Application of Classification and Regression Tree (CRT) Method for Predicting the Some Environmental Factors Affecting Weaning Weight of Awassi Lamb

Hatice HIZLI[1]* Çiğdem TAKMA[2] Şerife ERGÜL[1]

[1]Department of Animal Breeding, Eastern Mediterranean Agricultural Research Institute, Adana / Turkey
[2]Department of Animal Science, Faculty of Agriculture, Ege University Izmir / Turkey
*Corresponding author's email: hatice.hizlibostan@tarim.orman.gov.tr

Abstract: In this study, the effects of some environmental factors on the weaning weights of Awassi lambs, raised within the scope of the Awassi sheep breed sub-project of the Sheep Breeding National Project in the Hands of the People Project under the coordination of the General Directorate of Agricultural Research and Policies affiliated to the Ministry of Agriculture and Forestry, were investigated.

For this purpose, estimation was made by the regression decision tree (CRT) method, which is one of the machine learning algorithms. In the study, the effects of age of dam, gender, birth type, and flock type (elite and base flock), which are thought to influence weaning weight in Awassi lambs, were considered independent variables, while weaning weight was considered as dependent variable. According to the regression tree estimations, the most effective environmental factor for the weaning weight of Awassi lambs was found to be the birth type. While the effective factor for singleton lambs was gender, it was determined that the important factor for male lambs was the type of flock. The age of dam was found to be effective on the weaning weights of the lambs in the base flock. The results of this study revealed that the effects of various environmental factors on the healthy, efficient use and reproduction of sheep and goats can be defined with decision trees. As a result, it was concluded that regression decision trees are an important method and can be recommended as an alternative to traditional regression approaches in sheep breeding studies with both visual and predictive explanatory structure.

Keywords: *Machine learning, Regression decision tree, Awassi, Weaning weight*

İvesi Kuzularının Sütten Kesim Ağırlığını Etkileyen Bazı Çevresel Faktörlerin Tahmininde Sınıflandırma ve Regresyon Ağacı (CRT) Yönteminin Uygulanması

Öz: Bu çalışmada Tarım ve Orman Bakanlığına bağlı Tarımsal Araştırmalar ve Politikalar Genel Müdürlüğü koordinatörlüğündeki Halk Elinde Küçükbaş Hayvan Islahı Ülkesel Projesi'nin, Osmaniye İlinde yürütülen ve İvesi koyun ırkı alt projesi kapsamında yetiştirilen İvesi kuzularının sütten kesim ağırlıkları üzerine bazı çevresel faktörlerin etkisi incelenmiştir.

Bu amaçla makine öğrenmesi algoritmalarından biri olan regresyon karar ağaçları yöntemi ile tahminleme yapılmıştır. Araştırmada, 6326 kuzunun sütten kesim ağırlığına etkisi olduğu düşünülen ana yaşı, eşey, doğum tipi ve yetiştirme tipi (elit ve taban sürü) etkileri bağımsız değişken olarak ele alınırken, sütten kesim ağırlığı ise bağımlı değişken olarak dikkate alınmıştır. Regresyon ağacı tahminlemelerine göre İvesi kuzularda sütten kesim ağırlığı üzerine en etkili çevresel faktörün doğum tipi olduğu saptanmıştır. Tek doğumlu kuzular için etkili faktör eşey iken, erkek kuzular için önemli faktörün yetiştirme tipi olduğu belirlenmiştir. Taban sürüdeki kuzuların sütten kesim ağırlıklarına ise ana yaşı etkili bulunmuştur. Bu çalışmanın sonuçları ile küçükbaşların sağlıklı, verimli olarak kullanılması ve çoğalması üzerinde çeşitli çevresel faktörlerin etkilerinin karar ağaçları ile tanımlanabileceği gösterilmiştir. Sonuç olarak, regresyon karar ağaçlarının hem görsel hem de tahminsel olarak açıklayıcı yapısı ile koyun yetiştiriciliği çalışmalarında geleneksel regresyon yaklaşımlarına alternatif olarak önemli bir yöntem olduğu ve tavsiye edilebileceği kararına ulaşılmıştır.

Anahtar Kelimeler: *Makine öğrenmesi, Regresyon karar ağaçları, İvesi, Sütten kesim ağırlığı*

1. Introduction

Sheep is an animal species that has spread almost all over the world because it plays an important role in the utilization of low-yielding pastures, obtaining high-value-added products, creating income sources, ensuring development in rural areas, reducing poverty, creating employment, and providing household consumption. According to Food and Agriculture Organization (FAO) 2020 data, Turkey is among the countries with the highest number of sheep in the world (FAO, 2022). According to the 2022 data from the Turkish Statistical Institute (TUIK), the total number of sheep and goats in Turkey is 58,447,555 heads, 46,122,627 of which are sheep and 78.91% of the sheep's assets are domestic breeds (TUIK, 2022).

The Awassi breed is one of the most common domestic sheep breeds in Turkey, especially in Şanlıurfa, Gaziantep, Hatay, Kahramanmaraş, Adana and Osmaniye provinces (Kaymakçı, 2013). The homeland and spread area of the Awassi sheep are in the Mesopotamian region, and it is a sheep breed fully adapted to the semi-arid or arid countries of Southwest Asia such as Turkey, Syria, Iraq, Jordan, and Israel (Epstein, 1982; Kaymakçı, 2013). For this reason, it can walk long distances in hot and dry climatic conditions. It is grown in a nomadic system. Compared to other dairy sheep Awassi's ability to adapt to different environments is higher and their flock instinct is superior.

On the other hand, meat production constitutes the largest share of the income obtained from sheep breeding. For meat production, fertility should be increased, and lambs born should be resistant to environmental factors and have superior performance characteristics. Birth and weaning weight of lambs affect performance characteristics. In lambs, birth weight is an important determining factor in terms of weight gain in later periods and live weight in the weaning period. Weaning weight is also a factor that affects the further development of lambs sold or fattened after weaning (Boran & Torun, 2018).

The "Public Animal Breeding National Project" in Turkey was initiated by the General Directorate of Agricultural Research and Policies (TAGEM) in 2005 to increase the meat, milk, and reproductive efficiency of local breeds (Ertuğrul et al., 2014). Animal pedigree records have been started to be kept through breeder associations to meet the need for qualified breeders since 2006 (Cengiz et al., 2015). Thus, the flock consisting of animals with parental registration is called "elite" and the herds that are not registered as "base flock". Birth and weaning weight (90th day) yield records are taken as selection criteria in both herds and the best females from base herds are transferred to elite flock. For this reason, genetic parameter estimations are important in terms of economic characteristics in sheep flock.

Today, the genetic parameter estimation methods of meat and milk characteristics in sheep and goats have changed with the developing technology, and such estimations can now be made at a more advanced level with the help of molecular markers. In addition, the importance of making future predictions according to an appropriate model in the evaluation of complex and large numbers of records containing many variables in many disciplines is increasing. Especially today, very rapid innovations and advances in software suitable for computer technologies have made significant contributions to the solution of such problems.

For example, decision trees are a data mining approach that is visually easy to understand and interpret and used to solve classification and prediction problems. Classification and Regression Trees (CART, CRT), one of the decision tree algorithms, is a non-parametric statistical method that was first developed by Breiman et al. (1984), using both categorical and continuous variables to solve classification and regression problems. If the dependent variable is categorical, it is called a method classification tree (CT), and if it is continuous, it is called Regression Trees-RT (Chang & Wang, 2006). The use of decision trees for estimation in sheep farming is becoming more and more common. Balta & Topal (2018), the effects of year, flock type, age of dam, gender, birth type, and lamb color on birth weight of Hemşin lambs were examined by regression tree analysis method. According to the regression tree diagram, the most important variables affecting the birth weight of Hemşin lambs were found as year, birth type, gender, flock type, and age of dam, respectively. Olfaz et al. (2019) determined the effect of gender,

Birth and flock type on birth and weaning weight in Karayaka sheep breed with CART and found that birth type was effective on weaning weight. Akbulut et al. (2022) examined production effects of year of yield and age of dam on birth type in a herd of Awassi using ANOVA, Logistic Regression, and Classification Tree methods and found these effects to be significant.

This study aims to show the importance of the decision trees method, one of the data mining methods, and to determine the environmental factors affecting the weaning weights of Awassi lambs with the CRT the method.

2. Material and Methods
2.1. Material
The animal material of the study was provided by T.C. Awassi breed lambs, which are included in the project named "Breeding of Awassi Sheep by Public Hands", which is a sub-project of the "Animal Breeding at Hands of the People Project" project supported by the General Directorate of Agricultural Research and Policies (TAGEM) of the Ministry of Agriculture and Forestry and carried out in Osmaniye province. Weaning weights of a total of 6326 Awassi lambs born 2019-2021 from 4108 dams aged between

24-60 months were used in the analysis in elite and base flocks.

2.2. Methods

In this study, the regression decision tree algorithm was used to estimate the effect of environmental factors, since weaning weights are of the continuous variable type. The root node is at the top of the decision tree. While creating the decision tree, the root node includes all the observations of the relevant independent variable in the data. The root node is divided into two sub-branches, and testing and branching continue sequentially at each node. In the tree, a node that is split into child nodes is called the parent node of the child nodes, while the child nodes are the children of a parent node. Each subsequent split occurs at a child node and if the operation does not occur at the end of a branch, a child node (decision node, intermediary node, split node) occurs at this point. If the classification is possible in a node, this indicates that there is a leaf node, or in other words, an end node (leaf node, terminal node) at the end of that branch. This end node is one of the classes to be determined on the data set. Accordingly, the definitions of nodes belonging to a decision tree can be summarized in Figure 1 below.
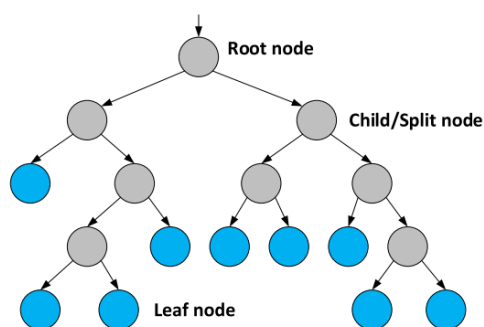

Figure 1. Illustration of a decision tree (Camana et al., 2019)
Şekil 1. Bir karar ağacı örneği (Camana et al., 2019)

In decision trees, some statistical approaches are used to place the data at the nodes of the tree. The regression tree uses least squares that minimize the residual sum of squares between the observation and the mean in each node (Breiman et al., 1984; De'ath & Fabricius, 2000).

In this study, CRT algorithm was preferred and calculations were made in this algorithm, since it was desired to reveal the more explanatory structure of this method compared to the traditional regression approach by using the decision tree model in the classification and regression problem.

In this study, the regression tree model was pruned using max depth parameter = 4 and all minimum cases were defined as 10 in the parent node and 5 in the child nodes. The CRT analysis was done using IBM SPSS Statistics Version 25 (IBM, 2020).

3. Results and Discussion

The regression tree was designed in the root (node 0), branch (nodes: 1, 4, and 6) and leaf (nodes: 2, 3, 5, 7, and 8) nodes, in this research (Figure 1).

In this study, the average weaning weight of Awassi lambs was found to be 24.471 ± 2.342 (Node 0). Similar findings were reported by Hızlı et al. (2022) with the comparison of different models for estimation of direct and maternal genetic parameters on body weights in Awassi sheep. However, it was found to be lower than the values reported by Gül & Ekici (2020), Özbeyaz et al. (2018), and Tunaz (2021), and higher than the results reported by Boran & Torun, (2018).

In this study, some environmental factors caused by this variation in weaning weight in Awassi were ranked according to their estimation importance with the regression tree analysis method. With the regression tree analysis method, the most effective environmental factor on the weaning age was estimated as the birth type. Therefore, the root node first branched out as a single (87.8%) and twin (12.2%) births according to birth type. The gender factor was found to be effective on the weaning weight of singleton Awassi lambs. The average weaning weight of singleton Awassi lambs was split into two child nodes according to gender, female (Node 3) and male (Node 4). The average weaning weights of female and male lambs were found almost the same as 24.30 ± 2.29 kg and 24.55 ± 2.33 kg, respectively.

There are many studies examining the effects of different environmental factors on growth characteristics in various sheep breeds. Behrem (2021), in Akkraman sheep; Supakorn & Pralomkarn (2012), on the other hand, found the effect of birth type and gender on the weaning weights of Thai native, Anglo-Nubian, Boer, and Saanen goats as significant, and they determined that the weights of singleton and male lambs were higher. Moreover, Sánchez-Dávila et al. (2015) the found lambing year had a significant effect on the weaning weight of Saint Croix hair sheep.

In the decision tree, weaning weights of male lambs were divided into two nodes (Node 5 and Node 6) according to flock type, elite and base flock. In this study, the average weaning weights of lambs for elite and base flock herds were approximately similar (Node

5:24.69 ± 2.37; Node 6: 24.29 ± 2.25). The cases in Node 6 (base flock) were then split according to age of the dam. It has been observed that the age of the dam is effective on the lambs in the base flock. Accordingly, the weaning weight of the lambs of dams older than 50 months and younger than 60 months of age (25.12 ± 2.32 kg) was 1 kg heavier than the average weight of lambs of dams 50 months of age and younger age (24.16 ± 2.21 kg). Babar et al., (2004) find out that higher average birth weight for Lohi lambs born by older dams. Balta & Topal (2018) also reported average birth weights for dams older than 3.5 years were found to be higher.
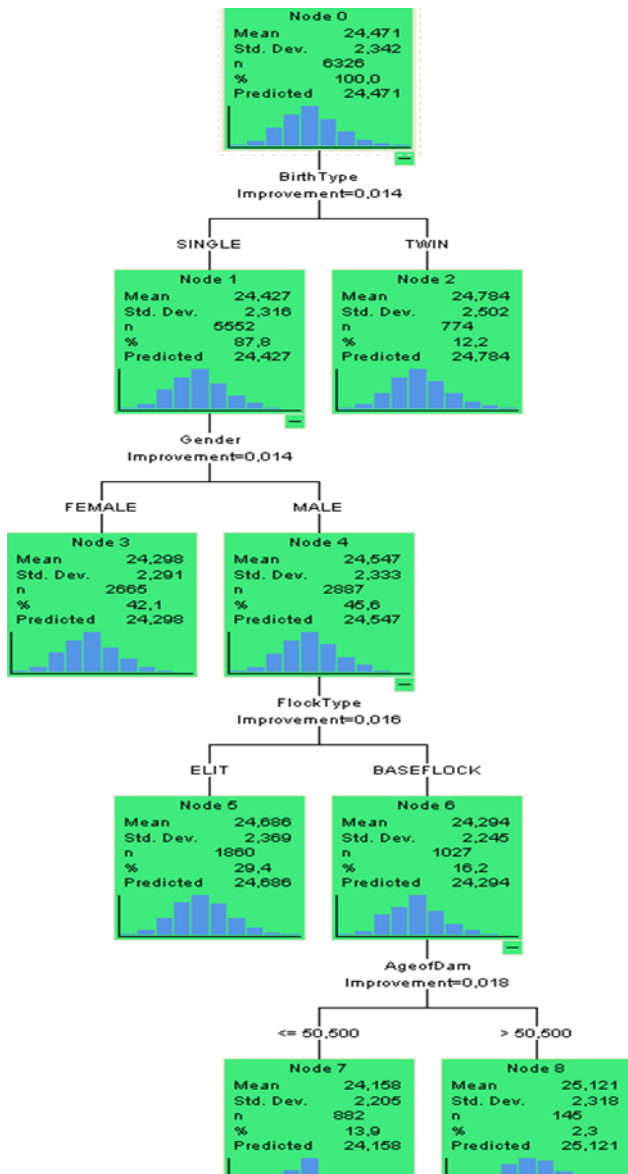

Figure 2. *Regression Tree Diagram for Weaning Weights*
Şekil 2. Sütten Kesim Ağırlıkları için Regresyon Karar Ağacı Diyagramı

The results of the importance of independent variables based on classification and regression trees (CRT) are presented in Table 1 and Figure 1,2,3.

Table 1. Importance of independent variables
Tablo 1. Bağımsız Değişkenlerin Önemleri

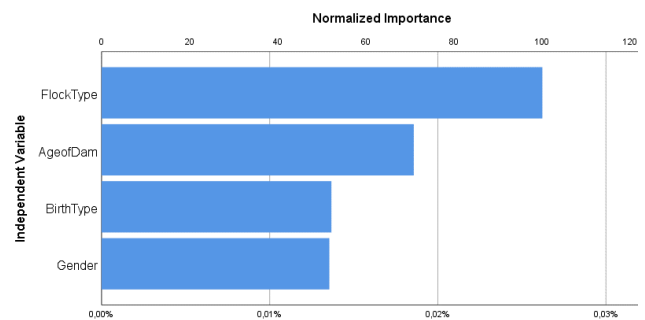| Variable	Importance	Normalized Importance	Risk Value	
			Estimate	Std. Error
Flock Type	0.026	100.0%	5.424	0.097
Age of Dam	0.019	70.8%		
Birth Type	0.014	52.2%		
Gender	0.014	51.7%		

Figure 3. Normalized importance levels of the factors
Şekil 3. Faktörlerin yüzde önem seviyeleri

As seen in Table 1 and Figure 1, the most effective variables for the weaning weight of Awassi lambs are flock type, age of the dam, birth type, and gender factors, respectively. It was determined that the least effective variable on weaning weight was gender, followed by the birth type. It was determined that the most effective factor was the type of rearing, and the second important factor was the age of the dam.

4. Conclusion

With the regression tree analysis method, the environmental factors affecting the weaning weight were estimated as birth type, gender, flock type, and age of dam according to their level of significant effect, respectively.

Environmental factors are as effective as genetic factors for quality meat and milk production in small ruminant farms. Meat production constitutes a large part of the income obtained from sheep breeding. For productivity survivability in sheep farms, high fertility, high viability, healthy and heavy lambs are needed in herds.

In sheep breeding studies, it is known that singleton, males, and lambs of older dams generally develop better than the others and their daily gain are higher. However, which of these factors is more

effective varies from breed to breed, care-feeding, and herd management. Awassi is breeds with higher milk yields than other domestic sheep breeds. Therefore, it can be thought that twin lambs of Awassi get enough colostrum Moreover, it is known that the birth weight of the lambs of the dams that gave birth more than once, increases according to the parity. For this reason, there is a positive relationship between parity and age of the dam.

On the other hand, accurate selection of sheep breeders is the most important condition increase the profit in the farms. Considering the results of this study; dams older than 50 months and younger than 60 months should be selected as breeders and transferred to elite flock. In conclusion, this study shows that regression decision trees are an important method for regression estimation in sheep breeding studies and are recommended.

References

Akbulut, Ö., Kaygısız, A., & Yılmaz, İ. (2022). A Comparative Research on Data Analysis with Factorial ANOVA, Logistic Regression and CHAID Classification Tree Methods. *Black Sea Journal of Agriculture*, 5(3), 314-322. http://doi.org/10.47115/bsagriculture.1087820

Babar, M.E., Ahmad, Z., Nadeem A. & Yaqoob, M. (2004). Environmental Factors Affecting Birth Weight in Lohi Sheep. *The Pakistan Veterinary Journal*, 24 (1): 5-8.

Balta, B., & Topal, M. (2018). Regression Tree Approach For Assessing The Effects of Non-Genetic Factors on Birth Weight of Hemşin Lamb. *Alinteri Journal of Agriculture Science*, 33(1), 65-73. http://doi.org/10.28955/alinterizbd.372671

Behrem, S. (2021). Effects of Environmental Factors Growth Traits of Akkaraman Sheep In Çankırı Province. *Livestock Studies* 61(1), 22-27. http://doi.org/10.46897/livestockstudies. 610104

Boran, Ö., & Torun, O. (2018). Çukurova Üniversitesi Ziraat Fakültesi Araştırma ve Uygulama Çiftliğinde yetiştirilen İvesi ve Çukurova Et Koyunlarının Döl Verimi İle Kuzularının Büyüme Performansının Araştırılması. *Çukurova Üniversitesi Fen ve Mühendislik Bilimleri Dergisi*, 36(9), 107-116.

Breiman, L., Friedman, J.H., Olshen, R., & Stone, A.C.G. (1984). Classification and Regression Trees. *Wadsworth International Group*, Belmont, California, USA.

Camana A. M. R., Ahmed, S., Garcia C. E., & Koo, I. (2020). Extremely Randomized Trees-Based Scheme for Stealthy Cyber-Attack Detection in Smart Grid Networks, The Institute of Electrical and Electronics Engineers, 8, 19921-19933, https://doi.org/10.1109/ACCESS.2020.2968934.

Cengiz, F., Karaca, S., Kor, A., Ertuğrul, M., Arık, İ. Z., & Gökdal, Ö. (2015). Küçükbaş Hayvan Yetiştiriciliğinde Değişimler ve Yeni Arayışlar. *Türkiye Ziraat Mühendisliği VIII. Teknik Kongresi Bildiriler Kitabı Sayı-2*, Ankara.

Chang, L.Y., & Wang, H.W. (2006) Analysis of Traffic Injury Severity: An Application of Non-Parametric Classification Tree Techniques. *Accident Analysis & Prevention*, 38, 1019-1027. http://dx.doi.org/10.1016/j.aap.2006.04.009

De'ath, G., & Fabricius, K.E. (2000). Classification And Regression Trees: A Powerful Yet Simple Technique for Ecological Data Analysis. *Ecology*, 81(11): 3178-3192. https://doi.org/10.1890/0012-9658(2000)081[3178:CARTAP]2.0.CO;2

Epstein, H. (1982). Awassi Sheep. *World Animal Review*. FAO, 44, 9-18.

Ertuğrul, M., Akın, A. O., Yıldırır, M., Dellal, G., Togan, İ., Pabuçcuoğlu, S., & Özder, M. (2014). Türkiye Çiftlik Hayvanları Genetik Kaynaklarının Korunması ve Sürdürülebilir Kullanımı. Türkiye Ziraat Mühendisliği *VIII. Teknik Kongresi Bildiriler Kitabı*, Sayı-1, 212.

FAO. (2022). Crops and Livestock Products, *https://www.fao.org/faostat/en/#data/QCL (Erişim tarihi:29.08.2022).*

Gül, S., & Ekici, H. (2020). İvesi Koyunlarında Farklı Yaşta Sütten Kesimin Kuzularda Büyüme ve Süt Verimi Üzerine Etkisi. *Hayvan Bilimi ve Ürünleri Dergisi*, 3(2), 95-103.

Hızlı, H., Takma, Ç., & Yazgan, E. (2022). Comparison of Different Models for Estimation of Direct and Maternal Genetic Parameters on Body Weights in Awassi Sheep, Archives Animal Breeding, 65, 121-128, https://doi.org/10.5194/aab-65-121-2022

IBM: SPSS Software v25 (2020). *https://www.ibm.com/analitics/spss-statistic-softwar*e

Kaymakçı, M. (2013). İleri Koyun Yetiştiriciliği. Genişletilmiş Dördüncü Baskı Bornova- İzmir 2013. *Meta Basım Matbaacılık*Olfaz, M., Tırınk, C., & Önder, H. (2019). Use Of CART and CHAID Algorithms in Karayaka Sheep Breeding. *Kafkas Universitesi Veteriner Fakültesi Dergisi*, 25(1): 105-110, http://doi.org/10.9775/kvfd.2018.20388

Özbeyaz, C., Bilgiç, Ö. F., & Kocakaya, A. (2018). Eskişehir de Yetiştirici Koşullarındaki İvesi Koyunlarında Bazı Özelliklerin İncelenmesi.

Lalahan Hayvancılık Araştırma Enstitüsü Dergisi, 58(1), 1-6.

Sánchez-Dávila, F., Bernal-Barragán, H., Padilla-Rivas, G., del Bosque-González, A.S., Vázquez-Armijo J.F. & Ledezma-Torres, R.A., (2015). Environmental Factors and Ram Influence Litter Size, Birth, And Weaning Weight in St Croix Hair Sheep Under Semi-Arid Conditions In Mexico. *Tropical Animal Health and Production*, 47(3), 825-31. https://doi.org/10.1007/s11250-015-0795-6.

Supakorn, C., & Pralomkarn, W. (2012). Genetic Parameter Estimates for Weaning Weight and Kleiber Ratio in Goats. *Songklanakarin Journal of Science & Technology*, 34(2), 165-172.

TUİK (2022). *Hayvansal Üretim İstatistikleri*, Haziran 2022 https://data.tuik.gov.tr/Bulten/Index?p=Hayvansal-Uretim-Istatistikleri-Haziran-2022-45594 (Erişim Tarihi: 29.08.2022)

Tunaz, A. T. (2021). Comparison of Selected Decision Tree Algorithms in The Prediction of Body Weight in Awassi Lambs. *JAPS: Journal of Animal & Plant Sciences*, 31(4), 944-953. https://doi.org/10.36899/JAPS.2021.4.0288