Federal Agency for Cartography and Geodesy

# Optimized Use of Engineered Features from GNSS Reflectometry for Prediction of Significant Wave Heights

Jan M. Becker, Ole Roggenbuck

## Introduction

- Evaluation of interference patterns in GNSS signal-to-noise ratio (SNR) observations allows estimation of Significant Wave Height (SWH) see e.g. [1,2]
- Refinement of method [1] evaluated here
- Outline of [1]: Prediction of SWH with supervised Machine Learning (ML) using engineered features derived from SNR interference patterns with
  i. kernel regression and clustering techniques
  ii. analysis of attenuation of oscillating pattern [2]

  Models used: Linear Regression (LinReg), Artificial Neural Network (ANN), Bagged Regression Tree (BaggedRT)
- Objective of refinements:
  I. Optimized extraction of information for SWH prediction from numerous redundant engineered features for use with LinReg
     ➡ Reduce model complexity
     ➡ Maintain predictive performance
  II. Refined use of engineered features: Apply grid search to Random Forest (RF)
     ➡ advancement over usage of BaggedRT

## Sensors and Data Sets

### Raw data acquisition for supervised ML as in [1]:

- Input data for feature engineering:
  – 1 Hz GNSS SNR observations of GPS L5 with JAVAD TRE 3 DELTA receiver and LEIAR25.R3 or LEIAR25.R4 antenna from FINO2 (Fig.1)
  – IGS precise orbits [3]
  – Meteorological data for elevation angle correction [4] from FINO2
- Ground truth values of SWH from radar sensor (1 minute sampling) from FINO 2

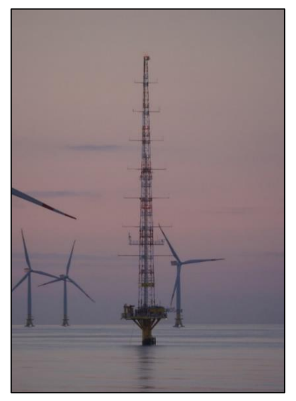

Figure 1: Research station FINO2 in the Baltic Sea. The GNSS antenna used for the acquisition of SNR observation data is mounted on the platform in the lower part of the station. Photo: Federal Agency for Cartography and Geodesy

### Arrangement of data sets for supervised ML as in [1]

| data set | periods used | cases |
|---|---|---|
| training | January 2021 - May 2021 | 4914 |
| testing | November 2020, August 2021, September 2021 | 3402 |

## Methods

### Feature Engineering [1]

1. i. Reflectometric analysis with kernel regression and clustering

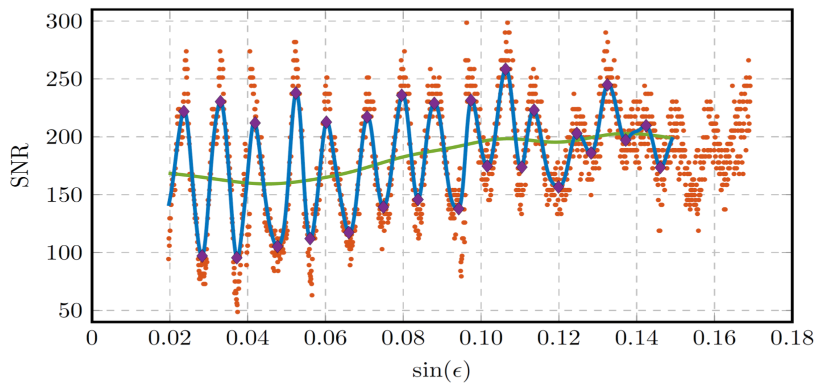


Figure 2: Interference pattern in scattered data evaluated with a kernel regression with appropriate bandwidth (blue). Taken from [1].

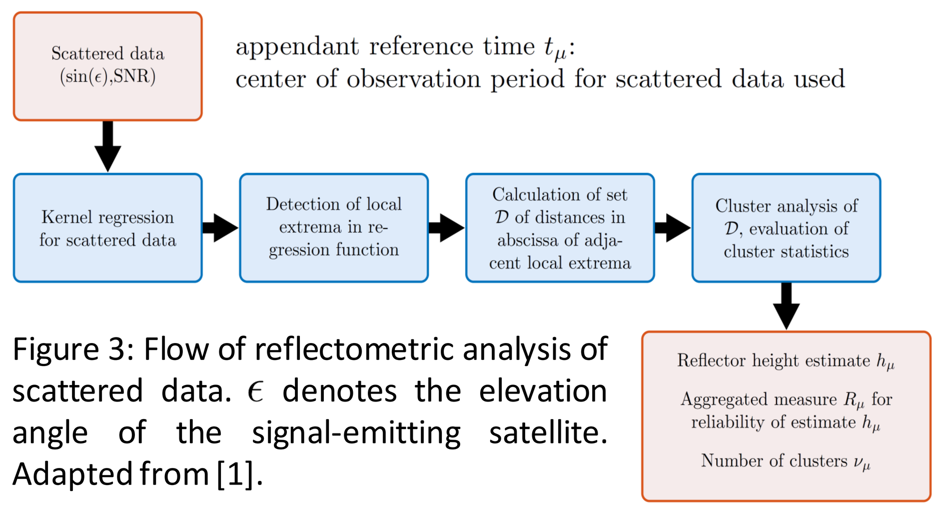


Figure 3: Flow of reflectometric analysis of scattered data. $\epsilon$ denotes the elevation angle of the signal-emitting satellite. Adapted from [1].

   ii. Reflectometric analysis with inverse modelling [2] yields damping coefficient $\delta_\mu$ for reference time $t_\mu$

2. Feature engineering: Use reflectometric analyses with $t_\mu$ in moving time window $MTW(t;T_w)$ of length $2T_w$ centered at time $t$

| Engineered Feature | Explanation |
|---|---|
| $\hat{R}(t;T_w)$ | Average of values $R_\mu$ with $t_\mu \in MTW(t;T_w)$ |
| $\hat{\nu}(t;T_w)$ | Average of values $\nu_\mu$ with $t_\mu \in MTW(t;T_w)$ |
| $\hat{h}(t;T_w)$ | Average of reflector height estimates $h_\mu$ with $t_\mu \in MTW(t;T_w)$ |
| $\hat{s}_h(t;T_w)$ | Standard deviation of reflector height estimates $h_\mu$ with $t_\mu \in MTW(t;T_w)$ |
| $\hat{m}_h(t;T_w)$ | Maximum of reflector height estimates $h_\mu$ with $t_\mu \in MTW(t;T_w)$ |
| $\hat{q}_h(t;T_w)$ | Upper quartile of reflector height estimates $h_\mu$ with $t_\mu \in MTW(t;T_w)$ |
| $\hat{\delta}(t;T_w)$ | Average of damping coefficients $\delta_\mu$ with $t_\mu \in MTW(t;T_w)$ |

Apply 5 different window lengths $T_w = \tau_1, ..., \tau_5$

**Settings of engineered features:**

$$\mathbf{\Pi}_a(t) := \left(\boldsymbol{\pi}_a(t;T_w=\tau_1), ..., \boldsymbol{\pi}_a(t;T_w=\tau_5)\right)$$ **setting *a***

$$\boldsymbol{\pi}_a(t;T_w) := \left(\hat{R}(t;T_w), \hat{\nu}(t;T_w), \hat{h}(t;T_w), \hat{s}_h(t;T_w), \hat{m}_h(t;T_w), \hat{q}_h(t;T_w)\right)$$

$$\mathbf{\Pi}_b(t) := \left(\boldsymbol{\pi}_b(t;T_w=\tau_1), ..., \boldsymbol{\pi}_b(t;T_w=\tau_5)\right)$$ **setting *b***

$$\boldsymbol{\pi}_b(t;T_w) := \left(\hat{\delta}(t;T_w), \hat{R}(t;T_w), \hat{\nu}(t;T_w), \hat{h}(t;T_w), \hat{s}_h(t;T_w), \hat{m}_h(t;T_w), \hat{q}_h(t;T_w)\right)$$

### Optimized extraction of information from engineered features for SWH prediction with Linear Regression

- Apply either forward selection scheme (FSS) similar to [5] or Principal Component Analysis (PCA) [6] to engineered features (setting *a*, *b*).
  a) FSS used in model training, see Fig. 4: rearrangement of features according to their importance

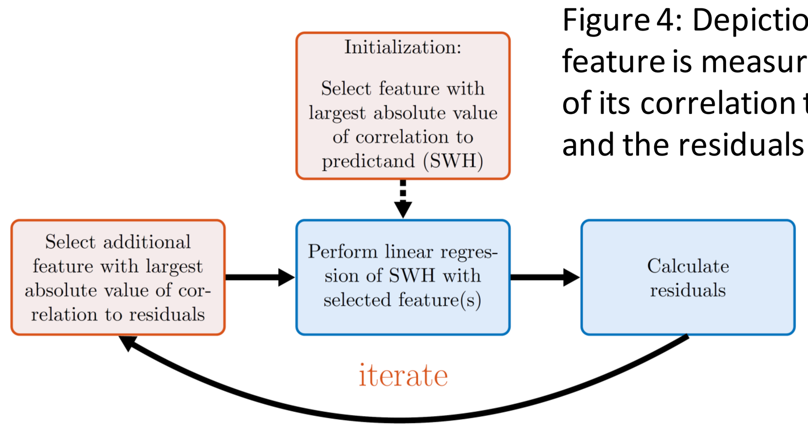


Figure 4: Depiction of FSS. Importance of a feature is measured in terms of the strength of its correlation to the predictand (initially) and the residuals (in further iteration).

  b) Application of PCA:
     – Particular orthogonal transformation of vector of normalized features yields vector of principal components („PCA features“)
     – Transformation (calculated with feature vectors from training data) diagonalizes correlation matrix
- Leading features from PCA and FSS respectively used for SWH prediction

### Grid search for SWH prediction with Random Forest and original engineered features

All combinations of following choices for hyperparameters evaluated in supervised ML

Number of trees: 500, 1500, 2500, 3500, 5000, 1000
Minimal number of samples in node allowing it to be split: 10, 20, 30, 40, 50, 100
Minimal number of samples in node allowing it to be considered as leaf : 10, 20, 30, 40, 50, 100
Maximal depth of trees: 4, 5, 6, 7, 8, 9, 10

## Results

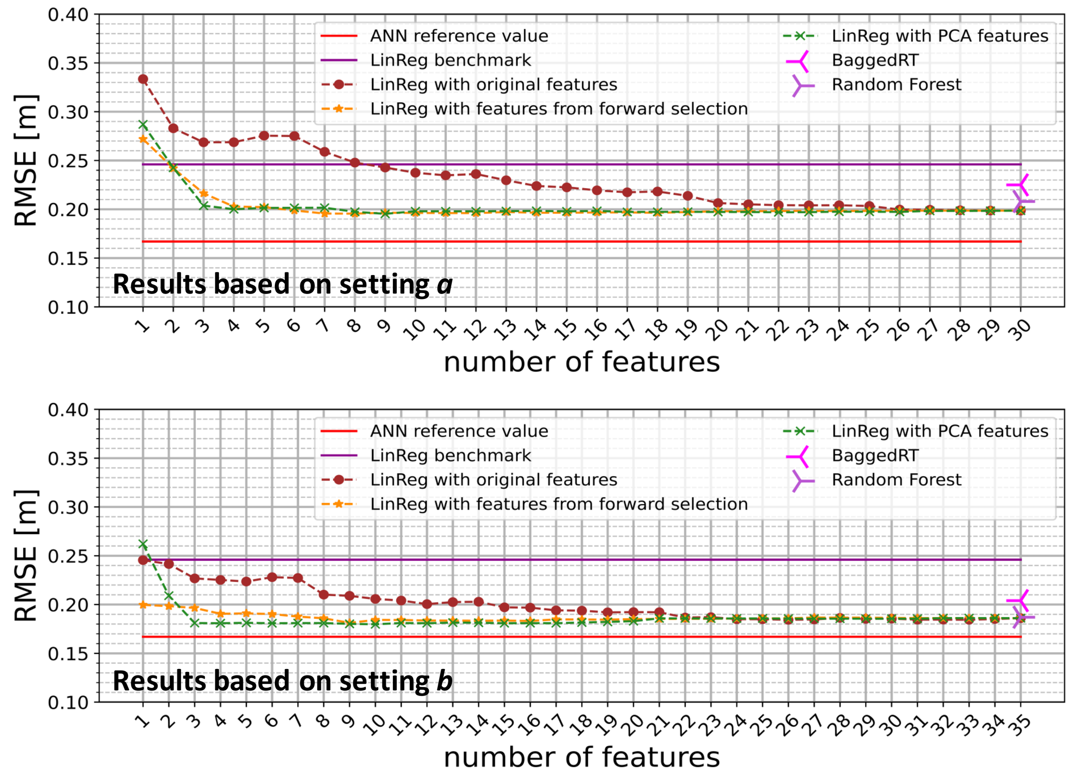


Figure 5: Testing RMSE of SWH predictions with LinReg for increasing number of features taken from the original setting (*a* or *b*) [1], the reordered setting from FSS and the principal components (PCA features), respectively. Additionally depicted testing RMSEs: Results of RF optimized with grid search and BaggedRT [1] (both for the full original feature settings *a*, *b*); best result obtained in [1] with an ANN as reference value; benchmark, i.e. result of LinReg with single feature $\hat{\delta}(t;\tau_1)$ .

| Setting of engineered features | First 9 features selected with FSS |
|---|---|
| $a$ | $\hat{R}(t;\tau_5)$ , $\hat{q}_h(t;\tau_4)$ , $\hat{\nu}(t;\tau_2)$ , $\hat{\nu}(t;\tau_5)$ , $\hat{h}(t;\tau_1)$ , $\hat{m}_h(t;\tau_5)$ , $\hat{h}(t;\tau_5)$ , $\hat{R}(t;\tau_1)$ , $\hat{h}(t;\tau_3)$ |
| $b$ | $\hat{\delta}(t;\tau_4)$ , $\hat{R}(t;\tau_2)$ , $\hat{h}(t;\tau_2)$ , $\hat{\nu}(t;\tau_1)$ , $\hat{R}(t;\tau_5)$ , $\hat{\delta}(t;\tau_1)$ , $\hat{s}_h(t;\tau_4)$ , $\hat{q}_h(t;\tau_5)$ , $\hat{\nu}(t;\tau_5)$ |

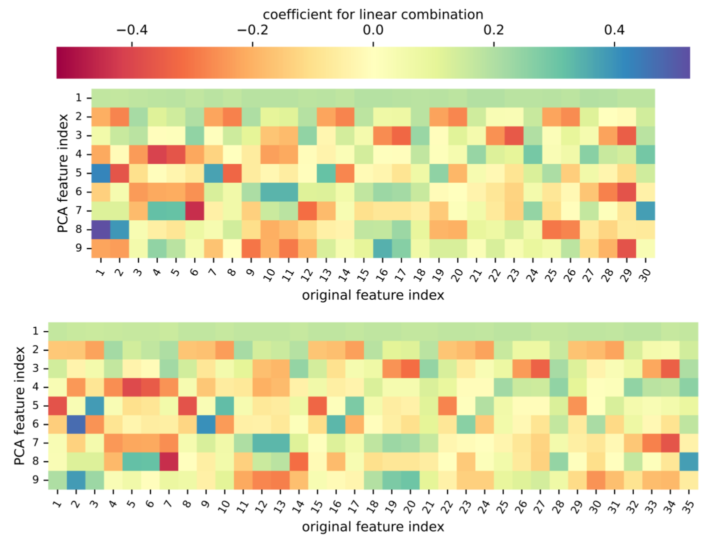


Figure 6: Visualization of the part of the PCA transformation matrix that is associated with the first 9 principal components (PCA features). Top: for setting *a*. Bottom: for setting *b*.

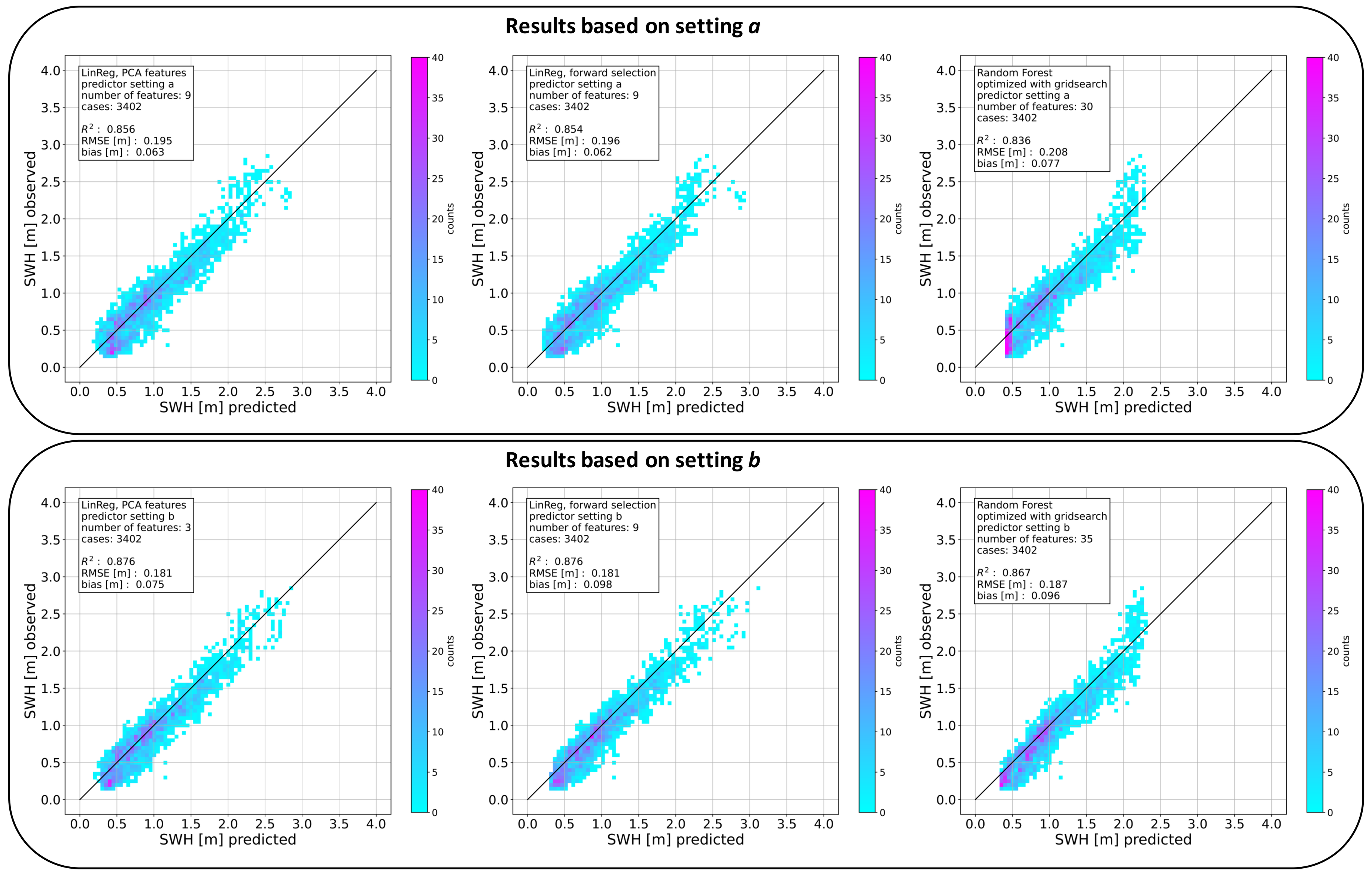


Figure 7: Observed SWH plotted against predicted SWH for refined predictions with LinReg and RF for the testing data. For the predictions with LinReg, only a limited number of leading features from FSS and PCA respectively was used, which suffices to reach the minimum testing RMSE (see Fig. 5).

## Concluding remarks

- With FSS/PCA, information from original engineered features is condensed into a few features that suffice to reach original accuracy of SWH prediction with LinReg
- New engineered features from setting *a* [1] contribute to enhanced SWH prediction with LinReg
- SWH prediction with RF shows improvement over BaggedRT and approaches accuracy of LinReg

### References


[1] Becker, J.M.; Roggenbuck, O. Prediction of Significant Wave Heights with Engineered Features from GNSS Reflectometry. *Remote Sens*. 2023, *15*, 822.
[2] Roggenbuck, O.; Reinking, J.; Lambertus, T. Determination of Significant Wave Heights Using Damping Coefficients of Attenuated GNSS SNR Data from Static and Kinematic Observations. Remote Sens. 2019, 11, 409.
[3] Dach, R.; Schaer, S.; Arnold, D.; Kalarus, M.S.; Prange, L.; Stebler, P.; Villiger, A.; Jäggi, A. CODE Final Product Series for the IGS; Astronomical Institute, University of Bern: Bern, Switzerland, 2020; Available online: http://www.aiub.unibe.ch/download/CODE (accessed on 29 June 2022)
[4] Bennett, G.G. The Calculation of Astronomical Refraction in Marine Navigation. J. Navig. 1982, 35, 255–259.
[5] Glahn, H.R.; Lowry, D.A. The Use of Model Output Statistics (MOS) in Objective Weather Forecasting. J. Appl. Meteorol. Climatol. 1972, 11, 1203–1211.
[6] Wilks, D.S. Statistical Methods in the Atmospheric Sciences, 3rd ed.; Academic Press Elsevier: Oxford, UK, 2011, 519-531.


Sharing is encouraged


### Acknowledgements

We gratefully acknowledge the provision of
- SNR observation data by our colleagues at BKG from Unit G3
- SWH radar sensor data by the FINO initiative (Research Platforms in the North and Baltic Seas), which was organized by the Federal Ministry for Economic Affairs and Climate Action (BMWK) on the basis of a resolution of the German Bundestag, by Project Management Jülich (PTJ) and coordinated by the German Federal Maritime and Hydrographic Agency (BSH).


### Further information

Jan Becker (jan.becker@bkg.bund.de) • Satellite Navigation • Federal Agency for Cartography and Geodesy • Richard-Strauss-Allee 11 • 60598 Frankfurt am Main, Germany • www.bkg.bund.de

BKG
Wir geben Orientierung.