SimTech Cluster of Excellence

University of Stuttgart Germany

Efficiently Integrating Environmental Data in Real-Time Simulations

Andreas Benzing, Boris Koldehofe, Kurt Rothermel

Motivation / Challenges

Environmental monitoring has gained increasing attention over recent years for a number of mostly simulation based tasks [?].

- Weather and climate observation and forecasting
- Real-time tracking of phenomena like hurricanes and floods
- Identification of impact of human behavior on nature

A huge and increasing number of sensors produce terabytes of data each day, posing new research challenges

- Scalability to a very large number of sensors and queries
- Provision of low latency despite transmission delays
- Unified access to sensor data from entirely different sources

In addition to these generic challenges, simulation applications commonly require special processing of sensor data before integration.

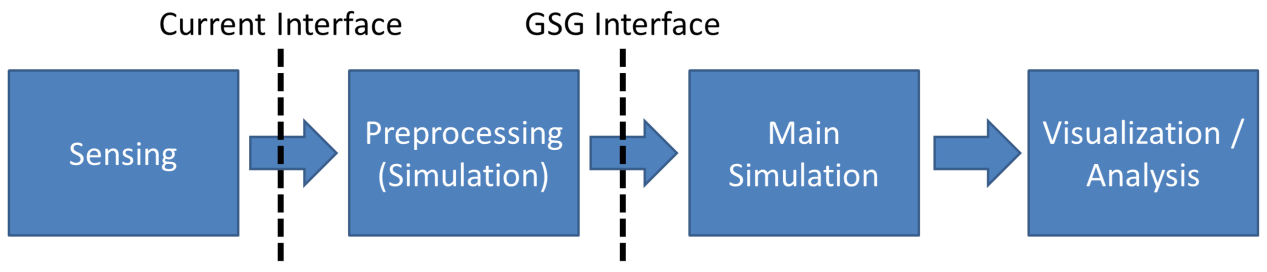

Figure: Common workflow for executing a simulation based on sensor data

For example, a dispersion simulation first has to compute a complete wind field from the measurements of available weather stations which is also done using a separate simulation.

Global Sensor Grid

Within SimTech, in collaboration with the IKE, IST, ISW, and IWS, we examine these special requirements. Especially the provision of sensor data at multiple resolutions is in the focus of our research.

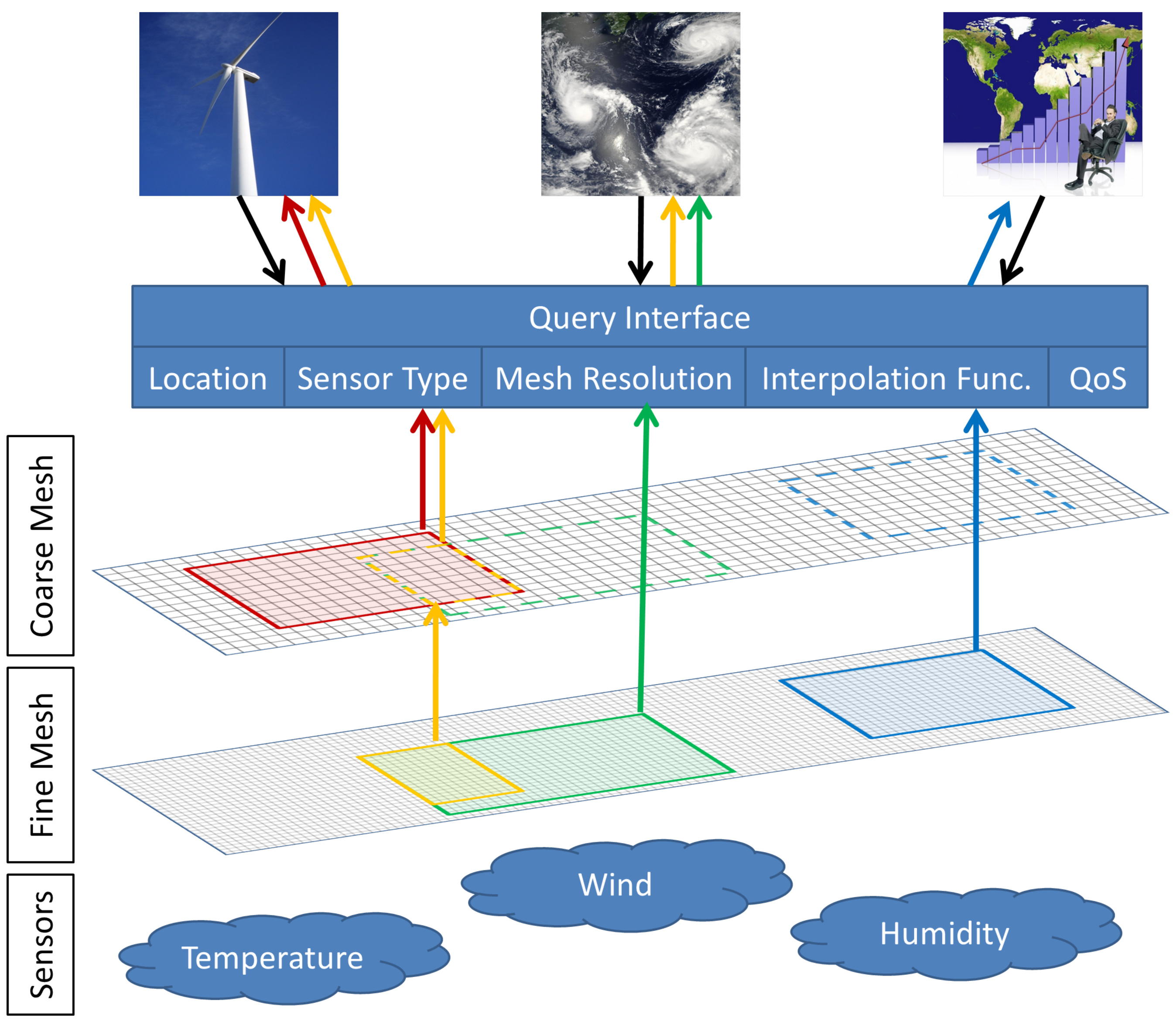

Figure: Overview of the Global Sensor Grid query interface and internal data

The global sensor grid supports the scalable integration of a very large number of sensors and provides a unified interface to sensor data. This helps developers of simulations by alleviating them from

- finding the sensors available for a certain region and parameter
- preprocessing tasks like filtering outliers and aggregating data

In addition, the GSG provides estimated values in real-time to compensate for network delays.

Distributed Data Stream Management

Sensors produce terabytes of data each day and therefore require efficient management of data streams [?, ?]. The goal is to minimize the management overhead and provide as much data as possible to clients without violating the constraints of any network node.

With sets of queries Q, connections C, and brokers B, where a query requires bandwidth $|q|$ and the available bandwidth is a:

$$\textbf{maximize} \sum_{q \in Q} |q| \textbf{subject to} \forall b \in B : \sum_{c \in C_b} |c| + \sum_{r \in Q_b} |r| \leq a_b$$

To minimize the overhead in stream management, similar queries have to be identified. This information can then be used to optimize the organization of streams. Spatial indexing is a method to

- quickly retrieve the sensors of interest and
- efficiently find intersections between queries.

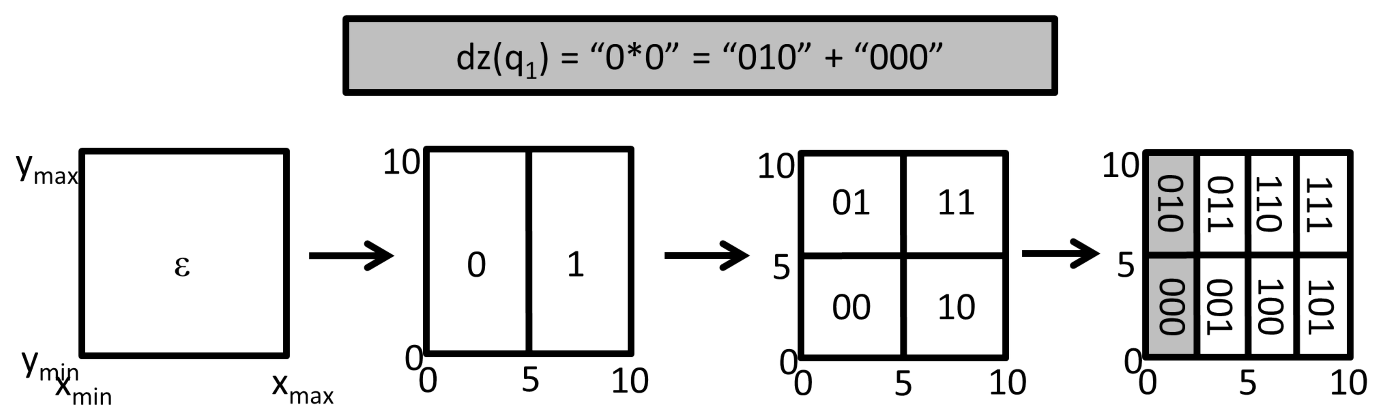

Figure: Indexing of data streams based on query area

For efficient support of preprocessed data, different resolutions have to be provided. By integrating resolution information with multidimensional indexing structures for query management, we significantly decreased the overhead of sensor data streaming.

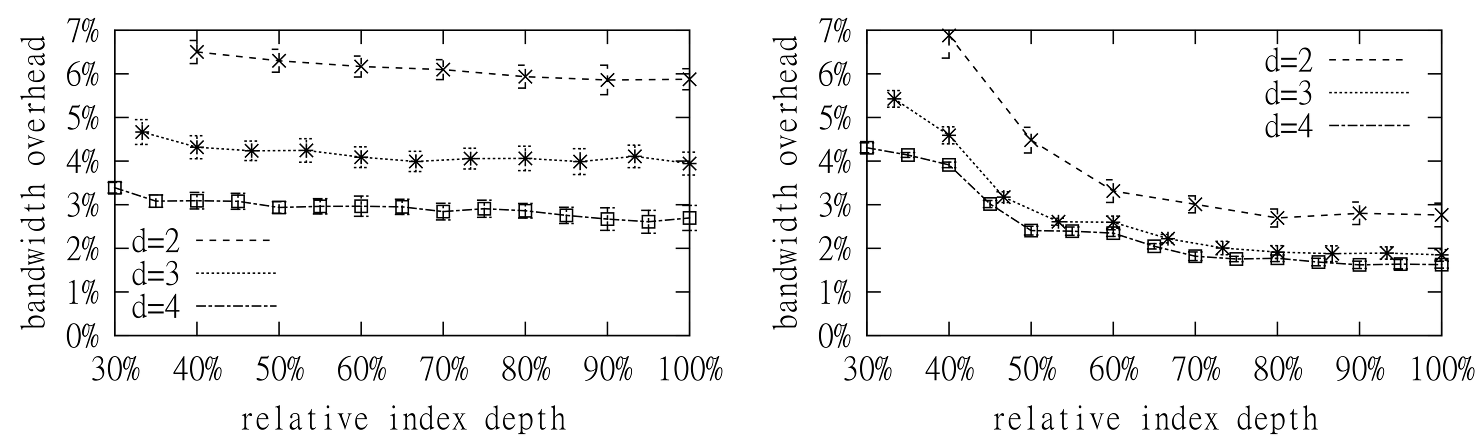

Figure: Overhead required to distribute preprocessed data streams in relation to the granularity of the query size for large (left) and small (right) areas of interest [?]

Future Work

The next step is to adapt situation detection mechanisms and integrate them into the GSG to further automate monitoring. Users can then specify a certain phenomenon and the GSG will automatically identify the region of interest and provide available data.

References

[1] M. Balazinska, A. Deshpande, M.J. Franklin, P.B. Gibbons, J. Gray, S. Nath, M. Hansen, M. Liebhold, A. Szalay, and V. Tao.
Data management in the worldwide sensor web.
Pervasive Computing, IEEE, 6(2):30–40, 2007.

[2] A. Benzing, B. Koldehofe, and K. Rothermel.
Efficient support for multi-resolution queries in global sensor networks.
In *COMSWARE 2011 Proceedings*, 2011 (to appear).

[3] A. Benzing, B. Koldehofe, M. Völz, and K. Rothermel.
Multilevel predictions for the aggregation of data in global sensor networks.
In *DS-RT '10 Proceedings*, pages 169–178. IEEE, 2010.

[4] C.-H. Chen-Ritzo, C. Harrison, J. Paraszczak, and F. Parr.
Instrumenting the planet.
IBM Journal of Research and Development, 53(3):1–16, 2009.