



Predicting Twist Condition by Bayesian Classification and Decision Tree Techniques

Ehsan Izadi Yazdan Abadi¹, Melody Khadem Sameni^{2*}, Masoud Yaghini³

¹ School of Railway Engineering, Iran University of Science and Technology, Tehran, Iran

² Assistant Professor of Iran University of Science and Technology

³ Associate Professor, Iran University of Science and Technology

ARTICLE INFO

Article history:

Received: 15.01.2022

Accepted: 08.04.2022

Published: 16.04.2022

Keywords:

Failure prediction

Geometric parameters

track twist

Chain failures

data mining

A B S T R A C T

Railway infrastructures are among the most important national assets of countries. Most of the annual budget of infrastructure managers are spent on repairing, improving and maintaining railways. The best repair method should consider all economic and technical aspects of the problem. In recent years, data analysis of maintenance records has contributed significantly for minimizing the costs. By studying each of line parameters, we can take steps towards condition based maintenance of that parameter. Chain faults as well as the cost of the track inspection will be reduced too. In this research, data mining techniques are used to investigate the relationship between geometric parameters of track and twist failure. By using other parameters of the track, twist can be predicted. Moreover, the parameter that has the highest impact on twist can be found. In this paper, Bayesian classification and decision tree techniques have been used. Finally, after conducting this study, it was found that alignment level (AL) and cross level (XLV) have the highest impact on causing track twist. This is the first research in the literature that have used all the above-mentioned methods to investigate twist failure in order to avoid chain failure and achieve condition-based maintenance.

1. Introduction

In railway lines, due to the high cost of construction, the issue of maintenance is usually of special importance. In order to reduce the cost of track failures, predictive based maintenances is suggested [1]. Analyzing the data collected from track allows us to examine the condition of the parameters. By obtaining the conditions of the geometrical parameters of track, condition-based maintenance can be achieved. Conditional maintenance can reduce the time required to maintain railway lines in addition to reducing costs.

In this study, we tried to determine the status of track geometric parameters in relation to twist failure to obtain the relationship between these failures. Using two data mining techniques

including the decision tree and Bayesian classification, each of which are predictive models, the relationship between geometric parameters and twist failure was determined. Finally, it was found that alignment (AL) and Cross level (VLX), which are the geometric parameters of the track, have the greatest effect on the failure of the twist parameter in the track.

By determining the parameters affecting the twist of the route, it is possible to understand the status of twist of the line by having other parameters before twist occurs. Therefore, before a failure occurs, it can be predicted and repair costs are significantly reduced. This study examines the Tehran-Mashhad railway route. This route is one of the main lines in the railway network of the Islamic Republic of Iran, which

*Corresponding author, Associate professor
Email: sameni@iust.ac.ir

is heavily under railway passenger traffic which put constraints on track maintenance.

2. Literature Review

Studies have been conducted in the field of maintenance of railway lines as well as the use of data mining in railways, which will be examined in this section. The most important issue for line maintenance and repair operations is full knowledge of the condition of the line and its deterioration rate [2].

For this reason, many studies emphasize that railroad breakdowns should be identified at the beginning of maintenance inspections [3].

Twist of the line can lead to horizontal and vertical roughness. The amount of this twist depends on factors such as train speed, dynamic load, build quality and line structure behavior [4].

Railways, as one of the most important assets in every country's railway, must be constantly inspected to increase its life cycle [5]. But the point that reduces the number of inspections along the lines is the cost of these inspections. Maintenance constraints have a large impact on costs, and by examining these constraints, costs will decrease [6].

That another study in this field shows how to reduce maintenance costs on railway by reducing existing constraints. For example, one of the most important constraints that increase time and cost is the environmental conditions at the time of maintenance, and the second factor is the constraints that are imposed on the repair team in some routes, such as tunnels. In this study, it was stated how to reduce the cost by eliminating these restrictions or minimizing them. Finally, each of the restrictions had a coefficient that showed its importance and it turned out that these restrictions have a great impact on the cost of repairing road breakdowns [7]. One of the techniques that can be used to examine the condition of geometric parameters in railways is data mining techniques, which can also be used to examine the limitations of maintenance. Data mining techniques have a significant efficiency in investigating rail failures where the data collected is very large, because they can process large volumes of data [8].

The most important point in examining the failures of railway routes is the issue of accurate

and timely diagnosis of failures. Over time, maintenance and repair costs increase because a breakdown in one part of the track, causes improper performance in that section, and this causes the breakdown to spread along the route [9]. Predictive models are one of the main components in data mining that can be used to identify failures before they occur [10].

A study conducted to predict geometrical parameters of track degradation states that the use of data mining techniques can help to improve track conditions by predicting failures [11]. Quality indicators can show the overall status of tracks [12].

With the help of data mining techniques, a clear relationship can be established between these quality indicators and geometric parameters [13].

For example, in another study which used neural network, the quality of the railway track was investigated and the conditions affecting the quality of the track were extracted [14]. Therefore, to reduce costs, we need to optimize the maintenance process of rail lines [15].

In addition to the high financial cost of repairs in the event of a track breakdowns not paying attention to failures of geometric parameters decreases safety [16]. On a railway track, the closer we get to the top of the track, the more intense the loads will be [17].

Therefore, the quality of the railway track is directly related to the status of the geometric parameters of the track [18].

In rail lines geometric parameters of the route should be properly examined to minimize maintenance and inspection costs of these parameters [19]. For this reason, many studies have been done recently to evaluate the geometry of track [20].

For example, in a study using data mining techniques, the causes of fault geometry failure were investigated. The passage of heavy loads as well as heavy traffic from this route was considered as the most important cause of breakdowns because the structure of the track was not compatible with the passage of this influx of traffic [21]. By obtaining the conditions of each of the geometric parameters in the railway tracks, condition-based maintenance can be achieved [22].

To summarize previous studies, it is observed that so far no study has been conducted to investigate the relationship between twist failure and other geometric parameters of the track and this study is one of the first studies in this field which uses Bayesian classification and decision tree.

3. Case study profile

The study data was obtained from the Tehran-Mashhad route. The route from Tehran to Mashhad is one of the main routes in the railway network of the Islamic Republic of Iran.

For this study, data of 40 km of the railway route in three consecutive periods between 2018 and 2020 have been collected using the EM120 machine.

The general specifications of the route are given in Table 1.

Table 1- General specification of the case study

Specifications under review	Specification type
The length of the route under consideration	40 Kilometers
Type of route measuring machine	EM120
Type of sleeper	Pre stressed B70
Type of rail	UIC60
Type of fastening system	Vossloh
Type of rail connection	welded
Type of transit traffic	Passenger and freight

In this route, tamping operations are performed almost once every 6 months, and the examined data are collected twice a year, and their examination determines the process of failure. Over time, the parameters deviate from their standard mode twice after the tamping operation is performed. Given that after each tamping the parameters return to their standard

state, in this study we only examined the effect of other failures on the torsion, meaning that the data were stored at specified intervals from the tamping operation until they deviate from their standard mode again.

In this study, the geometric parameters of the track are investigated. The number of geometric parameters of the track is five, the status of these parameters is monitored by a track measuring machine [23] such as EM120.

Five geometric parameters include Longitudinal level (LL), Cross level (VLX), Track gauge (GAU), Twist (TWS) and Alignment level (AL) [24]. Each of these parameters is others and before the occurrence of failure in a geometric parameter, signs of failure can be observed in other parameters. In this study, we investigated the relationship between the occurrence of twist failure and four other geometric parameters.

In this study, three situation is considered for parameters: Good (parameter is within acceptable range), +Bad (parameter is higher than acceptable level), -Bad (parameter is less than acceptable level). It must be mentioned that case study is an A1 line hence acceptable levels are defined accordingly.

As mentioned before, the data of this study were collected using EM120 along the railway track. Table 2 shows a small sample of the collected data. Descriptive statistics of the case study is shown in Table 3. Figure 1 shows the distribution of torsion along the route.

Table 2- Sample of data which is used in the case study

KM	LLL	LLR	GAU	XLV	ALL	ALR	TWS32	ACCH	ACCV	ACCL	ACCT
818.845	1.67	0.60	-0.116	-33.512	-2.930	-2.253	-0.215	-2.596	0.236	1.343	2.606
818.845	1.72	0.82	-0.463	-33.694	-3.097	-2.563	-0.128	-2.661	0.565	1.277	2.721
818.846	1.66	1.05	-0.843	-33.802	-3.358	-2.999	0.078	-2.596	0.894	1.277	2.745
818.846	1.59	1.07	-1.190	-33.852	-3.635	-3.423	0.088	-2.464	1.157	1.145	2.722
818.846	1.48	1.19	-1.633	-33.881	-3.659	-3.498	0.213	-2.464	1.617	1.080	2.947
818.846	1.43	1.27	-1.680	-33.858	-3.574	-3.361	0.311	-2.333	1.945	1.080	3.038
818.847	1.32	1.44	-1.664	-33.819	-3.348	-3.130	0.322	-2.136	2.274	0.948	3.120
818.847	1.17	1.40	-1.475	-33.721	-3.233	-3.041	0.175	-1.873	2.406	1.145	3.049
818.847	0.92	1.27	-1.380	-33.657	-3.022	-2.961	-0.066	-1.478	2.866	1.080	3.225
818.847	0.67	1.05	-1.269	-33.652	-2.770	-2.775	-0.280	-1.084	3.129	1.277	3.311
818.848	0.37	0.89	-1.080	-33.589	-2.467	-2.476	-0.628	-0.492	3.392	1.080	3.427
818.849	-1.13	-0.70	0.153	-33.703	-0.709	-1.307	-0.465	2.006	2.603	0.883	3.286
818.849	-1.31	-1.06	-0.021	-33.809	-0.567	-1.212	-0.338	1.940	2.340	1.014	3.039
818.850	-1.61	-1.20	-0.432	-34.179	-0.361	-1.012	-0.185	1.808	2.011	1.343	2.705

Table 3- Descriptive statistics of the case study

	LLL	LLR	GAU	XLV	ALL	ALR	TWS32
Mean	0.00	0.00	0.97	2.50	0.16	0.44	-0.28
Median	0.02	0.02	1.04	0.82	0.12	0.41	-0.29
Maximum	15.55	15.46	10.28	87.82	16.34	18.12	10.73
Minimum	-15.73	-16.19	-26.58	-125.72	-15.36	-19.16	-16.42
Std. Dev.	1.79	1.74	1.91	21.51	2.58	2.66	1.50
Skewness	-0.22	-0.19	-0.27	-0.57	-0.24	-0.27	-0.19
Kurtosis	8.38	8.19	5.68	16.30	7.91	7.24	6.49
Jarque-Bera	180488.20	168219.40	46258.35	1105681.00	150659.00	113074.20	76450.42
Sum	6.19	63.31	144441.60	372779.60	23909.72	65758.50	-42170.99
Sum Sq. Dev	475641.20	448443.70	542470.40	68900758.00	987919.30	1049650.00	333505.80
Observations	480000.00	480000.00	480000.00	480000.00	480000.00	480000.00	480000.00

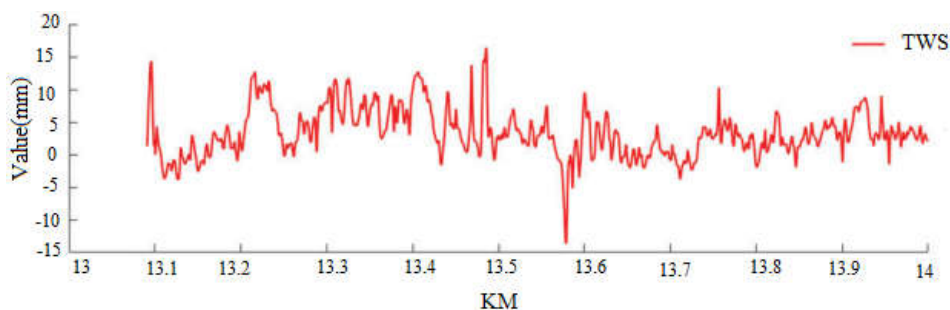


Figure 1-Variation of TWS in track

4. Developed Models

In this paper, two data mining techniques are applied to the collected data. The methods used are: classification and decision tree. By using these methods, the underlying rules have been extracted and future conditions of track are predicted. One of the models with which we could make predictions was the Markov model, which was less accurate than the Bayesian classification and decision tree models, and also Markov model, unlike the models used in this study, is memoryless which means that when used for one group of data, the model memory is cleared for the data of the other group and it performs the operations related to the next series of data from the beginning. Due to the stated reasons and the type of the data in this study, decision tree models and Bayes classifier are used.

4.1. Bayesian classification

In machine learning, a group of simple classifiers which are based on Bayes theorem and assume the independence of random variables are called Bayesian classification.

It is based on prediction of event A, based on probability of occurrence or non-occurrence of an event B [25]. Bayesian classification is used to predict the conditions of the twist parameter based on other geometric parameters. Also, Measurement of variables can lead to three possible outcomes: Good, + Bad (clockwise twist) and - Bad (counterclockwise twist).

4.2. Decision tree

One of the most widely used data mining techniques is the decision tree algorithm [26]. Decision tree is a predictive model which consists of branches, leaves and roots [25]. Leaves are the nodes where consecutive divisions end. The leaves are identified by a class. The root means the starting node of the tree [27]. In each internal node, there would be different branches as many as possible results. Each path from the root of a decision tree to a leaf expresses a rule. Finally, the leaf is labeled with the class that maximum occurrence (good, +bad or -bad). In the drawing decision tree, green indicates the geometric parameters and blue is the good predicted conditions and red

indicates the bad predicted conditions for the twist parameter. The results of the decision tree can be seen in Figure2.

5. Results

Obtained results of Bayesian classification are presented in Table 4 and decision tree in Figure3.

Table 4- Prediction of twist conditions based on the conditions of other geometric parameters

Attribute	Parameter	Good	+bad	-bad
.LLL	.good	1	0.997	0.997
.LLL	.-bad	0	0.003	0.002
.LLL	+.bad	0	0	0.001
.LLR	.good	1	0.999	0.999
.LLR	.-bad	0	0.001	0.001
.LLR	+.bad	0	0	0.001
.GAU	.good	0.933	0.913	0.918
.GAU	.-bad	0.059	0.076	0.073
.GAU	+.bad	0.007	0.01	0.009
.XLV	.good	0.846	0.728	0.751
.XLV	.-bad	0.05	0.122	0.054
.XLV	+.bad	0.104	0.15	0.194
.ALL	.good	0.946	0.889	0.901
.ALL	.-bad	0.019	0.075	0.009
.ALL	+.bad	0.035	0.036	0.09
.ALR	.good	0.94	0.905	0.893
.ALR	.-bad	0.017	0.059	0.006
.ALR	+.bad	0.043	0.035	0.1

It can be inferred that the most effective parameter for twist failure are XLV and the AL. This is because the percentage of failure of these parameters when the twist parameter is in bad conditions is higher than other parameters.

By using the decision tree it can be seen that that the failure of the longitudinal level of rail parameters alignment of the rail and super elevation occurred simultaneously with the failure of the twist. That is, when the conditions of these parameters were bad (+ or -), the conditions of bad twisting were predicted.

Decision tree predicts a parameter in good condition when most of the underlying parameters are in good condition. Below, parameters that lead to twist are identified and

twist situation is predicted in different conditions in text format.

Therefore, due to the characteristics of Bayesian classification and decision trees, the use of these two methods in this research is appropriate.

In the decision tree method, this model has used parameter ALL to start, and in the written form of this model, the three main branches are presented almost separately.

The first branch to be examined is for the situation that ALL is +Bad:

```
.ALL = +bad
| .XLV = +bad: good {good=4597, +bad=81, -bad=291}
| .XLV = -bad
| | .ALR = +bad
| | | .LLR = -bad: -bad {good=1, +bad=1, bad=3}
| | | .LLR = good
| | | | .GAU = -bad: -bad {good=2, +bad=0, -bad=4}
| | | | .GAU = good: good {good=19, +bad=2, -bad=17}
| | .ALR = -bad: +bad {good=0, +bad=3, -bad=0}
| | .ALR = good
| | | .GAU = -bad: +bad {good=3, +bad=6, -bad=0}
| | | .GAU = good: good {good=20, +bad=1, -bad=4}
| .XLV = good: good {good=364, +bad=3, -bad=130}
```

The second branch to be studied is for when the condition of parameter ALL is Good:

```
.ALL = good
| .XLV = +bad: good {good=10054, +bad=288,
```

```
-bad=672}
| .XLV = -bad
| | .ALR = +bad
| | | .LLR = -bad: +bad {good=1, +bad=3, -bad=0}
| | | .LLR = good: good {good=27, +bad=2, -bad=6}
| | .ALR = -bad: good {good=215, +bad=35, -bad=3}
| | .ALR = good: good {good=4802, +bad=201, -bad=204}
| .XLV = good: good {good=118413, +bad=1865, -bad=3615}
```

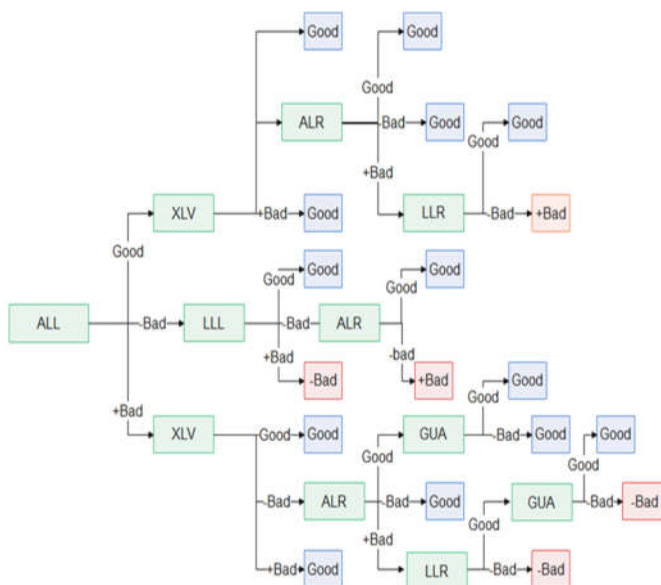


Figure 2-Decision tree to predict twist condition

```
| .LLL = +bad: -bad {good=0, +bad=0, -bad=3}
| .LLL = -bad
| | .ALR = -bad: +bad {good=5, +bad=8, -bad=2}
| | .ALR = good: good {good=2, +bad=0, -bad=0}
| .LLL = good: good {good=2644, +bad=195, -bad=42}
```

5. Discussion

From the relationship between the parameters, we found that any failure in one parameter may also affect other parameters and causes failure in them.

This research only studied the twist parameter and the effect of other parameters on this failure. This study can be very effective in improving the current maintenance status of the route and reduce financial costs to more reliability and availability.

6. Conclusions

The results of this research shows that alignment level and cross level have the highest relationship with twist failure along the line. Other geometrical parameters are less related to this failure. With Bayesian classification condition of different geometrical parameters of track can be analyzed at three defined conditions of twist (good, +Bad and -Bad). This can facilitate finding interrelationship between parameters. Decision tree which is one of the predictive models in data mining, predicts condition of twist in various conditions of other geometrical parameters. For instance, from the decision tree we can predict that if ALL is -Bad, LLL is -Bad and ALR is -Bad, twist would be in the condition of +Bad. For other conditions, rules can be extracted so that there would be no need to constantly monitor all geometrical parameters of track. Just by measuring some of them, condition of others can be predicted which will result in reducing maintenance teams and expensive machineries.

References

- [1] Khajehei, H., et al., Allocation of effective maintenance limit for railway track geometry. *Structure and Infrastructure Engineering*, 2019. 15(12): p. 1597-1612.
- [2] Kim, S.-S., et al., Parameter characteristics of rail inspection measurement system of HSR-350x. *Journal of mechanical science and technology*, 2009. 23(4): p. 1019-1022.
- [3] Wu, Q., et al., Identify severe track geometry defect combinations for maintenance planning.

International Journal of Rail Transportation, 2021: p. 1-19.

[4] Li, C., et al. Research on Dynamic Detection and Predictive System of Grooved Rail of Tram Twist Irregularity. in *2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP)*. 2021. IEEE.

[5] Andrews, J., D. Prescott, and F. De Rozières, A stochastic model for railway track asset management. *Reliability Engineering & System Safety*, 2014. 130: p. 76-84.

[6] Arasteh khouy, I., et al., Cost-effective track geometry maintenance limits. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 2016. 230(2): p. 611-622.

[7] Arasteh Khouy, I., Cost-effective maintenance of railway track geometry: a shift from safety limits to maintenance limits. 2013, Luleå tekniska universitet.

[8] Bai, L., et al., Classification-learning-based framework for predicting railway track irregularities. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 2016. 230(2): p. 598-610.

[9] Deng, J.-Y., et al. Dynamic Detection of Grooved Rail Irregularity Based on Inertial Reference Method. in *2020 IEEE International Conference on Networking, Sensing and Control (ICNSC)*. 2020. IEEE.

[10] Bai, L., et al., Markov-based model for the prediction of railway track irregularities. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of rail and rapid transit*, 2015. 229(2): p. 150-159.

[11] Khajehei, H., et al., Prediction of track geometry degradation using artificial neural network: a case study. *International Journal of Rail Transportation*, 2021: p. 1-20.

[12] Liu, R.-K., et al., Establishment of track quality index standard recommendations for Beijing metro. *Discrete Dynamics in Nature and Society*, 2015. 2015.

[13] Mehrali, M., M. Esmaili, and S. Mohammadzadeh, Application of data mining techniques for the investigation of track geometry and stiffness variation. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 2020. 234(5): p. 439-453.

- [14] Sadeghi, J. and H. Askarinejad, Application of neural networks in evaluation of railway track quality condition. *Journal of mechanical science and technology*, 2012. 26(1): p. 113-122.
- [15] Meier-Hirmer, C., et al., Maintenance optimization for a system with a gamma deterioration process and intervention delay: application to track maintenance. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 2009. 223(3): p. 189-198.
- [16] Pevzner, V., et al., Influence of long irregularities of the longitudinal profile on traffic safety under conditions of intensification of the transportation process. *VNIIZHT Scientific Journal*, 2020. 79(5): p. 271-275.
- [17] Prescott, D. and J. Andrews, A track ballast maintenance and inspection model for a rail network. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of risk and reliability*, 2013. 227(3): p. 251-266.
- [18] Kulkarni, R., et al. Monitoring of Alignment Level (AL) and Cross Level (CL) track geometry irregularities from onboard vehicle dynamics measurements using probabilistic fault classifier. in *The 27th IAVSD International Symposium on Dynamics of Vehicles on Roads and Tracks*. 2021.
- [19] Sadeghi, J., Development of railway track geometry indexes based on statistical distribution of geometry data. *Journal of Transportation Engineering*, 2010. 136(8): p. 693-700.
- [20] Sadeghi, J., M. Fathali, and N. Boloukian, Development of a new track geometry assessment technique incorporating rail cant factor. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 2009. 223(3): p. 255-263.
- [21] Sauni, M., et al., Investigating Root Causes of Railway Track Geometry Deterioration—A Data Mining Approach. 2020.
- [22] Xu, P., et al., Dynamic-time-warping-based measurement data alignment model for condition-based railroad track maintenance. *IEEE Transactions on Intelligent Transportation Systems*, 2014. 16(2): p. 799-812.
- [23] Sadeghi, J. and H. Askarinejad, Influences of track structure, geometry and traffic parameters on railway deterioration. *International Journal of engineering*, 2007. 20(3): p. 292-300.
- [24] Standard, B., Railway applications/Track-Track geometry quality. BS EN, 2003: p. 13848-1.
- [25] Yaghini, M. and R. Akhavan, Multicommodity network design problem in rail freight transportation planning. *Procedia-Social and Behavioral Sciences*, 2012. 43: p. 728-739.
- [26] De Oña, Rocío, Laura Eboli, and Gabriella Mazzulla., key factors affecting rail service quality in the Northern Italy: a decision tree approach. *Transport* 29.1,2014, 75-83.
- [27] Zheng, Z., Lu, P., & Tolliver, D. Decision tree approach to accident prediction for highway–rail grade crossings: Empirical analysis. *Transportation Research Record*, 2016, 2545(1), 115-122