NTNU
Norwegian University of
Science and Technology

Tolerance Limit Communication Schemes and Nowcasting for Resource Constrained Sensor Networks

Utilizing NWP, Temporal and Spatial Dependencies for a Temperature Sensor Network at Spitsbergen

Markus Mortensen

Master of Science in Physics and Mathematics
Submission date: July 2018
Supervisor: Ingelin Steinsland, IMF

Norwegian University of Science and Technology
Department of Mathematical Sciences

Abstract

In this thesis we consider resource constraint sensor networks that communicates with a fusion center. If communication and/or energy resources are limited, is it desirable to keep the number of transmissions from a sensor low. We suggest a communication scheme based on Temporal dependence and Tolerance limits and that is Trigger based (TTT). The TTT-communication scheme is based on, together with predictions about the process, tolerance limits and an updating scheme known to both a fusion center and a sensor. Through simulation studies we demonstrate that the number of observations a sensor monitoring a temporal process must transmit to a fusion center can be heavily reduced, without sacrificing much knowledge about the process of interest.

Further, we use the TTT-communication scheme for making hourly nowcasts of the temperature at 10 locations at Spitsbergen. In this case study we utilize weather forecasts from a physical weather model (numerical weather prediction, NWP) to improve the prediction in the TTT-communication scheme. The transmission rates for the sensors are reduced by $70-80\%$, while with a guaranteed maximum error of $\pm 1^{\circ}$C and RMSE below 0.45 for the nowcasted temperature.

Furthermore, by utilizing spatial dependence in the process, we extend the TTT- communication scheme to a spatial model for nowcasting, TTTS. Using the TTTS-model reduces the RMSE for the nowcasted temperature further, though with a guaranteed maximum error of $\pm 2^{\circ}$C. An algorithm for estimating the first and second moment of the truncated multivariate Gaussian variable is also presented.

Sammendrag

I denne studien ser vi på et ressursbegrenset sensornettverk som kan kommunisere med en sentral datamaskin. Det er ønskelig å begrense antall overføringer fra en sensor når den har begrensninger mht. kommunikasjon og/eller energiressurser. Vi foreslår et kommunikasjonsregime som er basert på tidsavhengighet, toleransegrense og er situasjonsstyrt (TTT). TTT-kommunikassjonsregimet er basert på prediksjoner for prosessen, toleransegrenser og et oppdaterende regime som er kjent for både en sentral datamaskin og for en sensor. Gjennom en simulasjonsstudie viser vi at antall overføringer som må overføres fra en sensor som overvåker en tidsavhengig prosess til en sentral datamaskin, kan reduseres kraftig uten å ofre mye kunnskap om prosessen.

Videre bruker vi TTT-kommunikasjonsregimet til å nå-predikere (nowcast) temperaturen hver time for 10 steder på Spitsbergen. I dette casestudie bruker vi værprediksjoner fra en fysisk værmodell (numerisk værprediksjon) til å forbedre prediksjonen fra TTT-kommunikasjonsregimet. Overføringsraten for sensorene er redusert med $70 - 80\%$, med en garantert maksimal feil på $\pm 1^{\circ}$C og RMSE under 0.45 for nå-prediksjonen av temperaturen.

I tillegg, ved å bruke romlig avhengighet i prosessen, utvider vi TTT- kommunikasjonsregimet til en romlig modell for nå-predikering, TTTS- Ved å bruke TTTS-modellen, reduserer vi RMSE for nå-prediksjonen mer, riktignok med en garantert maksimal feil på $\pm 2^{\circ}$C. Vi presenterer også en algoritme for estimering av de to første momentene av den trunkerte multivariate Gaussiske variabel.

Preface

This thesis is a part of the course TMA4905 Master's Thesis in Statistics at the Norwegian University of Science and Technology (NTNU), Department of Mathematical Sciences. It is the final part of the study program Industrial Mathematics. This work was carried out during the spring of 2018.

I would especially like to thank my supervisor, Professor Ingelin Steinsland. Our weekly meetings has in addition to being cheerful, provided me with excellent guidance. Her ideas and support have been invaluable for me during the work with this thesis. I would also like to thank the rest of the members in the Autonomous Adaptive Sensing research group at NTNU, for their contributions and ideas for this thesis.

Contents

Chapter 1

Introduction

In meteorology, nowcasting is a technique for describing the weather at the current state. The UK Met Office (2018) defines nowcasting as a very short-range forecasting that maps the current weather and uses the estimates for forecasting the weather a short period ahead. A detailed and sharp nowcast may significantly improve weather forecasts, as the forecaster has the possibility to correct for potential bias in the forecast and detect shifts in the weather systems. The weather models depends on accurate initial conditions in terms of the present state of the weather. The better and more detailed the available information is, the better prediction can be made by the forecaster. Traditional weather forecasts can be a good indicator for what the temperature for an area will be, but who hasn't checked their own thermometer and found that the forecast misses by several degrees? The forecasts models have a hard time fitting local conditions, and will never be able to perfectly predict the temperature at every location in the world. Nowcasting is therefore useful for both increasing the knowledge about the weather at the current state, and also for improving the forecasts for longer periods ahead.

Recently, the largest Norwegian weather forecasting service, The Norwegian Meteorological Institute (MET) and their online portal Yr.no, has opened up for private weather sensors to be included in their forecasts, MET Norway (2018). Yr reports that the forecasts are more accurate and that the forecasts are more frequently updated. Previously, the forecasts were updated every sixth hour. With the information from the private weather stations, the forecasts are now updated hourly. Netatmo, the provider of the weather stations that are included reports that the first winter these stations were incorporated in the forecasts, the rate of major temperature errors (over 3°C) are halved, NETATMO (2018). In addition, Netatmo claims that the geographic accuracy is multiplied by a factor by more than 6. Clearly, the introduction of private weather stations has improved the traditional weather forecasts. The weather stations described here are however dependent on both an energy source and internet access. They must be located closely to an indoor unit that communicates through the internet, and can be said to be on-the-grid, both on the energy grid and the communication grid.

In this thesis we consider weather stations that are located at places that can be char-

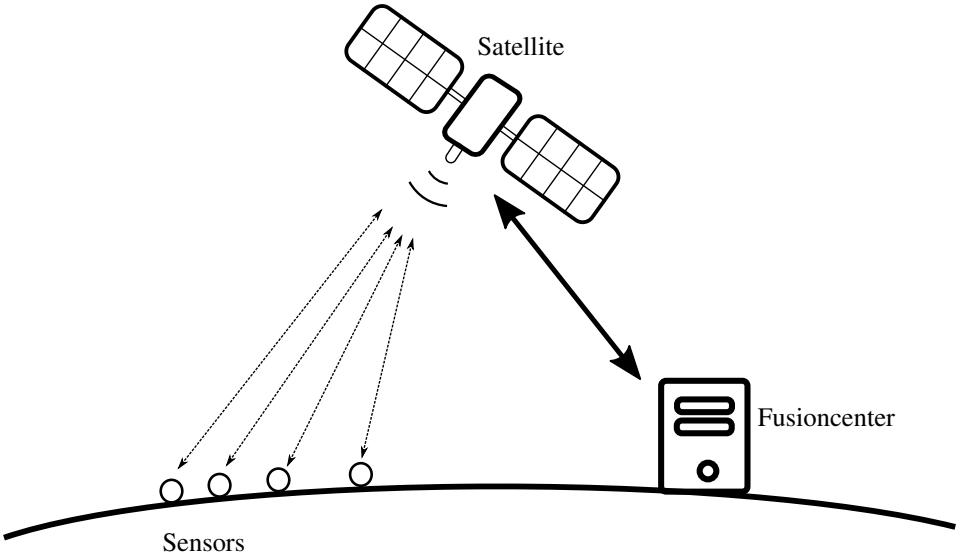


Figure 1.1: Sketch of setup. Communication between satellite and sensors is very energy-consuming and desirable kept at a low rate. Communication between satellite and fusion center has no restrictions.

acterized as off-the-grid. We consider stations that are not connected to a power grid, nor have coverage by mobile network or Wifi. This implies that the weather stations mentioned earlier are not possible to use, so we consider temperature stations (referred to as sensors from now) that are self-sufficient with energy and may communicate with a satellite. Specifically do we consider sensors located at Spitsbergen — the largest island of Svalbard. With its remote locality, situated about midway between the Norwegian mainland and the North pole, are most of Spitsbergen uninhabited and off both the energy and communication grid.

The sensor network is used in order to produce nowcasts for the temperature. Figure 1.1 shows a sketch of the setup, where the sensors may communicate with a satellite, which in turn can communicate with a central server (the fusion center). We assume these sensors make hourly observations which is used for making hourly nowcasts of the temperature. However, for several reasons, we wish to minimize the energy-use of the sensors. A network of such sensors may consist of numerous sensors, and manual maintenance (changing batteries) is neither desirable nor cost-effective. The large scale for such a network of sensors also implies that each sensor should be reasonable cheap, and reducing the battery capacity contributes to that.

One way to reduce the energy-consumption of the sensors is to reduce the number of transmissions from the sensors. According to Chapman et al. (2014), who has deployed sensors at locations with WiFi coverage, transmitting observations from the sensors is the biggest drain on the battery. Especially for sensors that are located in areas with weak or intermittent signal does transmissions consume a lot of energy. The sensors we consider in this thesis communicates via a satellite in stead of WiFi. If the sensors are equipped with

certain specification and communication protocols, Sanctis et al. (2016), the energy usage for communication with a satellite is not necessarily larger than communication through WiFi. We do however assume that the energy consumed for making observations is negligible compared to transmitting observations. The energy usage when transmitting is also investigated in Tamkittikhun et al. (2017), though the portion of energy used for transmissions are smaller here. The energy consumption is either way reduced by decreasing the number of transmissions. Reducing the number of transmissions means that some of the observations is only known to the sensors, and not available for the fusion center. This thesis presents a strategy for which observations that should be transmitted to the fusion center via the satellite. We refer to the strategy for when a sensor should transmit its observation as a communication scheme.

The idea behind our communication scheme is to utilize that the sensors make observations regardless of it is transmitted or not. This means that the sensors may be implemented with an event-triggered control, meaning that the sensors decide whether or not to transmit their observation based on the happening of an event, Heemels et al. (2012). We define the event to be that a sensor observes a value outside a tolerance interval. The tolerance interval is calculated in the fusion center, and transmitted to the sensor at a fixed rate. The tolerance interval is constructed by predicting the temperature ahead in time, and adding a tolerance level to the prediction. The prediction is calculated from a statistical model where the temporal dependency in the process we consider is utilized, together with a numerical weather prediction. Temperature for a location over time is often modelled as a time series, as it exhibits temporal dependence, i.e. the temperature at two time points close in time are likely to be quite similar, Adhikari and Agrawal (2013). A numerical weather forecast (NWP) provides predictions on many variables such as temperature, wind, rainfall and more, using powerful computers University of Illionois (2018). The communication scheme we present in this thesis is a Temporal dependent, Tolerance interval, Trigger based communication scheme, and is from here referred to as a TTT-communication scheme. The prediction found from the statistical model forming the TTT-communication scheme is referred to as the TTT-prediction.

Our setup is a two-way wireless sensor network with a fusion center, as the sensors in the network may also receive information from the fusion center. Receiving transmissions are as well energy consuming, as the sensors needs to wake up from sleep mode and turn on its radio. We therefore keep the communication from the satellite to the sensors at a low fixed rate. In this thesis we restrict this kind of communication to one transmission a day, in which each sensor receives the TTT-communication scheme for the following day. A part of our strategy is to let the sensors update their schemes when possible. Using event-triggered control for the sensors means that the fusion center gains knowledge about the process even though a sensor does not transmit an observation. As we have a fixed schedule for the sensors to make observations, the fusion center will either receive an exact value from a sensor, or know an interval in which the value lies at each time point t. A sensor may however only update its communication scheme when the observation is transmitted to the fusion center, as both the sensor and the fusion center must update their schemes identically.

The temperature in an area is a process that can be described by spatial-temporal statistical models. In this thesis we use a separable spatial-temporal model to make the commu-

nication schemes and for the nowcasting. We use the well-known AR(1)-process to make the TTT-predictions which the communication schemes are based on. Further, the temporal models are used as a base for the spatial model. In this thesis it becomes a truncated multivariate normal distribution, Horrace (2005). The spatial model is really a spatial correction of the TTT-prediction, hence we refer to the spatial model as a TTTS-model or TTTS-nowcast. The idea behind the TTTS-model is to utilize both the observations that have been transmitted to the fusion center, as well as the information we get by not receiving an observation. The TTTS-nowcast applies only to the locations where sensors exists. By modelling the temperature as a Gaussian Random Field, Bardeen et al. (1985), and by using the well known technique Kriging, Kleijnen (2009), one could nowcast the temperature for any points in the space. This is however not included in the scope of this thesis.

In order to reduce the number of transmissions we wish to make that the TTT- communication scheme so a sensor is likely to observe values in the interval. The TTT-scheme is made up by a prediction as well as a tolerance level. The tolerance level have a great impact on how often a sensor needs to transmit its observation. We will refer to how often a sensor transmit an observation as the transmission rate. We define the transmission rate for a sensor as the average number of transmissions each day. Smaller threshold value means that the sensor more often observes a value outside the interval. We therefore have a trade-off between smaller threshold values, which is preferable for nowcasting, and larger threshold values in order to reduce the transmission rate. By defining cost-functions for the trade-off one can evaluate this problem in a Bayesian decision setup and Value of Information approach, Eidsvik et al. (2015). In this thesis we leave this decision to the nowcaster.

We consider a twofold problem in this thesis. In order to reduce the transmission rate, we introduce the TTT-communication scheme. The TTT-communication scheme is built on a TTT-prediction, which comes from a statistical AR(1)-model which utilizes a NWP. The NWP we use for the case at Spitsbergen is provided by MET and named AROME-Arctic. AROME-Arctic, referred to as NWP from here, gives predictions for the temperature for each of the sensor locations we consider once a day. Secondly we utilize the spatial dependency in the process by using the information known to the fusion center to produce nowcasts for the temperature, using the model we name TTTS-nowcast.

To summarize, is the aim for this thesis is to

- utilize NWP and historical data
- benefit from the spatial and temporal dependency in the process

in order to

- reduce the transmission rate by using the TTT-communication scheme
- utilize all information known to the fusion center to produce nowcast using the TTTS-model.

We will present the findings through a simulation study and the case at Spitsbergen.

The thesis consist of seven chapters. In chapter 2 we present some background theory required for the development of the TTT-communication scheme as well as for the

TTTS-model. The chapter closes with a presentation of the evaluation methods used for the assessment of our models and an overview over the software used in the work of this thesis. Chapter 3 gives an introduction to our case study with a brief explanatory analysis to highlight some of the features in our data set. The TTT- and TTTS-model are derived and presented in chapter 4. The chapter ends with an adaption of the models to the case at Spitsbergen, including an algorithm for estimating the moments of the truncated multivariate Gaussian variable. A simulation study is performed in chapter 5. In chapter 6 the methods are applied on our case at Svalbard. The thesis closes with a conclusion and suggestions for further work in chapter 7.

Chapter 2

Background

This chapter presents background theory that is used for developing our models in chapter 4. We present some well-known statistical processes and results from them. Two evaluation methods for measuring the performance of our models is presented in section 2.3, before the chapter closes with a short presentation of the software used in the work of this thesis.

2.1 Autoregressive model

The autoregressive process of order 1 with parameters ϕ and σ_ε^2, denoted AR$(1; \mu, \phi, \sigma_\varepsilon^2)$, is defined by Wei (2006) as

$$Y_t = \mu + \phi Y_{t-1} + \varepsilon_t,$$

where ε_t is a white noise process with distribution $\varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$. This model is a Markov model, as future states only depend on the current state. This property is called the Markov property, and can be defined as $(Y_t | Y_{t-1} = y_{t-1}) \perp (Y_{t-k} = y_{t-k})$ for $k > 1$, where $\perp$ denotes independence. We use the notation $\{Y_t\}$ to denote that Y_t is a time series. The following results are from Wei (2006).

The AR(1) process $\{Y_t\}$ is Gaussian distributed, with the following moments,

$$\mathrm{E}[Y_t] = \frac{\mu}{1-\phi},$$
$$\mathrm{Var}[Y_t] = \frac{\sigma_\varepsilon^2}{1-\phi^2}.$$

Forecasting from the AR(1)-process is done by conditioning on an observation of the process,

$$(Y_{t+1} | Y_t = y_t) = \mu + \phi y_t + \varepsilon_{t+1}. \tag{2.1}$$

From equation (2.1) we see that the conditioned variable is also normally distributed, with

the parameters

$$\begin{aligned}\mathrm{E}[Y_{t+1}|Y_t = y_t] &= \mu + \phi y_t \\ \mathrm{Var}[Y_{t+1}|Y_t = y_t] &= \sigma_\varepsilon^2.\end{aligned}$$

The forecasted variables for n-step ahead is also normally distributed, where the parameters are given by the recursive formulas

$$\begin{aligned}\mathrm{E}[Y_{t+n}|Y_t = y_t] &= \mu + \phi\,\mathrm{E}[Y_{t+n-1}|Y_t = y_t] \\ \mathrm{Var}[Y_{t+n}|Y_t = y_t] &= \sigma_\varepsilon^2(1 + \phi^2 + \cdots + \phi^{2(n-1)}).\end{aligned}$$

A process is said to be stationary if $F_X(x_{t_1}, \ldots, x_{t_n}) = F_X(x_{t_1+k}, \ldots, x_{t_k+k})$ is fulfilled for all $t_1, \ldots, t_n$ and k. An AR(1)-process is weakly stationary if the following properties are fulfilled.

- The mean is independent of time, $\mathrm{E}[X_t] = \mu$ for all t
- The variance is independent of time, $\mathrm{Var}[X_t] = \sigma_\varepsilon^2$ for all t
- The correlation between two variables is only dependent on the time difference between them, $\mathrm{Cor}(X_t, X_{t+h}) = \mathrm{Cor}(X_{t+k}, X_{t+k+h})$ for all h, k and t.

Autocorrelation and partial autocorrelation function. The autocorrelation function (ACF) ρ_k is defined as

$$\rho_k = \frac{\mathrm{Cov}[X_t, X_{t+k}]}{\sqrt{\mathrm{Var}[Z_t]}\sqrt{\mathrm{Var}[Z_{t+k}]}},$$

when $\{X_t\}$ is a time series. The partial autocorrelation function (PACF) P_k is defined as

$$P_k = \mathrm{Cor}[X_t, X_{t+k}|X_{t+1}, \ldots, X_{t+k-1}].$$

The empirical ACF and PACF are commonly used for model selection for a time series.

Maximum Likelihood Estimation of parameters in the AR(1)-process. Given the observations $\boldsymbol{y} = [y_1, \ldots, y_n]$ from an $AR(1; \mu, \phi, \sigma_\varepsilon^2)$-process, one may estimate the parameters by the Maximum Likelihood Estimation method. The likelihood for the AR(1; $\cdot$)-process is defined as

$$\begin{aligned}\mathcal{L}(\mu, \phi, \sigma_\varepsilon^2; \boldsymbol{y_t}) = &-\frac{T}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_\varepsilon^2/(1-\phi^2)) - \frac{\left(y_1 - (\mu/(1-\phi))\right)^2}{2\sigma_\varepsilon^2/(1-\phi^2)} \\ &- \frac{T-1}{2}\log(\sigma_\varepsilon^2) - \sum_{t=2}^{T}\frac{(y_t - \mu - \phi y_{t-1})^2}{2\sigma_\varepsilon^2}.\end{aligned} \tag{2.2}$$

Expressions for exact maximum likelihood estimators is found by differentiating with respect to each of the parameters and setting the derivatives equal to zero. However, in practice this results in a system of nonlinear equations for which there are no simple solution. This results are therefore not presented, and we use numerical procedures to maximize equation (2.2).

2.2 Multivariate Gaussian random variables and truncated multivariate Gaussian random variables

This section provides some well-known, and some not so well-known results on basic probability theory and distributions. Unless otherwise states is the results in this section from Walpole et al. (2002).

The multivariate Gaussian random variable with n dimensions, denoted by $\boldsymbol{X} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, is defined by the following probability density function (pdf),

$$f(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right), \tag{2.3}$$

where $\mathrm{E}[\boldsymbol{X}] = \boldsymbol{\mu}$ and $\mathrm{Cov}[\boldsymbol{X}] = \boldsymbol{\Sigma}$. The cumulative distribution function (CDF) for the multivariate Gaussian random variable is defined as

$$F(\boldsymbol{x}) = P(\boldsymbol{X} \leq \boldsymbol{x}), \quad \boldsymbol{X} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

Since no closed forms exist for this expression, numerical methods are used for estimation.

A useful result from the multivariate Gaussian distribution is the conditional multivariate Gaussian distribution. When $\boldsymbol{X} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is partitioned so that one can write $\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_A \\ \boldsymbol{X}_B \end{pmatrix}$, $\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_A \\ \boldsymbol{\mu}_B \end{pmatrix}$ and $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_A & \boldsymbol{\Sigma}_{AB} \\ \boldsymbol{\Sigma}_{BA} & \boldsymbol{\Sigma}_B \end{pmatrix}$, then the conditional random variable $\boldsymbol{X}_A|\boldsymbol{X}_B = \boldsymbol{x}_b$ is as well Gaussian distributed with expectation $\boldsymbol{\mu}_{A|B}$ and covariance $\boldsymbol{\Sigma}_{A|B}$, defined by

$$\begin{aligned} \boldsymbol{\mu}_{A|B} &= \mathrm{E}[\boldsymbol{X}_A|\boldsymbol{X}_B = \boldsymbol{x}_B] = \boldsymbol{\mu}_A + \boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_B^{-1}(\boldsymbol{x}_B - \boldsymbol{\mu}_B) \\ \boldsymbol{\Sigma}_{A|B} &= \mathrm{Cov}[\boldsymbol{X}_A|\boldsymbol{X}_B = \boldsymbol{x}_B] = \boldsymbol{\Sigma}_A - \boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_B^{-1}\boldsymbol{\Sigma}_{BA}. \end{aligned} \tag{2.4}$$

Truncated distributions A truncated pdf is a conditional distribution that arises from restricting the domain of some other pdf. The truncated distribution for the stochastic variable X with probability density function $f(x)$ on the interval $[a, b]$ is defined as

$$f(x|a \leq X \leq b) = \begin{cases} \frac{f(x)}{F(b)-F(a)}, & x \in [a, b] \\ 0, & x \notin [a, b]. \end{cases}$$

Truncation of a Gaussian random variable, $X \sim \mathcal{N}(\mu, \sigma^2)$, leads to a truncated Gaussian distribution,

$$f(x|a \leq X \leq b) = \begin{cases} \frac{\phi(\frac{x-\mu}{\sigma})}{\sigma(\Phi(\frac{a-\mu}{\sigma})-\Phi(\frac{a-\mu}{\sigma}))}, & x \in [a, b] \\ 0, & x \notin [a, b], \end{cases} \tag{2.5}$$

where $\phi(\cdot)$ is the standard Gaussian density function, and $\Phi(\cdot)$ denotes the cdf for the standard normal distribution. From Johnson et al. (2005) we have that the first and second

moments for the univariate truncated Gaussian random variable is given by

$$\begin{aligned}
\mathrm{E}[X|a \leq X \leq b] &= \mu + \sigma \frac{\phi(\frac{a-\mu}{\sigma}) - \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} \\
\mathrm{Var}[X|a \leq X \leq b] &= \sigma^2 \left[1 + \frac{\frac{a-\mu}{\sigma}\phi(\frac{a-\mu}{\sigma}) - \frac{b-\mu}{\sigma}\phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} - \left(\frac{\phi(\frac{a-\mu}{\sigma}) - \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} \right)^2 \right].
\end{aligned} \tag{2.6}$$

The multivariate Gaussian random variable $\boldsymbol{X} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ may also be truncated, and is defined by Cartinhour (1990) as

$$f(\boldsymbol{x}|\boldsymbol{a} \leq \boldsymbol{X} \leq \boldsymbol{b}) = \begin{cases} \frac{f(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma})}{P(\boldsymbol{a} \leq \boldsymbol{X} \leq \boldsymbol{b})} & \boldsymbol{x} \in [\boldsymbol{a}, \boldsymbol{b}] \\ 0 & \boldsymbol{x} \notin [\boldsymbol{a}, \boldsymbol{b}], \end{cases} \tag{2.7}$$

where $f(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the multivariate Gaussian pdf defined in equation (2.3).

Horrace (2005) has shown that the marginal distributions from a truncated normal distribution are not truncated normal distributions in general. The conditional random variables are however truncated normal distributed.

Cartinhour (1990) has shown that the one-dimensional marginal pdf of a truncated multivariate Gaussian random variable may be written as

$$\begin{aligned}
f(x_n) = & \frac{1}{P\sqrt{2\pi c_{nn}}} \exp\left(-\frac{1}{2c_{nn}(x_n - \mu_n)^2} \right) \\
& \int_{a_{n-1}}^{b_{n-1}} \cdots \int_{a_1}^{b_1} \frac{1}{\sqrt{(2\pi)^{n-1}|A_1^{-1}|}} \\
& \exp\left(-\frac{1}{2}(\boldsymbol{x} - m(x_n))^T A_1 (\boldsymbol{x} - m(x_n)) \right) dx_1 \cdots dx_{n-1}.
\end{aligned} \tag{2.8}$$

The interested reader may find definitions of the parameters in the original article by Cartinhour.

Manjunath and Wilhelm (2012) have found that the first and second moments of the truncated multivariate normal distribution can be found by the following expressions,

$$\begin{aligned}
\mathrm{E}[Y_i] &= \sum_{k=1}^{d} \sigma_{i,k}(F_k(a_k) - F_k(b_k)) + \mu_i \\
\mathrm{E}[Y_i Y_j] &= \sigma_{i,j} + \sum_{k=1}^{d} \sigma_{i,k} \frac{\sigma_{i,j}(a_k F_k(a_k) - b_k F_k(b_k))}{\sigma_{k,k}} \\
&\quad + \sum_{k=1}^{d} \sigma_{i,k} \sum_{q \neq k} \left(\sigma_{j,q} - \frac{\sigma_{k,q}\sigma_{j,k}}{\sigma_{k,k}} \right) [(F_{k,q}(a_k, a_q) - F_{k,q}(a_k, b_q)) \\
&\quad - (F_{k,q}(b_k, a_q) - F_{k,q}(b_k, b_q))].
\end{aligned}$$

Again we refer the reader to the original paper for definitions in the expressions, though we choose to mention that $F_k(x_k)$ is the cumulative density function for the truncated variable x_k and may be found by integrating the expression in equation (2.8). Unfortunately, the required integral is not on closed form, which means that numerical methods must be used to estimate the parameters.

2.3 Evaluation methods

We present two methods for evaluating our methods and models. The models we present in section 4 will be evaluated in terms of these methods.

Root Mean Square Error (RMSE) is a commonly used method to measure the difference between point predictions and observations. It measures the root of the squared error of a point forecast, and represents the standard error of the difference between predictions and observations. When $\widehat{Y}_t$ is a point prediction for the observed values y_t for $t = 1, 2, \ldots, n$, the RMSE is defined as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\widehat{Y}_i - y_i)^2}{n}}.$$

The RMSE is strictly non-negative and obviously negative oriented, meaning that lower RMSE indicates better predictions. Perfect predictions gives a RMSE of zero.

Continuous Ranked Probability Score (CRPS) is a method for evaluating probabilistic forecasts. A probabilistic forecast differ from a point forecast by the fact that it provides a probabilistic distribution for the forecast, rather than just a specific value. For a probabilistic forecast with cdf $F_t(y_t)$ and corresponding observations y_t, the CRPS is defined by Gneiting and Raftery (2004) as

$$CRPS_t = \int_{-\infty}^{\infty} (F_t(x) - H(x - y_t))^2 dx,$$

where $H(x - y)$ is the Heaviside function,

$$H(x - y) = \begin{cases} 0, & \text{for } x < y, \\ 1, & \text{for } x \geq y. \end{cases}$$

For several probabilistic forecasts with cdf $F_t(y_t)$ and corresponding observations y_t for $t = 1, 2, \ldots, n$, is the CRPS given by

$$CRPS = \frac{\sum_{i=1}^{n} CRPS_i}{n}.$$

CRPS is also a strictly non-negative, negative oriented evaluation method that rewards both low variance and correct location for the probabilistic distribution. An illustration of a predictive CDF is shown in figure 2.1 together with an observation. The blue shaded area

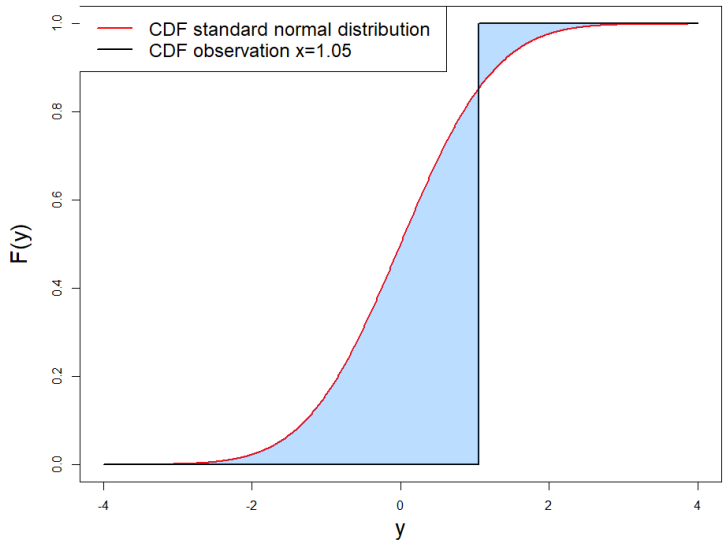


Figure 2.1: Illustration of CRPS. The red line represents a predictive distribution, whilst the black line illustrates the cdf of an observation. The CRPS is represented by the blue shaded area.

represents the value of the CRPS. For a point forecast, CRPS reduces to the mean absolute error (MAE), which is defined as

$$MAE = \frac{\sum_{i=1}^{n}(\widehat{Y}_i - y_i)}{n}. \tag{2.9}$$

2.4 Software

In this thesis, `R` is used for the analysis. Some computer-intensive programs have used the computation server Markov at NTNU in order to reduce the computational time. The AROME-Arctic forecast were loaded as a subset from .nc files and converted to a suitable file structure in `R` by the use of the library 'ncdf4'.

Chapter 3

Spitsbergen: Case Study and Explorative analysis

In this chapter we present the observations and forecasts used in the Spitsbergen case study. The first section will present the data used. The data consists of temperature observations from some observation stations, and a numerical weather prediction (NWP). To highlight some important features about the data, a brief explanatory analysis is presented in the Section 3.1.1. The chapter closes with an inspection of the performance of the NWP in the setup with a Tolerance limit, Trigger based communication scheme.

The area considered in this thesis is Spitbergen. Spitsbergen is the largest of the islands that make up the island group Svalbard, which is a Norwegian sovereign area. Svalbard is located north of the european main-land and were chosen as the area to consider due to this rural location, and because there are both high-quality observations and NWP available.

3.1 Temperature observations and NWP

This section presents the temperature observations and forecasts used in this study. The temperature observations are obtained from the online portal eKlima.no of Norwegian Meteorological Institute (MET). The data available are hourly observations of temperature at observation stations. A list of the stations, information about them, is presented in Table 3.1 in Section 3.1.1. The observations are done at 2 meters above the ground under controlled conditions. We consider these observations as exact observations of the temperature, i.e. no measurement error is assumed. We denote the observed temperature at time t and location s as y_t^s. Figure 3.1 shows a map of Spitsbergen with the available observation stations.

The NWP used in this study are obtained from the AROME-arctic weather model, provided by The Norwegian Meteorological Institute (2017). We have two NWPs available from the AROME-model: The raw NWP and a post-processed NWP. The post-processing is done by using statistical models and historical data from the observation stations. We

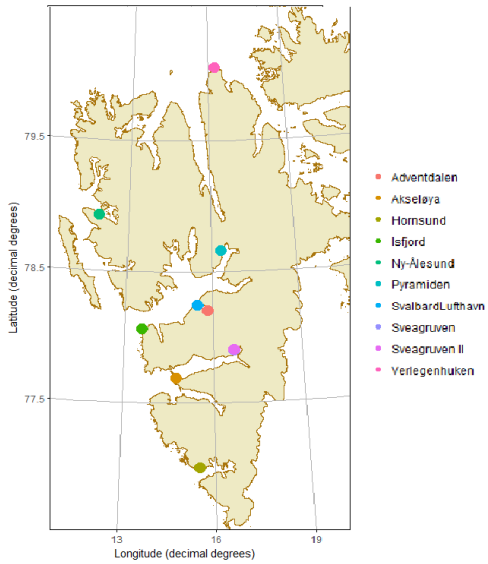


Figure 3.1: Map of Spitsbergen and observation stations.

aim to present a method for how sensors can be deployed in areas where historical data are not available. We therefore choose the raw model, as the post-processed model is fitted to the locations of the observation stations. We refer to the raw AROME-arctic weather model as NWP throughout this thesis. This NWP is a high-resolution forecasting system for the European Arctic. It includes predictions for several variables, but in this study we only use the temperature forecast. The forecasts are made on a grid with 2.5 km grid spacing, with each grid cell having a corresponding altidude. Figure 3.2 shows a map of the grid with the altitude of each grid cell. The temperature is forecasted 2 meters above the ground, where the altitude is calculated from the geopotential. This may produce negative values for the altitude. We set these to sea level, i.e. 0 masl.

The AROME-model produces several deterministic forecasts with lead times of 66 hours every day, some days every 3 hours, and other days with larger intervals. However, we only use one forecast each day, starting at 00:00 UTC. That gives a new forecast each day at midnight and maximum lead time of 23 at 23:00 UTC. When using the AROME-forecast in the following, we choose the grid cell that covers the observation station considered. In other words, we treat the observations from the stations as the real temperature, and the data from the AROME-model as the forecasted temperature at the location of the observation station. We denote the NWP at time t, lead time l and location s as $x^s_{t,l}$. For both the observations and NWP we use hourly data from January 1st 2017 to December 31st 2017 from 10 observation stations. That yields 8760 time points for us to consider. The AROME-forecast is unavailable some days, especially several days in June 2017. These days are omitted from the analysis.

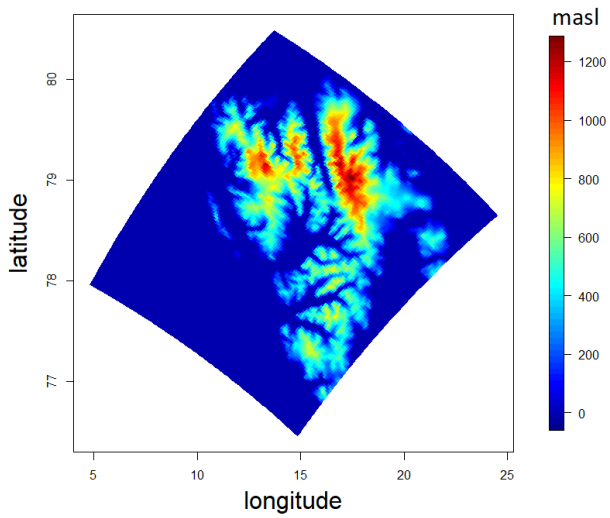


Figure 3.2: Altitude of grid cells in the NWP for Spitsbergen.

3.1.1 Explanatory analysis

The temperature at Svalbard is characterized by seasonal varieties, as well as large variations from between and within days. Figure 3.3 shows the observations made at Hornsund observation station, located at the north of Spitsbergen. The line shows the daily mean temperature, while the points shows the hourly observations. Temperature at the other observation stations have similar patterns. Figure 3.4 shows a close up for the temperature at Hornsund observation station at Nov 16. and Nov 17, with the corresponding NWP.

Table 3.1 summarizes some features of the data set we consider. We see from the table that there are differences in average between the observations and the NWP. According to the International Civil Aviation Organization (2017), does the temperature decrease with $6.5°$C per kilometer in altitude. The altitude difference is therefore likely to explain some of this discrepancy, as well as local conditions at the observation stations.

In the following, we will look at the difference between the observations and the NWP. We focus on the data from Hornsund observation station as this data is enough for the purpose of qualitative inspection.

Figure 3.5 shows the ACF and PACF for the discrepancy between the observed temperature and NWP at Hornsund observation station.

An important variable in weather forecasts is the lead time. The lead time is the time between a forecast is issued and the time it applies to. In general we excpect a forecast to perform better for shorter lead times. Figure 3.6a shows the MAE, defined in equation (2.9), as a function of lead time. Each line in the plot represents the error at different observation stations. Figure 3.6b shows the mean error as a function of the lead time. The figure shows that the mean error is lowest at the lead times around 12. The lead times also correspond to the time of the day, which implies that the forecast performs somewhat better

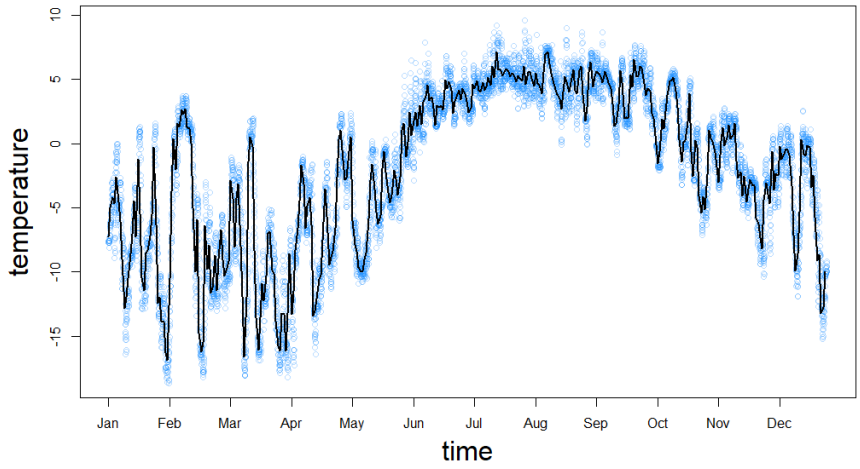


Figure 3.3: Hourly observations of temperature at Hornsund observation station in 2017. The black line shows the daily mean.

during daytime. Figure 3.6c shows the correlation between the forecast and observation as a function of the lead time. From the figure we see that the correlation gets smaller for longer lead times. Though, even for longer lead times are the forecast and observation strongly correlated.

3.1.2 NWP performance in case setting

In this section we present some simple statistical models to explore the performance of the NWP in our setting. We investigate the transmission rates for the sensors when the communication schemes are made directly from the NWP, and transmitted observations are not utilized, i.e. do not utilize temporal dependency. This could be the case if the process of interest lacks temporal dependency, or if the sensors are not able to do any calculations at all. That is, at time t, the sensor at location s transmits its observation if $y_t^s \notin [\widehat{y}_{i,t} - \tau, \widehat{y}_{i,t} + \tau]$, where $\widehat{y}_{i,t}$ is the prediction for y_t from model i and τ is the tolerance. We present 3 different models, and evaluate their performance with regards to the transmission rate. We present these models in this chapter as there are no new methods presented, but rather a presentation of how the NWP behaves in our setup.

Table 3.2 displays the models we consider in this section. Model 1 is simply the NWP alone. Table 3.1 shows that there is a bias between the observed temperature and NWP, likely due to the altitude difference. Model 2 is therefore the NWP with an intercept term. Figure 3.6b and 3.6c shows that the mean error and correlation changes with the lead time. We therefore let model 3 be linear regression model with lead-time dependent coefficients.

Figure 3.7 shows the transmission rates for the 3 models as a function of the tolerance. We define the transmission rate as the portion of time a sensor transmits its observation to the fusion center. We see that model 1 and 2 perform quite similar for most of the locations,

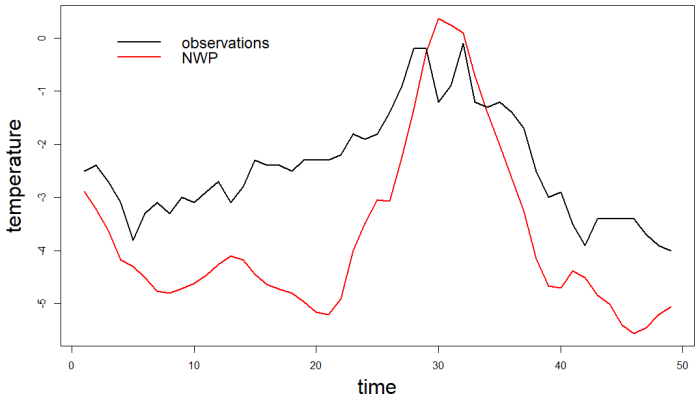
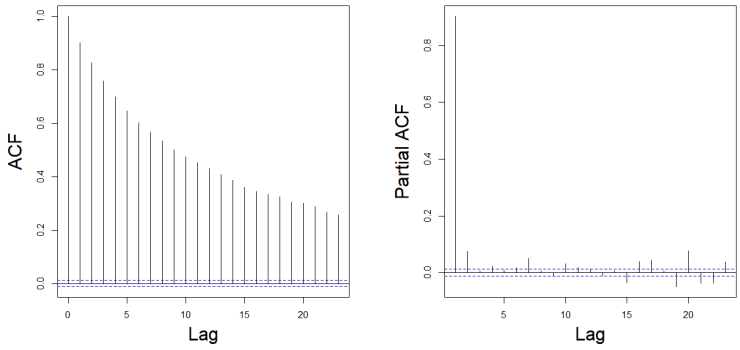


Figure 3.4: Observed temperature and NWP for Hornsund observation station Nov. 16-17.

s	Station name	masl station	masl NWP	avg. observed	avg. NWP
1	Hornsund	10	39.5	-2.19	-2.50
2	Sveagruva	9	32.4	-6.82	-3.06
3	Sveagruva II	50	439	-3.41	-6.41
4	Akseloya	20	0	-1.96	-2.18
5	Isfjord Radio	7	18.1	-1.87	-2.46
6	Svalbard Lufthavn	28	65.1	-2.55	-3.12
7	Adventdalen	15	95.6	-3.18	-4.36
8	Pyramiden	20	188	-3.38	-5.32
9	Ny-Aalesund	8	20.7	-3.09	-3.32
10	Verlegenhuken	8	2.7	-5.79	-5.40

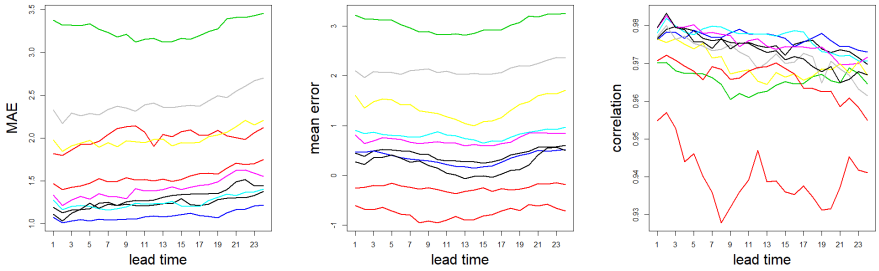
Table 3.1: Altitude for the observations stations and corresponding grid in the NWP. The average observed temperature and the average of NWP for each location. The locations is referred to by s and not their stations name in the rest of the thesis for simplicity.

but that model 2 outperforms model 1 clearly for others. Model 3 performs poorly. It is not surprising that we are unable to improve the NWP by a simple linear regression, as the NWP is made by a complex model by a professional weather forecast organization. The reason we are slightly able to improve the forecast by adding an intercept term is that the forecast from NWP are made at a different altitude than we observe the temperature. Local conditions around the observation stations may also contribute to our improvement. Based on the plots in Figure 3.6, and the results in Figure 3.7 do we disregard the lead-time as a variable in the rest of the thesis. We do however note that an intercept term to correct for the bias in the NWP improves predictions about the temperature.



(a) **(b)**

Figure 3.5: The figures show the empirical ACF and PACF of the discrepancy between the observed temperature and NWP at Hornsund observation station in 2017.

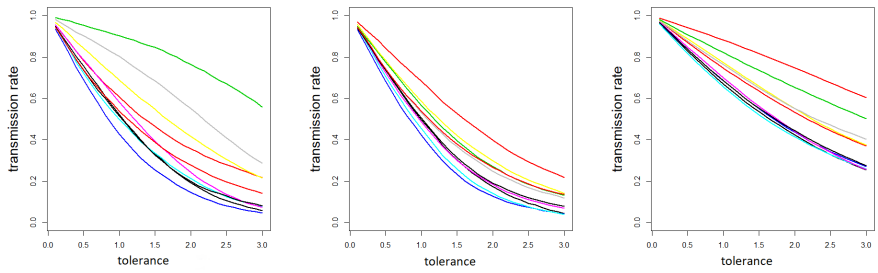


(a) **(b)** **(c)**

Figure 3.6: The figures shows respectively the MAE, mean discrepancy and correlation between the observed temperature and NWP. Each color represents one of the observation stations listed in Table 3.1.

	Model 1	Model 2	Model 3
$\widehat{y}_t =$	x_t	$\alpha + x_t$	$\alpha_l + \beta_l x_t$

Table 3.2: Caption



(a) Model 1 **(b)** Model 2 **(c)** Model 3

Figure 3.7: Transmission rate as a function of tolerance τ for 3 simple communication schemes not utilixing temporal dependency. Each color represents the transmission rate for a sensor.

Chapter 4

Models and Methods

This chapter presents the concepts behind our methods for prediction and nowcasting. Specifically are the TTT-communication scheme and TTTS-nowcast model presented in the first and second section. Section 4.3 considers the TTT-communication scheme in light of the case at Spitsbergen, before the chapter closes with the presentation of a Gibbs sampler used for estimating the moments of a truncated multivariate Gaussian random variable.

Respectively, we denote Y_t^s and x_t^s for a random variable that represents the temperature and NWP for the temperature at time t and location s. We assume that the NWP is available for both the fusion center and the sensors. Capital letters are used for stochastic variables, and lower case letters are used for values, e.g. observations or NWP.

4.1 TTT-communication scheme

In this section we present the strategy used for deciding whether or not a sensor should transmit its observation. We refer to the strategy we present as a Temporal dependence, Tolerance limit, Trigger based (TTT) communication scheme. The communication scheme is in reality an if-then statement, as the sensor transmits the observation if a condition is fulfilled. This setup is often referred to as event-triggered control. In the communication scheme we present, a sensor s transmits its observation y_t^s at time t if it deviates from a given value with more than a given tolerance. The given value, denoted $\widehat{y}_t^s$ from here, is a point prediction which utilized the temporal dependency of the process. We denote the tolerance τ. The condition for when an observation is transmitted is

$$y_t^s \notin [\widehat{y}_t^s \pm \tau]. \tag{4.1}$$

The TTT-communication scheme has the advantage that the fusion center gains information even when the sensor does not transmit an observation, as it then knows $y_t^s \in [\widehat{y}_t^s \pm \tau]$.

We aim to find methods that reduce the transmission rates for the sensors. Hence, we wish to find a value for $\widehat{y}_t^s$ such that the probability for the event in equation (4.1)

is minimized. We do so by estimating $\widehat{y}$ from the statistical process $\{Y_t^s\}$ by making a prediction about the process at time t. We consider $\{Y_t^s\}$ as a univariate time series for each location s. For simplicity we omit s in the notation, and consider the time series $\{Y_t\}$ for a generic location in the following.

In order to make predictions about the time series $\{Y_t\}$, we introduce a statistical model that utilizes the NWP and the temporal dependencies in the process. We define the discrepancy R_t as the difference between the NWP and the process $\{Y_t\}$ at time t as

$$R_t = Y_t - x_t. \tag{4.2}$$

$\{R_t\}$ is then also a time series. Figure 3.5 in section 3.1.1 shows the empirical autocorrelation and partial autocorrelation functions of the discrepancy time series $\{r_t\}$. The figure shows that $\{r_t\}$ seems to follow the well-known AR(1)-process, as the ACF is exponentially decaying and the PACF cuts off after 1 lag, Wei (2006). Therefore, we may utilize that succeeding errors are correlated, and introduce a model for the discrepancy,

$$R_t = \alpha + \phi R_{t-1} + \varepsilon_t, \tag{4.3}$$

where $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ is a white noise process. By combining equation (4.3) and (4.2) we have

$$Y_t = \alpha + x_t + \phi(Y_{t-1} - x_{t-1}) + \varepsilon_t, \tag{4.4}$$

which is easier to interpret. However, by examination of the moments of Y_t,

$$\begin{aligned} \mu_t &= \mathrm{E}[Y_t] = \frac{\alpha + x_t - \phi x_{t-1}}{1-\phi} \\ \sigma_t^2 &= \mathrm{Var}[Y_t] = \frac{\sigma_\varepsilon^2}{1-\phi^2}, \end{aligned} \tag{4.5}$$

we see that the mean is time-dependent and Y_t is therefore not a stationary time series. To avoid this challenge, we move further on with the model for discrepancy, defined in equation (4.3). It is irrelevant which of the two variables we predict, as the only thing that separates them is the x_t-term which is known both to the fusion center and the sensor. R_t is Gaussian distributed with moments defined as

$$\begin{aligned} \mathrm{E}[R_t] &= \frac{\alpha}{1-\phi}, \\ \mathrm{Var}[R_t] &= \frac{\sigma_\varepsilon^2}{1-\phi^2}. \end{aligned}$$

By conditioning on an observed value, e.g. $R_t = r_t$, we are able to make predictions on the process for yet unobserved steps. To ease the notation we denote the conditional variable $[R_{t+1}|R_t = r_t]$ as $R_{t+1|t}$. The 1-step and n-step ahead conditional variables are defined as

$$\begin{aligned} R_{t+1|t} &= \alpha + \phi r_t + \varepsilon_{t+1}, \\ R_{t+n|t} &= \alpha + \phi R_{t+n-1|t} + \varepsilon_{t+n} \end{aligned}$$

The maximum likelihood estimator for $\widehat{r}_{t+1|t}$, which we use as point predictors, are convenient enough the expected value of the variables,

$$E[R_{t+1|t}] = \alpha + \phi r_t. \tag{4.6}$$

As this is also used as our prediction, we denote the above expression as $\widehat{r}_{t+1|t}$. Following this notation, the n-step ahead prediction is found recursively by

$$E[R_{t+n|t}] = \widehat{r}_{t+n|t} = \alpha + \phi \widehat{r}_{t+n-1|t}. \tag{4.7}$$

$R_{t+n|t}$ is Gaussian distributed, with mean given in equation (4.7). The variance is

$$\mathrm{Var}[R_{t+n|t}] = \sigma_\varepsilon^2(1 + \phi^2 + \dots \phi^{2(n-1)}). \tag{4.8}$$

The condition given in equation (4.1) for when a sensor should transmit its observation at time t can now be written as

$$y_t \notin [\widehat{y}_{t|\cdot} \pm \tau],$$

where the sub script $|\cdot$ denotes that the last transmitted observation is conditioned on. The link between Y_t and R_t in equation (4.2) have been used in the above equation. We introduce I_t as notation for the interval in the above equation, that is $I_t = [\widehat{y}_{t|\cdot} \pm \tau]$. We assume that the sensor transmits an observation when initialized, hence there will always be at least one observation to condition on.

To summarize, the TTT-communication scheme we present consists of a tolerance τ and a prediction about the process $\widehat{y}_{t|\cdot}$. The prediction is a point prediction for the random variable $\widehat{Y}_{t|\cdot}$. We refer to the point prediction as the TTT-point prediction, and the random variable $\widehat{Y}_{t|\cdot}$ as the TTT-random variable. The distribution of $\widehat{Y}_{t|\cdot}$ is the probabilistic distribution for Y_t. The probabilistic distribution is Gaussian, with moments defined in equation (4.7) and (4.8).

In order to clarify the strategy we consider, we present a short example.

Example. *Assume the sensor transmits an observation to the fusion center at time t. By conditioning on that transmission both the fusion center and sensor calculates the communication scheme for the next time step, $t+1$. Then, if the observation at time $t+1$ is inside the interval, $y_{t+1} \in [\widehat{Y}_{t+1|t} \pm \tau]$, the observation is not transmitted to the fusion center. The fusion center still gains information, as it knows $y_{t+1} \in [\widehat{y}_{t+1|t} \pm \tau]$. For the next time step, the fusion center updates the communication scheme by calculating $\widehat{y}_{t+2|t}$. Even though it is known to the sensor, the sensor may not update its communication scheme based on the observation y_{t+1}, as the fusion center will have a different communication scheme. If however the observed value at time $t+1$ is outside the interval, $y_{t+1} \notin [\widehat{y}_{t+1|t} \pm \tau]$, the observation is transmitted and both the fusion center and sensor updates the communication scheme by conditioning on y_{t+1}.*

4.2 TTTS: Spatial model for nowcasting

Section 4.1 presented the TTT-communication scheme for reducing the transmission rate for a sensor using the temporal dependencies in the process. This section aims to utilize

the spatial dependencies in the process in order to produce a nowcast for the process at time t. One could of course let the TTT-point predictions, $\widehat{\boldsymbol{y}}_{t|\cdot}$, and observations, $\boldsymbol{y}_t$ form the nowcast, but then the spatial dependency in the process is not exploited. The model we present, referred to as TTTS, utilizes the information about both the observations that have been transmitted, and those which has not been transmitted in order to create nowcasts of the process.

At a time point t, we assume that $\widehat{\boldsymbol{Y}}_{t|\cdot} = [\widehat{Y}^1_{t|\cdot}, \ldots, \widehat{Y}^n_{t|\cdot}]$ follows a multivariate Gaussian distribution, where $\widehat{Y}^s_{t|\cdot}$ is the TTT-random variable for the location s. By defining the correlation matrix by the elements $\rho_{i,j} = \text{Cor}[\widehat{Y}^i_{t|\cdot}, \widehat{Y}^j_{t|\cdot}]$, is the distribution of $\widehat{\boldsymbol{Y}}_{t|\cdot}$ fully explained. However, a small adjustment must be made for the variance of $\widehat{Y}^i_{t|\cdot}$. The TTT-random variable is only conditioned on the last observations that was transmitted to the fusion center, and not the information we gain when an observation is not transmitted to the fusion center. We denote this information $\mathcal{H}_t$ and include the last transmitted observation y_k to the information set $\mathcal{H}_t$. That is $\mathcal{H}_t = [Y_{t|t-k} = y_t, Y_{t|t-k+1} \in I_t, \ldots, Y_{t|t-1} \in I_{t-1}]$. The additional information is irrelevant for the TTT-point prediction $\widehat{Y}_{t|t-k}$, as it can be shown that

$$\text{E}[Y_{t|t-k}] = \text{E}[Y_t|\mathcal{H}_t].$$

Note that this is only correct when the intervals I_t are symmetric around the mean of $\widehat{Y}_t$.

The variance is however not invariant to the additional information in $\mathcal{H}_t$. To demonstrate, we present an example.

Example. *Let $[Y_1, Y_2, Y_3]$ be multivariate Gaussian distributed with mean 0, variance 1 and correlation 0.8. We compare the variance of the conditional random variables $Y_1|Y_2 = y_2$ and $Y_1|Y_3 \in I_3, Y_2 = y_2$ to illustrate the effect of conditioning on the information set $\mathcal{H}$.*

We find the variance of $Y_1|Y_2 = y_2$ by using the formula in equation (2.4),

$$\text{Var}[Y_1|Y_2 = y_2] = \text{Var}[Y_1] - \frac{\text{Cov}[Y_1, Y_2]^2}{\text{Var}[Y_1]} = 0.36.$$

We find the variance of $Y_1|Y_3 \in I_3, Y_2 = y_2$ numerically as no closed form exist for the expression. A numerical routine gives $\text{Var}[Y_1|Y_3 \in I_3, Y_2 = y_2] = 0.22$, and we see that it is considerably smaller than the variance of $Y_1|Y_2 = y_2$.

To summarize, $\boldsymbol{Y}_t$ is multivariate normal distributed with parameters

$$\begin{aligned}\text{E}[\boldsymbol{Y}_t] &= [\widehat{y}^1_{t|\cdot}, \ldots, \widehat{y}^n_{t|\cdot}], \\ \Sigma_{i,j} &= \rho_{i,j}\sqrt{\text{Var}[Y^i_t|\mathcal{H}^i_t]\,\text{Var}[Y^j_t|\mathcal{H}^j_t]},\end{aligned}$$

where $\widehat{y}_{t|\cdot}$ is the TTT-point prediction.

For the remainder of this section we use a more general notation for our multivariate Gaussian random variables, namely $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$. $\boldsymbol{Y}$ represents the distribution for the temperature right before the sensors make observations and some of them transmit their observations. The idea behind nowcasting is to utilize the information available to give prediction about the process at the current state. We are therefore interested in the

distribtion of $\boldsymbol{Y}$ conditioned on the information we gain from the sensors. We denote this information as $\mathcal{I}$. $\mathcal{I}$ contains either an exact observation, or an interval for which the value lies in (observation not transmitted) for each sensor, and may look like this $\mathcal{I} = \{Y^1 = y^1, y^2 \in I^2, y_t^3 \in I^3, Y^4 = y^4 \ldots, y^n \in I^n\}$, where y represents observed values.

We are now interested in making a prediction for the random variable $\boldsymbol{Y}|\mathcal{I}$. We let the point-nowcast be the maximum likelihood estimate of this conditional distribution, i.e. $E[\boldsymbol{Y}|\mathcal{I}]$. The condition $\mathcal{I}$ may have three different forms, each of them requiring different methods in order to make a prediction. We summarize the different possibilities in a list,

1. All sensors transmits their observations, i.e. $\mathcal{I} = [Y_1 = y_1, \ldots, Y_n = y_n]$
2. None of the sensors transmits their observations, i.e. $\mathcal{I} = [y_1 \in I_1, \ldots, y_n \in I_n]$
3. Some of the sensors transmits their observations, i.e. $\mathcal{I} = [Y_1 \in y_1, y_2 \in I_2, \ldots, y_n \in I_n]$.

The first option gives the easiest nowcast to compute, as obviously $E[\boldsymbol{Y}|\boldsymbol{Y} = \boldsymbol{y}] = \boldsymbol{y}$. The nowcast for the second option is also easily computed, $E[\boldsymbol{Y}|\mathcal{I}^*] = \boldsymbol{\mu}$. In the univariate case, this is easily seen from equation (2.6). Although more complicated, the result should be clear for the multivariate truncated Gaussian random variable as well, from inspecting the pdf for the random variable given in equation (2.7).

The third option for the form $\mathcal{I}$ is both the most interesting and challenging condition. We introduce $\mathcal{I}^*$ and $\mathcal{I}'$ as notation in order to partition $\mathcal{I}$, where $\mathcal{I}^*$ contains the variables that are constrained by an interval and $\mathcal{I}'$ contains the variables that have been observed and transmitted to the fusion center. That is, $\mathcal{I} = [\mathcal{I}^*, \mathcal{I}']$ where $\mathcal{I}^* = [y^j \in I^j]$ for j such that $y^j \in I^j$ and $\mathcal{I}^* = [Y^j = y^j]$ for j such that $y^j \notin I^j$. In order to nowcast from the distribution $\boldsymbol{Y}|\mathcal{I}$ we start by realizing that we are only interested in the distribution of a subset of $\boldsymbol{Y}$, as some of them are already observed. Denote the variables of interest $\boldsymbol{Y}^*$, and the observed variable $\boldsymbol{Y}'$ so $\boldsymbol{Y}$ is partitioned,

$$\boldsymbol{Y} = [\boldsymbol{Y}^*, \boldsymbol{Y}'] \sim \mathcal{N}\left(\begin{bmatrix}\boldsymbol{\mu}^* \\ \boldsymbol{\mu}'\end{bmatrix}, \begin{bmatrix}\Sigma^* & \Sigma^{*\prime} \\ \Sigma'^* & \Sigma'\end{bmatrix}\right)$$

By using the formulas in equation (2.4), which defines the conditional mean and variance of a multivariate normal distribution we can find $\boldsymbol{Y}^*|\mathcal{I}'$,

$$\begin{aligned}\boldsymbol{\mu}^* &= \boldsymbol{\mu}^* + \Sigma^{*\prime}(\Sigma')^{-1}(\boldsymbol{y}' - \boldsymbol{\mu}'), \\ \Sigma^* &= \Sigma^* - \Sigma^{*\prime}(\Sigma')^{-1}\Sigma'^*.\end{aligned}$$

This means $\boldsymbol{Y}^* \sim \mathcal{N}(\boldsymbol{\mu}^*, \Sigma^*)$. By conditioning on $\mathcal{I}^*$ as well, we have a truncated multivariate normal distribution with known parameters. However, as mentioned in section 2.2 are the expressions for the moments of this distribution in general not on closed form. A method for estimating the moments is presented at the end of this chapter.

4.3 Inference and testing set-up

In this section we consider the TTT-communication scheme and TTTS-nowcast in light of the case we consider. We let the stochastic variable $y_{d,h}^s$ represent the temperature at

location s, day d and hour h, and similar for the NWP and discrepancy respectively, $x^s_{d,h}$ and $R^s_{d,h}$. We consider 10 locations at Svalbard, $s = 1, 2, \ldots, 10$. We consider the period from January 1st 2017 to December 31st 2017, $d = 1, 2, \ldots, 365$ and $h = 0, 1, \ldots, 23$. The motivation behind introducing t, h in the notation to replace t is to highlight that the NWP used in this thesis are issued once a day and we only use the forecast for that day. We assume that a NWP-ensemble, $\boldsymbol{x}_d = [x_{d,0}, \ldots, x_{23}]$, for a day d is issued at $00:00$ UTC[1], represented by $h = 0$. The ensemble is calculated from a numerical weather model which among other things takes observations of the weather from the previous days and hours as input. Different ensembles are therefore calculated from different models. The result is that even though we expect the same distributional properties for $R_{\boldsymbol{d}}$ and $R_{\boldsymbol{d+1}}$, is there a break in the autocorrelation structure between $R_{d,23}$ and $R_{d+1,0}$. In other words can we not except the same correlation between the pairs $(R_{d,23}, R_{d+1,0})$ as for pairs of discrepancies at the same day. This makes the time series $\{R_{d,h}\}$ non-stationary.

However, in stead of the non-stationary time series $\{R_t\}$ from $d = 1, h = 0$ to $d = 365, h = 23$, we rather have one stationary time series for each day, denoted $\{R_{d,h}\}$ for $d = 1, 2, \ldots, 365$. We assume these time series are independent and identical distributed. A consequence of having 365 independent time series for each day is that we need an observation at hour $h = 0$ each day in order to initialize the prediction model. In other words must the sensors transmit the first observation of the day to the fusion center.

Assuming identical time series means that all the time series can be used in the estimation of the parameters. The expression in equation (2.2) is the log-likelihood for one time series $\{Y_t\}$, but as we have several identical, independent time series, we get a sum of several log-likelihood functions,

$$\mathcal{L}(\alpha, \phi, \sigma^2_\varepsilon; \boldsymbol{r}_{d,h}) = \sum_{d=1}^{365} \mathcal{L}(\phi, \alpha, \sigma^2; \boldsymbol{r}_h). \tag{4.9}$$

We choose for simplicity to use all available data for estimating our parameters, as well as testing our models on the same data. For the TTT-prediction models we estimate 3 parameters for each sensor from 8760 data points. Hence is the number of parameters to be estimated very small compared to the size of data available, and overfitting is not likely to be an issue.

4.4 Estimating the moments for the truncated multivariate Gaussian random variable

The TTTS-nowcast is calculated from a truncated multivariate Gaussian distribution. More specifically are the point predictions for the nowcast the expected value of the truncated multivariate Gaussian random variable. However, as we stated in section 2.2, is the expression for the moments not on closed form. A routine for calculating the moments numerically exists in `R`, **tmvtnorm** by Manjunath and Wilhelm (2012), but in the work of this thesis have we found that this routine is unstable. We therefore use a Gibbs Sampler,

[1]This is admittedly a wrong assumption as they are issued a couple hours later due to being very computer intensive. Though, the assumption have no practical importance on the results or methods.

Givens and Hoeting (2012), to approximate samples from the truncated distribution which is used to estimate the moments. The Gibbs Sampler is a Markov Chain Monte Carlo algorithm for obtaining samples from multivariate pdfs. The samples can then be used to investigate properties of the distribution, such as the expected value and variance.

We present an algorithm that generates samples from the truncated multivariate Gaussian distribution. The algorithm is based on a Gibbs sampling procedure from Givens and Hoeting (2012), but adapted to our case. We denote $\boldsymbol{X} = [X_1, \ldots, X_n]$ and

$$\boldsymbol{X}_{-i}^{(t+1)} = [X_1^{(t+1)}, \ldots, X_{i-1}^{(t+1)}, X_{i+1}^{(t)}, \ldots, X_n^{(t)}].$$

The Gibbs sampler requires that the univariate conditional density of $X_i|\boldsymbol{X}_{-i} = \boldsymbol{x}_{-i}$, denoted $f(x_i|\boldsymbol{x}_{i-1})$, is easily sampled for $i = 1, 2, \ldots, n$. In our case is $f(x_i|\boldsymbol{x}_{i-1}$ truncated univariate conditional Gaussian distributed. The univariate random variable $X_i|\boldsymbol{X}_{i-1}$ without truncation is a conditional Gaussian random variable, i.e. $X_i|\boldsymbol{X}_{i-1} \sim \mathcal{N}(\mu_{i|i-1}, \Sigma_{i|i-1})$, where $\mu_{i|i-1}$ and $\Sigma_{i|i-1}$ can be found from equation (2.4). We then have that the truncated random variable $X_i|\boldsymbol{X}_{i-1}$ is a truncated univariate Gaussian random variable. The pdf for this distribution is found in equation (2.5), and is easily sampled from.

Set starting values $\boldsymbol{x}^{(0)}$;
for *i in 1:n* **do**
  generate sample from
  $(X_i^{(t+1)}|\boldsymbol{X}_{-i}^{(t+1)} = \boldsymbol{x}_{-i}^{(t+1)}) \sim f(x_i|\boldsymbol{x}_{-i}^{(t)})$,
  where $f(x_i|\boldsymbol{x}_{-i}^{(t)}$ is
end
Repeat the loop $T + k$ times;
Return $\boldsymbol{X}^{(t)}, t = T, T + 1, \cdots T + k$ as samples from the desired distribution.

Algorithm 1: Gibbs sampler

Using Algorithm 1, we generate samples from the truncated multivariate distribution that are used for estimating the moments which is used in the TTTS-nowcast.

Chapter 5

Simulation Study and Evaluation

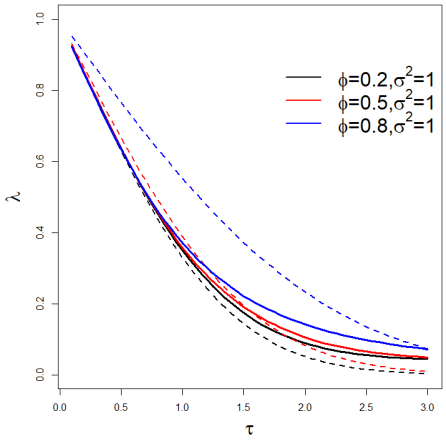
In this chapter we test the TTT-communication scheme and the TTTS-nowcasting model on a simulation study. Section 5.1 presents the simulation study for the communication scheme. A simulation study for nowcasting is performed in Section 5.2. The aim of the simulation studies are to gain understanding about how the performance of the models. In the case study at Spitsbergen which is presented in the next chapter, the temperature is fitted to the statistical models we use in the TTT-scheme and TTTS-nowcasting. In this simulation study we study the performance on simulated data from theoretical processes, which eliminates the effect of potential poor model choice and estimation of parameters. Furthermore does the simulation study enable us to test our strategy for several processes, so we can investigate under which conditions the TTT-communication scheme and TTTS-nowcast model performs the best.

Section 5.1 provides approximated theoretical results for the transmission rate using the TTT-communication scheme. The study provides us with an idea of how the models will perform for the case at Spitsbergen.

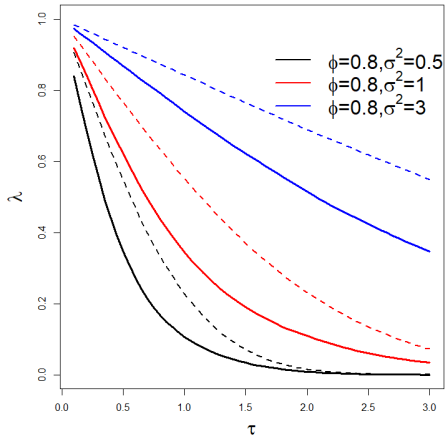
5.1 Simulation study: Communication scheme

In this section we investigate the performance of the TTT-communication scheme on simulated time series. We simulate $AR(1; \phi, \sigma_\varepsilon^2)$-series with different choices of the autocorrelation ϕ and white noise variance parameter σ_ε^2. By investigating the empirical transmission rates of these series, we get an approximation of the theoretical transmission rates. These results can therefore be used to evaluate the performance of the TTT-communication scheme for different temporal dependencies. These results provide insight on which parameters our strategy performs well.

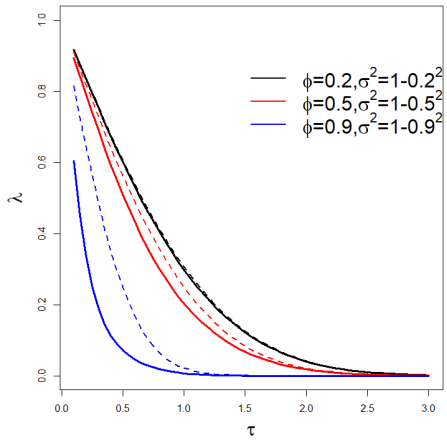
Figure 5.1 show the transmission rate λ as a function of the tolerance τ for several AR(1)-series with different parameters using the TTT-communcation scheme. The dashed lines show the transmission rates for a simpler, constant communication schemes. That is, for all t are the intervals kept constant, $I = [-\tau, \tau]$. The time series simulated have a length of $N = 100,000$, and the transmission rate is the portion of points were outside



(a) Constant white-noise variance parameter, $\sigma_\varepsilon^2 = 1$.



(b) Constant autocorrelation parameter, $\phi = 0.8$



(c) Constant variance, $\mathrm{Var}[Y_t] = 1$

Figure 5.1: Transmission rates for a simulated AR-series with different values for the parameters using the TTT-communication scheme shown in solid lines. Dashed lines shows the transmission rates for the corresponding AR-series when using a constant communication scheme, $I = [-\tau, \tau]$.

their corresponding intervals. Figure 5.1a shows the transmission rate for 3 series with different values for autocorrelation ϕ and constant white-noise variance σ_ε^2. We see from the figure that the there is only a minor difference in the transmission rates for the 3 series. This is somewhat counter - intuitive, as stronger autocorrelation should imply that intervals based on predictions of the process are beneficial. However, the variance of the process is a function of ϕ, as $\mathrm{Var}[Y_t] = \sigma_\varepsilon/(1-\phi^2)$. Larger variance means that the predictions is less sharp, and the process is more spread. From this plot it seems that the larger variance and stronger autocorrelation more or less outweighs each other.

In Figure 5.1b the autocorrelation is kept constant for the 3 series, whilst the white-noise variance σ_ε^2 is varied. The transmissions rates in this plot behave more like expected: Larger variance leads to higher transmissions rates. Also the transmissions rates in Figure 5.1c are as expected. In this figure the variance of the process is held constant, that is $\mathrm{Var}[Y_t] = 1$, whilst ϕ and σ_ε^2 are varied. As expected, stronger autocorrelation leads to smaller transmission rates.

The TTT-communication scheme outperforms the constant communication scheme for all the time series except for two. For the time series with weak and medium autocorrelation, the black and red lines in Figure 5.1a, the constant communication scheme gives a lower transmission rates for tolerances greater than 1.

We conclude from Figure 5.1 that the strategy of constructing the communication scheme based on predictions of the process — compared to keeping a fixed communication scheme fixed— is beneficial in most cases. The exceptions are when the process exhibits weak autocorrelation.

5.2 Simulation study: Nowcasting

In this section we do a simulation study in order to demonstrate and test the TTTS-nowcast model. The TTTS-model is based on the distribution $\boldsymbol{Y}|\mathcal{I}$, where $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ and $\mathcal{I}$ contains some information about $\boldsymbol{Y}$. Remember from section 4.2 that $\mathcal{I}$ may take three different forms, dependent on what information is known about $\boldsymbol{Y}$. The first option, in which $\mathcal{I}$ contains exact observations for all variables in $\boldsymbol{Y}$ is obviously not interesting. The second option is a bit more interesting. The point prediction is not adjusted, as $\mathrm{E}[\boldsymbol{Y}|\mathcal{I}] = \mathrm{E}[\boldsymbol{Y}] = \boldsymbol{\mu}$. The variance is however decreased. Using the Gibbs sampler presented in section 4.3, we compare the covariance matrix for an unconditional random variable $\boldsymbol{Y}$, with the corresponding conditional random variable $\boldsymbol{Y}|\mathcal{I}$. Table 5.1 summarizes the covariance for some different choices of unconditional covariance matrices Σ of different dimensions. The covariance matrices we consider are on the form

$$\Sigma_n = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix},$$

where the variance is 1 for all variables, and the correlation between all pairs of variables are equal.

We see from Table 5.1 that the conditional correlation is significantly reduced when conditioning on $\mathcal{I}$. The conditional variance is also reduced, and the decrease is greater for more dimensions, though the number of dimensions seems to be irrelevant when the covariance for the unconditional variable is low.

	$\mathrm{Var}[Y]$	$\mathrm{Cov}[Y]$	$\mathrm{Var}[Y\|\mathcal{I}]$	$\mathrm{Cov}[Y\|\mathcal{I}]$
$n = 1$	1		0.29	
$n = 2$	1	0.8	0.26	0.12
n =2	1	0.2	0.29	0.02
n =5	1	0.8	0.23	0.08
n =5	1	0.2	0.29	0.01
n =10	1	0.8	0.21	0.05
n =10	1	0.2	0.29	0.01

Table 5.1: Overview over the effect on the covariance of conditioning on $\mathcal{I}$, when $\mathcal{I}$ contains information about intervals for which the observations are in.

The remainder of this section focus on the third option, that is when some observations are known, and some variables are truncated. We continue with the notation $\mathcal{I} = [\mathcal{I}^*, \mathcal{I}']$ from section 4.2. For comparison, we introduce three more models for nowcasting,

$$\mathrm{E}[\boldsymbol{Y}|\mathcal{I}'], \tag{5.1}$$

$$\mathrm{E}[Y_i|\mathcal{I}', Y_i \in I_i], \quad i = 1, 2, \ldots, n, \tag{5.2}$$

and

$$\mathrm{E}[\boldsymbol{Y}]. \tag{5.3}$$

Only the expressions for the point-nowcasts from these models are presented here. We will also evaluate the models in terms of CRPS, which requires a predictive distribution for the nowcast. The distributions are derived using the methods presented in Chapter 4, and is not repeated here. The model in equation (5.1) calculates the nowcast by conditioning only on the observations that are known. Note that this model in equation (5.1) may give nowcasts outside the interval we know for the variable, as the truncation is not applied. The model in equation (5.2) conditions on the observations that are known, and also use the truncation for the given variable Y_i. The model in eqation (5.2) is a truncated version of equation (5.1) for the variable Y_i. The model in equation (5.3) corresponds to using the TTT-point predictions, as no spatial dependence are utilized. All models have in common that they do not utilize all the information available. We therefore expect these models to perform worse than the TTTS-nowcast,

$$\mathrm{E}[\boldsymbol{Y}|\mathcal{I}]. \tag{5.4}$$

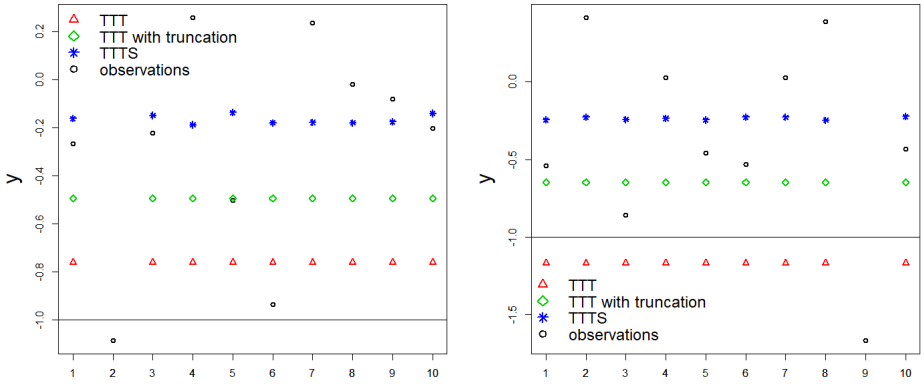
To illustrate the properties of the different models, an example is presented.

Example. *We simulate two samples from the* 10*-dimensional normal distribution*

$$\boldsymbol{Y} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right), \tag{5.5}$$

where $\rho = 0.7$*. Figure 5.2 shows the samples together with the nowcasts from equations* (5.1)*,* (5.2) *and* (5.4)*. The black points are the samples, blue points are the nowcast*

from TTTS-model, (5.4)*, and the red and green points are respectively the nowcasts from equations* (5.1) *and* (5.2)*. The TTT-point nowcast is not included in the figure, as it simply equals the mean of* $\boldsymbol{Y}$ *for the observations not transmitted. The horizontal lines represent the intervals,* $I = [-1, 1]$ *for where the observations are not transmitted to the fusion center. Note that in figure 5.2b is the red points outside the interval, as the truncation is not used in equation* (5.1)*. The blue points, the TTTS-nowcast, are closer to the center of the interval than the green points. Conditioning on* $\mathcal{I}^*$ *has the effect that the nowcast is pulled towards the unconditional mean of* $\boldsymbol{Y}$*.*



(a) 1 sample with point-nowcasts from the distribution given in equation (5.5).

(b) 1 sample with point-nowcasts from the distribution given in equation (5.5).

Figure 5.2: Overview of different nowcast models in action. All the nowcast-models condition on the observations that have been transmitted to the fusion center. In addition to observations shown in black points, is the point-nowcasts from the model in equation (5.1), (5.2) and (5.4) shown in respectively red, green and blue points. The horixontal lines represents the interval $I = [-1, 1]$ for which the observations are not transmitted to the fusion center.

We simulate 1000 10-dimensional samples for the random variable defined in equation 5.5 with different values for the correlation ρ. The mean CRPS for the different models are summarized in table 5.2. We see from the table that the TTTS-nowcast, $\boldsymbol{Y}|\mathcal{I}$, performs best for all choices for ρ. The difference in CRPS are greater for more correlated processes. Table 5.3 shows the RMSE for the different models. The TTTS-nowcast outperforms the other models here as well.

The MSE and CRPS are calculated from the same simulations. We see from the tables 5.2 and 5.3 that the TTTS-nowcast model performs slightly better than the other models presented in this section for comparison. However, the gain of using the TTTS-nowcast is greater when $\boldsymbol{Y}$ is more correlated.

In addition to testing the TTTS-nowcast for the random variable in equation (5.5), where all the elements in the random vector have equal variance and equal correlation between all variables, we use the TTTS-nowcast on a different random variable. The

	$\boldsymbol{Y}\|\mathcal{I}$	$\boldsymbol{Y}$	$\boldsymbol{Y}\|\mathcal{I}'$	$Y_i\|\mathcal{I}', Y_i \in I_i$
$\rho = 0.2$	0.203	0.205	0.22	0.203
$\rho = 0.5$	0.166	0.187	0.213	0.167
$\rho = 0.8$	0.091	0.147	0.13	0.095

Table 5.2: Mean CRPS for the different nowcast models tested on 3 different simulations with different correlation. Each simualtion contains 1000 simulations of the 10-dimensional variable defined in equation 5.5.

	$\boldsymbol{Y}\|\mathcal{I}$	$\boldsymbol{Y}$	$\boldsymbol{Y}\|\mathcal{I}'$	$Y_i\|\mathcal{I}', Y_i \in I_i$
$\rho = 0.2$	0.431	0.442	0.534	0.431
$\rho = 0.5$	0.371	0.418	0.666	0.384
$\rho = 0.8$	0.243	0.383	0.517	0.269

Table 5.3: RMSE for the different nowcast models tested on 3 different simulations with different correlation. Each simualtion contains 1000 simulations of the 10-dimensional variable defined in equation 5.5.

random vector resulting from the TTT-model will rarely be on the form of equation (5.5), so we sample a random covariance matrix in order to test the TTTS-model. That is, we simulate 1000 10-dimensional samples from the distribution $\mathcal{N}(\mathbf{0}, \Sigma)$, where

$$\Sigma = \begin{bmatrix} 3.85 & 2.7 & 3.42 & 2.26 & 2.87 & 2.47 & 1.51 & 1.92 & 2.62 & 1.45 \\ 2.7 & 4.15 & 2.45 & 1.65 & 2.87 & 2.06 & 2.23 & 2.33 & 2.5 & 2.19 \\ 3.42 & 2.45 & 3.55 & 2.45 & 2.8 & 2.64 & 1.67 & 2.06 & 2.7 & 1.7 \\ 2.26 & 1.65 & 2.45 & 2.56 & 1.77 & 1.84 & 1.88 & 1.41 & 1.35 & 1.63 \\ 2.87 & 2.87 & 2.8 & 1.77 & 3.41 & 2.02 & 2.02 & 1.82 & 2.74 & 1.81 \\ 2.47 & 2.06 & 2.64 & 1.84 & 2.02 & 2.71 & 1.16 & 2.01 & 2.17 & 1.74 \\ 1.51 & 2.23 & 1.67 & 1.88 & 2.02 & 1.16 & 2.47 & 1.23 & 1.4 & 1.76 \\ 1.92 & 2.33 & 2.06 & 1.41 & 1.82 & 2.01 & 1.23 & 1.96 & 1.71 & 1.57 \\ 2.62 & 2.5 & 2.7 & 1.35 & 2.74 & 2.17 & 1.4 & 1.71 & 3.1 & 1.6 \\ 1.45 & 2.19 & 1.7 & 1.63 & 1.81 & 1.74 & 1.76 & 1.57 & 1.6 & 1.88 \end{bmatrix} \tag{5.6}$$

The RMSE and mean CRPS are summarized in table 5.4. We see that also here does the TTTS-model slightly outperform the other models.

	$\boldsymbol{Y}\|\mathcal{I}$	$\boldsymbol{Y}$	$\boldsymbol{Y}\|\mathcal{I}'$	$Y_i\|\mathcal{I}', Y_i \in I_i$
RMSE	0.254	0.367	0.570	0.286
CRPS	0.091	0.138	0.143	0.103

Table 5.4: RMSE and mean CRPS for the different nowcast models tested on a simulation containing 1000 samples from the distribution in equation (5.5) with covariance matrix given in equation (5.6).

The simulation study has shown that the TTTS-nowcast model is preferable among those we have compared. However, for weakly correlated spatial processes, is the TTT-nowcast almost as good as the TTTS-nowcast.

Chapter 6

Spitsbergen: Communication Scheme and Nowcasting

In this chapter, the TTT-communication scheme and TTTS-nowcast model is used for the case study at Spitsbergen. In Section 6.1 is the results from using the TTT-communication scheme for the sensors presented. Section 6.2 presents the results from nowcasting using the TTTS-model for a given tolerance limit $\tau = 1$.

6.1 Communication scheme for a sensor network at Spitsbergen

In this section we use the TTT-communcation scheme for the sensor network at Spitsbergen. Our aim is to reduce the transmission rates for the sensors, whilst still being able to produce good nowcasts for the temperature. As in Section 4.1, we model the discrepancy independently for each location, resulting in one univariate time series for each location.

Table 6.1 shows the maximum likelihood estimators for the parameters $\alpha, \phi, \sigma_\varepsilon^2$ for the locations considered. The parameter estimation for the time series $\{R_t\}$ is discussed in section 4.3.

Figure 6.1 shows the TTT-communication scheme presented in action on an excerpt of the period considered. The green line represents the NWP, while the black dots is the observed temperatures. The blue shaded area is the observed value with a tolerance of $\tau = 1$, that is $y_t \pm 1$. The blue dots represent the observations transmitted to the fusion center, as the prediction missed the blue area. The red dots show the TTT-predictions that missed the interval.

Figure 6.2 shows in red lines the transmission rate as a function of the tolerance τ for each of our 10 sensors. The black lines show the transmission rates for a simulated AR(1)-series with parameters corresponding to the estimated parameters in Table 6.1. The blue lines are the same as in figure 3.7b, which show the transmission rate if the NWP in addition to a constant term is used as prediction only, $Y_t = \alpha + x_t$. The figure shows that

Location	α	$SD(\alpha)$	ϕ	$SD(\phi)$	σ_ε^2
1	-0.515	0.017	0.893	0*	0.54
2	1.173	0.077	0.913	0*	1.786
3	-3.382	0.043	0.913	0*	1.007
4	-0.804	0.031	0.936	0*	0.446
5	-0.724	0.022	0.913	0*	0.52
6	-0.599	0.025	0.892	0*	0.789
7	-2.419	0.044	0.884	0*	1.589
8	-2.591	0.044	0.894	0*	1.392
9	-0.098	0.038	0.905	0*	0.978
10	0.78	0.061	0.935	0*	0.904

Table 6.1: Estimated parameters for the discrepancy time series $\{R_t^s\}$ for each location s with standard error for the estimated values. * indicates that the value is smaller than 10^{-4}.

τ	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10
0.5	0.43	0.52	0.43	0.34	0.38	0.42	0.51	0.46	0.48	0.45
1	0.22	0.31	0.23	0.16	0.19	0.21	0.29	0.26	0.25	0.25
1.5	0.13	0.21	0.15	0.1	0.12	0.14	0.18	0.17	0.16	0.16

Table 6.2: Transmission rate for each sensor s using the TTT-communication scheme for some values of the tolerance τ.

the model for discrepancy, defined in equation (4.3), clearly outperforms the more simple models investigated in section 3.1.2. Note also that the transmission rate is bounded from below at $\lambda_{\text{min}} = 1/24$ in this communication scheme, as we require each sensor to transmit the observation made at $00:00$ UTC each day.

The plots in Figure 6.2 show the transmission rate as a function of the tolerance. Table 6.2 shows the transmission rate for some chosen values of τ. The transmission rate is considerably smaller than 1, which would be the transmission rate if all observations were transmitted. A nowcaster may now choose a tolerance level, and equip the sensors with batteries that can handle the transmission rates we have found. For example we see in the table that sensor 9 and 10 on average transmits an observation 1 in 4. In our setup that means that on average they transmit 6 observations a day, and must have a energy resources scaled according to this.

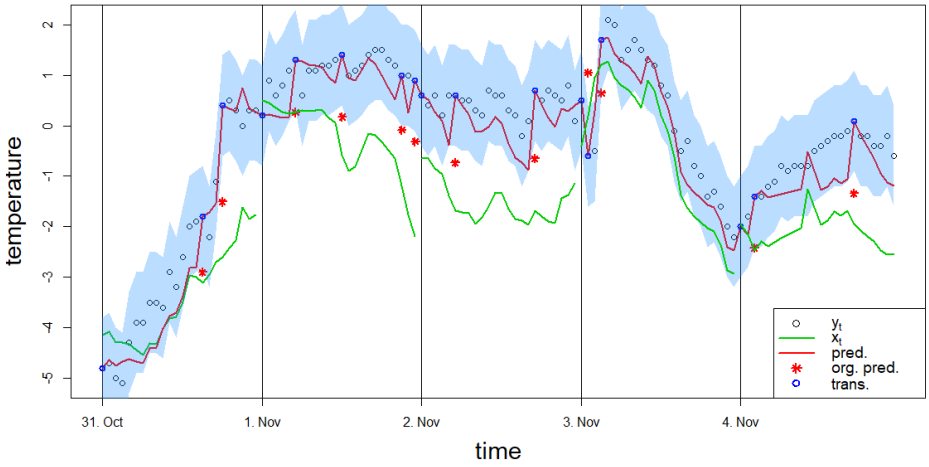


Figure 6.1: Observed values, NWP and the resulting TTT-communication scheme on an excerpt of the period considered at Hornsund observation station. The red points indicates the TTT-predictions that led the sensor to transmit its observation, as the observed value was outside the interval for that time step. The red line shows the TTT-point prediction after updating for transmitted observations.

6.2 Nowcasting the temperature of Spitsbergen

By using the approach presented in section 4.2, we produce nowcasts for the temperature at Svalbard. Note that we really produce nowcasts for the discrepancy $\boldsymbol{R}$. However, as the NWP is known for all sensors and the fusion center is it equivalent to nowcasting the temperature $\boldsymbol{Y}$. The nowcasts is therefore presented for the temperature. In this section we set the tolerance limit to be 1, that is $\tau = 1$. As presented in section 4.2, we use the univariate distributions from the TTT-model as the marginal distributions for $\boldsymbol{Y}$. Remember, $\boldsymbol{Y}$ is really conditioned on $\boldsymbol{\mathcal{H}}$, where $\mathcal{H}^s$ represents the information about Y^s known to the fusion center at previous time points at location s. We omit this notation, and let $\boldsymbol{Y}$ represent the conditional distribution $\boldsymbol{Y}|\boldsymbol{\mathcal{H}}$. Using an empirical correlation matrix and the parameters estimated in section 6, the distribution $\boldsymbol{Y}$ is fully explained. The empirical correlation matrix is estimated from historical data for the discrepancy r. Hence, can we produce the nowcast for each time point sequentially from $\boldsymbol{Y}_t|\mathcal{I}$.

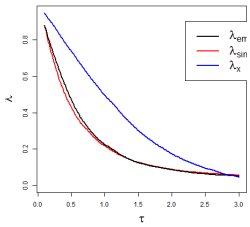
For all time points t we have the distribution

$$\boldsymbol{Y}_t \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma),$$

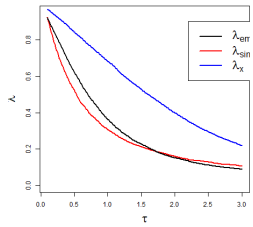
where

$$Y_t^s \sim \mathcal{N}(\widehat{Y}_{t|\cdot}^s, \sigma_{t|\mathcal{H}_t}^{s2}), \tag{6.1}$$

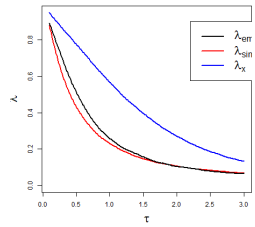
and $\Sigma_t^{i,j} = \sqrt{\sigma_{t|\mathcal{H}_t}^{s2}}\sqrt{\sigma_{t|\mathcal{H}_t}^{s2}}\,\mathrm{Cor}[Y^i, Y^j]$. The following example and figure 6.3 illustrates the TTTS-nowcast and TTT-prediction for the time point $t = 259$.



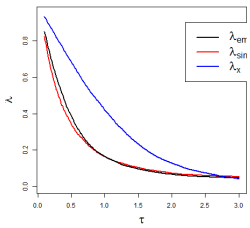
(a) Hornsund



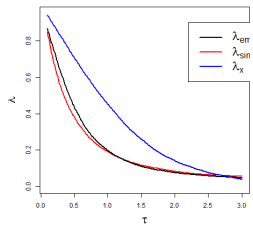
(b) Sveagruva



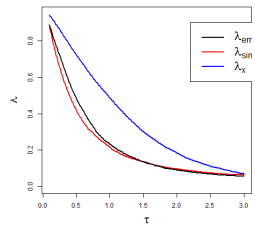
(c) Sveagruva II



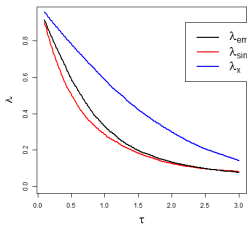
(d) Akseløya



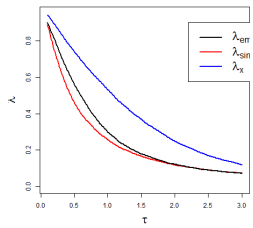
(e) Isfjord Radio



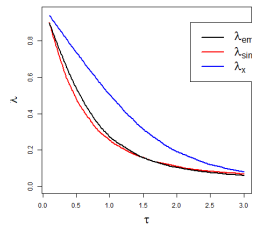
(f) Svalbard Lufthavn



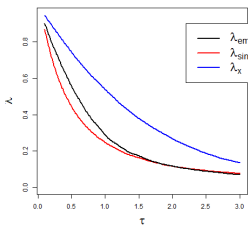
(g) Adventdalen



(h) Pyramiden



(i) Ny-Ålesund



(j) Verlegenhuken

Figure 6.2: Transmission rate for each sensor as a function of tolerance limit τ using the TTT-communication scheme are shown in black lines. The red lines shows the approximated theoretical transmission rate for a sensor monitoring an AR(1-process with corresponding parameters as estimated for the discrepancy at each location. The parameters are found in Table 6.1. The approximated theoretical results are found by simulations as done in Section 5.1. The blue lines are the transmission rates when using a constant communication scheme as done in Section 3.1.2.

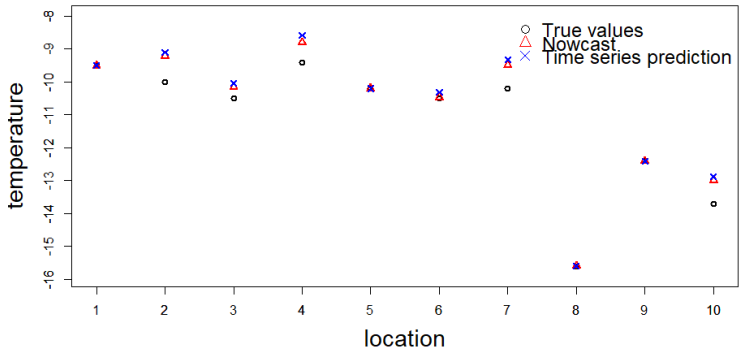


Figure 6.3: Visualization of TTT-point prediction, TTTS-nowcast and observed values at the time point $t = 259$ for all locations.

Example. *Figure 6.3 shows the nowcast for time point* $t = 259$ *in red triangles, the temperature in black circles and the TTT-predictions in blue crosses. At this time point observations from stations* 1, 5, 8 *and* 10*, that is* $y^1_{t=259}$, $y^5_{t=259}$, $y^8_{t=259}$, $y^9_{t=259}$, *are available for the fusion center. The last known observations for the other stations, i.e.* $t - k^i$ *are from time points* $[257, 257, 254, 256, 258, 258]$ *for respectively stations* 2, 3, 4, 6, 7 *and* 10.

The correlation matrix used for the case study is the empirical correlation between the discrepancies between the TTT-point predictions and the observed temperatures for the different locations[1]. It is given in equation (6.2).

$$\widehat{\mathrm{Cov}[\boldsymbol{R}]} = \begin{bmatrix} 1.00 & 0.22 & 0.20 & 0.21 & 0.18 & 0.14 & 0.15 & 0.16 & 0.10 & 0.06 \\ 0.22 & 1.00 & 0.47 & 0.21 & 0.16 & 0.22 & 0.25 & 0.24 & 0.11 & 0.12 \\ 0.20 & 0.47 & 1.00 & 0.26 & 0.19 & 0.27 & 0.27 & 0.24 & 0.13 & 0.12 \\ 0.21 & 0.21 & 0.26 & 1.00 & 0.29 & 0.29 & 0.25 & 0.20 & 0.20 & 0.12 \\ 0.18 & 0.16 & 0.19 & 0.29 & 1.00 & 0.25 & 0.19 & 0.15 & 0.19 & 0.10 \\ 0.14 & 0.22 & 0.27 & 0.29 & 0.25 & 1.00 & 0.40 & 0.25 & 0.21 & 0.17 \\ 0.15 & 0.25 & 0.27 & 0.25 & 0.19 & 0.40 & 1.00 & 0.23 & 0.21 & 0.15 \\ 0.16 & 0.24 & 0.24 & 0.20 & 0.15 & 0.25 & 0.23 & 1.00 & 0.16 & 0.14 \\ 0.10 & 0.11 & 0.13 & 0.20 & 0.19 & 0.21 & 0.21 & 0.16 & 1.00 & 0.17 \\ 0.06 & 0.12 & 0.12 & 0.12 & 0.10 & 0.17 & 0.15 & 0.14 & 0.17 & 1.00 \end{bmatrix} \tag{6.2}$$

Table 6.3 summarizes the results from the TTTS-model in terms of CRPS and RMSE respectively. For comparison are the CRPS and RMSE for the TTT-predictions included. The difference between the results from the two models can be interpreted as the increase of accuracy of utilizing the spatial dependency in the process. The TTTS-nowcast performs only slightly better than the TTT-nowcast. Considering the correlation in the spatial

[1] Again are we using the same data set for estimation and testing, but due to the large size of our set is this ignored.

	RMSE		CRPS	
	TTTS	TTT	TTTS	TTT
1	0.439	0.441	0.250	0.257
2	0.433	0.436	0.239	0.257
3	0.437	0.443	0.246	0.259
4	0.440	0.443	0.258	0.256
5	0.435	0.437	0.252	0.254
6	0.437	0.443	0.249	0.258
7	0.432	0.437	0.235	0.257
8	0.430	0.433	0.238	0.255
9	0.441	0.444	0.246	0.259
10	0.439	0.440	0.245	0.258
mean	0.436	0.440	0.246	0.257

Table 6.3: Results from the case study at Spitsbergen. The TTTS-nowcast is compared with the nowcast consisting of the TTT-point predictions and TTT-random variables.

process, given in equation (6.2) are these results coherent with the results from the simulation study in section 5.2, which showed that the gain of using the nowcast from the TTTS-model compared with the TTT-nowcast is small when the spatial process is weakly correlated.

Chapter 7

Discussion and conclusion

In this study, we have considered a resource-constained sensor network that communicates with a fusion center. We have proposed the TTT-communication scheme for the sensors in order to reduce the number of transmissions. It is desirable to limit the number of transmissions from the sensors in the network to the fusion center, as communication is a big drain on the battery sensors. The communication scheme we propose is a Temporal dependence, Tolerant limit, Trigger based communication scheme that drastically reduces the number of transmissions from the sensors. The TTT-model utilizes temporal dependence in a process in order to predict the process ahead in time. Based on this prediction a tolerance limit is set for the process, which leads to the trigger based scheme for when the sensor should transmit its observations. Based on the TTT-scheme we have developed a spatial extension, named TTTS. It aims to nowcast the process using the information available in the fusion center. The approach have been tested in a simulation study, as well as on a case study of the hourly temperature at Spitsbergen in 2017.

We have shown that using the TTT- and TTTS-model for random variables results in truncated random variables. Specifically we have considered Gaussian random variables, which resulted in truncated Gaussian random variables. We present both point nowcast and probabilistic nowcast in this thesis, where both rely on the first and second moments of the truncated multivariate Gaussian random variable. However, no closed form exists for these moments. Additionally, available software by Manjunath and Wilhelm (2012) for calculation of these moments has been experienced to be unstable. We have therefore presented an algorithm for estimating these moments. The algorithm we have presented is based on the Gibbs sampler, but as creating an efficient sampler was not in the main scope of this thesis, is it very likely room for improvements.

The TTT-communication scheme have been tested on a process that follows the AR(1)-model. The simulation study shows, not surprisingly, that the TTT-scheme is more effective when the process considered has strong temporal dependency. Further work on this topic may include deriving a theoretical expression for the transmission rate for a sensor observing an AR(1)-process using the TTT-communication scheme. A simulation study has also been performed for the TTTS-model. In terms of RMSE and CRPS, have we

shown that utilizing the spatial model TTTS is preferable to only using the purely temporal model TTT. When nowcasting a process with low spatial dependency, does the results admittedly show that the TTTS model only performs negligible better than the TTT-model. A possible extension of this work may be to nowcast the temperature for locations without a sensor. In such an extension is it natural to rely on the TTTS-model, as the TTT-model not utilizes any spatial dependency.

The TTT-communication scheme and its spatial extension TTTS can be interpreted as a separable spatial-temporal model, Cressie and Wikle (2011). The communication scheme is kept separate from the spatial dependence, as it is vital that the fusion center and each sensor operates with the same communication scheme, as that is necessary for being able to gain information when an observation is not transmitted. In order to include the spatial dependence in the communication scheme, each sensor is dependent on information about the process at other locations. In our case have we assumed that transmitting information to a sensor from the fusion center is very energy-consuming.

In the case study of the temperature at Spitsbergen, we chose a maximum error tolerance for our nowcast at $\pm 1^{\circ}$C. This yielded in a reduction in the number of transmissions from the sensors by $70 - 80\%$. The TTT-model were also evaluated in terms of RMSE and CRPS. Further work on this topic may include regarding this setup in a Value of Information regime, Eidsvik et al. (2015). Both the simulation study and case at Svalbard has shown that there is a trade-off between the transmission rate and nowcast quality, e.g. tolerance limit, CRPS, RMSE. Further work on the TTT-communication scheme may be to adapt the scheme to fit into the applications described by Chapman et al. (2014) and Chapman and Thornes (2011), which uses a sensor network for winter road maintenance. The communication scheme in this setting should perhaps depend on the value of the process as well, e.g. that higher accuracy is desired when the temperature is near the freezing point. This also opens up for introducing self-triggered communication sensors, Heemels et al. (2012).

Bibliography

Adhikari, R., Agrawal, R. K., 2013. An introductory study on time series modeling and forecasting. arXiv preprint arXiv:1302.6613.

Bardeen, J. M., Szalay, A., Kaiser, N., Bond, J., 1985. The statistics of peaks of gaussian random fields. Astrophys. J. 304 (FERMILAB-PUB-85-148-A), 15–61.

Cartinhour, J., 1990. One-dimensional marginal density functions of a truncated multivaiate normal density function. Communications in Statistics-Theory and Methods 19, 197–203.

Chapman, L., Thornes, J., 2011. What spatial resolution do we need for a route-based road weather decision support system? Theoretical and applied climatology 104 (3-4), 551–559.

Chapman, L., Young, D., Muller, C., Rose, P., Lucas, C., Walden, J., 01 2014. Winter road maintenance and the internet of things.

Cressie, N., Wikle, C. K., 2011. Statistics for spatio-temporal data. hoboken.

Eidsvik, J., Mukerji, T., Bhattacharjya, D., 2015. Value of Information in the Earth Sciences: Integrating Spatial Modeling and Decision Analysis. Cambridge University Press.

Givens, G. H., Hoeting, J. A., 2012. Computational statistics. Vol. 710. John Wiley & Sons.

Gneiting, T., Raftery, A. E., 2004. Strictly proper scoring rules, prediction. Tech. rep., and estimation. Technical Report 463, Department of Statistics, University of Washington.

Heemels, W., Johansson, K., Tabuada, P., 2012. An introduction to event-triggered and self-triggered control. In: 51st IEEE Conferance on Decision and Control. Maui, Hawaii, USA, pp. 3270–3285.

Horrace, W. C., 2005. Some results on the multivariate truncated normal distribution. Journal of Multivariate Analysis 94, 209–221.

International Civil Aviation Organization, 2017. INTERNATIONAL STANDARD ATMOSPHERE.

Johnson, N. L., Kemp, A. W., Kotz, S., 2005. Univariate discrete distributions. Vol. 444. John Wiley & Sons.

Kleijnen, J. P., 2009. Kriging metamodeling in simulation: A review. European journal of operational research 192 (3), 707–716.

Manjunath, B., Wilhelm, S., 2012. Moments calculation for the doubly truncated multivariate normal density.

MET Norway, 2018. Private weather observations improve temperature forecasts on Yr. https://www.met.no/en/archive/private-weather-observations-improve-temperature-forecasts\-on-yr.

NETATMO, 2018. Thanks to Netatmo, Yr, the Norwegian Weather public service, can give more accurate forecasts for 5 countries! https://www.netatmo.com/blog/en/weather/understanding-the-impact-of-weather-on-our\-daily-lives-for-better-tomorrows-thanks-to-the\-netatmo-community/.

Sanctis, M. D., Cianca, E., Araniti, G., Bisio, I., Prasad, R., Feb. 2016. Satellite Communications Supporting Internet of Remote Things. IEEE Internet of Things Journal 3 (1), 113–123.

Tamkittikhun, N., Hussain, A., Kraemer, F. A., 2017. Energy consumption estimation for energy-aware, adaptive sensing applications. In: Mobile, Secure, and Programmable Networking. pp. 222–235.

The Norwegian Meteorological Institute, 2017. The AROME-Arctic weather model. https://www.met.no/en/projects/The-weather-model-AROME-Arctic.

UK Met Office, 2018. Nowcasting. https://www.metoffice.gov.uk/learning/making-a-forecast/hours-ahead/nowcasting.

University of Illionois, 2018. Numerical Weather Predition. http://ww2010.atmos.uiuc.edu/(Gh)/wwhlpr/numerical_weather_prediction.rxml.

Walpole, R. E., Myers, R. H., Myers, S. L., Ye, K. E., 2002. Probability Statistics for Engineers Scientists.

Wei, W., 01 2006. Time Series Analysis: Univariate and Multivariate Methods, 2nd edition, 2006.