Journal of Image Processing Theory and Applications (2026)
Clausius Scientific Press, Canada

DOI: 10.23977/jipta.2026.090103
ISSN 2560-6239 Vol. 9 Num. 1

Automatically Detecting Zokor Molehills in Forests Using Machine Learning

Keqin Song[1,a], Zhiwei Wang[1,b], Limin Wang[1,c], Jihui Hao[1,d], Lijie Wang[1,e], Tianliang Zhang[2,3,f,*]

[1]Weichang Manchu and Mongol Autonomous County State-Owned Luanhe Forest Farm, Weichang County, Chengde, China
[2]Institute of Applied Mathematics, Hebei Academy of Sciences, No. 46 South Youyi Street, Shijiazhuang, China
[3]Hebei Information Security Certification Technology Innovation Center, No. 46 South Youyi Street, Shijiazhuang, China
[a]skq202505@163.com, [b]250383166@qq.com, [c]769651419@qq.com, [d]13832450358@163.com, [e]550326365@qq.com, [f]tianliangzn@126.com
**Corresponding author*

Keywords: UAV remote sensing; YOLOX; Object detection; Zokor mounds; Forest monitoring

Abstract: Traditional monitoring of zokor mounds relies primarily on manual ground surveys, which suffer from low efficiency, high costs, significant limitations due to terrain and vegetation cover, high rates of missed observations, and strong subjectivity in data collection. These methods struggle to meet the demands of modern forestry for precise, large-scale, and dynamic monitoring. This study focuses on the Saihanba Forest Area in Hebei Province, China, and develops an automated detection system for zokor mounds based on UAV visible-light imagery and the YOLOX deep learning model. A DJI Mavic 3 drone was used to acquire visible-light remote sensing imagery at a flight altitude of 30 m. A dedicated dataset for zokor mounds in the forest area was constructed, comprising 862 valid images and 1,286 annotated samples, covering targets of varying habitats and sizes. YOLOX was selected as the core detection model, and transfer learning based on COCO pre-trained weights was employed for model training and parameter optimization to enhance convergence speed and generalization ability. Experimental results show that the optimized YOLOX model achieved an average precision (AP) of 99.90% on the test set, enabling accurate identification of zokor mounds against complex forest backgrounds with excellent resistance to false positives and false negatives; the model's precision-recall (PR) curve lies predominantly in the upper-right quadrant of the graph, maintaining stable and efficient detection performance across different confidence thresholds. This study demonstrates the feasibility and advantages of using visible-light imagery from consumer-grade drones combined with deep learning methods for the rapid, large-scale monitoring of zokor mounds in forested areas. This approach provides efficient, low-cost technical support for the precise prevention and control of rodent damage, the dynamic assessment of forest ecosystems, and the scientific evaluation of the ecological impacts of zokors, and holds significant theoretical value and promising prospects for engineering applications.

1. Introduction

Moles are subterranean rodents endemic to forest and grassland ecosystems. Their burrowing activities create numerous exposed mounds on the ground (known as mole mounds), and this mound-building behavior has a dual effect on the ecosystem: On the one hand, molehills destroy surface vegetation, bury seedlings, and sever plant root systems, leading to a decline in tree survival rates and posing a biological hazard to the regeneration of planted forests and agricultural production[1-3]; on the other hand, moderate mound-building by zokors alters soil physicochemical properties, creates heterogeneous habitats, enhances plant community species diversity, and promotes material cycling and regeneration in alpine meadow and forest ecosystems, serving as a key driver of natural ecosystem succession[4-6]. Whether for the precise prevention and control of zokor damage or the scientific assessment of zokor ecological effects, data on the spatial distribution, density, and dynamic changes of zokor mounds serve as the core basis. Traditional monitoring of zokor mounds primarily relies on manual ground surveys, with data collected through plot counting and visual estimation. This approach is not only time-consuming and labor-intensive, with high survey costs, but is also constrained by terrain and vegetation cover. It suffers from high rates of missed detection, strong subjectivity in data collection, and difficulties in achieving large-scale synchronized monitoring, failing to meet the needs of modern forestry and grassland management for precision and dynamism. Therefore, there is an urgent need to develop a large-scale, rapid, high-precision, and low-cost automated monitoring method for zokor mounds.

Drone remote sensing technology offers the advantages of mobility, flexibility, low cost, high spatial and temporal resolution, and minimal terrain constraints, and has become a core technical tool for monitoring agricultural pests and diseases and surveying ecological resources. In recent years, the integration of UAV remote sensing with deep learning-based object detection algorithms has yielded research results in fields such as crop disease identification, forestry resource surveys, and wildlife monitoring. In rodent monitoring research, existing studies have primarily focused on macro-ecological analyses of rodent population distributions or habitat suitability assessments based on ground-based spectral data; research on the automatic detection of zokor mounds in UAV imagery remains relatively scarce. As an improved model in the YOLO series, YOLOX employs an anchor-free design, decouples the detection head from the SimOTA label matching strategy[7-8] and offers significant advantages in small-object detection and interference resistance in complex backgrounds. To date, no studies have applied YOLOX to the detection of zokor mounds in UAV imagery.

In response to the practical needs of zokor mound monitoring, this study focuses on the Saihanba Forest Area in Hebei Province, China, to develop a technical framework for the automatic detection of zokor mounds based on UAV visible-light imagery. The main research components include: (1)methods for UAV image data acquisition and preprocessing, establishing aerial photography technical specifications and image correction workflows tailored for zokor mound detection; (2)Constructing a dedicated dataset for zokor mound detection in forest areas, including sample annotation and dataset segmentation; (3)Developing a zokor mound detection model based on the YOLOX architecture, and performing model training and parameter optimization via transfer learning; (4)Systematically evaluating the model's detection performance to validate the feasibility of the methodology.

2. Materials and Methods

2.1. Overview of the Study Area

The study area is located around the Saihanba Mechanical Forest Farm within Weichang Manchu and Mongol Autonomous County, Chengde City, Hebei Province, China (116°51′–117°39′E, 42°02′–

42°36′N). Situated in the transitional zone between the Inner Mongolia Plateau and the Northern Hebei Mountains, it ranges in elevation from 1,010 to 1,940 m and has a cold-temperate continental monsoon climate. with an annual average temperature of -1.3 ℃ and annual precipitation of approximately 450 mm, primarily concentrated between June and August. The study area is dominated by planted forests, with dominant tree species including Mongolian pine, larch, and white birch. Extensive grasslands and shrublands are distributed throughout the forest, making it a core region for the conservation and ecological restoration of planted forests in northern China. This region is a concentrated habitat for the steppe zokor and the Daurian zokor, where rodent damage is widespread. Young stands are particularly severely affected by steppe zokors, and zokor mounds are commonly found throughout the forest, providing a typical experimental setting for this study.

2.2. Data Collection

2.2.1. Collection Equipment

For this experiment, a DJI Mavic 3 drone was used as the data collection platform. It was equipped with a visible-light imaging sensor featuring 20 million effective pixels and a 24mm equivalent focal length, supporting image storage in both JPG and RAW formats. This device is lightweight, highly maneuverable, and delivers consistent image quality, making it well-suited for the practical demands of forestry fieldwork.

2.2.2. Collection Plan

Data collection took place from June to October 2025, a period during which vegetation in the study area was thriving and zokors were highly active. The spectral and textural differences between exposed zokor mounds and the surrounding vegetation were significant, facilitating target identification. To ensure uniform image illumination and avoid interference from backlighting and shadows, data collection was scheduled for sunny days between 10:00 a.m. and 2:00 p.m., when the sun's altitude angle exceeded 45°. For this experiment, the flight altitude was set at 30 m, with a forward overlap rate of 80%, a side overlap rate of 70%, and a flight speed of 5 m/s, corresponding to a ground resolution of approximately 1.0 cm per pixel. A total of three flight missions were completed, yielding a cumulative total of 1,248 raw aerial images, the actual image capture results are shown in Figure 1.

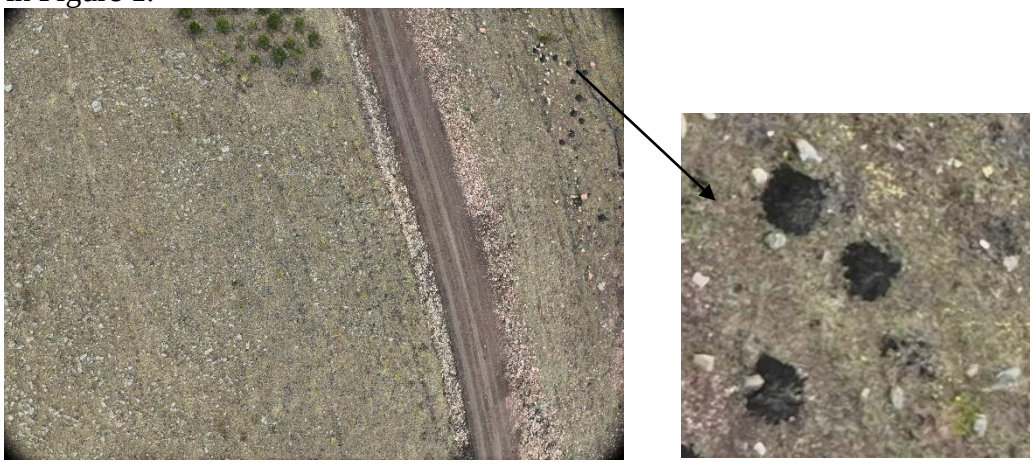

Figure 1: Single image and ground resolution display

2.3. Image Preprocessing

Raw aerial images captured by UAVs are subject to vignetting, radial distortion, and tangential distortion due to the optical characteristics of the lens and the shooting orientation. These issues result in uneven brightness at the edges of the images and geometric distortion, which can compromise the accuracy of subsequent target feature extraction and detection. Therefore, this study first performs preprocessing on the raw images, primarily consisting of two core steps: vignetting compensation and lens distortion correction.

2.3.1. Vignetting Correction

Lens vignetting causes higher brightness in the center of the image and a decrease in brightness toward the edges, resulting in differences in grayscale values for the same object at different locations within the image. In this study, based on a lens optical attenuation model, a polynomial fitting method is used to correct vignetting in the image. The correction formula is as follows:

$$I'_{(x,y)} = I_{(x,y)} \times (k[5] \cdot r^6 + k[4] \cdot r^5 + \cdots + k[0] \cdot r + 1) \quad (1)$$

In the equation: $I_{(x,y)}$ represents the grayscale value of pixel (x,y) in the original image, and $I'_{(x,y)}$ represents the corrected grayscale value; r is the pixel distance from pixel (x,y) to the optical center of the image; k_0-k_5 are the lens vignetting correction coefficients, which are directly retrieved from the EXIF metadata of the UAV image.

2.3.2. Lens Distortion Correction

Manufacturing and installation errors in UAV lenses can cause radial and tangential distortion in images, resulting in deviations in the geometric positions of ground features. This study corrects both types of distortion; the formula for the radial distortion correction model is as follows:

$$\begin{cases} x_{corrected} = x(1 + k_{1r^2} + k_{2r^4} + k_{3r^6}) \\ y_{corrected} = y(1 + k_{1r^2} + k_{2r^4} + k_{3r^6}) \end{cases} \quad (2)$$

The formula for the tangential distortion correction model is as follows:

$$\begin{cases} x_{corrected} = x + [2p_1xy + p_2(r^2 + 2x^2)] \\ y_{corrected} = y + [p_1(r^2 + 2y^2) + 2p_2xy] \end{cases} \quad (3)$$

In the equation: (x,y) represents the normalized pixel coordinates of the distorted image, and ($x_{corrected}$, $y_{corrected}$) represents the corrected pixel coordinates; k_1、 k_2, and k_3 are the radial distortion coefficients, and p_1 and p_2 are the tangential distortion coefficients, all of which are extracted from the camera calibration file and image metadata.

After completing the geometric correction, a 5×5 Gaussian filter is applied to the image for denoising, which smooths out image noise while preserving the edge texture features of the hamster pile, thereby completing the image preprocessing workflow.

2.4. Construction of the Detection Model

This study employs YOLOX as the core detection model. YOLOX is an anchor-free single-stage object detection model improved upon YOLOv3. By incorporating the Focus module, the PANet bidirectional feature fusion network, decoupled detection heads, and the SimOTA label matching strategy, it achieves an excellent balance between detection accuracy and inference speed. It exhibits strong robustness for detecting small objects in complex backgrounds and is well-suited for the

detection of zokor nests in drone imagery.

2.4.1. Overall Architecture of the YOLOX Model

The YOLOX model consists of four core components: the input, the backbone network, the neck feature fusion network, and the detection head. The overall architecture is as follows:

Input: Adaptive image resizing is employed to uniformly convert input images of varying sizes into a standard 640×640-pixel format. The Mosaic data augmentation method is introduced, which fuses four images through random scaling, cropping, and stitching. This enriches the background and target scale distribution of the dataset, thereby enhancing the model's generalization ability and training efficiency.

Backbone Network: Composed of the Focus module, the CBS (Conv_BN_SiLU) module, the Residual Block (ResBlock_body), and the SPP (Spatial Pyramid Pooling) module. The Focus module uses skip-pixel sampling to convert the height and width information of the input image into channel information, achieving downsampling while avoiding feature loss; The CBS module, comprising convolutional layers, batch normalization layers, and the SiLU activation function, performs convolutional feature extraction; the ResBlock module increases network depth through residual connections, addresses the vanishing gradient problem, and effectively extracts deep semantic features from the images; the SPP module fuses multi-scale features via pooling layers with different kernel sizes, enhancing the network's ability to extract features from mouse piles of varying sizes.

Neck Network: The PANet path aggregation network is adopted. Through a bidirectional feature fusion structure combining top-down and bottom-up approaches, it effectively integrates shallow-layer textural details with deep-layer semantic features. This addresses the issue of easily lost small-scale features in UAV imagery and significantly improves the detection performance of small targets.

Detection Head: By adopting a decoupled head architecture, the model separates the object classification and bounding box regression tasks into two independent branches for computation. This resolves the conflict between classification and regression tasks found in traditional coupled heads, significantly improving detection accuracy and convergence speed while adding only a minimal computational overhead. Additionally, the model employs an anchor-free design, eliminating the need for predefined anchor boxes. This reduces computational complexity and enhances the model's adaptability to targets of varying scales.

2.4.2. Model Training Strategy

This study employs a transfer learning approach for model training. We initialize the model network using weights pre-trained on the COCO public dataset, followed by global fine-tuning on the study's own data set. This effectively addresses the issues of slow convergence and poor generalization in small-scale datasets, accelerating model convergence and improving detection accuracy. The core model training parameters are set as follows: the Stochastic Gradient Descent with Momentum (SGDM) algorithm is used to update model parameters, with momentum set to 0.937 and a maximum gradient threshold of 30; The initial learning rate is set to 5e-4, with a learning rate decay factor of 0.99 and a weight decay coefficient of 5e-4; the learning rate is adjusted once after every training epoch; the total number of training epochs is set to 100, with a batch size of 16; the dataset order is shuffled before each training session, and the model weights with the lowest validation set loss are selected as the final detection model.

2.5. Model Evaluation Metrics

This study employs evaluation metrics commonly used in the object detection field, including Precision (P), Recall (R), and Average Precision (AP), to comprehensively assess the model's

detection performance. The formulas for each metric are as follows:

Precision: The proportion of true positive samples among the results predicted as positive by the model; it reflects the model's ability to avoid false positives. The formula is:

$$P = \frac{TP}{TP+FP} \times 100\% \tag{4}$$

Recall: The proportion of true positive samples correctly predicted by the model, reflecting the model's ability to avoid false negatives. The formula is:

$$R = \frac{TP}{TP+FN} \times 100\% \tag{5}$$

Average Precision: Refers to the area under the ROC curve. It comprehensively reflects the model's precision and recall performance at different confidence thresholds and is a core evaluation metric for object detection models. The formula is:

$$AP = \int_0^1 P(R)dR \tag{6}$$

Where: TP represents true positives, the number of mole hill samples correctly detected by the model; FP represents false positives, the number of non-target samples mistakenly identified as mole hills by the model; FN represents false negatives, the number of actual mole hill samples not detected by the model.

3. Results and Analysis

3.1. Dataset Construction and Split

In this study, based on preprocessed UAV orthophotos, the LabelImg annotation tool was used to perform bounding box annotations on zokor mounds, with the annotation class designated as "zokor_mound." To accommodate the model's input dimensions, the images were cropped into 640×640-pixel tiles, and blank tiles without targets were removed, ultimately resulting in the construction of a dedicated dataset for zokor mound detection. The dataset comprises a total of 862 valid images, with 1,286 labeled zokor mounds, covering samples of varying habitats and sizes. The smallest target measures 8×10 pixels, while the largest is 320×350 pixels, ensuring the dataset's diversity and representativeness. The dataset was randomly split into training, validation, and test sets in a 7:1.5:1.5 ratio, comprising 603 images in the training set, 130 in the validation set, and 129 in the test set. There is no overlap between the subsets, ensuring the objectivity of model evaluation.

3.2. Analysis of the Model Training Process

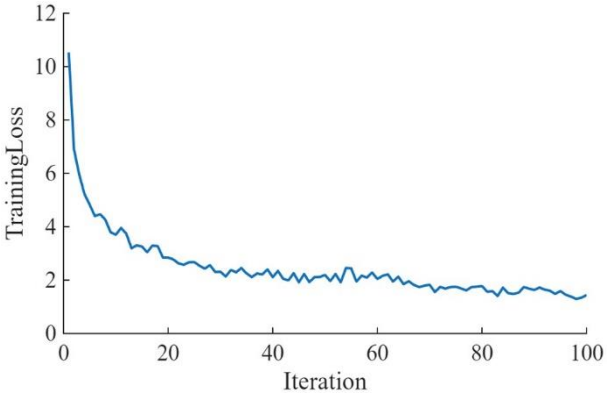

Figure 2: Model Training Loss Curve

The loss function curve for the model training process is shown in Figure 2, where the horizontal axis represents the number of training epochs and the vertical axis represents the loss value. The curve shows that as the number of training epochs increases, the model's training loss continues to decrease. The rate of decrease is fastest during the first 20 epochs, indicating rapid convergence. By approximately 85 epochs, the loss curve levels off, and the training loss eventually converges to 1.30, demonstrating good convergence.

3.3. Analysis of Model Detection Performance

The performance of the trained YOLOX model was evaluated using the test set, which comprised a total of 129 annotated images of groundhog mounds. The trained YOLOX model demonstrated excellent detection performance for groundhog mounds in drone imagery, achieving an average precision (AP) of 99.90% and a precision rate exceeding 99%. This indicates that the model has an extremely low false positive rate and can effectively distinguish groundhog mounds from similar background objects such as clumps of soil, rocks, and bare ground in the forest. The model's ROC curve is shown in Figure 3; the closer the ROC curve is to the upper-right corner of the graph, the better the model's overall performance. As shown in the figure, the model's ROC curve is generally located in the upper-right quadrant of the graph, with an area under the curve (AUC) of 0.990. Within the range where recall increases from 0 to 0.9, the model's precision consistently remains above 0.9, with only a slight decline when recall approaches 1.0. This fully demonstrates that the model maintains stable and excellent detection performance across different confidence thresholds. The primary objective of this paper is to achieve accurate identification of mole mounds, and the trained model has met this objective. The specific identification results are shown in Figure 4, where all mole mounds were successfully identified.

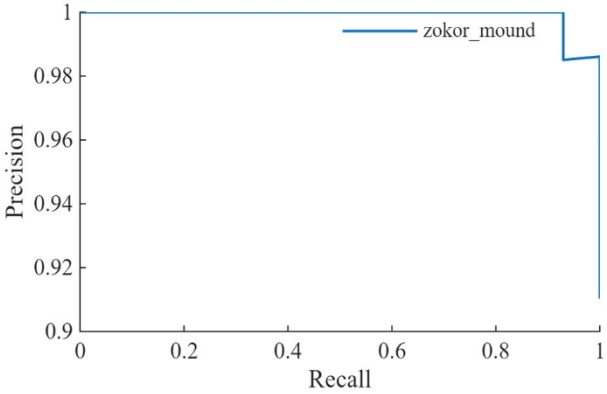

Figure 3: Model Test PR Curve

Figure 4: Results of Zokor Mound Recognition

4. Discussion

This study developed an automated detection method for zokor mounds based on UAV visible-light imagery and the YOLOX deep learning model, enabling high-precision, rapid, and large-scale automated monitoring of zokor mounds in forested areas. The innovations and advantages of this study are primarily reflected in three aspects: the application of the YOLOX model to the task of detecting zokor mounds in UAV imagery; the construction of a zokor mound detection dataset for forested areas; and the validation of the superior performance of the anchor-free single-stage detection model in this scenario, with a model detection AP of 99.90%. This method achieves high-precision detection using only visible-light imagery from consumer-grade UAVs, eliminating the need for expensive professional equipment such as multispectral or hyperspectral sensors. With low data acquisition costs and a low operational threshold, it is easily adoptable by grassroots forestry management departments. The optimized model balances detection accuracy with inference speed, enabling deployment on UAV edge computing platforms and portable terminals to facilitate real-time detection of zokor mounds, thereby greatly enhancing the convenience of field operations.

At the same time, this study has certain limitations that require further refinement in future research: First, the dataset used in this study only covers imagery from the vicinity of the Saihaba Mechanical Forest Farm. The model's generalization ability under different seasons and vegetation cover conditions requires further validation. In future work, we will construct multi-temporal and multi-regional datasets covering periods of sparse vegetation, such as spring and autumn, to enhance the model's adaptability to various scenarios; Second, the model's detection accuracy for extremely small zokor mounds completely obscured by vegetation still has room for improvement. Future work will incorporate attention mechanisms and multi-scale feature fusion modules to refine the model and further enhance its small-object detection capabilities; Finally, this study has only achieved object detection of mole mounds. In future work, we will integrate multi-object tracking algorithms with digital elevation models to enable automatic counting, volume estimation, and spatial distribution pattern analysis of mole mounds, thereby providing more comprehensive data support for rodent pest risk assessment.

5. Conclusions

Guided by the practical needs of monitoring zokor mounds in forested areas, this study developed an automated detection method for zokor mounds based on UAV visible-light imagery and the YOLOX deep learning model. The study systematically completed data collection, image preprocessing, dataset construction, model training, and performance validation. The main conclusions are as follows: The specialized dataset for detecting zokor mounds in forest areas constructed in this study comprises 862 images and 1,286 annotated samples, covering zokor mounds of various sizes across different habitats. This provides a reliable data foundation for training deep learning models and serves as a data reference for subsequent related research. The zokor mound detection model, built based on the YOLOX architecture, achieved an average precision (AP) of 99.90% on the test set. It can accurately identify zokor mound targets in complex backgrounds and demonstrates excellent resistance to false positives and false negatives.

This study validated the feasibility and superiority of combining visible-light imagery from consumer-grade drones with deep learning methods for large-scale, rapid, and automated monitoring of zokor mounds in forest areas. It effectively addresses the numerous limitations of traditional manual survey methods, providing an efficient and low-cost technical approach for precise pest control and dynamic monitoring of forest and grassland ecosystems in China's forest regions, with broad application prospects.

Acknowledgements

This research was supported by the Special Science and Technology Project for the Construction of the Chengde National Innovation Demonstration Zone for the Sustainable Development Agenda under Grant 202302F007.

References

[1] Li S C.(2025). Assessment of Forest Tree Damage Caused by Myospalax baileyi and Screening of Rodenticides in the Changling Mountain Forest Area. Forest Investigation Design,54(05),28-32.
[2] Song X N, Li L, Lu S H, et al.(2025). The Impact of Rodents in Heilongjiang's Forest Areas on Cash Crops and Forest Regeneration. Rural Economy and Science-Technology,36(15),81-84.
[3] Liu J X, Xie S Q, Zhang Z H, et al.(2020). Study on spatial sidtribution and sampling technique of major rodent pest in Saihanba area. Foretry and Ecological Sciences,35(02),186-190.
[4] Niu, Y, Yang, S, Zhu, H et al.(2020). Cyclic formation of zokor mounds promotes plant diversity and renews plant communities in alpine meadows on the Tibetan Plateau. Plant Soil,446, 65–79.
[5] Xiang, Z, Bhatt, A, Tang, Z, et al. (2021) Disturbance of plateau zokor-made mound stimulates plant community regeneration in the Qinghai-Tibetan Plateau, China. J. Arid Land 13, 1054–1070.
[6] Su, F, Wang, F, Li, D, et al(2025). Plateau zokor mounds decrease soil respiration by shifting microenvironment. CATENA 258.
[7] Ge, Z, Liu, S, Wang, F, et al(2021). Yolox: exceeding yolo series in 2021. ArXiv, abs/2107.08430.
[8] Han Y, Liu H R, Lin W S.(2025). Tree Species Identification Based on Improved YOLOv10 and UAV RGB Imagery. FOREST ENGINEERING,41(05),922-935.