Modular Multi-Sensor Fusion for Underwater Localization for Autonomous ROV Operations

Martin Scheiber*, Alexandre Cardaillac†, Christian Brommer*, Stephan Weiss* and Martin Ludvigsen†
*Institute of Smart Systems Technologies, Group Control of Networked Systems
University of Klagenfurt, Klagenfurt, Austria
{martin.scheiber,christian.brommer,stephan.weiss}@ieee.org
†Faculty of Engineering, Department of Marine Technology
Norwegian University of Science and Technology (NTNU), Trondheim, Norway
{alexandre.cardaillac,martin.ludvigsen}@ntnu.no

***Abstract*—Localization filters for underwater vehicles are mostly tailored for specific sensor suites, environments, or missions. It is also well known that the underwater environment can evolve over time and throughout the mission, affecting the vehicle's sensors, e.g., tide, currents, and vehicle proximity to structures, especially in harbor areas. In this paper, the Modular and Robust Sensor-Fusion Framework (MaRS) is extended to work with underwater vehicles and their environment. It enables efficient use of asynchronous sensors and handles measurement outliers and outages. Sensor-frame initialization and online extrinsic calibration methods are also explored. Tests are performed in real harbor-like environments using a small remotely operated vehicle (ROV) and show improved handling of sensors and state estimation results.**

***Index Terms*—ROV Navigation, State-Estimation, EKF, Sensor-Fusion**

I. Introduction

Usage of remotely operated vehicless (ROVs) in underwater operations has increased over the last decade. They are quickly deployed and minimize the risks for humans. With the growing demand for complex missions such as exploration or inspection tasks relying on higher-level autonomy, accurate and robust localization of such vehicles becomes essential. Underwater navigation is a difficult task, and many challenges exist affecting different sensor modalities. These challenges range from visual disturbances for cameras due to the water quality and light conditions, acoustic reflections and shadow zones for sonars and Doppler sensors or other acoustic-based sensors such as acoustic positioning systems, to magnetic distortions caused by metal structures for magnetometers.

Therefore, there is a need for a reliable combination of sensors leveraging the strength of each sensor type while detecting outliers and failures in other sensors currently adversely affected by the surroundings. Especially in harbor-like environments, where signal reflections and obstructions are present, robust and reliable state estimation is necessary. Past work mainly considered disturbance effects in simulations [1] with multiple sensing modalities. However, the environmental effects causing the sensors to fail are challenging to simulate realistically. Thus, as presented in this work with a modular fusion method, real-world tests are crucial to evaluate a multi-sensor fusion approach for underwater sensor suites thoroughly.

Kalman Filters (KF) and their various variants, such as the extended Kalman filter (EKF), the Unscented Kalman Filter (UKF), and the Error-State Kalman filter (ESKF), which are commonly used in underwater multi-sensor fusion applications for localization [2], [3]. However, they are usually pre-compiled and fixed for a specific sensor configuration prior to the mission. [1] proposed a scalable and modular framework for ROV localization as an adaptation of the localization framework called Vind [4] by modifying a 2D mobile robot localization to 3D ROV localization. It is EKF-based and designed to meet four main requirements: modularity, scalability, reconfigurability, and performance. However, tests were performed for simulated effects, and verification with real data is very limited. The authors also show that their approach is sensitive to a loss or disturbed position signal, even in simulation.

[5] presented a non-EKF framework that fuses body-velocity and position measurements in a cascaded, complementary filter and tests it in a shallow environment. While this work fuses noisy measurements from these sensors, an additional outlier rejection scheme is needed to account for disturbed measurements. Further, compared to EKF-based approaches, no (un-)certainty of the state estimate can be given with this approach, and unreliable sensors, i.e., loss-of-signal, are not sufficiently handled.

In this work, we leverage a recently developed framework for unmanned aerial vehicles (UAVs), the Modular and Robust Sensor-Fusion (MaRS) [6] framework, and extend it to sensors commonly used on real underwater ROVs. These extensions within MaRS allow us to focus on real-world data with its

This work was supported in part by the BugWright2 EU H2020-Project under the Grant agreement No. 871260 and in part by the Army Research Office under Cooperative Agreement Number W911NF-21-2-0245.

Pre-print version, accepted Jul./2022 at OCEANS22,
DOI: 10.1109/OCEANS47191.2022.9977298 ©IEEE.

Fig. 1. Blueye X3 ROV used for the experiments.

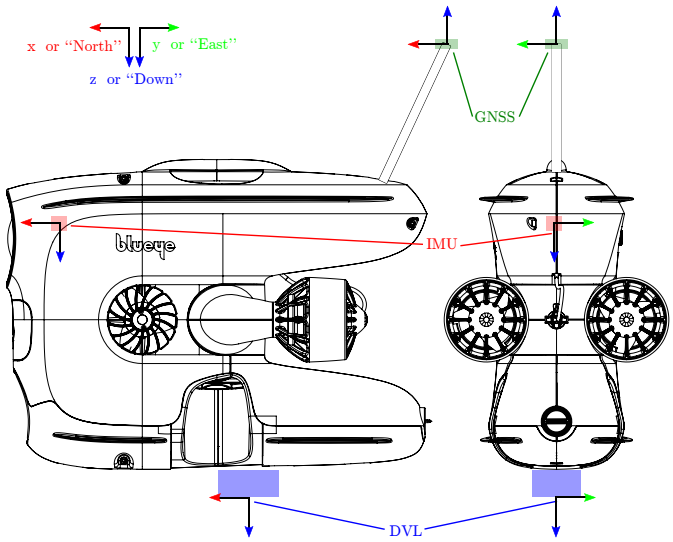


Fig. 2. Sensors and their reference frames used within MaRS for state estimation of our X3 ROV, as viewed from the left and back.

inherent polluting elements like asynchronous sensor measurements, multi-rate readings, and out-of-sequence measurements as we design our algorithm to run onboard a real ROV in a harbor-like environment for usage in ship-hull inspections. We further include methods for sensor frame initialization and online extrinsic calibration, given the information provided by the currently available sensors. Further, we include and show the capability of handling unreliable sensors and measurements during mission time and compare our method to industrial underwater position sensors (outdoor in the fjord or harbor environments) or image-based ground truth (indoor test basin).

II. Setup and Method

While we design the following framework for our Blueye X3 small ROV[1] as shown in Fig. 1, please note, that MaRS' modularity would allow it to run with any number of different sensors and setups as shown in [6]. The X3 features two sets of instruments: internal sensors and external sensors that were manually added. The drone itself contains two inertial measurement units (IMUs) with an integrated inclinometer. Both IMUs provide 3 degrees of freedom (DoF) gyroscope, accelerometer, and magnetometer readings. A pressure sensor is used to measure the depth and a forward-looking camera so that the ROV is vision-enabled. It has a manually fixed tilt and is located behind a glass dome. As for the external sensor payload, a Doppler velocity log (DVL) sensor is mounted and provides 3 DoF body velocity and the distance to the ground. The sensor also features its own IMU with an inclinometer. The manufacturer provides a processed 6 DoF pose relative to a start position through a proprietary method. A standard global navigation satellite system (GNSS) receiver mounted on a stick is also employed. This allows a fix positioning correction every time the ROV goes back to the surface and also to be aligned with a global GNSS frame. Finally, the underwater drone is also equipped with a multi-beam forward-looking sonar which is not used in this work.

[1]Blueye: https://www.blueyerobotics.com

A. State-Estimation

The presented state-estimation module extends our previous work, the Modular and Robust Sensor-Fusion (MaRS) [6] framework, to account for the aforementioned underwater sensors, namely the DVL and pressure sensor. MaRS is an error-state EKF that provides modularity for adding and removing sensors at run-time while being computationally highly efficient.

Our goal is to estimate the vehicle's IMU position ${}_N\mathbf{r}_{NI}$, orientation $\mathbf{q}_{NI}$, and velocity ${}_N\mathbf{v}_{NI}$ within our gravitationally-aligned navigation frame N. Additionally, we estimate the biases of the EKF's main propagation sensor, the IMU, represented by ${}_I\mathbf{b}_\mathbf{a}$ and ${}_I\mathbf{b}_\omega$ for acceleration and angular velocity, respectively. Thus the vehicle's core state can be summarized to

$$\mathbf{x}_{core} = \begin{bmatrix} {}_N\mathbf{r}_{NI}^\mathsf{T}, & \mathbf{q}_{NI}^\mathsf{T}, & {}_N\mathbf{v}_{NI}^\mathsf{T}, & {}_I\mathbf{b}_\mathbf{a}^\mathsf{T}, & {}_I\mathbf{b}_\omega^\mathsf{T} \end{bmatrix}^\mathsf{T} \in \mathbb{R}^{16\times 1}. \quad (1)$$

In general, the core state and covariance can be propagated using the Newton-Euler dynamics model for a vehicle in 3D (2-6) and the IMU measurements. $\Omega(\boldsymbol{\omega})$ is the right side quaternion multiplication matrix (c.f. [7]), ${}_I\breve{\boldsymbol{\omega}}_{NI}$ and ${}_I\breve{\mathbf{a}}_{NI}$ the measured angular velocity and linear acceleration, respectively. ${}_N\mathbf{g}$ is the gravity vector expressed in the navigation frame, which due to the definition of the navigation frame is generally ${}_N\mathbf{g} = \begin{bmatrix} 0, & 0, & -9.81 \end{bmatrix}^\mathsf{T} \mathrm{m/s^2}$.

$$ {}_N\dot{\mathbf{r}}_{NI} = {}_N\mathbf{v}_{NI} \quad (2)$$

$$ \dot{\mathbf{q}}_{NI} = \frac{1}{2}\Omega({}_I\breve{\boldsymbol{\omega}}_{NI} - {}_I\mathbf{b}_\omega - {}_I\mathbf{n}_\omega)\mathbf{q}_{NI} \quad (3)$$

$$ {}_N\dot{\mathbf{v}}_{NI} = \mathbf{R}_{NI}\left({}_I\breve{\mathbf{a}}_{NI} - {}_I\mathbf{b}_\mathbf{a} - {}_I\mathbf{n}_\mathbf{a}\right) + {}_N\mathbf{g} \quad (4)$$

$$ {}_I\dot{\mathbf{b}}_\mathbf{a} = {}_I\mathbf{n}_{\mathbf{b}_\mathbf{a}} \quad (5)$$

$$ {}_I\dot{\mathbf{b}}_\omega = {}_I\mathbf{n}_{\mathbf{b}_\omega} \quad (6)$$

Further, upon receiving a sensor measurement, the core states are corrected using the following sensor measurement

equations. ${}_I\mathbf{r}_{ID}$, ${}_I\mathbf{r}_{IP}$ and ${}_I\mathbf{r}_{IG}$ are calibration states that describe the translation between the vehicle's IMU and the sensors DVL, pressure, and position, respectively. Similarly $\mathbf{R}_{ID}$ and $\mathbf{R}_{IA}$ describe the orientation offsets between the IMU and DVL or attitude sensor.

$$\mathbf{z}_{\text{DVL}} = \mathbf{R}_{ID}^{\mathsf{T}}\mathbf{R}_{NI\,N}^{\mathsf{T}}\mathbf{v}_{NI} + \mathbf{R}_{ID}^{\mathsf{T}}\lfloor {}_I\breve{\boldsymbol{\omega}}_{NI} - {}_I\mathbf{b}_{\omega}\rfloor_{\times}\,{}_I\mathbf{r}_{ID} \tag{7}$$

$$\mathbf{z}_{\text{Pressure}} = \|{}_N\mathbf{g}\| \cdot \rho_{\text{water}} \cdot \begin{bmatrix}0,0,1\end{bmatrix}\left({}_N\mathbf{r}_{NI} + \mathbf{R}_{NI}\,{}_I\mathbf{r}_{IP}\right) \tag{8}$$

$$\mathbf{z}_{\text{Pos}} = {}_{G_0}\mathbf{r}_{G_0N} + \mathbf{R}_{G_0N}\left({}_N\mathbf{r}_{NI} + \mathbf{R}_{NI}\,{}_I\mathbf{r}_{IG}\right) \tag{9}$$

$$\mathbf{z}_{\text{Att}} = \mathbf{R}_{A_0N}\mathbf{R}_{NI}\mathbf{R}_{IA} \tag{10}$$

$\lfloor\bullet\rfloor_{\times}$ is the skew-symmetric matrix (as defined in [7]) and represents the cross-product, $\|{}_N\mathbf{g}\| = 9.81\,\mathrm{m/s^2}$ is the norm of the gravitational acceleration, and $\rho_{\text{water}} = 997\,\mathrm{kg/m^3}$ is the density of water. Additionally, the reference frame of the position sensor (G_0) and attitude sensor (A_0) might be different than the navigation frame (N) used. Therefore, the calibration states ${}_{G_0}\mathbf{r}_{G_0N}$, $\mathbf{R}_{G_0N}$, and $\mathbf{R}_{A_0N}$ are added to account for a transformation between these frames (e.g., east-aligned East-North-Up (ENU) GNSS frame and north-aligned North-East-Down (NED) navigation frame).

B. State-Initialization

Initially, all initial extrinsic calibration states, i.e., ${}_I\mathbf{r}_{ID}$, ${}_I\mathbf{r}_{IP}$, ${}_I\mathbf{r}_{IG}$, $\mathbf{R}_{ID}$, and $\mathbf{R}_{IA}$, are calibrated by measuring the individual sensor translation and rotation offset manually or through software tools [8], [9].

We then assume the ROV to be static on startup such that we can initialize the navigation frame N gravitationally aligned using the direction of gravity from the past n accelerometer measurements. Further, as an initial guess, we initialize the yaw using the magnetometer. It should be noted that due to metallic structures in harbors, this initial guess might be significantly different from our global yaw. However, if another global sensor such as GNSS is used, the state $\mathbf{R}_{G_0N}$ is updated to correct for the yaw-misalignment. In an ideal scenario $\mathbf{R}_{G_0N} = \mathbf{I}_3$, otherwise magnetic interference detection methods can be explored [10]. The initial attitude can then be used to provide an initial guess for $\mathbf{R}_{A_0N}$ upon receiving the first attitude measurement $\mathbf{z}_{\text{Att}}$.

Furthermore, the navigation frame's origin is set to the startup position, i.e., ${}_N\mathbf{r}_{NI} = 0$, and the initial state covariance is associated. Due to the static assumption on startup, also the velocity is set to $\mathbf{0}$. The biases are initialized using the mean values provided by the manufacturer (given in Tab. I). Finally, upon receiving the first position measurement, we can then derive the position sensor reference frame ${}_{G_0}\mathbf{r}_{G_0N}$.

C. Measurement Rejection

In the real world, sensor measurements are rarely perfect and highly accurate but rather noisy due to environmental impacts. Reflections and distortions caused by structures tend to influence the accuracy, especially of the DVL and GNSS sensors. The GNSS sensor, for example, is influenced heavily by the water, as we will show below. As the low-power signal cannot penetrate water (easily) and is reflected by it, the GNSS measurements are noisy if the receiver is close or just covered by water. Also, as can be expected, upon diving, the signal can be lost completely. Similarly, the DVL sensor expects a flat seafloor, and structures present (e.g., ship hulls and harbor walls) can reflect the signal and thus distort the measurement.

Therefore, there is a need to detect and reject faulty and noisy measurements properly. In our framework we achieve this by performing a likelyhood test on every measurement received, in this case a χ^2-test using the prior sensor states and covariance.

χ^2-tests are a useful addition to EKFs but they have a drawback: If a sensor measurement is not received and the core state not corrected for some time, e.g., GNSS while performing a dive, the core state might drift from the true value. In that case, the χ^2-test can trigger for every new measurement, rejecting it and thus not correcting the core states. To circumvent this, our framework performs a complete sensor re-initialization (c.f. Sec. II-B) if ten consecutive χ^2-tests have failed, taking advantage of MaRS' ability to add and remove sensors dynamically.

Fig. 3. Test site for the experiments: outdoor pool located at the Trondheim NTNU's Biological Station.

III. Experiments and Results

We perform various experiments in our main testing site, which is located at the Trondheim NTNU's Biological Station in Norway[2] in a 7x10 meter outdoor pool as shown in Fig. 3. We chose this pool as it emulates the space available at a harbor when inspecting ship hulls[3] well. The platform presented in Sec. II equipped with a WaterLinked A50 DVL sensor, a water-proof GNSS sensor, a TE connectivity MS5837-30BA pressure sensor, and an InvenSense MPU-9250 IMU is used. Additionally, the attitude provided by the proprietary complementary filter of the DVL driver is added as a measurement in MaRS. The sensor measurement rates and their measurement uncertainty, as provided by the manufacturer, are given in

[2]TBS: https://www.ntnu.edu/biology/research/tbs

[3]See project BugWright2: https://www.bugwright2.eu

TABLE I
SENSOR RATES AND MEASUREMENT UNCERTAINTY

Sensor	Observed Rate		Measurement Uncertainty	
accelerometer	100.0	Hz	0.041 62	m/s^2
gyroscope	100.0	Hz	0.002 46	rad/s
DVL	5.5	Hz	0.04	m/s
pressure	43.0	Hz	7.5	kPa
GNSS	1.0	Hz	0.5	m
attitude	4.5	Hz	2.0	°

Tab. I. Further, MaRS and the sensor drivers run onboard the platform on a Raspberry-Pi3-equivalent board.

a) Rectangular Motion: In our first experiment, we perform several low-speed rectangular motions in the pool and compare our position estimates with the measurements and positions obtained from the GNSS sensor and DVL software (c.f. Fig. 4). The position measurements of the DVL are aligned with the MaRS estimate using the first pair of poses. Please note that the GNSS measurements are not required to be aligned due to our initialization method using the magnetometer (c.f. Sec. II-B).

At low speed, the low-rate GNSS measurements are a good reference to compare with our state estimates. As can be clearly seen in Fig. 4a, our state estimation is able to follow the low-rate GNSS closely for the North and East coordinates. In comparison, the DVL positioning software accumulates a North-drift within the first motion (seconds 0-10), which it can never correct again afterward.

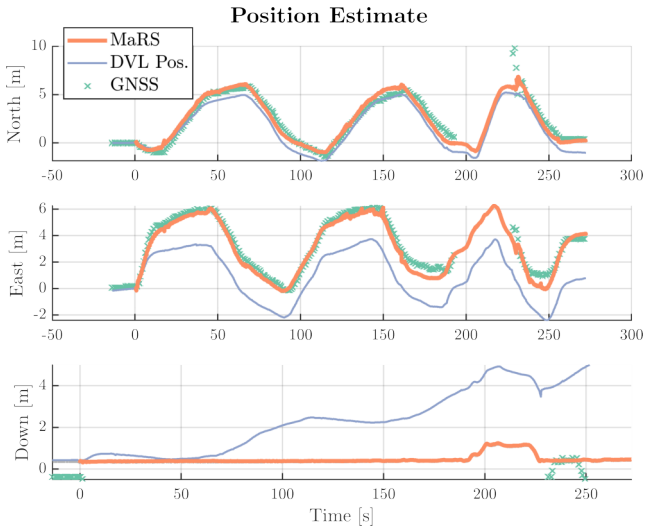
An additional advantage of our sensor fusion is the improved depth estimation. Compared to the DVL and GNSS sensors whose signals are disturbed, MaRS is able to provide a good depth estimate by relying mostly on the pressure sensor as shown in detail in Fig. 4b. Minor differences ($< 10\,\mathrm{cm}$) are present due to the offset and online calibration estimation between the IMU and pressure sensor, i.e. state ${}_I\mathbf{r}_{IP}$.

Furthermore, while we do stay close to the surface for most of the trajectory to acquire good GNSS measurements, we also perform a dive between seconds 190 and 225. Within this time, no GNSS measurements are received. Yet, MaRS is able to continue to estimate the position correctly fusing the other sensor measurements. Because they are inaccurate, upon regaining GNSS, the first couple of measurements are rejected through the χ^2-test. After successfully passing the test at approximately second 230, the position is corrected accordingly. This shows that our framework is able to handle measurement outages and reuse them upon regaining reliable measurements.

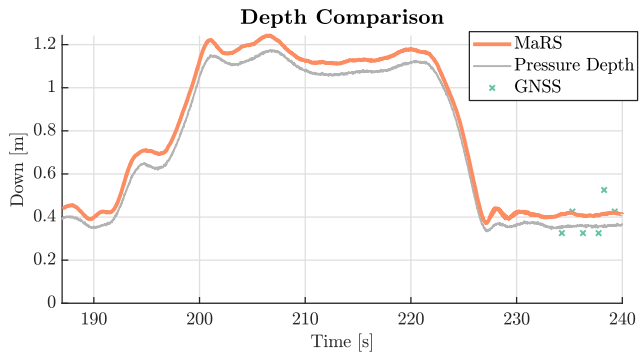
b) Pattern Motion: In another test, we pilot a "MaRS" pattern with fast motions. As a result, the low-rate GNSS measurements tend to be less accurate and are sparser distributed. Fig. 5 shows the 2D position comparison between MaRS, GNSS, and the DVL position estimate. Again the DVL estimate is aligned with the MaRS estimate using the first pair of corresponding poses within the sequence.

Similar to the previous experiment, our estimation framework MaRS is able to follow the shape of the pattern. However, due to the faster motion with this trajectory, the GNSS is less accurate and the distance between two consecutive measurements increased. Compared to the previous experiment where most GNSS measurements are used, in this experiment, these less-accurate GNSS measurements can result in failed χ^2-tests, prohibiting the state update. Therefore, a position difference between our estimates and the raw measurements (noticeably for the "M" and "R" of the pattern). Further, as seen previously, the DVL position estimate accumulates a minor yaw and position drift at the start of the trajectory, which it can never correct for, thus showing the need for fusing multiple sensor measurements.

MaRS' trajectory is provided at the combined rate of all sensors, i.e., at approximately 150 Hz (c.f. Tab. I). While the higher rate compared to the positional output from either GNSS or DVL position is useful for control, this implies that also the intermediate propagation estimates between two



(a) North-East-Down comparison for position estimates



(b) Depth comparison when diving

Fig. 4. Aligned comparison between MaRS (orange), GNSS measurements (green), and the position obtained through the proprietary method of the DVL software (blue) for a rectangular trajectory with slow motion. MaRS is able to follow the GNSS measurements for the North and East axis correctly, while the DVL estimator accumulates a minor drift in the beginning (a). Neither the GNSS nor the DVL provide good depth estimates. Thus MaRS follows the measurements of the more accurate pressure sensor as shown when a dive is performed between seconds 190 - 225 (b).

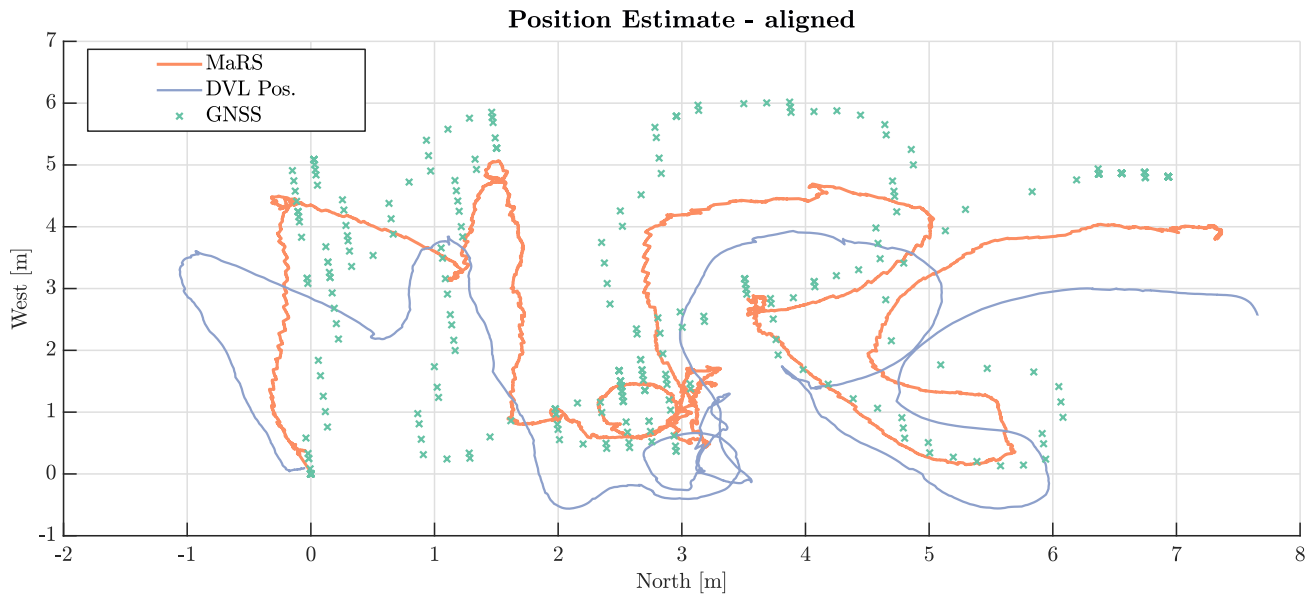


Fig. 5. Estimation comparison between MaRS, GNSS, and the DVL-provided positions when swimming a fast-motion "MaRS" pattern. The DVL positions are aligned using the first pair of corresponding poses while the GNSS measurements are presented "raw". GNSS is less accurate in this experiment due to the faster motions, which can result in the receiver being closer to (or under) the water. MaRS is able to detect some of these faulty measurements through its χ^2-tests and rejects them. In comparison, the DVL-provided position accumulates some drift in the beginning, which cannot be corrected for.

consecutive and corrective position updates are provided. Therefore, in faster motions MaRS' trajectory tends to be more jittery as seen in Fig. 5. If desired in ROV or autonomous applications, the estimates' publication rate could be lowered, and hence the trajectory smoothed.

IV. Conclusion

With this work, we improved and presented a modular state estimation framework for ROVs using common sensors such as a DVL, an IMU, a pressure sensor, and a GNSS sensor (when above the water surface). We compared our estimation technique with the estimates provided by the integrated DVL sensor and saw that through our modular estimation MaRS is able to handle sensor outages as well as measurement outliers due to environmental disturbances.

Overall we presented and extended a state-of-the-art estimation framework usable for reliable ROV navigation within limited environments such as harbors by taking advantage of multiple different sensors.

Acknowledgement

The authors would like to thank Borja Serra and Andre Marquardt from Blueye Robotics for their continuous technical support within this project.

References

[1] E. I. Grøtli, J. Tjønnås, J. Azpiazu, A. A. Transeth, and M. Ludvigsen, "Towards more autonomous rov operations: Scalable and modular localization with experiment data," *IFAC-PapersOnLine*, vol. 49, no. 23, pp. 173–180, 2016, 10th IFAC Conference on Control Applications in Marine SystemsCAMS 2016.

[2] Y. Xincun, O. Yongzhong, S. Fuping, and F. Hui, "Kalman filter applied in underwater integrated navigation system," *Geodesy and Geodynamics*, vol. 4, no. 1, pp. 46–50, 2013.

[3] A. Gómez-Espinosa, E. Cuan-Urquizo, and J. González-García, "Autonomous underwater vehicles: Localization, navigation, and communication for collaborative missions," *Applied Sciences*, vol. 10, p. 1256, 02 2020.

[4] J. Azpiazu, M. Bjerkeng, and E. I. Grøtli, "Vind: A robot self-localization framework," in *Proceedings of the 4th International Conference on Control, Mechatronics and Automation*, ser. ICCMA '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 1–6.

[5] A. Vasilijevic, B. Borovic, and Z. Vukic, "Underwater Vehicle Localization with Complementary Filter: Performance Analysis in the Shallow Water Environment," *Journal of Intelligent and Robotic Systems: Theory and Applications*, vol. 68, no. 3-4, pp. 373–386, 2012.

[6] C. Brommer, R. Jung, J. Steinbrener, and S. Weiss, "MaRS: A Modular and Robust Sensor-Fusion Framework," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 359–366, 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9286578/

[7] J. Solà, "Quaternion kinematics for the error-state Kalman filter," *arXiv preprint arXiv:1711.02508*, p. 95, 11 2017. [Online]. Available: http://www.billion.uk.com/downloads/usermanual/annex_m.pdfhttp://arxiv.org/abs/1711.02508

[8] P. Furgale, J. Rehder, and R. Siegwart, "Unified Temporal and Spatial Calibration for Multi-Sensor Systems," *IEEE International Conference on Intelligent Robots and Systems (IROS)*, pp. 1280–1286, 2013.

[9] J. Rehder, J. Nikolic, T. Schneider, T. Hinzmann, and R. Siegwart, "Extending kalibr: Calibrating the extrinsics of multiple IMUs and of individual axes," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, vol. 2016-June. IEEE, 5 2016, pp. 4304–4311. [Online]. Available: http://ieeexplore.ieee.org/document/7487628/

[10] C. Brommer, C. Böhm, J. Steinbrener, R. Brockers, and S. Weiss, "Improved State Estimation in Distorted Magnetic Fields," in *2020 International Conference on Unmanned Aircraft Systems (ICUAS)*. Athens, Greece: IEEE, 9 2020, pp. 1007–1013. [Online]. Available: https://ieeexplore.ieee.org/document/9213913/