Systemy Logistyczne Wojsk
Zeszyt 63(2025)
ISSN 1508-5430, s. 151-164
DOI: 10.37055/slw/218680

Instytut Logistyki
Wydział Bezpieczeństwa, Logistyki i Zarządzania
Wojskowa Akademia Techniczna
w Warszawie

Military Logistics Systems
Volume 63(2025)
ISSN 1508-5430, pp. 151-164
DOI: 10.37055/slw/218680

Institute of Logistics
Faculty of Security, Logistics and Management
Military University of Technology
in Warsaw

Modelling Military Equipment Losses with Open-Source Visual Intelligence: Evidence from the War in Ukraine

Olimpia Wiktoria Sobczyk
olimpia.sobczyk@wat.edu.pl; ORCID: 0009-0001-4206-5377
Military University of Technology, Faculty of Security, Logistics and Management, Poland

Abstract. The research niche of this article is the use of open-source visual intelligence and automated computer vision, combined with classical time-series modelling, to analyse and forecast military equipment losses in the Russo-Ukrainian war. The purpose of the research was to test whether visually confirmed data (the Oryx repository) can be algorithmically transformed into a reliable weekly series and used for short-term forecasting. Two hypotheses were examined: (H1) open visual data yield a series with stable trend and seasonality; (H2) ARIMAX with Fourier terms outperforms seasonal ARIMA at short horizons. The methodology comprised a Python pipeline in which YOLOv8 localized date stamps and EasyOCR read them; STL decomposition characterized the trend and seasonal structure. Forecasting employed SARIMA and ARIMAX with sine/cosine pairs for the annual period. Results confirm pronounced annual seasonality (peaking in March–April) and a trend cresting in early spring 2023; relative to SARIMA, ARIMAX reduced errors by 14.6–23.7% (MAE) and 23.8–25.5% (RMSE) in both in-sample fit and rolling validation. The conclusions indicate that, despite limitations (a lower bound on true losses), public visual data provide a robust, verifiable analytical basis with strong predictive potential; future work should incorporate machine learning, exogenous covariates, and probabilistic forecasting.
Keywords: data mining; military equipment losses; war in Ukraine; neural networks; forecasting

Introduction

This article evaluates the analytical potential of openly available data on military equipment losses in armed conflicts. Recent literature highlights the growing role of OSINT communities and open-source investigation in documenting the Russo-Ukrainian war (Charlton et al., 2024; Limonier, 2022; Kotisova, Velden, 2023). The central research question asks whether such data can be effectively used to model and forecast future equipment losses. Accordingly, three detailed questions were

formulated: (1) Is it possible to derive a time series of equipment losses from visual data originating in an armed conflict? (2) What statistical properties might such a time series exhibit? (3) Are classical statistical algorithms adequate as predictive tools for loss analysis? Two hypotheses are examined: (H1) open visual data yield a series with stable trend and seasonality; (H2) ARIMAX with Fourier terms outperforms seasonal ARIMA at short horizons. Questions (1)– (2) operationalise H1, while question (3) operationalises H2. The research employed analysis and synthesis, comparative methods, abstraction, inference and generalization, as well as mathematical modelling and simulation. The article is structured as follows. Section 1 presents the data-processing pipeline for constructing the time series. Section 2 visualizes the resulting series using the example of Russian BMP-1 losses. Section 3 details the predictive-model design based on statistical algorithms, evaluates model quality, and indicates directions for further development.

Deriving a time series from Oryx visual materials

This study restricts itself to visual materials owing to the wealth of content available on the Oryx analytical portal (kaggle and Oryx datasets). The author contends that these visuals constitute an objective information source that can be converted into a time series using automated content-recognition methods. An example image and a possible date-annotation scheme are presented in Fig. 1

Fig. 1. A photograph documenting the loss of an infantry fighting vehicle posted on TikTok
Source: TikTok

A baseline attempt to detect dates using stand-alone optical character recognition (OCR) within the infantry fighting vehicles (IFV) category proved insufficient. Consequently, a processing pipeline was designed in which a convolutional neural network (CNN) first detects the likely location of a date in an image (McCord, Illingworth, 1994; Agrawal, Srikant, 1995; Bigus, 1996; Burges, 1998; Morzy, 2013). OCR is then applied only to the detected region, and the extracted dates are exported to a spreadsheet in a normalized format. For training-data preparation, images depicting Russian tanks were used. Manual annotation was performed with MakeSense.ai by adding rectangular bounding boxes around visible dates and labelling each box as "date". Annotations were exported in the widely used YOLO format (class, x_center, y_center, width, height; all normalized). The dataset structure comprised an /images/ directory with graphics files and a /labels/ directory with the corresponding YOLO text files. A data.yaml configuration specified paths, the number of classes (nc: 1), and class names (names: [date]), ensuring compatibility with YOLO-based object-detection models. The learning curves and validation metrics (YOLOv8) are shown in Fig. 2.

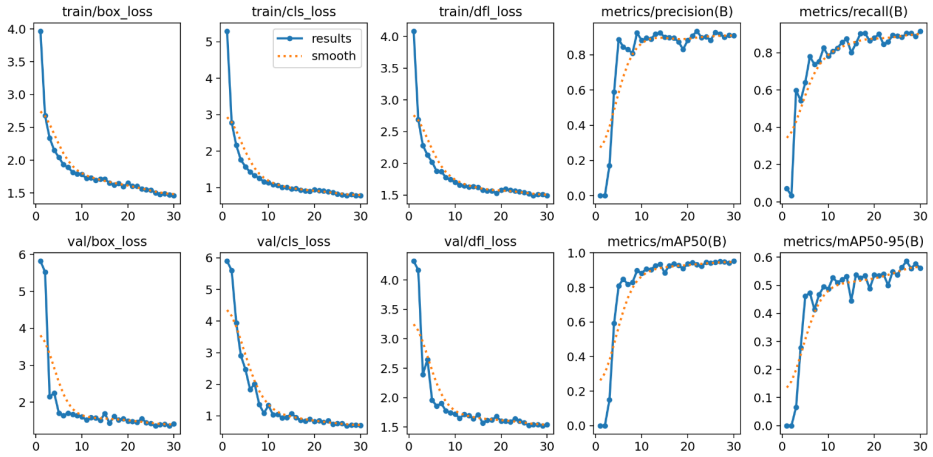

Fig. 2. Learning curves and validation metrics (YOLOv8)
Source: Own study

Training curves show steady decreases in box loss, classification loss, and Distribution Focal Loss (DFL), indicating effective learning of both localization and class assignment. Validation losses mirror training behaviour, suggesting good generalization with no material overfitting. Precision rises rapidly to ~0.90 and recall to ~0.85; mAP@0.5 approaches 0.90, while mAP@0.5:0.95 stabilizes around 0.55–0.60, reflecting stricter localization under higher Intersection over Union (IoU) thresholds. Confusion-matrix analysis on the validation set yielded TP = 222, FN = 17, FP = 37, TN = 0; hence recall ≈ 0.93, precision ≈ 0.86, specificity = 0, and accuracy ≈ 0.80.

The F1–confidence curve (Fig. 3) peaks at F1 = 0.90 for a confidence threshold of 0.373, indicating a balanced operating point between precision and recall.

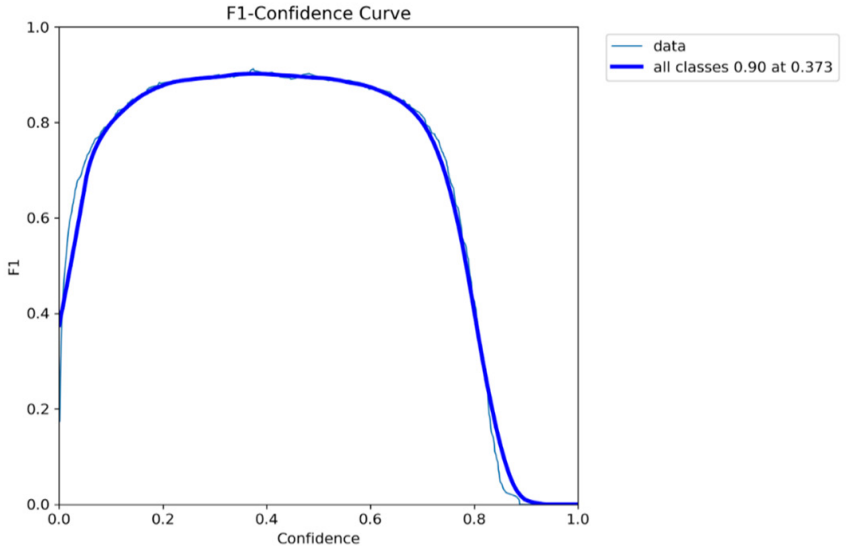

Fig. 3. The F1–Confidence curve
Source: Own study

The precision–confidence curve (Fig. 4) attains 1.00 at a confidence threshold of approximately 0.877, reflecting near-perfect precision but a concomitant increase in missed detections (reduced recall).

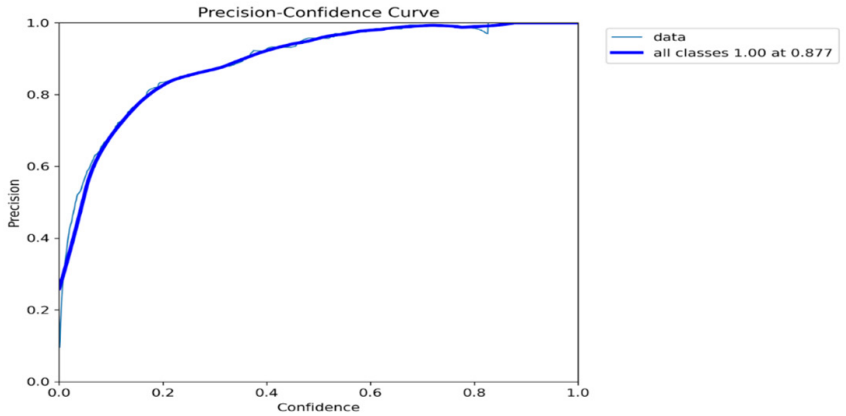

Fig. 4. The Precision–Confidence curve
Source: Own study

The precision–recall curve (Fig. 5) reports mAP@0.5 = 0.952 for the date class, indicating high detection quality at the 0.5 IoU thresholds.

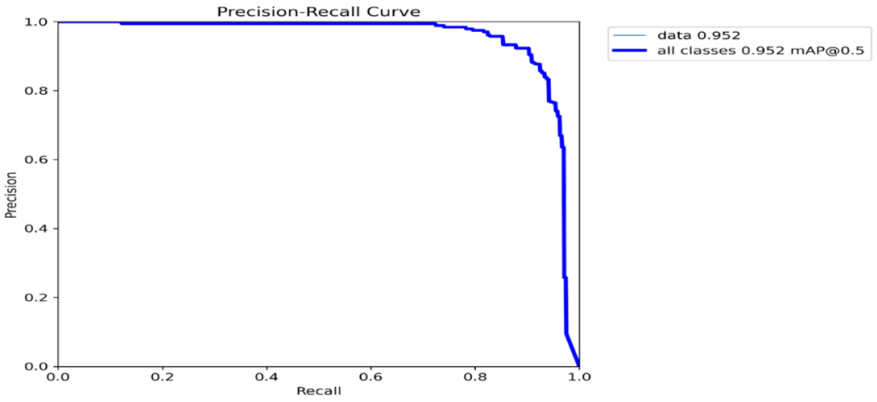

Fig. 5. The Precision–Recall curve
Source: Own study

The recall–confidence curve (Fig. 6) exhibits the expected decrease in recall as the confidence threshold increases, consistent with more conservative decision thresholds.

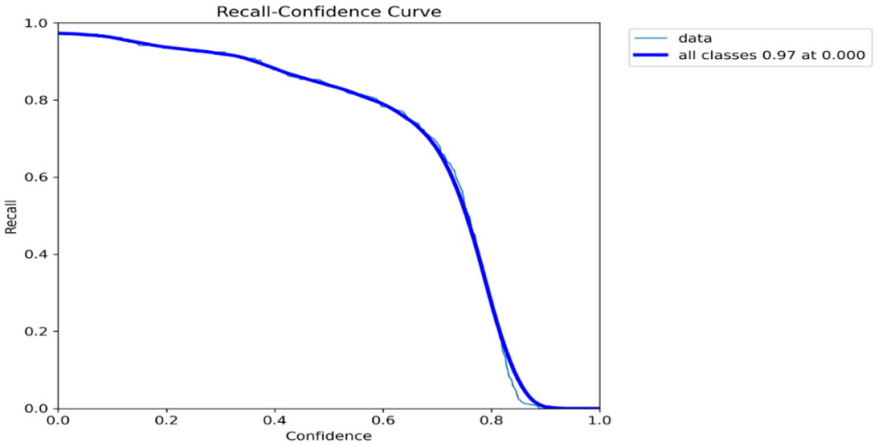

Fig. 6. The Recall–Confidence curve
Source: Own study

Exploratory analysis of bounding-box centres and dimensions (x, y, width, height)—Fig. 7—suggests a mild concentration of date stamps near the left/right image edges and, vertically, predominantly toward the bottom. Widths and heights are typically small, consistent with short, single-line date strings. No strong correlations were observed between position and size, suggesting balanced coverage in the training set and limited positional bias. Increasing the input resolution (e.g., from 640×640 to 960×960 or 1280×1280) could improve detection of small, thin date glyphs and enhance OCR performance on faint, low-contrast inscriptions.

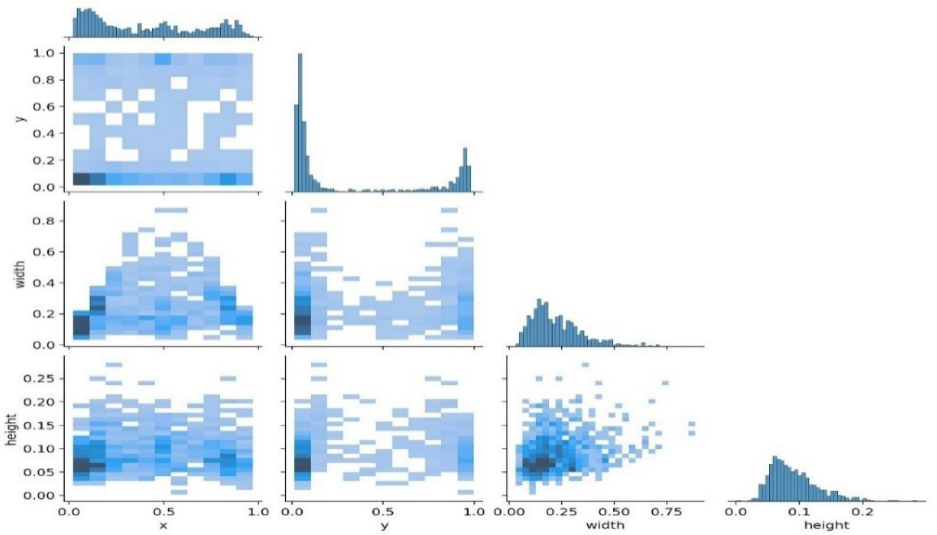

Fig. 7. Spatial statistics of annotations
Source: Own study

The complete Python pipeline proceeds as follows: (i) parse dates from filenames when available; (ii) apply the YOLOv8 detector to localize candidate date regions (Ultralytics, 2023); (iii) run OCR exclusively on the cropped regions; and (iv) normalize and export detected dates to a text or tabular (e.g., CSV) format. To better handle light-coloured glyphs (e.g., white), EasyOCR was employed and proved effective (JaidedAI, 2020). Out of 3,577 images in the infantry-vehicles subset, specific dates were automatically extracted from 3,384 images, with 193 failures due to absent or unreadable dates. Thus, ~94.6% of the dataset was successfully processed for subsequent time-series construction and forecasting.

Exploratory Visualization of Russian Infantry Fighting Vehicles

The time series was decomposed using STL (Seasonal–Trend decomposition using Loess). STL separates a series into three components: a trend (slow, long-run change in level), a seasonal pattern (systematic variation repeating at a fixed period), and a remainder (variation not explained by trend or season, i.e., noise/short-term fluctuations). Because STL employs local smoothing (LOESS), it is flexible (seasonality need not be perfectly sinusoidal), handles time-varying trends well, and offers a robust variant that is resistant to spikes/outliers. The decomposition can be specified as additive ( ) or as "near-multiplicative" after a monotonic transformation (e.g., log1p), thereby avoiding some limitations of classical approaches (Fayyad, 1996; Friedman, 1997, Nielsen 2020). For the present analysis, the seasonal period was set to and the robust option was enabled. The data consist of weekly counts with many zeros and occasional spikes. Under these conditions, an additive specification with robust = True provides outlier resistance without requiring a log transformation (which would be problematic in the presence of zeros). The seasonal period reflects an annual cycle in weekly resolution, consistent with operational rhythm and the calendar. The resulting decomposition is shown below Fig. 8.

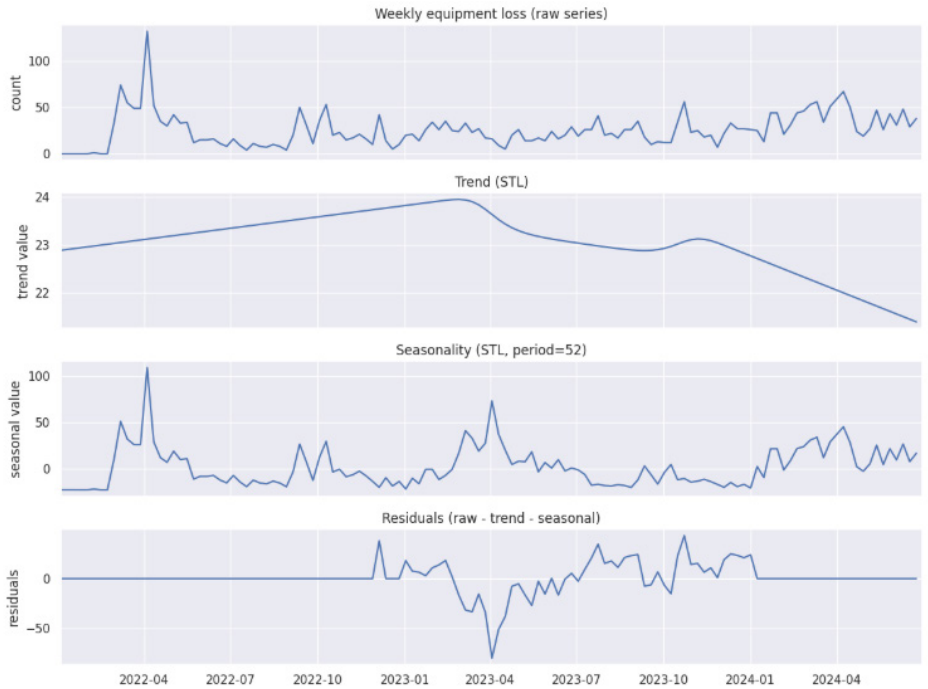

Fig. 8. STL decomposition of the weekly series
Source: Own study

Raw series layer. High, short-lived spikes are visible at the beginning of 2022, followed by moderate fluctuations thereafter. After 2023, the amplitude of local changes increases, though without the extreme surges observed at the start of the sample.

Trend (STL). The trend rises from the beginning of the sample to around Q1 2023, then gradually declines through 2023, with a slight flattening at the turn of 2023/2024 and a renewed decrease in 2024. This indicates that the long-run level of losses in infantry vehicles peaked in early spring 2023 and subsequently trended downward.

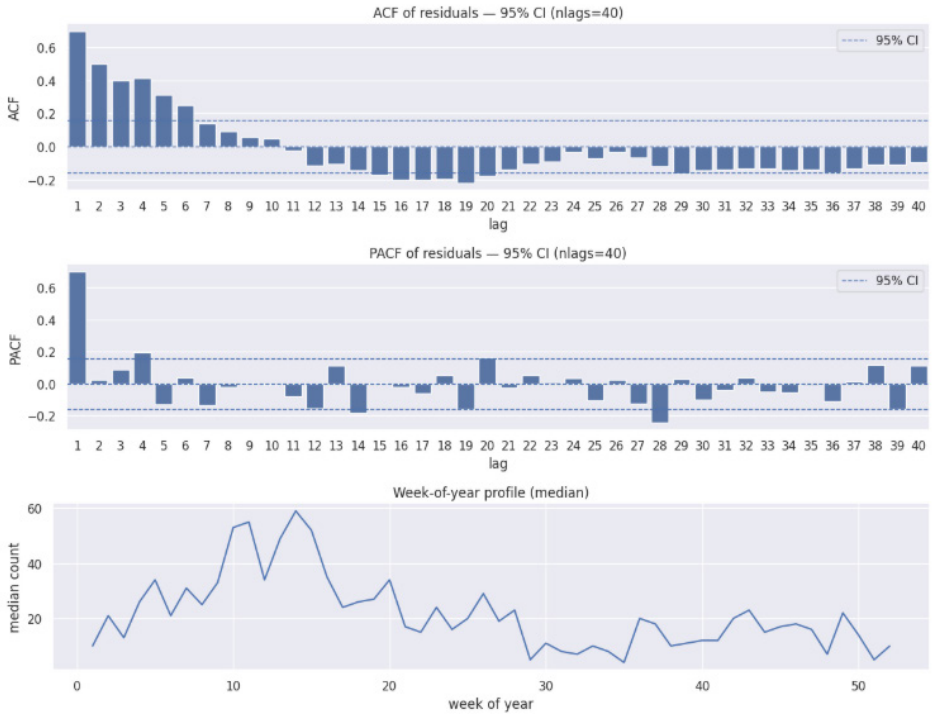

Fig. 9. Residual autocorrelation (ACF) and partial autocorrelation (PACF)
Source: Own study

Seasonality (STL, period = 52). The seasonal component exhibits recurring annual patterns: the highest values occur in March–April (weeks ~10–15), with weaker "humps" visible in late autumn/early winter (weeks ~42–50). The seasonal amplitude is moderate (it does not scale proportionally with the series level), which supports an additive rather than a multiplicative specification. The median profile of counts by week-of-year confirms seasonality: a pronounced peak in March–April (weeks ~10–15), with smaller elevations in autumn and around the year's turn. This

provides a visual confirmation of the annual operational rhythm. Remainder (raw – trend – seasonal)**.** After removing trend and seasonality, short-term fluctuations remain—among them a single large negative deviation around spring 2023 (a strong one-off anomaly) and a sequence of small oscillations in subsequent months. This suggests that short-horizon forecasting should include AR/MA components (e.g., ARIMA), because decomposition alone does not fully whiten the series. The Fig. 9 presents the residuals' autocorrelation (ACF) and partial autocorrelation (PACF).

Residual ACF. Strong positive autocorrelation at short lags (1–5), followed by gradual damping and a negative tail at intermediate lags. This is a typical pattern of persistent short-term memory. Residual PACF. A pronounced spike at lag 1 (with a weaker one around lag 3), which further points to an AR(1) component. Over the first 40 lags, no strong seasonal traces are visible, indicating that the model has handled seasonality effectively.

Short-Horizon Forecasting of Weekly Equipment Losses

To evaluate short-horizon predictive performance with annual seasonality, a seasonal ARIMA model with weekly frequency was estimated and assessed using rolling one-step-ahead validation (Nielsen, 2020). The specification includes regular differencing to remove low-frequency drift and a seasonal difference at period 52 to address the annual cycle. The Fig. 10 below contrasts the observed series, the in-sample one-step fit, and rolling forecasts over the final validation window, together with a one-week-ahead point forecast and its 95% confidence interval.

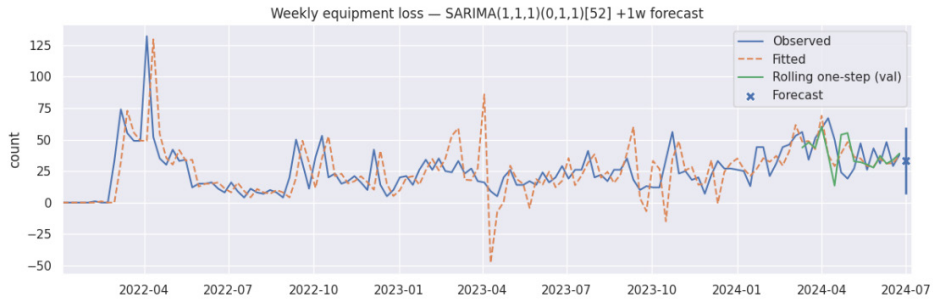

Fig. 10. Weekly equipment losses—SARIMA(1,1,1)(0,1,1)[52] with +1-week forecast (16-week rolling validation)

Source: Own study

The blue line shows weekly observations; the orange dashed line shows the in-sample one-step fit; and the green line shows rolling one-step forecasts over the last 16 weeks. The fit reproduces the level and overall shape of the series, though extreme

spikes remain partially flattened and some troughs are slightly overestimated. The validation line tracks the observations closely and does not exhibit the large overshoots associated with longer forecast windows; short oscillations around the trend are visible, which is typical of an MA component following abrupt changes. At the end of the sample, a +1-week point forecast with a 95% confidence interval is displayed; the interval width reflects short-term uncertainty. Overall, the plot confirms annual seasonality (m = 52) and a mild upward drift in 2024.

To assess whether residuals behave as approximate white noise and to identify any remaining short-term or seasonal structure not captured by the SARIMA specification, autocorrelation (ACF) and partial autocorrelation (PACF) diagnostics were examined with 95% confidence bounds (Fig. 11).

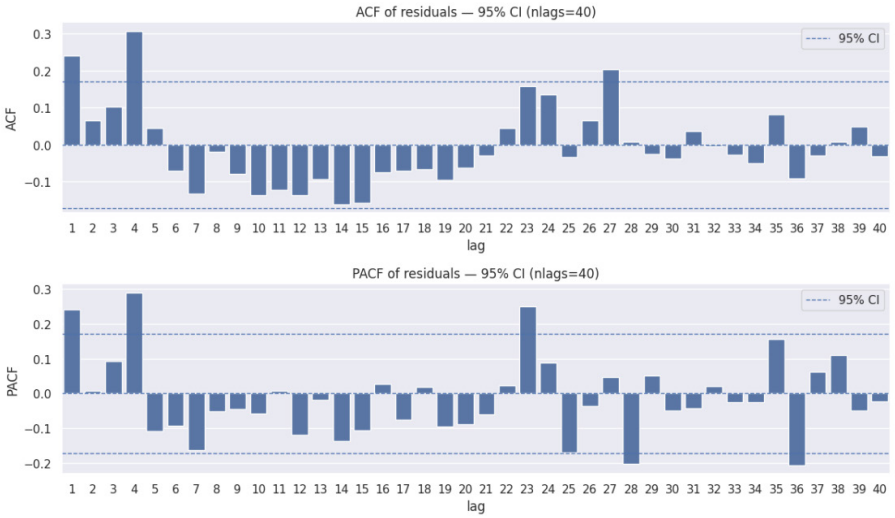

Fig. 11. Residual ACF/PACF for SARIMA(1,1,1)(0,1,1)[52]

Source: Own study

The residual ACF indicates significant autocorrelation at short lags (≈ 3–4 weeks) and around 24–26 weeks, suggesting remaining short-term structure and a half-seasonal pattern (≈ m/2). The PACF corroborates these findings (notable spikes at lags ~1 and ~3 and around 23–24), implying that the residuals are not ideal white noise—the seasonal MA(1) at 52 does not fully capture the amplitude/phase of the seasonal cycle. For this reason, although a previously tested AR(2) term did not improve information criteria (AIC/BIC) or error metrics, a ARIMAX specification with Fourier terms was subsequently employed to better represent multi-harmonic seasonality and reduce residual ACF/PACF peaks.

To benchmark a seasonal ARIMA baseline against an exogenous-terms alternative, an ARIMAX specification was estimated with Fourier harmonics for the annual cycle (,) (Nielsen, 2020). The model retains the ARIMA(1,1,1) backbone for short-term dynamics while representing seasonality via sine/cosine pairs rather than seasonal AR/MA terms. The Fig. 12 contrasts observed weekly counts, the in-sample one-step fit, rolling one-step validation over the final window, and a +1-week point forecast with its 95% confidence interval.

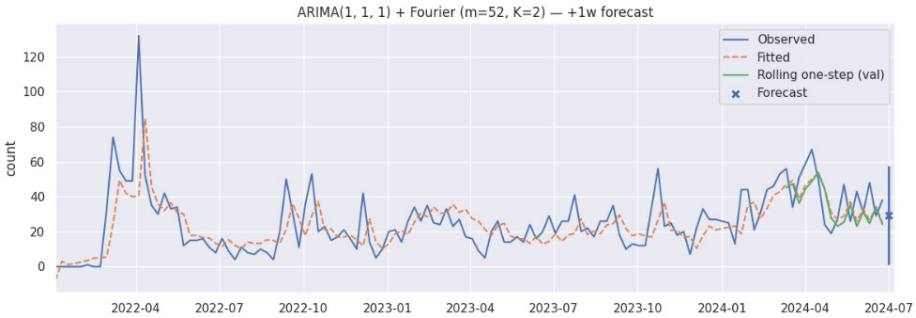

Fig. 12. ARIMA(1,1,1) + Fourier (m = 52, K = 2) — +1w forecast
Source: Own study

The fitted line closely tracks the series level and local shape, with visibly reduced over smoothing of peaks relative to a plain SARIMA and fewer overestimates at troughs. Rolling one-step validation aligns tightly with observations, indicating stable short-horizon performance; residual oscillations are low-amplitude and consistent with remaining MA-type adjustments after abrupt changes. The terminal +1-week forecast is accompanied by a 95% confidence interval whose width reflects short--term uncertainty. To verify whether the Fourier terms effectively removed seasonal structure and to inspect remaining short-lag dependence, residual autocorrelation (ACF) and partial autocorrelation (PACF) were examined with 95% confidence bounds (Fig. 13).

Residual ACF shows a pronounced short-lag spike around lag 4 and otherwise small, quickly damping correlations; no dominant multiples of 52 are present, suggesting that the annual cycle has been captured by the Fourier regressors. The PACF exhibits a clear spike at lag 4 with minor neighbours, consistent with a weak residual AR component. Overall, diagnostics indicate substantially reduced seasonal structure and only limited short-term dependence.

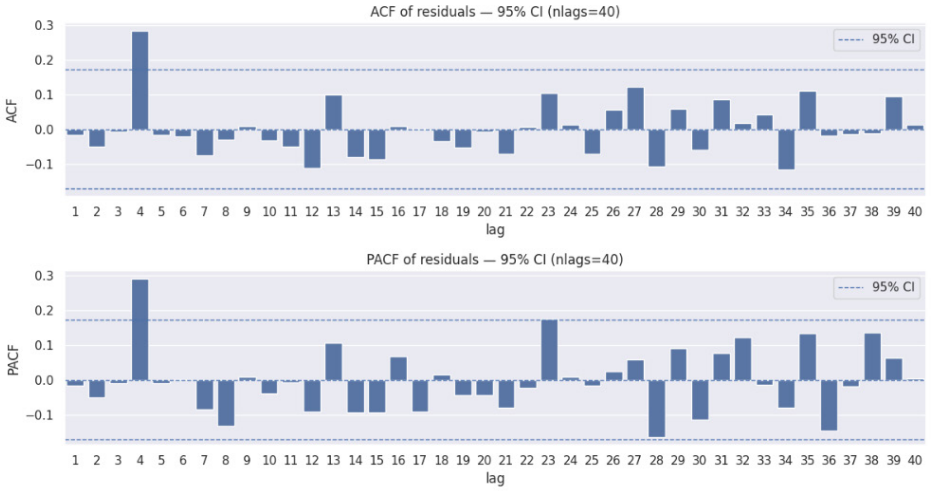

Fig. 13. ACF/PACF of ARIMAX residuals
Source: Own study

In the training sample (in-sample), the ARIMAX model achieved MAE = 9.8211 and RMSE = 14.4389, compared with MAE = 12.8667 and RMSE = 19.3704 for SARIMA, corresponding to error reductions of 23.7% (MAE) and 25.5% (RMSE). In the rolling holdout, ARIMAX obtained MAE = 12.5968 and RMSE = 14.3181, versus MAE = 14.7452 and RMSE = 18.7809 for SARIMA, i.e., reductions of 14.6% (MAE) and 23.8% (RMSE). Overall, ARIMAX delivers materially lower errors in both fit and short-horizon validation, consistent with the diagnostics indicating that Fourier terms capture the multi-harmonic seasonality more effectively than a purely seasonal ARIMA structure.

Conclusions

This article assessed the analytical potential of openly available visual data on military equipment losses in the Russo-Ukrainian war and examined the extent to which such data can support modelling and forecasting of future losses. The point of departure was the Oryx dataset, which records only visually confirmed cases (photographs or video). While the credibility of each recorded loss is high, the scope of the dataset should be regarded as a lower bound on actual losses, given the exclusion of events lacking visual documentation and possible asymmetries in media availability across the belligerents. The analysis indicates that visual evidence can provide a solid basis for estimating real losses. Because every event is strictly validated, the Oryx dataset is suitable for studying the structure and scale of losses;

however, a more comprehensive picture would benefit from automated extraction of image metadata, integration of operational reports, and a simplified, harmonized classification scheme.

The predictive potential of SARIMA and ARIMAX models was verified. For further model development, the deployment of gradient-boosted decision-tree algorithms and selected neural-network approaches is recommended, preferably on a larger, heterogeneous dataset that spans multiple time series for different equipment categories. In sum, despite inherent limitations, publicly available visual data can be effectively leveraged for modelling and forecasting military equipment losses, while methodological extensions and data integration promise additional gains in accuracy and robustness.

BIBLIOGRAPHY

[1] Agrawal, R., Srikant, R., 1995. Mining sequential patterns, Proceedings of the 11th International Conference on Data Engineering, Taipei, Taiwan.

[2] Bigus, J.P., 1996. Data mining with neural networks, McGraw-Hill.

[3] Burges, C.J.C., 1998. A tutorial on support vector machines for pattern recognition, Data Mining and Knowledge Discovery, 2(2).

[4] Charlton, T., Mayer, A.T., Ohme, J., 2024. A common effort: new division of labor between journalism and OSINT communities on digital platforms, The International Journal of Press/Politics, OnlineFirst.

[5] Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., 1996. From data mining to knowledge discovery in databases, AI Magazine, 17(3).

[6] Friedman, J.H., 1997. Data mining and statistics: what's the connection? Proceedings of the 29th Symposium on the Interface: Computing Science & Statistics, Houston, Texas.

[7] Kotisova, J., Velden, I., 2023. The affective epistemology of digital journalism: emotions as knowledge among on-the-ground and OSINT media practitioners covering the Russo-Ukrainian war, Digital Journalism.

[8] Limonier, K., 2022. The war in Ukraine, open-source investigation and the potential for "digital fieldwork" in geopolitics, Political Geography, 99.

[9] McCord, N.M., Illingworth, W.T., 1994. Practical guide to neural nets, Addison-Wesley.

[10] Mincewicz, W., 2023. Internetowy OSINT jako źródło wiedzy o wojnie w Ukrainie, Zeszyty Naukowe Pro Publico Bono, 1(1).

[11] Morzy, M., 2013. Eksploracja danych – przegląd dostępnych metod i dziedzin zastosowań, VI edycja: Hurtownie Danych i Business Intelligence, Centrum Promocji Informatyki, Warszawa.

[12] Nielsen, A, 2020. Szeregi czasowe. Praktyczna analiza i predykcja z wykorzystaniem statystyki i uczenia maszynowego, Helion, Gliwice

[13] Ivaniuk, P, 2026. Equipment Losses During Russian Invasion of Ukraine [online]. Available at: https://www.kaggle.com/datasets/piterfm/2022-ukraine-russia-war-equipment-losses-oryx. [Accessed: 01 February 2026].

[14] Janovsky, J, 2026. Attack On Europe: Documenting Russian Equipment Losses During The Russian Invasion Of Ukraine [online]. Available at: https://www.oryxspioenkop.com/2022/02/attack-on-europe-documenting equipment.html [Accessed: 01 February 2026].