Cooperative Information Augmentation in a Geosensor Network

Malte Jan Schulze, Claus Brenner, Monika Sester
Institute of Cartography and Geoinformatics, Leibniz Universität Hannover, Germany
Appelstraße 9a, 30167 Hannover
{maltejan.schulze, claus.brenner, monika.sester}@ikg.uni-hannover.de

Keywords: geosensor network, simulation environment, cooperation, data acquisition, distributed processing, mobile, mapping quality, meteorology

This paper presents a concept for the collaborative distributed acquisition and refinement of geo-related information. The underlying idea is to start with a massive amount of moving sensors which can observe and measure a spatial phenomenon with an unknown, possibly low accuracy. Linking these measurements with a limited number of measuring units with higher order accuracy leads to an information and quality augmentation in the mass sensor data. This is achieved by distributed information integration and processing in a local communication range.

The approach will be demonstrated with the example where cars measure rainfall indirectly by the wiper frequencies. The a priori unknown relationship between wiper frequency and rainfall is incrementally determined and refined in the sensor network. For this, neighboring information of both stationary rain gauges of higher accuracy and neighboring cars with their associated measurement accuracy are integrated. In this way, the quality of the measurement units can be enhanced.

In the paper the concept for the approach is presented, together with first experiments in a simulation environment. Each sensor is described as an individual agent with certain processing and communication possibilities. The movement of cars is based on given traffic models. Experiments with respect to the dependency of car density, station density and achievable accuracies are presented. Finally, extensions of this approach to other applications are outlined.

1. INTRODUCTION

Geosensor networks are composed of a possibly large number of individual sensors with measuring, positioning and communication capabilities. Through local cooperation of neighboring sensors the whole network is able to perform actions that go beyond an individual sensor's capabilities and achieve a common global goal. In this way the geosensor network is able to acquire information about the environment in an unprecedented detail.

Geosensor networks mark a paradigm shift in measuring systems in two ways: from centralized to decentralized data acquisition, and from a separation of measurement and processing to integrated acquisition and analysis.

The advantages of geosensor networks lie in their scalability and also in their fault tolerance, as the role of individual sensors is not crucial - due to the high redundancy. These properties lead to a large number of applications of geosensor networks e.g. in environmental monitoring or in military.

From a computational and geoinformatics point of view, the challenge is to devise algorithms that are able to work locally and still achieve a common global solution. There are many spatial algorithms that operate in a centralized manner, presuming access to all the information; however, in the case where a local processing unit only has a limited view of the surrounding information, existing algorithms have to be adapted or new ones have to be devised to achieve a decentralized processing.

1.1 Prerequisites of our approach

Sensors can have different capabilities. In our approach, we start with the assumption that the cooperation of a large number of sensors of similar, but limited, quality and a few sensors with higher quality can lead to an enrichment of the poor quality measurement of the limited sensors. The measurements are integrated and accumulated in a Kalman Filter and thus – over time – lead to a higher accuracy of the sensed information.

1.2 Problem statement

Rainfall is the most important information source for hydrological planning and water resources management. Especially the modelling of high dynamic processes like floods and erosion rely on high resolution rainfall information. For this measurement, non-recording stationary gauges exist, which measure with a daily observation interval. These instruments are typically available in a high density (e.g. in Germany 1 station per 90 km^2). The density of recording rain stations is still inadequate (e.g. in Germany 1 station per 1800 km^2).

The idea of our approach is to densify the number of stations using unconventional sensors, which are massively available and can measure rainfall (at least approximately), namely cars: when it rains, car drivers start their wipers in order to clean the windshields. Thus, starting the wipers is an indication for liquid on the windshield; the frequency of the wiper is related to the amount of rainfall. The exact relation between wiper frequency and rainfall is unknown, however, it can be calibrated on-the-fly using measurements from the environment: on the one hand, if a car passes by a recording rain station; on the other hand, if a car passes by another car, which has been calibrated at a rainfall station recently. Thus, by locally exchanging and accumulating the measurements, the quality of the a priori unknown information, namely the amount of rainfall, can incrementally be determined and refined.

1.3 Approach

We simulate traffic and rainfall using a real road network. Traffic is simulated by generating random routes on the road network; the rainfall is simulated by generating a raincloud. Cars move in this environment and measure rainfall with their wipers. The initial coarse rainfall measurement quality of each car is iteratively improved through local cooperation of moving cars and rainfall stations.

1.4 Overview of the paper

After a description of related work, we will introduce our approach to the above described problem in section 3. We describe our simulation environment and the implementation of the Kalman filter. In section 4, examples are shown which verify the results. Section 5 gives a brief summary and an outlook on future work.

2. RELATED WORK

A general overview of wireless sensor networks is given in (Akyildiz et al., 2002). Geosensor networks for the observation and monitoring of environmental phenomena are a recent trend in GIScience. Traditional geodetic networks consist of a fixed set of dedicated sensors with a given configuration and measurement regime. The processing of the data is usually done in a centralized fashion. The advent of geosensor networks brings about the chance to move from a centralized approach to an approach using distributed sensors with computation and communication capabilities (Stefanidis & Nittel 2004).

The advantages as opposed to a centralized system are its scalability, and its high spatial and temporal resolution. In order to fully exploit a geosensor network in the way described, methods for local information aggregation have to be devised. Such methods have to take the neighbourhood and the communication range of the individual sensors into account. There are many application areas for geosensor networks, e.g. environmental observations (Duckham & Reitsma, 2009), surveillance, traffic monitoring and new multimodal traffic (Raubal et al., 2007).

Decentralized algorithms for geosensor networks have been investigated by several researchers and for different applications. Laube et al. (2008) describe an algorithm to detect a moving point pattern, namely a so-called flock pattern. A flock is described as a group of objects that moves in a certain distance over a certain time. In a similar spirit, Laube & Duckham (2009) present a method for the detection of clusters in a decentralized way. Depending on the communication range, clusters of a certain size (radius) can be detected.

Walkowski (2008) presents an approach for the optimal arrangement of geosensor nodes in order to correctly describe an underlying temporally varying phenomenon, like a toxic cloud. He assumes to have sensors that are able to move; however, the determination of the locations of lacking information has to be determined in a centralized fashion. Zou & Chakrabarty (2004) describe an approach to optimally cover an area with a given set of sensors. Sester (2009) presents an approach for cooperative detection of a boundary of a spatial phenomenon using a mobile geosensor network.

For traffic simulation there are programs that simulate not only the movements of the traffic objects on the infrastructure, but also the behaviour and the decisions of the users. For a consistent modelling of these aspects agent based approaches are used, where each traffic participant is modelled individually (Raney & Nagel, 2006).

In terms of fusing measurements in an optimal way, Kalman filtering is a widely employed technique, which is described in standard textbooks (Brown & Hwang, 1997, Simon, 2006).

The principle applicability and suitability of our approach has been investigated earlier by Haberlandt & Sester (2009). There, the main focus was to explore the quality of the interpolation taking different traffic densities and given wiper-rainfall-relationships into account.

3. APPROACH

3.1 Basic concept of simulation environment

The main objective in our work is to describe the quality of rain measurement using cars as rain gauges. In opposition to rain measurement stations that can record the rainfall data directly by using dedicated rainfall sensors, the cars in our approach do not have such sensors. We consider the wiper frequencies of a car as correlated to the rainfall intensity. When the intensity is high, one would switch the wiper frequency of the car to a high value in order to have a better visibility. When there is no rainfall at all, the wipers of the car would not be used.

The cars are considered as sensor nodes that can measure their position (for example via GPS) and their wiper frequency. In addition, they can perform calculations based on the locally collected data and share them with other cars using a wireless communication device (see Fig. 1).

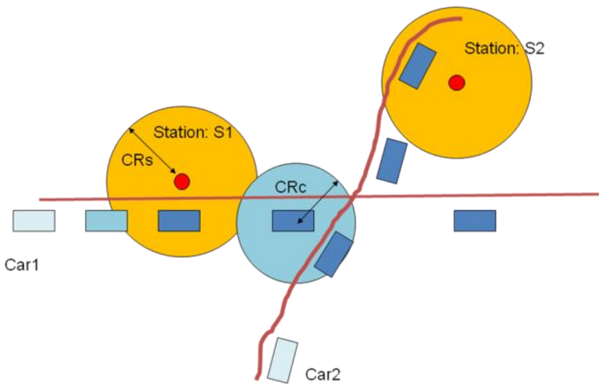

Fig. 1: Communication between cars and stations, with communication ranges CRc and CRs, respectively.

In order to determine the intensity of the rainfall from the wiper frequency information, we need a functional relationship between the wiper frequency and the rainfall intensity, otherwise the collected wiper frequency data of a car leads to a very uncertain estimation for the rainfall intensity. To simulate this case, we give cars without any information about the functional relationship a high standard deviation.

To provide high quality rain measurement data, a few weather stations, that can measure the rainfall intensity with a very high certainty, are distributed across our road network. The cars can use those high quality data, to improve their own certainty about the rainfall measurement.

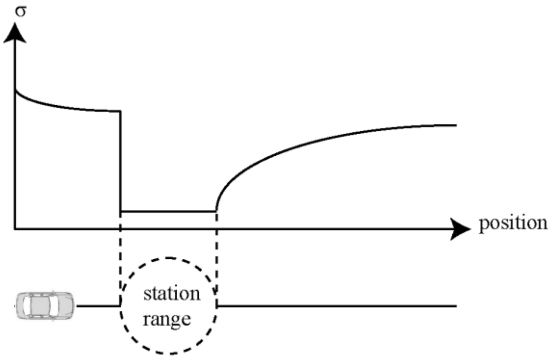

Fig. 2: Improvement of the certainty of a car by communication with a weather station.

As shown in Fig. 2, the standard deviation decreases rapidly, when the car enters the communication range of a weather station, leading to a high certainty of rainfall measurements from the car. When the car leaves the communication range, the certainty gently decreases until it reaches the original level. While decreasing, the car can still share its information with other cars that are not in the range of a weather station, helping to improve their level of certainty.

3.2 Implementation of simulation environment

Car movement

The simulation environment describes an agent based system, where each car is considered as an agent that follows a certain trajectory through a road network. We determine the movement of the cars by randomly selecting start- and endpoint of each trajectory. The movement through the road network is calculated using the A*- algorithm to determine the shortest path. The visited nodes of the road network are saved together with a timestamp. The simulation itself is based on a central start- and end time with constant time steps of 10s. For each step, the position of all cars is calculated by using a linear interpolation between two nodes.

Rainfall simulation

The rainfall intensity in our simulation environment is modelled by a mixed Gaussian with randomly distributed centers. The calculated field is normalized. The calculation of the Gaussian is based on (1). The result for the simulated raincloud is shown in Fig. 3. For this simulation, the rainfall intensity is considered to be stationary.

$$cloud(x,y) = \frac{1}{2\pi\sigma} \cdot e^{-\frac{(x-x_0)^2+(y-y_0)^2}{2\sigma^2}} \tag{1}$$

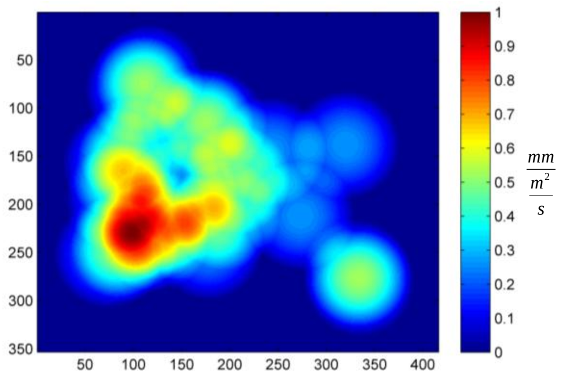

Fig. 3: Simulated distribution of rainfall intensity.

Observation of rainfall and communication strategy

For each car, a Kalman filter is implemented to describe the system state $\mathbf{x}$ and its quality $\mathbf{\Sigma}^{+}_{xx,k}$ (2).

$$\begin{aligned}
&\mathbf{x}^{+}_{k} = \begin{bmatrix} \dot{x}^{+}_{k} \\ \ddot{x}^{+}_{k} \end{bmatrix} \rightarrow \mathbf{\Sigma}^{+}_{xx,k} = \begin{bmatrix} \sigma^{+\ 2}_{\dot{x},k} & 0 \\ 0 & \sigma^{+\ 2}_{\ddot{x},k} \end{bmatrix}, \mathbf{\Sigma}_{ww} = \begin{bmatrix} \sigma^{2}_{w\dot{x}} & 0 \\ 0 & \sigma^{2}_{w\ddot{x}} \end{bmatrix} \\
&\mathbf{x}^{-}_{k+1} = \dot{x}^{+}_{k} + \triangle t^{k+1}_{k} \cdot \ddot{x}^{+}_{k} \quad \rightarrow \quad \mathbf{\Phi}^{\mathbf{k+1}}_{\mathbf{k}} = \begin{bmatrix} 1 & \triangle t^{k+1}_{k} \\ 0 & 1 \end{bmatrix}
\end{aligned} \tag{2}$$

The system state consists of two variables $\dot{x}$ and $\ddot{x}$. The rainfall intensity is described by $\dot{x}$, which can be considered as the rainfall speed, having the unit $mm/m^2/s$. It can be determined from the wiper frequencies of a car and is directly observed by a weather station. As the cars move underneath the stationary rainclouds, a second parameter $\ddot{x}$ is estimated, which describes the change of the rainfall intensity, having the unit $mm/m^2/s^2$. The certainty of the system state is described by the covariance matrix $\mathbf{\Sigma}^{+}_{xx,k}$. The covariance increases with the time passed, as the system noise $\mathbf{\Sigma}_{ww}$ accumulates. To make a statement about the quality of the rainfall measurement, we focus on the standard deviation $\sigma^{+}_{\dot{x},k}$ of the rainfall intensity. To predict the system state in the next epoch k+1, the transition matrix $\mathbf{\Phi}^{\mathbf{k+1}}_{\mathbf{k}}$ is used. This is a standard transition matrix usually employed for the estimation of object positions using the assumption of constant speed. To update the system state with observations, three different cases of communication are taken into account:

1. The car is located outside the communication range of other cars and stations. In this case, there is no data exchange. The car determines the rainfall intensity l^{own}_{k+1} with a high standard deviation $\sigma^{own}_{l,k+1}$ due to the uncertainty of the wiper-rainfall relationship. Only one observation is used to update the system state.

2. The car is located inside the communication range of a weather station. The weather station determines the rainfall intensity and transmits the data to the car. Once the data exchange is done, the car uses the observation $l^{station1}_{k+1}$ and its small standard deviation $\sigma^{station1}_{l,k+1}$ to update its own system state. The weather station does not update its measurements with the car measurements, because the weather station is measuring with highest accuracy and the improvement by the cars is not significant. The small standard deviation helps to improve the certainty of the system state (as shown in Fig. 2). If the car is in communication range of two or more stations, the observations are put together in a vector and their standard deviations are used to build a covariance matrix for the observation vector.

3. The car is located outside the communication range of a weather station, but inside the communication range of another car. It receives the rainfall intensity and its standard deviation from the system state of the other car and uses it as an

observation together with its own observation, if its system state is more uncertain than the system state of the other car. The rainfall intensities can be considered as equal, as the communication range of a car is very small. If more cars with a smaller standard deviation are in communication range, all observations are put together in an observation vector $\mathbf{l}_{k+1}$ and its covariance matrix $\mathbf{\Sigma}_{ll,k+1}$.

$$\mathbf{l}_{k+1}=\begin{bmatrix} l_{k+1}^{own} \\ l_{k+1}^{car,1} \\ \vdots \\ l_{k+1}^{car,n} \end{bmatrix} \rightarrow \mathbf{\Sigma}_{ll,k+1}=\begin{bmatrix} \sigma^2_{l_{k+1}^{own}} & & & \mathbf{0} \\ & \sigma^2_{l_{k+1}^{car,1}} & & \\ & & \ddots & \\ \mathbf{0} & & & \sigma^2_{l_{k+1}^{car,n}} \end{bmatrix} \tag{3}$$

Mapping of the rainfall

In order to map the rainfall data, the area of the road network is converted from vector to raster data. Each cell from the road network is a possible candidate to receive information about the rainfall once a car passes by. We consider two factors that will have influence on the quality of the mapped data. The quality of the information in a cell is decreasing with the elapsing time, but it will increase with the number of cars that pass this cell. In order to model this fact, a second Kalman filter for each cell that can be passed by a car is implemented. Its system state is described as follows:

$$\begin{aligned} &\dot{x}_k^+ \rightarrow \mathbf{\Sigma}_{xx,k}^+ = \sigma_{\dot{x},k}^{+\ 2},\ \mathbf{\Sigma}_{ww} = \sigma_{w\dot{x}}^2 \\ &\mathbf{\Phi}_{\mathbf{k}}^{\mathbf{k+1}} = 1 \end{aligned} \tag{4}$$

As in our case the simulated raincloud is static, we do not need the parameter $\ddot{x}_k^+$, which was implemented in the Kalman filter for the cars (2). The decay in quality is modelled with the system noise $\mathbf{\Sigma}_{ww}$, which is added to the system state at every time step as a part of the prediction.

Once a car passes by, the system state of the cell is updated, using the system state of the car about the rainfall intensity and its standard deviation as an observation.

After the simulation run, we are able to make a statement about the quality of the rainfall mapping by looking at the following statistics:

- The difference between the mapped data and the simulated values.
- The standard deviation of each cell.
- The number of times a cell has been visited.
- The coverage of the area.

They will be presented in the following chapter, where we discuss the first experiments that we have done in the presented simulation environment.

4. EXPERIMENTS

We took road data as well as the locations of the weather stations from a study area of approx. 3300 km^2 in the Bode river basin located in the Harz Mountains in Northern Germany (Haberlandt & Sester, 2009). Our results are based on a given car density and station distribution. Some parameters are chosen identical for every run of the simulation: The communication range for a car-car system is set to 200 m and for a station-car system to 2000 m. The simulation time is 1.5 h for each run and the cars are driving with an average speed of 70 km/h. The size for each cell is set to 200 m. The relation between the system noise and the measurement uncertainty, which controls the abatement of the car's certainty, is identical for each run.

Simulation run with 50 cars

As a result of this run, we reached a standard deviation based on differences between the mapped rain values and the given ones of 6 %, which is acceptable. In total, 25 % of all reachable cells were mapped during the simulation. As some of them were visited twice or more often, an average visiting rate of 0.69 for each of them was reached.

An example for the improvement of the system state of a single car is given exemplarily in Fig. 4. It shows that the certainty of the system state of a car improves rapidly when it communicates with a weather station. After the communication range is left, it decreases slightly until it reaches the initial value again. Similarly, the communication with a car leads to an improvement of the quality, although it is not as high as in comparison with the weather station. An interesting fact is shown in the third break of the curve. The system state can improve even more, when two cars communicate several times in a row.

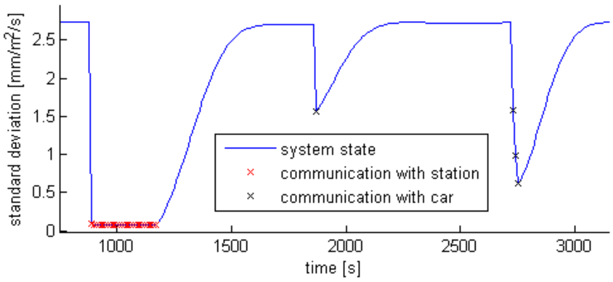

Fig. 4: Improvement of the system state by communication with other participants.

The quality of the mapped data is shown in Fig. 5. It gives an overview over the simulation area, the distribution of the weather stations and shows the standard deviation of each mapped cell.

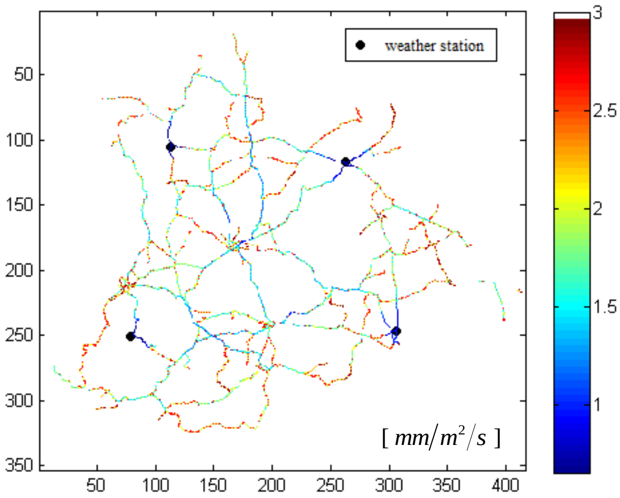

Fig. 5: Standard deviation of each reached cell with a distribution of four stations and 50 cars.

It confirms the statement of Fig. 4, as it shows dark blue areas around the station, which stands for a low standard deviation. The standard deviation on roads, that are chosen more often, seems to be on a lower level than on other roads that fork from them. This effect can be explained by the number of visits, as shown in Fig. 6.

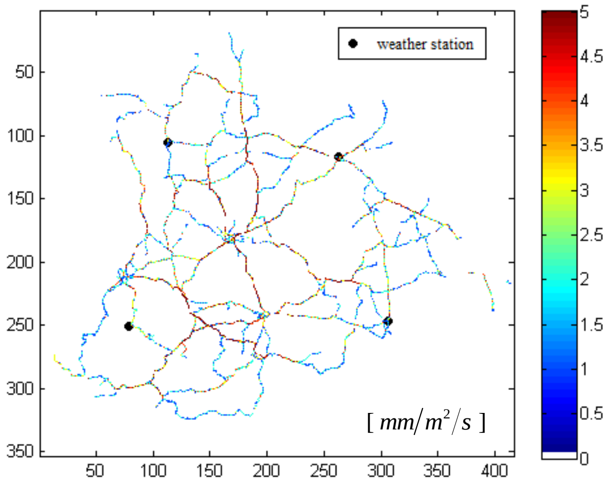

Fig. 6: Number of times a cell has been visited using 50 cars.

It shows that these roads are more often visited, than the other ones. In fact, the correlation between visiting time and variance of a cell is calculated to -0.73, which means, that the quality of mapping is not only affected by the weather station information, but also by the number of visits.

The following example shows the mapping quality results with the original station distribution. The original station distribution leads to a better mapping in the area where they are placed, although some of them are never reached by a car. It confirms the dependency of the mapping quality on the number of visiting times, because the standard deviation between Fig. 5 and Fig. 7 is nearly identical for roads, which have been chosen more often, and therefore nearly independent from the station distribution.

In order to improve the mapping quality, we did another simulation run with 100 cars. The results of this run are presented in the next section.

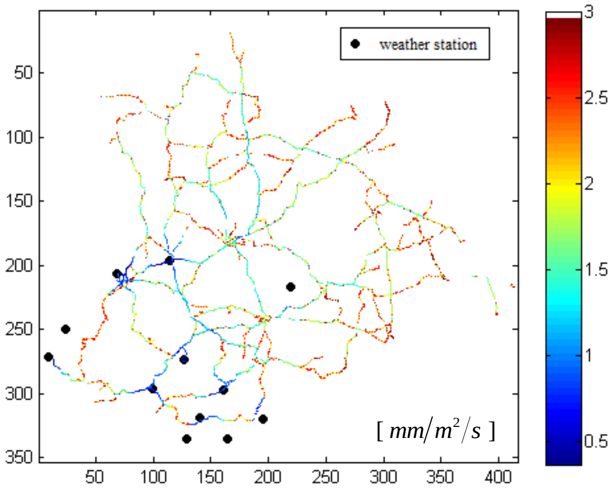

Fig. 7: Standard deviation of each visited cell, using the original distribution of stations.

Simulation run with 100 cars

As a result of this test run, we reached a standard deviation based on differences to the original rain values of about 7 %, which is the same order as the simulation above has shown. The coverage of the area is slightly higher with about 35% of all reachable cells. On average, each cell was visited 1.4 times. The standard deviation of each mapped cell is shown in Fig. 8.

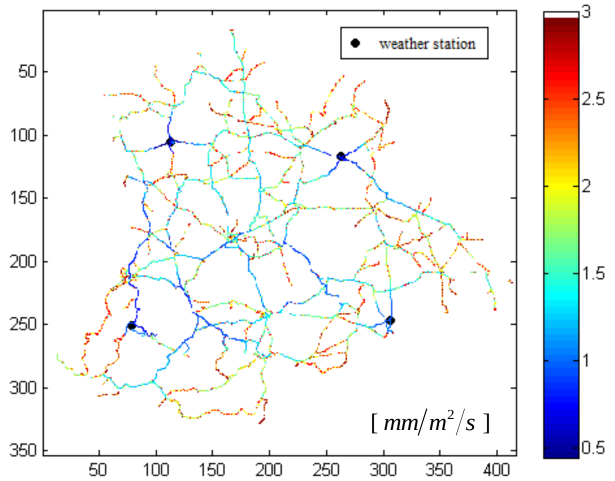

Fig. 8: Standard deviation of each visited cell with a ficticious distribution of four stations and 100 cars.

The main roads of a low standard deviation are much the same as in the tests runs that are described before, but they reached a higher level of system certainty, which can be even at the same level as the area, that is covered by the rain stations. According to the results already mentioned, this indicates that a small number of roads are chosen more often than others, these are the main roads in the network which connect the towns. This leads to the conclusion that weather stations to improve the system state of a car are much more needed at roads that are not so highly frequented, as the main roads. As the chance is high that a car, which receives information from a weather station on a low frequented road, will continue its journey on a main route is much higher than the other way around, the whole area will be mapped with a higher quality.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we presented an approach to use a sensor network in order to predict rainfall intensities over a large area. Our sensor network is made of two different sensor types – highly accurate, but stationary, rain stations, and moving cars, which measure the rainfall only indirectly (and inaccurately) via their wiper frequencies. Although we concentrate on the rainfall application here, the basic principle can be easily adapted to other scenarios which involve moving low-budget sensors which improve their accuracy by communication with other (possibly more accurate) sensors.

In order to evaluate our approach, we used a real street network and real weather station locations. We then simulated rainfall intensity using a mixture of Gaussians as well as the positions of cars over time. From this, we derived results regarding the standard deviation of the estimated rainfall intensity, which is considered to be a measure of the system's certainty about the estimated state.

There are a number of improvements possible, which we will consider in future work. First, we assumed some constants in our simulation, especially the system and measurement noise in the Kalman filters. These constants should be verified using real data. Second, we used a rather simple model for the relationship between the wiper frequency and the rainfall intensity. However, ideally, this relationship should be more complicated and the filter should include calibration parameters, such as an offset and bias. Finally, the assumption of static rainfall could be replaced by a moving rain field and simulated traffic could be replaced by real (measured) traffic frequencies and speeds.

6. REFERENCES

Akyildiz, I. F., Su, W., Sankarasubramaniam, Y. & Cayirci, E., 2002. Wireless sensor net-works: a survey. Comput. Netw. 38(4), 393–422.

Brown, R. G., Hwang, P. Y. C., 1997. Introduction to random signals and applied Kalman Filtering. John Wiley & Sons.

Duckham, M. & Reitsma, F., 2009. Decentralized environmental simulation and feedback in robust geosensor networks. Computers, Environment and Urban Systems, vol. 33(4), 25-268.

Haberlandt, U., 2007. Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event. J. of Hydrol., 332, 144-157, 2007.

Haberlandt, U. & Sester, M., 2009. Areal rainfall estimation using moving cars as rain gauges – a modelling study. Hydrology and Earth System Sciences Discussions, vol. 6, no. 4, p. 4737-4772, 2009.

Laube, P. & Duckham, M., 2009. Decentralized spatial data mining in distributed systems. in H. Miller & J. Han, eds, ‘Geographic Knowledge Discovery, Second Edition’, CRC, Boca Raton, FL, pp. 211–220.

Laube, P., Duckham, M. & Wolle, T., 2008. Decentralized movement pattern detection amongst mobile geosensor nodes. in T. Cova, K. Beard, M. Goodchild & A. Frank, eds, ‘Lecture Notes in Computer Science 5266’, Springer, Berlin, pp. 211–220.

Raney, B. & Nagel, K., 2006. An improved framework for large-scale multi-agent simulations of travel behaviour, in: Rietveld, P., Jourquin, B., Westin, K. (Editors) Towards better per-forming European Transportation Systems, p. 42, London: Routledge.

Raubal, M., S. Winter, S.Teβmann, Ch. Gaisbauer, 2007. Time geography for ad-hoc shared-ride trip planning in mobile geosensor networks, ISPRS Journal of Photogrammetry and Remote Sensing Volume 62, Issue 5, October 2007, Pages 366-381.

Sester, M., 2009. Cooperative Boundary Detection in a Geosensor Network using a SOM, Proceedings of the International Cartographic Conference, Chile, 2009.

Simon, D., 2006. Optimal state estimation. John Wiley & Sons.

Stefanidis, A. & Nittel, S., eds., 2004. Geosensor Networks, CRC Press.

Walkowski, A. C., 2008. Model based optimization of mobile geosensor networks, in L. Bernard, A. Friis-Christensen & H. Pundt, eds, ‘AGILE Conf.’, Lecture Notes in Geoinformation and Cartography, Springer, pp. 51–66.

Zou, Y. & Chakrabarty, K., 2004. Sensor deployment and target localization in distributed sensor networks. ACM Trans. Embed. Comput. Syst. 3(1), 61–91.