k}}} + J^T_{\mu_{T+k}}(X_{T+k} - \mu_{X_{T+k}}) \right] p(X_{T+k}) dx_{T+k}. \tag{20}$$

This gives,

$$E_x[E_y[y_{T+k} \times X_{T+k}]] = \mu_{T+k}\Big|_{X=\mu_{X_{T+k}}} \mu_{X_{T+k}} + J^T_{\mu_{T+k}}\{\Sigma_{X_{T+k}}\}. \tag{21}$$

So that the cross-covariance terms of the $\Sigma_{X_{T+k}}$ are given by

$$cov(y_{T+k}, X_{T+k}) = J^T_{\mu_{T+k}}\{\Sigma_{X_{T+k}}\}. \tag{22}$$

Now, that we have all components needed 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 $\mathbf{X}$, **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. (22)

  5. Predict following time step $\hat{y}_{T+k}$ using non-corrected models predictions and Eq. 13

end for

Appendix B: Gaussian Conditional Random Fields

Gaussian Conditional Random Fields (GCRF) [5] is a structured regression model. The model 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 (not only chains or trees), and can represent the structure as a function of time, space, or any other user-defined structure. It models the structured regression problem as estimation of a joint continuous distribution over all nodes and takes the following log-linear form:

$$P(y|X) = \frac{1}{Z} exp(-\sum_{i=1}^{K}\sum_{k=1}^{K} \alpha_k (y_i - R_k)^2 - \sum_{i\sim j}\sum_{l=1}^{L} \beta_l {S_{ij}}^{(l)} (y_i - y_j)^2) \tag{23}$$

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 functions are any *pre–trained* unstructured predictors that map $X \rightarrow y_i$ independently, and might also be used to incorporate domain specific models. Similarity *matrix* S^l is used to define the weighted undirected graph structure between labels.

This choice of quadratic feature functions enables representation of this distribution as a multivariate Gaussian [5] to ensure efficient and convex optimization:

$$P(y|X) = \frac{1}{(2\pi)^{\frac{N}{2}} \mid \Sigma \mid^{\frac{1}{2}}} exp\left(-\frac{1}{2}(y-\mu)^T \Sigma^{-1} (y-\mu)\right) \tag{24}$$

where Σ^{-1} represents the inverse covariance matrix:

$$\Sigma^{-1} = \begin{cases} 2\sum_{k=1}^{K} \alpha_k + 2\sum_g \sum_{l=1}^{L} \beta_l S_{ig}^{(l)}, i = j \\ 2\sum_{l=1}^{L} \beta_l S_{ij}^{(l)}, i \neq j \end{cases} \tag{25}$$

The posterior mean is given by

$$\mu = \Sigma \mathbf{b}, \tag{26}$$

where b is defined as

$$b_i = 2\left(\sum_{k=1}^{K} \alpha_k R_k\right). \tag{27}$$

This specific way of modeling will allow 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 and inference: The learning task is to optimize parameters α and β by maximizing the conditional log–likelihood,

$$(\hat{\alpha}, \hat{\beta}) = \underbrace{argmax}_{\alpha,\beta} logP(y|X;\alpha,\beta). \tag{28}$$

Parameters α and β are learned by a gradient-based optimization. Gradients of the conditional log-likelihood are

$$\frac{\partial \mathcal{L}}{\partial \alpha_k} = -\frac{1}{2}(y-\mu)^T \frac{\partial \Sigma^{-1}}{\partial \alpha_k}(y-\mu) + (\frac{\partial b^T}{\partial \alpha_k} - \mu^T \frac{\partial \Sigma^{-1}}{\partial \alpha_k})(y-\mu) + \frac{1}{2} Tr(\Sigma \frac{\partial \Sigma^{-1}}{\partial \alpha_k}) \tag{29}$$

$$\frac{\partial \mathcal{L}}{\partial \beta_l} = -\frac{1}{2}(y+\mu)^T \frac{\partial \Sigma^{-1}}{\partial \beta_l}(y-\mu) + \frac{1}{2} Tr(\Sigma \frac{\partial \Sigma^{-1}}{\partial \beta_l}) \tag{30}$$

Maximizing the conditional log–likelihood is a convex objective, and can be optimized using standard Quasi-Newton optimization techniques. Note that there is one constraint that is needed to assure that the distribution is Gaussian, which is to make the Σ^{-1} matrix positive-semidefinite. To ensure this and make the optimization unconstrained, the exponential transformation of parameters $\alpha_k = e^{u_k}$ and $\beta_l = e^{v_l}$ is used [5]. 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.

Appendix C: Comparison models

Gaussian Processes as an iterative model

A Gaussian process regression is a powerful framework which is a generalization of a multivariate Gaussian distribution over finite vector space to a function space of infinite dimension [6]. Assumption of a Gaussian prior is present over functions that map $x \rightarrow y$. Gaussian processes are defined with

$$f(x) \sim GP(m(x), k(x, x')), \tag{31}$$

where $m(x)$ is the mean function and $k(x, x')$ is the covariance function in the form of a kernel that is required to be positive definite. In our experiments, we are using a Gaussian kernel

$$S(x_i, x_j, \psi) = \psi_0 exp\left(-\frac{1}{2}\sum_{d=1}^{D}\frac{(x_d^i - x_d^j)^2}{\psi_d^2}.\right). \tag{32}$$

If we denote covariance of the training part as $C = K + \sigma_y^2 I_N$, $K_{ij} = k(x_i, x_j)$, the joint density of the observed outputs $\mathbf{y}$ and test output y_* is presented as

$$\begin{pmatrix}\mathbf{y}\\ y_*\end{pmatrix} = \mathcal{N}\left(0, \begin{bmatrix} C & \mathbf{k}_* \\ \mathbf{k}_*^{\mathbf{T}} & c_* \end{bmatrix}\right), \tag{33}$$

where $\mathbf{k}_*$ is the covariance vector for the new test point $\mathbf{x}_*$. The posterior predictive density is given by

$$p(y_*|\mathbf{x}_*, \mathbf{X}, \mathbf{y}) = \mathcal{N}(y|0, C), \tag{34}$$

$$\mu_* = \mathbf{k}_*^{\mathbf{T}} C^{-1}\mathbf{y}, \qquad \sigma_*^2 = c_* - \mathbf{k}_*^{\mathbf{T}} C^{-1}\mathbf{k}_*. \tag{35}$$

The σ_*^2 is the predictive variance or uncertainty at test point x_* [6]. In order to take account of the uncertainty of future predictions which provide the 'inputs' for estimating further means and uncertainties, test points are considered as random inputs $\mathbf{x}_* \sim \mathcal{N}(\mu_{x_*}, \Sigma_{x_*})$ and a Gaussian approximation shown in [2, 4, 7]. Then, the predictive distribution of this iterative uncertainty propagation method has mean and variance

$$\mu_* = \mathbf{k}(\mu(x_*))^T C^{-1}\mathbf{y}, \tag{36}$$

$$\sigma_*^2 = \sigma^2(\mu(x_*)) + \frac{1}{2}Tr\left(\Sigma_{x_*}\frac{\partial^2\sigma^2(x_*)}{\partial x_* \partial x_*}\right) + \frac{\partial\mu(x_*)}{\partial x_*}^T \Sigma_{x_*}\frac{\partial\mu(x_*)}{\partial x_*} \tag{37}$$

This approach has been successfully applied in the past in a model–based predictive control framework for control of pH process benchmark [8].

Linear regression as an iterative and direct model

From the family of direct uncertainty propagation models we will use a linear parameterized model (DLR) [3]. Linear regression form representation is:

$$y = Xw^T + \varepsilon, \varepsilon \sim \mathcal{N}(0, \sigma_y^2) \tag{38}$$

where w is an unknown set of weights. The weight and noise variance are estimated by $\hat{w} = (X^T X)^{-1} X^T y$ and

$$\sigma_y^2 = \frac{(y - X\hat{w}^T)^T (y - X\hat{w}^T)}{N - d - 1}, \tag{39}$$

where X is matrix representation of all data available for training, N is the number of training examples and d is the number of attributes. For the auto–regressive representation of Linear model we have variance estimation, given prediction y_{T+k} [3]:

$$\sigma_{T+k}^2 = \sigma_y^2 \left(1 + X_{T+k}(X^T X)^{-1} X_{T+k}^T\right). \tag{40}$$

Then, the construction of confidence intervals for the new prediction y_{T+k} are given by T–distribution with $n - d - 1$ degrees of freedom for $(1 - \alpha) \times 100\%$ interval estimator

$$\left[y_{T+k} \pm t_{n-d-1,\alpha/2} \hat{\sigma} \sqrt{1 + x_{T+k}(X^T X)^{-1} x_{T+k}^T}\right]. \tag{41}$$

In the experimental section, this model will be noted as DLR. We will also apply it in an iterative setup, and call it ILR.

GCRF with parameters sensitive to uncertainty of unstructured predictors (DGCRF)

To enable the GCRF model to propagate directly, we could allow it to be sensitive to the uncertainty of the unstructured predictors, as in [9]. The idea is to observe parameters α of the GCRF model described in Appendix B as functions rather than scalars. In order to allow GCRF to incorporate uncertainty, α can be treated as non parametric function

$$\alpha_{k,p} = \frac{e^{u_{k,p}}}{\sigma_{k,p}^2} ci_{k,p}. \tag{42}$$

where $\sigma_{k,1}^2$ represents the uncertainty estimation of unstructured predictor k for the p'th time step. Additionally, percentage of nodes that fall into the 95% confidence interval ($ci_{k,p}$) is used as a quality index to augment this approach. For the direct approach to uncertainty propagation we will use the Direct Linear Regression model described in Section as unstructured predictor.

Appendix D: Additional Experimental Results for Iterative Uncertainty Propagation Models

In this Appendix we provide additional experimental results to the ones that are presented in the main paper. As in there we show results in terms of predictive error (Mean Squared Error– MSE) and plots of predictions and propagating uncertainty.

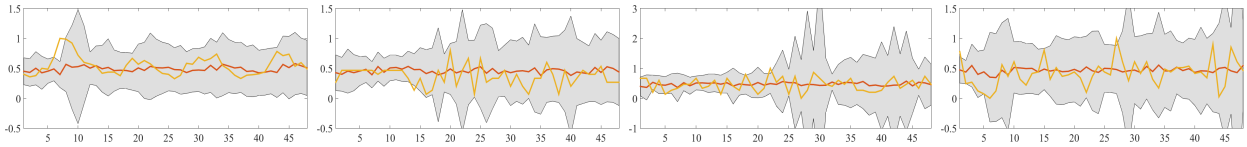
First, we show the additional results of the iterative methods obtained on HCUP data, and afterwards results of three types of experiments are shown:

- iterative models with lagged predictions as inputs - on disease networks,
- iterative models with input variables in addition to lagged predictions - on precipitation network,
- and direct models - on precipitation network.

Additional Experiments on Healthcare Data

Additional Experimental Results on Admission Rate Prediction

In addition to the Figures 2 and 3 from the main paper, we provide the following few figures in Figure 1. We show predictions and confidence intervals of the extended GCRF model for several different diseases. We can observe that there is no propagation if the model is doing a good job, since the correction term does not increase excessively as distribution of input features remains relatively unvaried. However, as soon as the model starts making errors in prediction and distribution of input variables shifts, the confidence interval widens.



(a) Respiratory Failure (b) Male Genital Organs Cancer (c) Disorder of Lipid Metabolism (d) Eye Disorders

Figure 1: Predictions (red lines) and uncertainty estimates (gray area) of GCRF model for **admission rates** of four different diseases (orange line) for 48 months (4 years) ahead (x–axis).

Additional Experimental Results on Mortality Rate Prediction

In addition to the top 6 killing diseases in California, shown in Figure 4 in the main paper, here we present prediction of mortality rate for top twelve killing diseases in the state of California for the extended GCRF model in Figure 2.

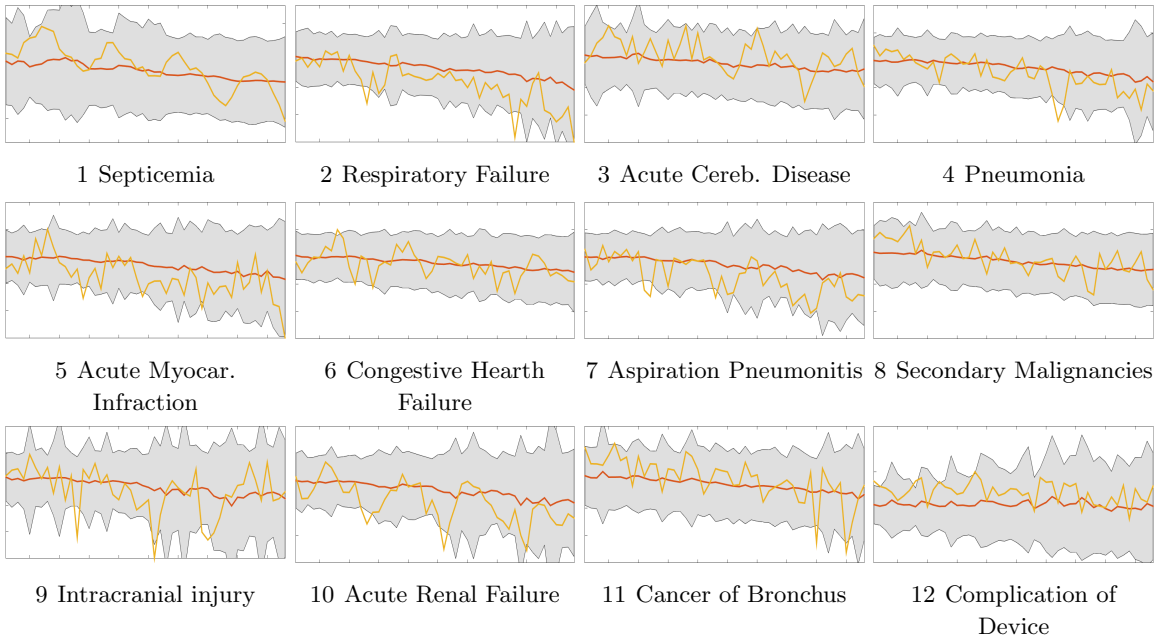


Figure 2: Predictions (red lines) and uncertainty estimates (gray area) of GCRF for **mortality rate** of top twelve killing diseases in California (orange lines) for 48 months (4 years) ahead (x–axis).

Experiments on precipitation network

Adding inputs in iterative predictions

In some applications, the input variables might be available in the future. In our climate application, in addition to precipitation, there are 6 more variables at each node which we use as input attributes for each station. These variables are acquired from the NCEP/NCAR Reanalysis 1 project [10], which is using a state-of-the-art analysis/forecast system to predict climate parameters using past data from 1948 to the present (data available on NOAA website: http://www.esrl.noaa.gov/psd/). These 6 variables are omega (Lagrangian tendency of air pressure), precipitable water, relative humidity, temperature, u-wind, and w-wind (zonal and meridional components of the wind, respectively).

Thus, iterative models can potentially include these input variables in their predictions. However, this way of modeling leads to larger input dimensionality of the models and results in more progressive uncertainty propagation, as shown in Figure 4. The results of this, iterative multiple steps ahead prediction with input variables in terms of predictive accuracy are shown in Figure 3.

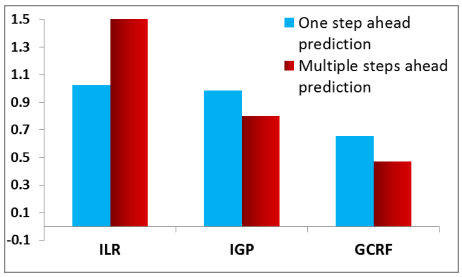
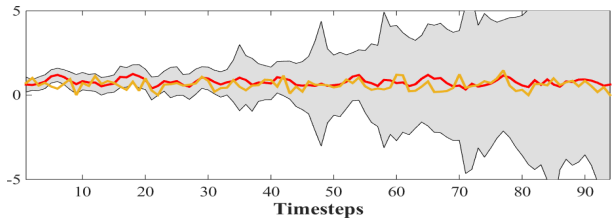
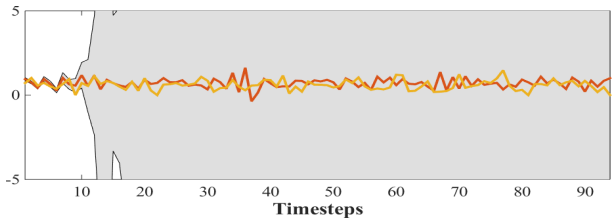


Figure 3: MSE of one (blue) and multiple (red) steps ahead predictions of **precipitation** on all stations using **iterative** methods **with included input variables**



(a) GCRF Station 1



(b) IGP Station 1

Figure 4: Predictions (red lines) and uncertainty estimates (gray area) of GCRF and IGP **iterative** models **with included inputs** for **precipitation** (true values– orange line) for 96 months (8 years) ahead

Appendix E: Experimental Results for Direct Uncertainty Propagation Models

Direct models of uncertainty propagation on precipitation data

In our climate application, input variables are available over entire predictive horizon as outputs from climate models, as mentioned in Appendix D. Thus, we were able to apply direct methods of uncertainty propagation (DLR and DGCRF described in Appendix C).

Note that this way of modeling does not take into account lagged predictions as inputs: only features of the nodes are used as inputs. The accuracy results are shown in Figure 5.

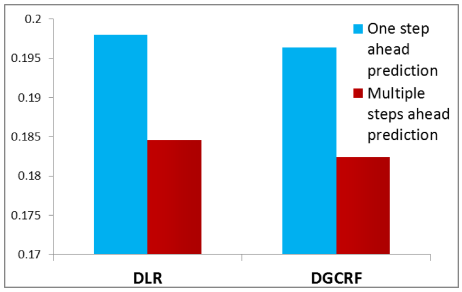
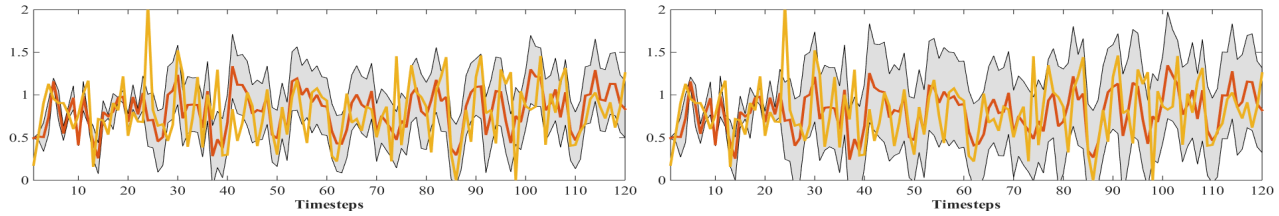


Figure 5: MSE of one (blue) and multiple (red) steps ahead predictions of **precipitation** on all stations using **direct** methods

In Figure 6 we show training and testing time steps to demonstrate the level of uncertainty propagation

for direct uncertainty propagation models. The DGCRF's uncertainty propagation is completely dependent on uncertainty propagation of the DLR model, as it is used as an unstructured predictor to the DGCRF model. We can see that uncertainty increases rapidly and then stabilizes after just a few time steps and remains relatively large over the entire prediction period.



(a) DGCRF (b) DLR

Figure 6: Predictions (red lines) and uncertainty estimates (gray area) of DGCRF and DLR **direct** models of **precipitation** (true values– orange line) for 96 months (8 years) ahead.

References

[1] Agathe Girard. *Approximate methods for propagation of uncertainty with Gaussian process models*. PhD thesis, 2004.

[2] Agathe Girard, Carl Edward Rasmussen, Joaquin Quinonero-Candela, and Roderick Murray-Smith. Gaussian process priors with uncertain inputs – application to multiple-step ahead time series forecasting. In *Neural Information Processing Systems*, 2003.

[3] Ralph C Smith. *Uncertainty Quantification: Theory, Implementation, and Applications*, volume 12. SIAM, 2013.

[4] Joaquin Quinonero Candela, Agathe Girard, Jan Larsen, and Carl Edward Rasmussen. 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, pages II–701. IEEE, 2003.

[5] Vladan Radosavljevic, Slobodan Vucetic, and Zoran Obradovic. Continuous conditional random fields for regression in remote sensing. In *Proc. 19th European Conf. on Artificial Intelligence*, 2010.

[6] C.E. Rasmussen and C.K.I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[7] Agathe Girard and Roderick Murray-Smith. Gaussian processes: Prediction at a noisy input and application to iterative multiple-step ahead forecasting of time-series. In *Switching and Learning in Feedback Systems*. Springer, 2005.

[8] Juš Kocijan, Roderick Murray-Smith, Carl Edward Rasmussen, and Agathe Girard. Gaussian process model based predictive control. In *American Control Conference*. IEEE, 2004.

[9] Djordje Gligorijevic, Jelena Stojanovic, and Zoran Obradovic. 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*, 2015.

[10] Eugenia Kalnay, Masao Kanamitsu, Robert Kistler, William Collins, Dennis Deaven, Lev Gandin, Mark Iredell, Suranjana Saha, Glenn White, John Woollen, et al. The ncep/ncar 40-year reanalysis project. *Bulletin of the American meteorological Society*, 1996.