Bulgarian Journal of Meteorology and Hydrology

Национален институт по метеорология и хидрология

National Institute of Meteorology and Hydrology

Bul. J. Meteo & Hydro 26/1 (2022) 1-23

Fog and horizontal visibility forecasting with stochastic models

Neyko Neykov*, Anastassia Stoycheva, Ilian Gospodinov, Orlin Gueorguiev, Plamen Neychev, Kiril Slavov

National Institute of Meteorology and Hydrology,
Tsarigradsko shose 66, 1784 Sofia, Bulgaria

Abstract: The paper discusses stochastic models for forecasting horizontal visibility in categories and fog occurrence at the NIMH central meteorological station (CMS), Sofia. Data for the period from 01.01.2001 to 31.12.2016 of the following meteorological predictors such as temperature, dew point, relative humidity, wind direction and speed, pressure, fog and visibility measured at CMS of NIMH, Sofia and at the high-mountain MS on Cherni Vruh, their lags for the past 8 synoptic terms, as well as radiosonde observations measured at 12:00 GMT at CMS of NIMH, Sofia such as temperature, relative humidity, wind direction and speed at standard pressure levels 925, 850 and 700 hPa, are used to predict fog events and visibility in categories given in meters. The probabilities of fog occurrence and visibility are modeled using the binary and ordinal logistic regression models. The occurrence probability forecasts are turned into binary forecasts of fog event using an optimal cutoff value based on the standard ROC curve analysis. The maximum posterior principle is used in converting the probabilistic horizontal visibility forecasts into categorical forecasts. The cross-validation technique is used in order to assess the models performances. The results show that 3, 6 and 12 hours fog and visibility forecasts based on test data sets are comparable with those based on the training data sets.

Keywords: fog forecast, horizontal visibility forecast, binary and ordinal logistic regression.

* Neyko.Neykov@meteo.bg

Прогнозиране на мъгла и намалена хоризонтална видимост със стохастични модели

Нейко Нейков, Анастасия Стойчева, Илиан Господинов, Орлин Георгиев, Пламен Нейчев, Кирил Славов

Национален институт по метеорология и хидрология
бул. Цариградско шосе 66, 1784 София

Резюме: В статията са разгледани стохастични модели за прогнозиране на хоризонтална видимост, дефинирана в категории и поява на мъгла в централната метеорологична станция (ЦМС) на НИМХ, гр. София. Използвани са данни за периода от 01.01.2001 до 31.12.2016 г. на следните метеорологични предиктори - температура, точка на оросяване, относителна влажност, скорост и посока на вятъра, налягане, мъгла, хоризонтална видимост, измерени в основните синоптични срокове в ЦМС на НИМХ, гр. София и във високо-планинската метеорологична станция на Черни връх, и техни лагове за последните 8 синоптични срока, както и данни от метеорологичния сондаж, провеждан в 12:00 GMT в ЦМС, гр. София като температура, относителна влажност, скорост и посока на вятъра на стандартните нива 925, 850 и 700 hPa. Вероятностите за поява на мъгла и хоризонталната видимост в категории са моделирани с бинарна и ординална логистични линейни регресии. Предсказаните стойности на вероятността за поява на мъгла се превръщат в бинарна прогноза на събитието поява на мъгла във всеки срок с помощта на оптимална прагова стойност, определена със стандартния ROC анализ. Принципът на максималната апостериорна вероятност е използван за трансформирането на оценените вероятности за прогноза на хоризонтална видимост в категорийна прогноза във всеки синоптичен срок. Техниката на крос-валидирането е използвана за оценка на ефективността на моделите. Резултатите от моделите показват, че прогнозите за мъгла за 3, 6 и 12 часа напред и хоризонтална видимост, основани на тестовите извадки от данни са сравними с тези, на базата на данните от обучаващите извадки.

Ключови думи: прогноза на мъгла, прогноза на хоризонтална видимост, бинарна и ординална логистични регресии.

1. ВЪВЕДЕНИЕ

1.1. Актуалност на темата

Прогнозирането на мъглата като атмосферно явление в приземния слой е актуална задача с определено социален и икономически характер. Издаването на точни прогнози за появата и продължителността на мъглата е от изключително значение за безопасността на въздушния и наземния транспорт. Безспорно, най-заинтересуваните в това отношение са летищните метеорологични служби

поради значителните загуби около затрудненията на въздушния транспорт при продължителни обстановки с мъгла, както и закъсненията, поради краткотрайни, но гъсти радиационни мъгли. От съществено значение за безопасността на сухоземния транспорт е установяване на районите с мъгла, както и свръхкраткосрочна прогноза, която би могла да се използва в популярните и полезни приложения за състоянието на пътищата, трафика в големите градове и по магистралите. В допълнение към транспортните задачи, мъглата има своето влияние и върху замърсяването на градската среда с фини прахови частици с различен размер. Поради това ранното информиране на обществеността чрез прогнози за поява, продължителност на мъгла и намалена видимост е актуална задача, важна за обществото и има ясно изразена социална и икономическа полза.

1.2. Състояние на изследванията

За прогнозиране на мъгла в оперативната дейност на НИМХ се използва псевдопотенциалната температура (ППТ), като критерий за статична устойчивост на атмосферата и при влажен въздух с наситена водна пара. Големите положителни вертикални градиенти на псевдопотенциалната температура в даден слой означават силна устойчивост на стратификацията в него. Мъглите в Софийско поле се характеризират с покриваща температурна инверсия. Тази инверсия, съвместно с разпределението на влажността, създава един силно устойчиво стратифициран, покриващ мъглата, задържащ слой. Този слой би трябвало да се идентифицира ясно във вертикалния профил на псевдопотенциалната температура, а с това и вертикалната мощност на мъглата. Съществените промени във времевите редици на ППТ дават много добра информация за поведението на мъглата, особено за нейното разсейване.

Съществуват числени модели на граничен слой с висока разделителна способност, които имат за цел да симулират процесите на образуване и развитие на мъгла и ниска слоеста облачност. През последните години, във връзка с постоянното увеличаване на разделителната способност, в числените модели за прогноза на времето се внедряват числени схеми с явно описание на процесите на формиране и еволюция на хидрометеорите, с което числените модели прогнозират все по-успешно параметрите за наличие на мъгла и хоризонтална видимост (Bengtsson et al., 2017). Въпреки това развитието на стохастични модели за краткосрочна прогноза с използване на локални предиктори остава подход за постигане на висока успеваемост.

Географското разположение на София и други големи български градове и наличието на високопланински метеорологични станции позволява бърз анализ на условията на инверсия като причина за образуването на мъгла и може да служи за основа на аналитични и стохастични модели за свръхкраткосрочна прогноза на този процес. Например, Стойчева и Евтимов (2014, 2015) и Stoycheva and Evtimov (2014), предлагат използването на индекс на устойчивост Черни връх-София.

Отчитайки степента на статична устойчивост на атмосферния слой над София, този индекс зависи само от две температури - температура на 2 m в централната метеорологична станция на НИМХ в Младост 1, София и температурата, измерена във високопланинската обсерватория Черни връх, което го прави полезен за оперативната практика.

Моделирането на вероятностите за поява на мъгла и намалена видимост са предмет на интензивни изследвания през последните години. В публикациите на Grace and Ferriere (2001), Stern and Parkyn (2001), Forthun et al (2006), Zazzaro et al. (2010), Fabbian et al. (2007), Chmielecki and Raftery (2011), Bartok et al. (2011), Hacker et al. (2013), Boneh et al. (2015), Colabone et al. (2015) и на други автори се разглеждат модели на вероятността за поява на мъгла, основани на логистичната регресия, но с различни метеорологични предиктори. Например, Hacker et al. (2013) използват за предиктор индекса на устойчивост, пресметнат с помощта на приземната температура и точката на оросяване, температурата и скоростта на вятъра на 850 hPa от сондаж, докато Stern and Parkyn (2001) използват градиенти на налягането, приземна температура и точка на оросяване на летището в Мелбърн.

Ординалната логистична регресия е използвана от Kneringer et al. (2019) и Dietz et al. (2019) за моделиране на вероятностите за прогноза на намалената хоризонтална видимост в четири категории до 350 m, до 650 m, до 1000 m и над 1000 m на летището във Виена. Прогнозата чрез тези вероятности е от 30 минути до 2 часа напред, като се обновява на всеки 30 минути. За предиктори са използвани стандартни приземни метеорологични елементи като относителна влажност, точка на оросяване, височина на долната граница на облачността, скорост на вятъра, северна и югоизточна посоки на вятъра, всяка от тях дефинирана като фактор с две нива, съответно 1 и 0, валеж преди 12 часа, дефиниран като фактор с две нива, съответно 1 и 0, съответни на регистриран и нерегистриран валеж, разликата между температура на повърхностната на почвата и температурата на 2 m, разликата между температурата на 2 m и температурата на височината на наблюдателната кула на летището във Виена и слънчев зенитен ъгъл.

Други автори използват голям набор от метеорологични параметри за модели на вероятностите на видимост от станции и измрвателни уреди в района на техните изследвания. Например, около летище Мюнхен, Müller et al. (2007) предлагат модел на видимостта, основан на ординална логистична регресия със следните метеорологични предиктори - хоризонтална видимост от инструментални и окомерни измервания от летището, среден интензитет на валеж, вид на валеж, температура на почвата на различни дълбочини, влагосъдържание в почвата, температура на земната повърхност, температура на въздуха на определена стандартна височина, данни за вертикален профил на граничния атмосферен слой от микровълнов уред или по данни от станции от съседни хълмове: температура, вятър, влажност, височина на долна граница на облачността, количество облачност, мощност на дълговълновата радиация,

достигаща земната повърхност, мощност на късовълновата радиация, достигаща земната повърхност, мощност на дълговълновата радиация, излъчвана от земната повърхност и данни от радиосондаж на атмосферата от най-близка аерологична станция. Като предиктори са използвани още вертикален профил на граничния слой на няколко нива, конструиран по данни от наземни станции, разположени по околни на летището хълмове.

От литературната справка се вижда, че в световен мащаб е натрупан положителен опит в прогнозирането на мъгла и намалена видимост със стохастични модели. Тези модели се основават на методологията на бинарната и ординалната логистични регресии, в които се включват подходящи метеорологични предиктори. Оценяването и проверката на хипотези за значимост на предиктори в тези модели се основава на метода на максималното правдоподобие Agresti (2001) и Wood (2017). Наличието на софтуер за моделиране на бинарни и ординални данни, съответно с Бернулиевото и полиномното (мултиномното) разпределния, предоставят възможности за рутинното им приложение в практиката.

1.3. Цел и задачи

Основни цели на настоящата статия са създаване на стохастични прогностични модели за поява на мъгла за период от 3, 6 до 12 часа напред и прогностичен модел на хоризонтална видимост за период от 3 часа напред, като се използва метеорологичната информация в синоптичните срокове в 00, 03, 06, 09, 12, 15, 18 и 21 GMT, и от аерологичен сондаж в 12 GMT. От вероятностна гледна точка: а) явлението мъгла се интерпретира като бинарна случайна величина, приемаща стойност 1, ако е регистрирана мъгла в даден срок и 0 в противен случай; б) хоризонталната (намалената) видимост се интерпретира като ординална (наредени) категорийна случайна величина със следните условни категории на видимост за модели с 3 и 5 категории, дефинирани както следва: а) видимост "до 2000 m", „между 2000 и 10000 m" и „над 10000 m"; б) видимост „до 50 m", „между 50 и 500 m", „между 500 и 2000 m", „между 2000 и 20000 m" и „над 20000 m". Моделирането на разпределенията на времевите редици от бинарни и ординални данни се основава на динамичните бинарна и ординална логистични регресии с използването на метеорологични предиктори от измерванията в наблюдателната мрежа на НИМХ в синоптичните срокове, техни лагове, включително лагове на мъглата и видимостта, за да бъде отчетена динамиката на процесите.

2. ОПИСАНИЕ НА ДАННИТЕ

В изследването са включени само дните от студените полугодия на периода 2001-2016 г. Използвани са данни от: 1) централната метеорологична станция на НИМХ-София и високопланинската метеорологична станция на Черни връх: температура (T) на 2 m, температура на точката на оросяване (Td), относителна влажност на

въздуха (Rh), скорост на вятъра (ws), атмосферно налягане (p), мъгла (fog) и хоризонтална видимост (hvis) в основните и междинните синоптични срокове (00, 03, 06, 09, 12, 15, 18 и 21 GMT), и техни лагове от предходните 8 синоптични срока; 2) аерологичен сондаж в 12 GMT на стандартните нива (925, 850 и 700 hPa): температура (T), относителна влажност (Rh), скорост (ws) и посока (wd) на вятъра, и техния лаг от предходния ден. Общият брой на сроковете е 23328.

Да означим с $T_t^{Sf}, Td_t^{Sf}, Rh_t^{Sf}, ws_t^{Sf}, p_t^{Sf}, fog_t^{Sf}$ и $hvis_t^{Sf}$ приземните метеорологични предиктори в текущия срок t в станцията на НИМХ-София, докато съответните им аналози в станция Черни връх означим с $T_t^{Ch}, Td_t^{Ch}, Rh_t^{Ch}, ws_t^{Ch}, p_t^{Ch}, fog_t^{Ch}$ и vis_t^{Ch}. Лаговете на метеорологичните предиктори в предходните на текущия срок t са формирани по техните стойности в сроковете t-1, t-2 ,…, t-8, съответно. За предикторите от аерологичния сондаж на ниво 925 hPa са използвани означенията $T_t^{925}, Rh_t^{925}, u_t^{925}$ и v_t^{925}, където u_t^{925} и v_t^{925} са компонентите на вятъра, пресметнати по стойностите на ws и wd. Аналогични означения са въведени за предикторите на нивата 850 hPa и 700 hPa. Към вектора на приземните метеорологични предиктори в срок t и техните лагове се добавят стойностите на предикторите от аерологичните сондажи от предходния ден, ако текущят срок t е по-малък от 15 часа, в противен случай се добавят стойности на тези предиктори от аерологичния сондаж в 12 GMT. По този начин, броят на потенциалните предиктори и техните лагове, включени в изследването за прогноза на мъгла и хоризонтална видимост със стохастични модели е 124.

3. МОДЕЛИРАНЕ НА ВЕРОЯТНОСТИТЕ ЗА ПОЯВА НА МЪГЛА

3.1. Моделиране на вероятността за поява на мъгла

Нека J_t е случайна бинарна величина, приемаща стойност $j_t = 1$ или $j_t = 0$, ако е регистрирана или не е регистрирана мъгла в синоптичния срок t, за t=1,2,...,L, където L е броят на сроковете, включени в изследването. Нека $Z_t = (z_{t1}, \dots, z_{tk})^T$ е вектор от метеорологични елементи, характеризиращи времето в срок *t*, докато

$$x_t^T = \left(1, j_{t-1}, \dots, j_{t-l}, z_{t1}, \dots, z_{tk}, z_{t-1,1}, \dots, z_{t-l,1}, \dots, z_{t-1,k}, \dots, z_{t-l,k}\right)^T$$

е вектор, състоящ се от лагове на j_t и z_t от ред 1,2,...,*l*, характеризиращи синоптичните обстановки в предходните *l* срока.

Моделирането на вероятността $p_t(x_t) = Pr(J_t = 1|x_t)$ за поява на мъгла при метеорологичните условия x_t се основава на методологията на динамичната логистична авторегресия. За модел на тази вероятност най-често се използва стандартното логистично разпределение, което се дефинира чрез израза

$$p(x_t^T\beta) = \frac{exp(x_t^T\beta)}{1+exp(x_t^T\beta)}, \quad (1)$$

където x_t^T е вектора от метеорологичните предиктори, дефиниран по-горе, $\beta = \left(\beta_0, \beta_1 \cdots, \beta_p\right)^T$ е вектор от неизвестните коефициенти, отчитащи влиянието (приноса, тежестта) на съответните предиктори, скаларното произведение $x_t^T\beta$ се нарича линеен предиктор.

Пример за линеен предиктор, включващ температурата на въздуха на 2 m, точката на оросяване и относителната влажност в станцията на НИМХ-София в синоптичните срокове t-1 и t-2, които са част от общия брой предиктори, използвани в модела на вероятността за поява на мъгла, се дефинира с израза:

$$x_t^T\beta = \beta_0 + \beta_1 \mathrm{T}_{t-1}^{Sf} + \beta_2 T_{t-2}^{Sf} + \beta_3 Td_{t-1}^{Sf} + \beta_4 Td_{t-2}^{Sf} + \beta_5 Rh_{t-1}^{Sf} + \cdots$$

където t=3,…,L, а L=23328 е броят на синоптичните срокове през студеното полугодие за периода 2001-2016 г.

Понеже времевият ред на появата на мъгла за всеки срок t е Бернулиево разпределен, то оценката на неизвестния коефициент β се определя по метода на максималното правдоподобие (ММП). За целта се търси максимум на функцията на правдоподобие (ФП), дефинирана по данните

$$\max_{\beta} \prod_{t=l+1}^{L} \left(1 - p(x_t^T\beta)\right)^{1-j_t} \left(p(x_t^T\beta)\right)^{j_t},$$

където j_t е индикатор, дефиниран по-горе и приемащ стойност 1 или 0, l е максималният времеви лаг, използван в предиктора x_t^T.

За максимизирането на ФП могат да се използват пакетите glm, mgcv и много други от безплатната програмна среда R (R Core Team 2021). За избор на конкуриращи се модели върху едни и същи данни се използват информационните критерии AIC и BIC на Акайке или Шварц, съответно. Статистическият тест, основан на отношението на правдоподобие, се използва за определяне на значими предиктори в йерархично вложени конкуриращи се модели за определяне на оптимален модел за $p(x_t)$. Един предиктор е статистически значим, ако оценката на съответния му коефициент в линейния предиктор е статистически значим.

3.2. Разкриване на зависимости в бинарна логистична регресия за поява на мъгла

Стандартен подход за моделиране на вероятността за поява на мъгла е чрез използването на логистична линейна регресия. Вместо използването на линеен предиктор като аргумент на логистичната вероятност е прието да се използва по-обща методология, която позволява разкриването и отчитането на нелинейни

зависимости на предикторите в модела на вероятността. За целта се използват кубични сплайни с неизвестни коефициенти от предикторите, B-, P- или други сплайн-трансформации, Wood (2017). По този начин се постига една по-добра и гъвкава апроксимация на моделите към данните. За разкриването на нелинейни зависимости в модела на вероятността за поява на мъгла със сплайн функции сме използвали процедурата gam от пакета mgcv на програмната среда R, предложен от Wood (2017). Резултатите от анализа са дадени в Таблица 1, която съдържа информация за качеството на модела. В първите 3 реда на тази таблица са дадени оценките на свободния коефициент (intercept) и коефициентите пред лаговете на предикторите за поява на мъгла в предходните срокове, за да бъде отчетена динамиката на процеса. Оценките на сплайн функциите от съответните метеорологични предиктори са статистически значими със стандартните нива на съгласие, според теста на отношенията на правдоподобие. По технически причини не се разпечатват оценките на коефициентите на сплайн функциите от предикторите.

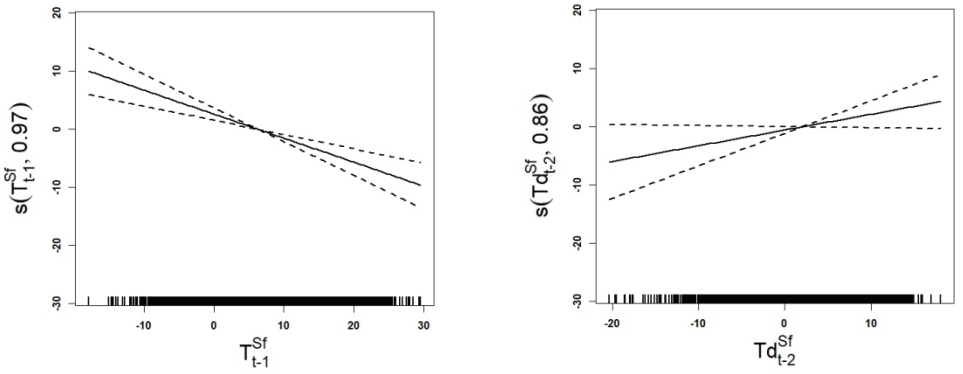
Table 1. GAMs binary logistic spline regression model for 3 hours ahead fog forecast based on training data set within the cold half period 2001-2016

Таблица 1. Бинарен логистичен регресионен модел за 3-часова прогноза на мъгла през студените полугодия на периода 2001-2016 г., базиран на сплайн функции от метеорологични предиктори.

	Estimate	Std.Error	z-value		Pr(>|z|)	Signif
(Intercept)	-11.4553	1.8082	-6.335		2.37e-10	***
fog_{t-1}^{Sf}	3.0222	0.2629	11.497		<2e-16	***
fog_{t-2}^{Sf}	1.1934	0.2788	4.280		1.87e-05	***
	Approximate significance of smooth terms:					
	edf	Ref.df	Chi.sq		p-value	
$s(T_{t-1}^{Sf})$	2.635549	9	60.217		<2e-16	***
$s(Td_{t-1}^{Sf})$	2.478318	9	62.076		<2e-16	***
$s(Td_{t-2}^{Sf})$	2.693737	9	12.175		0.000713	***
	…					
$s(u_{t-1}^{925})$	1.761568	9	9.067		0.001925	**
	Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1					
	Deviance explained = 78.3%					

На Фигура 1 с непрекъсната черна линия са представени емпиричните функционални зависимости в логистичния модел на вероятността за поява на мъгла и сплайн-функции от температурата на 2 m (a) и относителната влажност

(b). Вертикалните чертички на оста Ox съответстват на стандартизираните стойностите на тези метеорологични предиктори. От графиката на плот (a) се вижда, че зависимостта на сплайн предиктора от температурата при 2 m е обратно пропорционална. Следователно вероятността за поява на мъгла нараства с понижаване на температурата на 2 m. От графиката на плот (b) се вижда, че зависимостта на вероятността за поява на мъгла е право пропорционална на нарастването на температурата на точката на оросяване. Аналогични резултати са получени за останалите предиктори от Таблица 1, поради което няма да бъдат разглеждани. Този модел на вероятността на 3-часова прогноза за поява на мъгла със сплайн функции от метеорологичните предиктори обяснява 78.3% от дисперсията на данните. Подобни резултати са получени за моделите на 6- и 12- часова прогноза на вероятността за мъгла, които обясняват 69.3% и 58.1% от дисперсията на данните.



(a) (b)

Fig. 1. Empirical relationship between GAMs binary logistic spline regression model for fog forecast of standardaized (a) temperature of 2 m and (b) dew point at NIMH- Sofia station for the cold half period 2001-2016

Фиг. 1. Емпирични зависимости в логистичния регресионен модел за прогноза на мъглата от стандартизираните стойности на (a) температурата на 2 m и (b) температура на точка на оросяване в студените полугодия през периода 2001-2016 г. в станция НИМХ-София.

3.3. Бинарна логистична регресия за поява на мъгла с линейни предиктори

В този параграф ще бъдат разгледани стохастичните модели на вероятността за поява на мъгла за 3-, 6- и 12-часова прогноза чрез използване на линейна логистична регресия. Основание за това ни дават резултатите за линейни емпирични зависимости на големия брой предиктори от предишния параграф. За

моделиране на вероятността за поява на мъгла за целите на 3-часовата прогноза са включени само метеорологични предиктори с лагове от 1 до 8 от станциите на НИМХ-София и на Черни връх, както и измерванията от аерологичните сондажи на трите стандартни нива от предходния ден, описани в параграф 2. Видимостта ($hvis$) не е включена като предиктор в този модел. Аналогично за моделиране на вероятността за поява на мъгла за 6- и 12-часова прогноза са използвани само метеорологичните предиктори с лагове от 2 до 8 и от 4 до 8, съответно. Следователно, борят на потенцилните предиктори в 3-, 6- и 12-часовите модели на вероятността за поява на мъгла е 96, 84 и 72, съответно.

Голям брой от включените метеорологични предиктори в 3-часовия модел на вероятността за поява на мъгла се оказаха статистически незначими, според тестовете за отношението на правдоподобие (Таблица 2), при стандартните нива на съгласие. В резултат на използването на стъпковата процедура stepAIC от библиотеката MASS за избор на статистически значими предиктори с помощта на информационния критерий на Бейс (BIC) се достигна до редуциран модел със следния линеен предиктор:

$$\begin{aligned} x_t^T\hat{\beta} = -68.03 &- 0.99\,\mathrm{T}_{t-1}^{Sf} + 0.57\mathrm{T}d_{t-1}^{Sf} + 0.41\mathrm{T}d_{t-2}^{Sf} - 0.04\mathrm{T}d_{t-3}^{Sf} - 0.11\mathrm{T}d_{t-4}^{Sf} \\ &- 0.26p_{t-1}^{Sf} + 0.12p_{t-2}^{Sf} - 0.54ws_{t-1}^{Sf} - 0.33ws_{t-2}^{Sf} - 0.12ws_{t-3}^{Sf} \\ &- 0.81ws_{t-3}^{Sf} + 0.01p_{t-1}^{Ch} + 0.26p_{t-2}^{Ch} + 0.53fog_{t-1}^{Sf} + 0.60fog_{t-2}^{Sf} \\ &+ 0.67fog_{t-3}^{Sf} + 0.05Rh_{t-8}^{925} + 0.02Rh_{t-8}^{850} - 0.01Rh_{t-8}^{700} + 0.02T_{t-8}^{925} \\ &+ 0.30T_{t-8}^{850} - 0.21T_{t-8}^{700} + 0.11u_{t-8}^{925} - 0.03v_{t-8}^{925} \end{aligned}$$

Стойността на информационните критерии BIC на модела, включващ всички предиктори в изследването, е равна на 1914.17, докато стойността на редуцирания модел с минимален BIC, достигнат със стъпковата процедура stepAIC е 1108.4. Стойностите на функциите на отклоненията на модел без предиктори (NULL deviance), съдържащ само свободен коефициент, и на редуцирания модел са равни на 3682.5 и 962.34, съответно. Тези стойности са дадени в първи и предпоследен ред на пета колона от Таблица 2, така наречената таблица на отклоненията. В трета колона на таблицата са дадени стойностите на отклоненията (Deviance), с които всеки предиктор на модела редуцира, стойността на функцията на отклонение на модела без предиктори. В шестата колона на таблицата са дадени стойностите на вероятностите (tail probability), съответни на квантилите на тестовете на отношенията на правдоподобие за проверка на хипотезата за значимост на оценките на коефициентите, съответни на метеорологичните предиктори. Понеже тези стойности са твърде малки, то хипотезата за незначимост на съответните оценки на коефициентите на модела се отхвърля, с изключение на предикторите относителна влажност, Rh_{t-8}^{700}, и v компонентата на вятъра, v_{t-8}^{925}, на нива 700 и 925 hPa в предходния ден t-8. Останалите метеорологични предиктори в редуцирания модел са статистически значими. Частта на обяснената стойност на отклоненията, от метеорологичните предиктори е равна на 1-(962.34/10598)/(3682.5/10621)=0.74,

т.е. предикторите в редуцирания модел обясняват 74% от стойността на функцията на отклонение на модела без предиктори.

Table 2. Deviance table of 3 hours ahead fog forecast binary logistic reduced regression model based on training data set within the cold half period 2001-2016.
Таблица 2. Таблица на отклоненията на 3-часов бинарен логистичен линеен редуциран регресионен модел за поява на мъгла, базиран на обучаващата извадка от данни за студените полугодия на периода 2001-2016 г.

	Df	Deviance	Resid.Df	Resid.Dev	Pr(>χ^2)	Signif.
Null			10621	3682.50		
T_{t-1}^{Sf}	1	667.82	10620	3014.68	<2.2e-16	***
Td_{t-1}^{Sf}	1	878.39	10619	2136.29	<2.2e-16	***
Td_{t-2}^{Sf}	1	143.63	10618	1992.66	<2.2e-16	***
Td_{t-3}^{Sf}	1	122.29	10617	1870.37	<2.2e-16	***
Td_{t-4}^{Sf}	1	86.57	10616	1783.80	<2.2e-16	***
p_{t-1}^{Sf}	1	53.36	10615	1730.44	2.771e-13	***
p_{t-2}^{Sf}	1	20.23	10614	1710.21	6.880e-06	***
ws_{t-1}^{Sf}	1	91.09	10613	1619.12	<2.2e-16	***
ws_{t-2}^{Sf}	1	14.10	10612	1605.02	0.0001736	***
ws_{t-3}^{Sf}	1	38.23	10611	1566.79	6.300e-10	***
p_{t-1}^{Ch}	1	124.82	10610	1441.97	<2.2e-16	***
p_{t-2}^{Ch}	1	17.30	10609	1424.67	3.192e-05	***
fog_{t-1}^{Sf}	1	355.49	10608	1069.18	<2.2e-16	***
fog_{t-2}^{Sf}	1	12.77	10607	1056.41	0.0008063	***
fog_{t-3}^{Sf}	1	9.92	10606	1046.49	0.0028142	**
Rh_{t-8}^{925}	1	31.22	10605	1015.27	4.892e-07	***
Rh_{t-8}^{850}	1	3.50	10604	1011.77	0.0613528	.
Rh_{t-8}^{700}	1	0.29	10603	1011.48	0.5881897	
T_{t-8}^{925}	1	5.50	10602	1005.98	0.0189702	*
T_{t-8}^{850}	1	10.76	10601	995.22	0.0010376	**
T_{t-8}^{700}	1	16.53	10600	978.69	4.780e-05	***
u_{t-8}^{925}	1	15.76	10599	962.93	0.0003542	***
v_{t-8}^{925}	1	0.59	10598	962.34	0.4439174	
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1						
Deviance explained = 74%						

Аналогични резултати са получени за 6- и 12-часовите модели на вероятността за прогноза на мъгла, отклоненията на които са дадени в Таблици 3 и 4, съответно. Частта на обяснените стойности на отклоненията, вследствие на метеорологичните предиктори на тези модели, е съответно равна на

1-(1219.03/10603)/(3682.5/10621)=0.67 и 1-(1617.2/10604)/(3682.5/10621)=0.56.

Table 3. Deviance table of 6 hours ahead fog forecast binary logistic reduced regression model based on training data set within the cold half period 2001-2016.

Таблица 3. Таблица на отклоненията на 6-часов бинарен логистичен линеен редуциран регресионен модел за поява на мъгла, базиран на обучаващата извадка от данни за студените полугодия на периода 2001-2016 г.

	Df	Deviance	Resid.Df	Resid.Dev	Pr(>Chi)	
NULL			10621	3682.2		
T_{t-2}^{Sf}	1	632.86	10620	3049.34	<2.2e-16	***
T_{t-8}^{Sf}	1	19.83	10619	3029.51	8.465e-06	***
ws_{t-3}^{Sf}	1	358.40	10618	2771.11	<2.2e-16	***
p_{t-2}^{Sf}	1	16.27	10617	2754.84	5.488e-05	***
p_{t-4}^{Sf}	1	1.06	10616	2753.78	0.304239	
p_{t-5}^{Sf}	1	3.62	10615	2750.16	0.057107	.
fog_{t-2}^{Sf}	1	1073.58	10614	1676.58	<2.2e-16	***
p_{t-3}^{Ch}	1	136.56	10613	1540.02	<2.2e-16	***
ws_{t-2}^{Ch}	1	20.19	10612	1519.83	6.997e-06	***
Td_{t-8}^{Ch}	1	51.24	10611	1468.59	8.156e-13	***
T_{t-8}^{850}	1	23.05	10610	1445.54	1.575e-06	***
T_{t-8}^{700}	1	0.77	10609	1444.77	0.381334	
Rh_{t-8}^{700}	1	3.38	10608	1441.39	0.066166	.
Rh_{t-8}^{850}	1	53.47	10607	1387.92	4.311e-11	***
Rh_{t-8}^{925}	1	127.73	10606	1260.19	<2.2e-16	***
u_{t-8}^{925}	1	34.35	10605	1225.84	8.028e-07	***
u_{t-8}^{700}	1	6.81	10604	1219.03	0.009061	**
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1						
Deviance explained = 67%						

Table 4. Deviance table of 12 hours ahead fog forecast binary logistic reduced regression model based on training data set within the cold half period 2001-2016.
Таблица 4. Таблица на отклоненията на 12-часов бинарен логистичен линеен редуциран регресионен модел за поява на мъгла, базиран на обучаващата извадка от данни за студените полугодия на периода 2001-2016 г.

	Df	Deviance	Resid.Df	Resid.Dev	Pr(>Chi)	
NULL			10621	3682.2		
T_{t-4}^{Sf}	1	456.32	10620	3225.9	<2.2e-16	***
Td_{t-4}^{Sf}	1	507.73	10619	2718.1	<2.2e-16	***
fog_{t-4}^{Sf}	1	496.87	10618	2221.3	<2.2e-16	***
p_{t-5}^{Ch}	1	77.82	10617	2143.4	<2.2e-16	***
fog_{t-5}^{Sf}	1	32.44	10616	2111.0	1.229e-08	***
T_{t-4}^{Sf}	1	18.40	10615	2092.6	1.789e-05	***
Td_{t-8}^{Ch}	1	15.05	10614	2077.5	0.0001045	***
slp_{t-8}^{Ch}	1	2.89	10613	2074.7	0.0892453	.
T_{t-7}^{Sf}	1	145.90	10612	1928.8	<2.2e-16	***
T_{t-8}^{850}	1	38.62	10611	1890.1	5.145e-10	***
Rh_{t-8}^{700}	1	56.34	10610	1833.8	6.083e-14	***
Rh_{t-8}^{925}	1	160.29	10609	1673.51	<2.2e-16	***
T_{t-8}^{700}	1	24.74	10608	1648.77	1.854e-06	***
u_{t-8}^{925}	1	18.47	10606	1630.3	1.726e-05	***
u_{t-8}^{700}	1	12.80	10604	1617.5	0.0003462	***
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1						
Deviance explained = 56%						

3.3.1. Анализ на ROC кривите

С помощта на оценките на коефициентите на модела $\hat{\beta}$ и данните от тестовата извадка x_t се пресмятат стойностите на линейния предиктор $x_t^T\hat{\beta}$ за всяко t и съответните стойности на вероятностите $\hat{p}(x_t) = p\left(x_t^T\hat{\beta}\right)$, дефинирани с (1). За целите на прогнозирането на мъгла и проверката на качеството на модела е необходимо стойностите на пресметнатите вероятности да бъдат трансформирани в бинарен ред от 0 и 1, който ред да бъде сравнен с бинарния времеви ред на мъглата в данните на тестовата извадка. Определянето на оптимална стойност, праг, $cutoff$, в зависимост от който вероятността $\hat{p}(x_t)$ за поява на мъгла се

трансформира в бинарна величина, приемаща стойност 1, ако $\hat{p}(x_t) \geq cutoff$ или 0 в противен случай, се основава на методологията на ROC (Receiver Operating Characteristic) кривата (графиката), чрез която се оценява качество на бинарната класификация, т.е. сходството между двата бинарни реда при прогноза, Fawcett (2006).

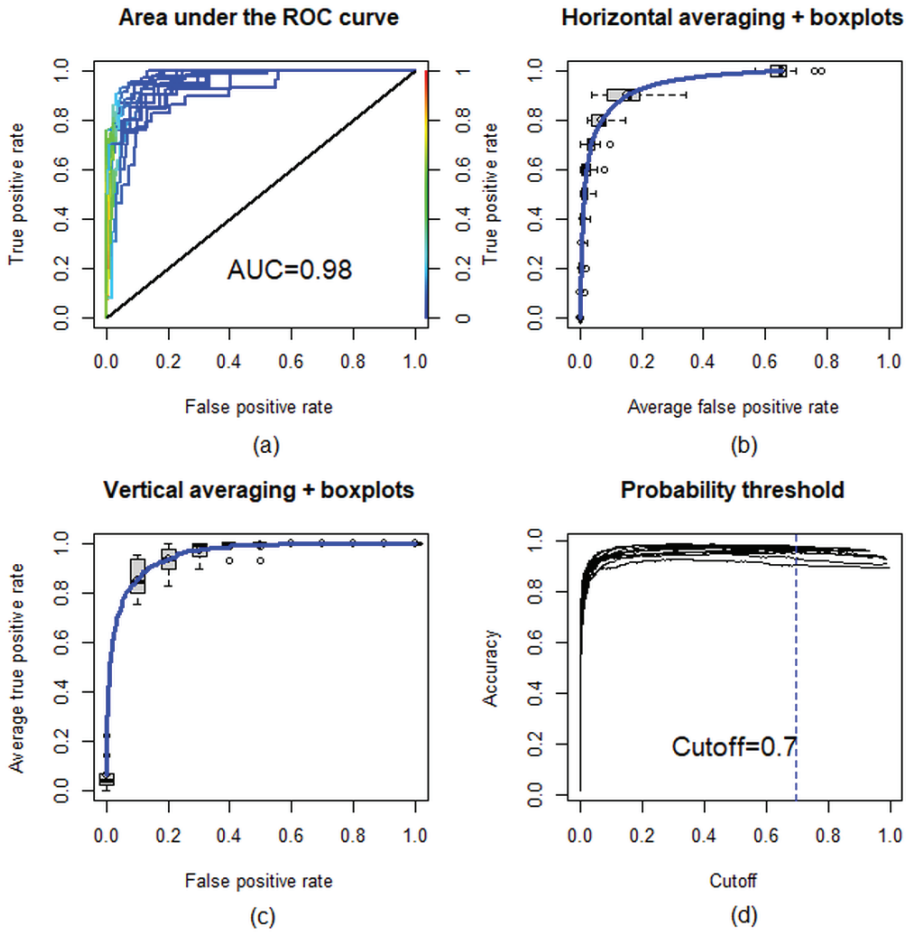


Fig. 2. Graphical forecast performance: (a) ROC curves; (b) and (c) horizontal and vertical distributions of AUC curves; (c) threshold (cutoff) value.

Фиг. 2. Графична оценка на прогнозата: (a) ROC криви; (b) и (c) хоризонтално и вертикално разпределение на лицата под AUC; (d) прагова стойност (cutoff).

На Фиг. 2(a) със заглавия AUC (Area under the ROC curve) са дадени фамилии от криви, получени с техниката на крос-валидирането при формирането на

различните обучаващи и тестови извадки. Всяка крива съответства на различна обучаваща извадка от данни чрез последователно изключване на всяко студено полугодие за формиране на съответна тестова извадка. Мярката за качеството на сходството представлява средна стойност на лицата под ROC кривите равна на 0.98, като стойността на перфектния модел е равна на 1. Високата стойност на това лице съответства на високо сходство между двата бинарни реда. На Фиг. 2(d) е дадена оптималната прагова стойност $cutoff$, равна на 0.7. От графиките на ROC кривите се забелязва значително вариране на лицата под AUC кривите при формирането на различните обучаващи извадки, което се потвърждава също от хоризонталните и вертикални бокс-плотове, характеризиращи тяхното разпределение. Аналогични резултати са получени за 6- и 12-часова прогноза, като AUC=0.97 и $cutoff$=0.66, и AUC=0.89 и $cutoff$=0.52. За провеждането на ROC анализа е използвана библиотеката ROCR, Sing et al. (2020).

3.3.2. Анализ и интерпретация на резултатите

Резултатите за точността и грешката на 3-часовите прогностични модели за поява на мъгла от сравняването на предсказаните бинарни времеви редове с времевите редове на тестовите извадки за някои студени полугодия, формирани по метода на крос-валидирането, са дадени в Таблица 5. Например, точността на прогнозите за поява на мъгла по данни на тестовите извадки за студените полугодия на 2001-2002 г., 2006-2007 г. и 2011-2012 г. са (0.993+0.871)/2=0.932, (0.960+0.829)/2=0.894 и (0.988+0.765)/2=0.877, съответно. Резултатите за останалите полугодия се характеризират с различна точност, поради вариациите в отделните извадки. Поради това точността и грешката на даден модел се характеризират със средната стойност на резултатите от тестовите извадки, получени по метода на крос-валидирането, съответно.

Table 5. Cross-validation confusion table for the 3 hours ahead fog forecasts based on the binary logistic regression model about some cold half periods within 2001-2016 yrs.
Таблица 5. Таблица с точността и грешките на прогноза за мъгла на 3-часов бинарен логистичен регресионен модел по данни на някои студени полугодия през периода 2001-2016 г. по метода на крос-валидирането.

2001-2002		мъгла		2003-2004		мъгла	
		не	да			не	да
прогноза	не	0.993	0.007	прогноза	не	0.995	0.005
	да	0.129	0.871		да	0.062	0.938
2004-2005		мъгла		2006-2007		мъгла	
		не	да			не	да
прогноза	не	0.984	0.016	прогноза	не	0.960	0.040
	да	0.125	0.875		да	0.171	0.829

2007-2008		мъгла		2011-2012		мъгла	
		не	да			не	да
прогноза	не	0.960	0.040	прогноза	не	0.988	0.012
	да	0.171	0.829		да	0.235	0.765
2012-2013		мъгла		2014-2015		мъгла	
		не	да			не	да
прогноза	не	0.985	0.015	прогноза	не	0.988	0.012
	да	0.316	0.684		да	0.160	0.840

Table 6. Cross-validation confusion table for the 6 and 12 hours ahead fog forecasts based on the binary logistic regression models about some cold half periods within 2001-2016.
Таблица 6. Таблица на точността и грешките на прогноза за поява мъгла на 6- и 12-часови бинарни логистични регресионни модели по данни на някои студени полугодия през периода 2001-2016 г. по метода на крос-валидирането.

2007-2008		мъгла		2012-2013		мъгла	
		не	да			не	да
6-часова прогноза	не	0.967	0.033	6-часова прогноза	не	0.983	0.017
	да	0.165	0.835		да	0.385	0.615
12-часова прогноза		не	да	12-часова прогноза		не	да
	не	0.97	0.03		не	0.92	0.08
	да	0.21	0.79		да	0.25	0.75

В Таблица 6 са дадени аналогични резултати на тези от Таблица 3 за точността и грешката на прогнозата за 6- и 12-часови прогностични модели на поява на мъгла за студените полугодия на 2007-2008 и 2012-2013 години по метода на крос-валидирането.

В Таблица 7 са дадени средните стойности на точността и грешката на 3-, 6- и 12-часовите прогностични модели за поява на мъгла за студените полугодия за периода 2001-20016 г. От тази таблица се вижда, че точността на прогнозата намалява пропорционално с увеличаване на срока на нейното издаване. Средните стойности на вярно издадените прогнози на тези модели са равни на (0.985+0.865)/2=0.925, (0.978+0.806)/2=0.892 и (0.975+0.691)/2=0.833. От Таблица 5 се вижда, че 3-часовият модел няма да предсказва вярно в 1.4% (false alarm ratio) и 13.5% (false negative ratio), съответно. Бихме могли да очакваме, че в 1.4% от случаите моделът ще предсказва, че ще има мъгла, но реално няма да се наблюдава мъгла и в 13.5% от случаите моделът ще предсказва, че няма да има мъгла, но реално ще бъде наблюдавана мъгла. Очаква се моделите по-често да прогнозират срокове без мъгла, когато в действителност в сроковете ще бъде регистрирана мъгла (false negative ratio) в сравнение с прогнозата на срок с мъгла,

когато в действителност няма да бъде регистрирана мъгла (false positive), понеже 0.135 > 0.014.

Table 7. Cross-validation average confusion table for the 3, 6 and 12 hours ahead fog forecasts based on the binary logistic regression models for the cold half periods of 2001-2016 yrs.
Таблица 7. Усреднени точност и грешка на прогнозата за поява на мъгла на 3-, 6- и 12-часови бинарни логистични регресионни модели по данни от студено полугодие за периода 2001-2016 г. по метод на крос-валидирането.

	мъгла				мъгла				мъгла			
3-часова прогноза		не	да	6-часова прогноза		не	да	12-часова прогноза		не	да
	не	0.985	0.014		не	0.978	0.022		не	0.975	0.025
	да	0.135	0.865		да	0.194	0.806		да	0.309	0.691

4. МОДЕЛИРАНЕ НА ХОРИЗОНТАЛНА ВИДИМОСТ

4.1. Моделиране на хоризонтална видимост чрез ординална логистична регресия

Нека *Vt* е полиномно (мултиномно) разпределена случайна величина, приемаща някоя от числовите стойности 1, 2, 3,..., *J*, съответстващи на краен брой състояния на случаен процес в срок *t*. В частност, тези *J* категории (състояния) можем да свържем с интервалите, дефиниращи хоризонталната видимост в метри. Вероятността за принадлежност на величината *Vt* в срок *t* към *j*-та категория в ординалната логистична регресия се дефинира като

$$\pi_{tj}(\alpha_j, \beta, x_t) = P(V_t = j|x_t) = \frac{exp(\alpha_j + x_t^T\beta)}{1+exp(\alpha_j + x_t^T\beta)} - \frac{exp(\alpha_{j-1} + x_t^T\beta)}{1+exp(\alpha_{j-1} + x_t^T\beta)} \quad (2)$$

където α_j са неизвестни коефициенти за $j = 1, \dots, J$, удовлетворяващи уловието $\alpha_1 < \alpha_2 < \cdots < \alpha_J$, β е вектор от неизвестни коефициенти, чиято *i*-та координата β_i характеризира степента на влияние на предиктора x_{ti} от x_t Agresti (2002).

От дефиницията на тези вероятности, наричани апостериорни, се вижда, че линейният предиктор $x_t^T\beta$ е общ за всички вероятности, докато наредбата на вероятностите се определя от коефициентите α_j за $j = 1, \dots, J$. Ако оценим вероятностите $\pi_{tj}(\alpha_j, \beta, x_t)$ за $j = 1, \dots, J-1$ тогава *J*-та апостериорна вероятност $\pi_{tJ}(\alpha_J, \beta, x_t)$ се определя от равенството $\pi_{tJ}(\alpha_J, \beta, x_t) = 1 - \sum_{j=1}^{J-1} \pi_{tj}(\alpha_j, \beta, x_t)$, което следва от дефиницията на полиномното разпределение.

Оценките на неизвестните коефициенти $(\alpha_1, \cdots, \alpha_{J-1}, \beta)$ се определят по метода на МП като се максимизира функцията на правдоподобие на полиномното разпределение

$$\max_{(\alpha,\beta)} L(\alpha,\beta) = \prod_{t=l+1}^{L} \prod_{j=1}^{J-1} \pi_{tj}^{v_{tj}}(x_t^T\beta) \text{ с ограничение } \alpha_1 < \alpha_2 < \cdots < \alpha_J,$$

където $v_{tj} = 1$ ако $V_t = j$ или $v_{tj} = 0$ ако $V_t \neq j$, l е максималния времеви лаг на вектора от предикторите x_t^T, използван в модела на вероятността $\pi_{tJ}(\alpha_J, \beta, x_t)$.

След определянето на оценките $\hat{\alpha}_j$ и $\hat{\beta}$ на неизвестните коефициенти α_j и β на вероятностите $\pi_{tj}(\alpha_j, \beta, x_t)$ наблюденията се класифицират на принципа на максималната апостериорна вероятност. Наблюдението x_t за $t = 1, \ldots, L$ се класифицира към h-та категория, ако $\pi_{th}(\hat{\alpha}_h, \hat{\beta}, x_t) > \pi_{tj}(\hat{\alpha}_j, \hat{\beta}, x_t)$ за $j = 1, \ldots, J$. В случай на равни апостериорни вероятности, наблюдението се класифицира произволно към някоя от равновероятните категории.

За максимизирането на ФП могат да се използват библиотеки като nnet (neural net), mgcv и други от безплатната програмна среда R. Ще отбележим, че бинарната и ординална линейни логистични регресии са основни модели в задачите за класификация и дискриминация в литературата по приложна статистика, невронни мрежи и машинно обучение (machine learning).

4.2. Анализ и интерпретация на резултатите от модели на хоризонтална видимост

В този параграф са дискутирани модели на вероятността за принадлежност към дадена категория за целите на 3-часовата прогноза на видимост в станцията на НИМХ-София, основани на ординалната логистична регресия. Ще бъдат разгледани модели с 3 и 5 категории, дефинирани както следва: а) видимост "до 2000 m", „между 2000 и 10000 m" и „над 10000 m"; б) видимост „до 50 m", „между 50 и 500 m", „между 500 и 2000 m", „между 2000 и 20000 m" и „над 20000 m". За моделирането на тези вероятности са използвани само стойностите на метеорологичните предиктори с лагове от 1 до 8 срока от станциите на НИМХ-София и Черни връх, както и измерванията от аерологичните сондажи на трите стандартни нива от предходния ден, описани в параграф 2. По този начин ще бъде отчетена динамиката в процеса на промяна на видимостта за периода 2001-2016 г. В моделите не са включени мъгла и видимост като предиктори, нито техни лагове, регистрирани в станцията на Черни връх. Броят на потенциалните предиктори в моделите е 94.

Table 8. Table of coefficient estimates of the ordinal logistic regression model for 3- hours horizontal visibility forecast in 3 categories for the period 2001-2014.

Таблица 8. Таблица с оценките на коефициентите на ординален логистичен регресионен модел на 3-часова прогноза на хоризонтална видимост в 3 категории за периода 2001-2014 г.

Predictors	Coeff. Estimate	Std.Error	zvalue	Pr(>|z|)	Signif
Intercept 1	$\hat{\alpha}_1$=16.334239	4.739872	3.446	0.000569	***
Intercept 2	$\hat{\alpha}_2$=21.392172	4.741546	4.512	6.43e-06	***
T_{t-1}^{Sf}	$\hat{\beta}_1$= 0.037495	0.001614	23.237	<2e-16	***
p_{t-1}^{Sf}	$\hat{\beta}_2$=-0.015711	0.004888	-3.214	0.001308	**
fog_{t-1}^{Sf}	$\hat{\beta}_3$=1.859119	0.917691	2.026	0.042779	*
fog_{t-1}^{Sf}	$\hat{\beta}_4$= 0.094460	0.018596	5.080	3.78e-07	***
ws_{t-2}^{Sf}	$\hat{\beta}_5$= 0.058681	0.013631	4.305	1.67e-05	***
Td_{t-2}^{Sf}	$\hat{\beta}_6$= 0.015112	0.004955	3.050	0.002291	**
$\mathrm{T}d_{t-3}^{Sf}$	$\hat{\beta}_7$=-0.096448	0.018429	-5.233	1.66e-07	***
fog_{t-2}^{Sf}	$\hat{\beta}_8$=-1.826776	0.562705	-3.246	0.001169	**
Td_{t-1}^{Ch}	$\hat{\beta}_9$= 0.007884	0.001293	6.099	1.07e-09	***
t_{t-2}^{Ch}	$\hat{\beta}_{10}$= 0.074445	0.011125	6.692	2.21e-11	***
ws_{t-3}^{Ch}	$\hat{\beta}_{11}$=-0.018706	0.004922	-3.801	0.000144	***
ws_{t-5}^{Ch}	$\hat{\beta}_{12}$=-0.016947	0.004866	-3.483	0.000496	***
Td_{t-8}^{Ch}	$\hat{\beta}_{13}$=-0.004835	0.001204	-4.016	5.91e-05	***
t_{t-8}^{925}	$\hat{\beta}_{14}$=-0.120284	0.010567	-11.383	<2e-16	***
t_{t-8}^{850}	$\hat{\beta}_{15}$=0.055381	0.013702	4.042	5.30e-05	***
t_{t-8}^{700}	$\hat{\beta}_{16}$=-0.035339	0.009823	-3.598	0.000321	***
u_{t-8}^{925}	$\hat{\beta}_{17}$= 0.019610	0.007308	2.683	0.007291	**
u_{t-8}^{850}	$\hat{\beta}_{18}$= 0.019050	0.008610	2.213	0.026890	*
v_{t-8}^{850}	$\hat{\beta}_{19}$=-0.024586	0.007094	-3.466	0.000529	***
$visbS_{t-1}^{Sf}$	$\hat{\beta}_{20}$=-3.034797	0.038019	-79.823	<2e-16	***
$visbS_{t-2}^{Sf}$	$\hat{\beta}_{21}$=-0.374101	0.053387	-7.007	2.43e-12	***
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1					
Deviance explained = 55.7%					

Първо ще бъде разгледан стохастичният модел на хоризонтална видимост с 3 категории за станцията на НИМХ-София, чиито неизвестни коефициенти са оценени по данни на обучаваща се извадка от наблюдения за периода 2001-2014 г. с методологията на ординалната логистична регресия. Оценките на значимите коефициенти на този модел са дадени в Таблица 8. Оценките $\hat{\alpha}_1$ и $\hat{\alpha}_2$ удовлетворяват условието $\hat{\alpha}_1 = 16.33 < \hat{\alpha}_2 = 21.39$. Чрез този модел се обяснява 55.7% от стойността на отклонението на модела без предиктори.

Table 9. Predicted posterior probabilities using the ordinal logistic regression model with 3 horizontal visibility categories based on 3 consecutive observations from test data set.
Таблица 9. Предсказани апостериорни вероятности чрез ординален логистичен регресионен модел с 3 категории на хоризонтална видимост за 3 последователни наблюдения от тестовата извадка.

срок	$\pi_{t1}(\hat{\alpha}_1, \hat{\beta}, x_t)$	$\pi_{t2}(\hat{\alpha}_2, \hat{\beta}, x_t)$	$\pi_{t3}(\hat{\alpha}_3, \hat{\beta}, x_t)$
t=1	0.145	0.505	0.350
t=2	0.587	0.353	0.060
t=3	0.944	0.050	0.005

В Таблица 9 са дадени оценките на апостериорните вероятности на този модел, пресметнати за 3 поредни срока от тестовата извадка за периода 2015-2016 г., състояща се от 4183 срока. Съгласно принципа на максималната апостериорна вероятност, наблюдението в срок t=1 е класифицирано (предсказано) към 2-ра категория, докато за наблюденията в срок t=2 и t=3 са класифицирани към 1-ва категория. Аналогично са класифицирани останалите наблюдения от тестовата извадка. Вследствие на това е формиран категориен времеви ред, приемащ стойности 1, 2 и 3, съответни на категориите видимост “до 2000 m”, “между 2000 и 10000 m” и “над 10000 m”. Честотното разпределение на времевите категорийни редове на хоризонтална видимост на тестовата извадка и предсказания по модела е представено в Таблица 10, като честотите да дадени в скоби. Предсказаните честоти на хоризонатална видимост “до 2000 m”, “между 2000 и 10000 m” и “над 10000 m” са съответно равни на 535, 2409 и 1239, дадени в последния стълб на таблицата, докато в последния ред на таблицата са дадени честотите на срещане на дадена категория на видимост по наблюденията на тестовата извадка, съответно равни на 553, 2328 и 1302. Честотите, дадени в последния стъб и ред формират маргиналните разпределения на таблицата.

По главния диагонал на таблицата са дадени честотите, в скоби, на правилно предсказаните случаи към съответните три категории, докато неправилно предсказания брой наблюдения, в скоби, са поместени в останалите клетки на таблицата. По тези честоти са пресмятат оценките на вероятностите за точната и грешната прогноза по модела, съответно равни на 385/535=0.72, 147/535=0.275, 3/535=0.006, 162/2409=0.067 и т.н. Точността на модела за прогноза на хоризонтална

видимост с 3 категории е равна на (0.720+0.804+0.799)/3=0.774. Вижда се, че прогностичните качества на този модел са натоварени със сравнително големи грешки. Така например в 27.5% от случаите моделът ще предсказва, че видимостта е до 2000 m, но реално видимостта ще бъде между 2000 и 10000 m. Аналогично в 19.6% от случаите моделът ще предсказва, че видимостта ще бъде между 2000 и 10000 m, но реално видимостта ще бъде над 10000 m.

Table 10. Confusion table for the horizontal visibility ahead forecast in 3 categories based on ordinal logistic regression model and test data set for the period 2015-2016.
Таблица 10. Точност и грешки на модел на прогноза на хоризонтална видимост в 3 категории по данни на тестова извадка за период 2015-2016 г.

		хоризонтална видимост			честота
прогноза		до 2000 m	2000-10000 m	над 10000 m	
	до 2000 m	0.720(385)	0.275(147)	0.006(3)	535
	2000-10000 m	0.067(162)	0.804(1938)	0.128(309)	2409
	над 10000 m	0.005(6)	0.196(243)	0.799(990)	1239
	честота	553	2328	1302	4183

Подобни резултати са получени за модел на видимост с 5 категории, дадени в Таблица 11. Точността на прогнозата на този модел е равна на (0.800 + 0.048 + 0.701 + 0.841 + 0.750)/5 = 0.628. Вижда се, че ниската успеваемост на прогнозата се дължи на факта, че броят (честотата) на случаите с видимост 50-500 m е само 9, докато честота на случаите на другите категории е много по-висока, поради което оценяването на неизвестните коефициенти на вероятността за видимост между 50 и 500 m в модела не е надеждно. Интерпретацията на грешките на модела са аналогични на тези като модела за хоризонтална видимост в 3 категории.

Table 11. Confusion table for the horizontal visibility forecast in 5 categories based on ordinal logistic regression model and test data set for the period 2014-2016.
Таблица 11. Точност и грешки на ординален логистичен регресионен модел на прогноза на хоризонтална видимост в 5 категории по данни на тестова извадка за период 2014-2016

		хоризонтална видимост					
		до 50 m	50-500 m	500-2000 m	2000-20000 m	над 20000 m	честота
прогноза	до 50 m	0.800(44)	0.018(1)	0.164(9)	0.018(1)	0.000(0)	55
	50-500 m	0.333(7)	0.048(1)	0.571(12)	0.048(1)	0.000(0)	21
	500-2000 m	0.007(4)	0.010(6)	0.701(405)	0.275(159)	0.007(4)	578
	2000-20000 m	0.000(0)	0.000(1)	0.061(143)	0.841(1965)	0.097(227)	2336
	над 20000 m	0.000(0)	0.000(0)	0.002(3)	0.248(323)	0.750(977)	1303
	честота	55	9	572	2449	1208	6473

5. ЗАКЛЮЧЕНИЕ

В настоящата статия се предлагат стохастични модели за прогноза на мъгла и хоризонтална видимост, използващи метеорологични предиктори в основните синоптични срокове и данни от аерологичен сондаж на стандартни нива 925, 850 и 700 hPa. Получените резултати показват, че тези модели, основани на бинарната и ординална логистични регресии, са адекватно средство за вероятностна прогноза за поява на мъгла и хоризонтална видимост. Чрез 3- часовите прогностични модели за поява на мъгла се обяснява 74%, докато за принадлежност към дадена категория на хоризонтална видимост се обяснява 56% от разсейването на данните. Необходимо е привличането на допълнителни предиктори към тези модели, като температурата на земната повърхност, влажност и температура на почвата на стандартни нива, както и прогностични стойности от числените модели на времето, достъпни в НИМХ, като ALADIN, WRF, IFS, GFS, за да бъде изследван техния потенциал. В резултат на изследването е разработена методика и програмно обезпечаване за прогнозиране на поява на мъгла и хоризонтална видимост със стохастични модели, които биха могли да бъдат използвани в други населени места на територията на България.

БЛАГОДАРНОСТИ

Авторите изказват благодарност на синоптиците д-р Б. Маркова и И. Василев за критичните забележки и конструктивни препоръки на предишни версии на статията.

ЛИТЕРАТУРА

Agresti, A. (2002). Categorical Data Analysis. 2nd, John Wiley, NY.

Bartok, J., Babic, F., Bednar, P., Paralic, J., Kovac, J., Bartokova, I. and Hluchy, L. (2011). Data Mining for Fog Prediction and Low Clouds Detection. *Computing and Informatics*, 30, 1441-1464.

Bengtsson, L., Andrae, U., Aspelien, T., Batrak, Y., Calvo, J., de Rooy, W., Gleeson, E., Hansen-Sass, B., Homleid, M., Hortal, M., Ivarsson, K., Lenderink, G., Niemelä, S., Nielsen, K. P., Onvlee, J., Rontu, L., Samuelsson, P., Muñoz, D. S., Subias, A., Tijm, S., Toll, V., Yang, X., & Køltzow, M. Ø. (2017). The HARMONIE–AROME Model Configuration in the ALADIN–HIRLAM NWP System, *Monthly Weather Review*, 145(5), 1919-1935

Boneh,T., Weymouth, G.T., Newham, P., Potts, R., Bally, J., Nicholson, A. and Korb, K.B. (2015). Fog Forecasting for Melbourne Airport Using a Bayesian Decision Network. *Weather and Forecasting*, 30, 1218-1233.

Chmielecki, R. M., Raftery, A. E. (2011). Probabilistic Visibility Forecasting Using Bayesian Model Averaging. *Monthly Weather Review*., 139, 1626-1636.

Colabone, Ferrari, A. L., Vecchia, F. and Tech, A. R. B. (2015). Application of Artificial for Fog Forecast Neural Networks, *J. Aerosp. Technol. Manag*., Sao Jose dos Campos, 7(2), 240-246.

Dietz, S.J., Kneringer, P., Mayr, G.J. and Zeileis, A. (2019). Forecasting Low-Visibility Procedure States with Tree-Based Statistical Methods. *Pure Appl. Geophys*. 176, 2631-2644.

Fabbian, D., de Dear, R. and Lellyett, S., (2007). Application of Artificial Neural Network Forecasts to Predict Fog at Canberra International Airport, *Weather and Forecasting*, 22(2), 372-381.

Fawcett, T. (2006). An introduction to ROC analysis. Pattern Rec. Lett., 27, 861-874.

Forthun, G. M., Johnson, M. B., Schmitz, W. G., Blume, J. and Caldwell, R. J., (2006). *Phys. Geography*, 27, 206-222.

Grace, W. and Ferriere, P. (2001). Statistical-empirical forecasting guidance for the occurrence of fog at Mount Gambier Airport. *Aust. Met. Mag*. 50, 15-27.

Hacker, M., Masbou, M. and Bentzien, S. (2013). FOGCAST Probabilistic fog forecasting based on operational (high-resolution) NWP models. EMS-2013 poster.

Kneringer, P., Dietz, S.J., Mayr, G.J. and Zeileis, A. (2019). Probabilistic Nowcasting of Low-Visibility Procedure States at Vienna International Airport During Cold Season. *Pure Appl. Geophys*. 176, 2165-2177.

Müller, M.D., Schmutz, C. and Parlo, E. (2007). A one-dimensional ensemble forecast and assimilation system for fog prediction. *Pure Appl. Geophys*. 164, 1241-1264.

R Core Team (2021). R: A language and environment for statistical computing, R Development Core Team, R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Stoycheva, A. and Evtimov, S. (2014). Studying the fogs in Sofia with Cherni vrah-Sofia stability index. *Bulgarian Geophysical Journal*, 40, 23-32.

Стойчева, А. и Евтимов, С., (2014). Диагностика на мъглите в София за периода 1992-2012 г., *Год. на СУ „Св. Кл.сОхридски"*, ФзФ, 107, стр. 132-145.

Стойчева, А. и Евтимов, С., (2015). Мъглата в София 03-10/01/2014 анализ чрез псевдопотенциалната температура, *Год. на СУ „Св. Кл. Охридски*", ФзФ, 108, 25-47.

Sing, T., Sander, O., Beerenwinkel, N., Lengauer, Th., Unterthiner, Th. and Ernst, F.G.M. (2020). ROCR: Visualizing the performance of scoring classifiers in R. R package version 1.0-11. https://CRAN.R-project.org/package=ROCR

Stern, H. and Parkyn, K. (2001). Predicting the likelihood of fog at Melbourne Airport. 8th Conference on aviation, range and aerospace meteorology, Amer. Meteor. Soc., Dallas, Texas, 10-15 Jan., 1999.

Wood, S.N. (2017). Generalized Additive Models: An Introduction with R. 2nd edition, Chapman and Hall, CRC Press.

Zazzaro, G., Pisano, F.M., Mercogliano, P., (2010). Data Mining to Classify Fog Events by Applying Cost-Sensitive Classifier. In: *International Conference on Complex, Intelligent and Software Intensive Systems 2010*, 1093-1098.