

# Epidemic intelligence data of Crimean-Congo haemorrhagic fever, European Region, 2012 to 2022: a new opportunity for risk mapping of neglected diseases

Angela Fanelli<sup>1,2</sup>, Johannes Christof Schnitzler<sup>3</sup>, Marco De Nardi<sup>4</sup>, Alastair Donachie<sup>3</sup>, Iliaria Capua<sup>2</sup>, Gianvito Lanave<sup>1</sup>, Domenico Buonavoglia<sup>1</sup>, Paula Caceres-Soto<sup>5</sup>, Paolo Tizzani<sup>5</sup>

1. Department of Veterinary Medicine, University of Bari, Bari, Italy

2. One Health Center of Excellence, University of Florida, Gainesville, Florida, United States

3. Intelligence Innovation and Integration unit, World Health Organization, Berlin, Germany

4. SAFOSO AG, Liebefeld, Switzerland

5. World Animal Health Information and Analysis Department, World Organisation for Animal Health, Paris, France

**Correspondence:** Angela Fanelli (angela.fanelli@uniba.it)

## Citation style for this article:

Fanelli Angela, Schnitzler Johannes Christof, De Nardi Marco, Donachie Alastair, Capua Iliaria, Lanave Gianvito, Buonavoglia Domenico, Caceres-Soto Paula, Tizzani Paolo. Epidemic intelligence data of Crimean-Congo haemorrhagic fever, European Region, 2012 to 2022: a new opportunity for risk mapping of neglected diseases. *Euro Surveill.* 2023;28(16):pii=2200542. <https://doi.org/10.2807/1560-7917.ES.2023.28.16.2200542>

Article submitted on 05 Jul 2022 / accepted on 12 Jan 2023 / published on 20 Apr 2023

**Background:** The Epidemic Intelligence from Open Sources (EIOS) system, jointly developed by the World Health Organisation (WHO), the Joint Research Centre (JRC) of the European Commission and various partners, is a web-based platform that facilitates the monitoring of information on public health threats in near real-time from thousands of online sources. **Aims:** To assess the capacity of the EIOS system to strengthen data collection for neglected diseases of public health importance, and to evaluate the use of EIOS data for improving the understanding of the geographic extents of diseases and their level of risk. **Methods:** A Bayesian additive regression trees (BART) model was implemented to map the risk of Crimean-Congo haemorrhagic fever (CCHF) occurrence in 52 countries and territories within the European Region between January 2012 and March 2022 using data on CCHF occurrence retrieved from the EIOS system. **Results:** The model found a positive association between all temperature-related variables and the probability of CCHF occurrence, with an increased risk in warmer and drier areas. The highest risk of CCHF was found in the Mediterranean basin and in areas bordering the Black Sea. There was a general decreasing risk trend from south to north across the entire European Region. **Conclusion:** The study highlights that the information gathered by public health intelligence can be used to build a disease risk map. Internet-based sources could aid in the assessment of new or changing risks and planning effective actions in target areas.

## Introduction

The Epidemic Intelligence from Open Sources (EIOS) [1] platform is a unique collaboration between various public health stakeholders around the globe.

The initiative is led by the World Health Organization (WHO) and brings together new and existing networks and systems to create a unified all-hazards, One Health approach to early detection, verification and assessment of public health risks and threats. The EIOS system, jointly developed under the EIOS initiative by WHO, the Joint Research Centre (JRC) of the European Commission and partners is a web-based platform that enables monitoring of information in near real-time from around 15,000 publicly accessible sources worldwide (e.g. news media, social media, scientific papers), using specific search algorithms. EIOS promotes and catalyses new and innovative collaborative development, appealing and connecting several actors, institutions and countries [2].

Information collated from open sources is automatically processed, analysed and enriched with metadata including classifications based on more than 600 categories representing different concepts (for example: diseases, variants, public health measures). Other meta information relevant for this study is the detection of mentioned geolocations. Geolocations are identified using named entity recognition with disambiguation (NEROne) [3], mapped to the geographical database GeoNames [4].

In order to facilitate detection of new relevant threats or monitoring of ongoing threats, the EIOS system provides the user with several filter options to focus on, for the given user, most relevant information. Specifically, data can be filtered by time period, mentioned countries, geolocations, region, language, source of the data, as well as categories. All these filters can be combined to create a tailored search and filter capabilities

## KEY PUBLIC HEALTH MESSAGE

### What did you want to address in this study?

Epidemic Intelligence from Open Sources (EIOS) initiative provides a means of disseminating information among public health agencies. In this study, we investigated whether online information retrieved from the EIOS can be used to produce a risk map for Crimean-Congo haemorrhagic fever (CCHF) within the European Region. CCHF is a tick-borne disease caused by a virus belonging to the *Nairoviridae* family.

### What have we learnt from this study?

Using CCHF as a test study, we learnt that epidemic intelligence tools like EIOS can be used to automatically extract information on outbreak locations, which may, in turn, be used to identify the areas at high risk of disease occurrence.

### What are the implications of your findings for public health?

Online data on the risk of disease in different locations can be used by decision makers, public health authorities as well as veterinary services to identify areas where surveillance is most needed.

which can be saved. These characteristics positioned the EIOS as one of the most relevant systems for event-based surveillance [5].

EIOS collects information from a wide range of unofficial and official sources (e.g. official channels from national authorities) on animal and human diseases. The World Organisation for Animal Health (WOAH) and the WHO, organisations that disseminate officially confirmed information on animal and human diseases respectively, have different mechanisms and systems for collecting and disseminating this official information. In the field of animal health, the World Animal Health Information System (WAHIS) is the WOAH reference global database that gathers information on the official animal disease situation as reported by the National Veterinary Services [6]. This database has been proved to be particularly useful to understand the evolution of animal diseases at global and regional level [7-12]. Information on human health is available through different mechanisms, with varying coverage, completeness, and update frequency (accessible for example through the Global Health Observatory [13], the Global Influenza Programme [14], Disease Outbreak News [15] among others). Data from media sources, expert networks and social media can be pivotal to identify significant disease outbreak information in a near real-time way [16], but they also represent critical sources to understand the spatio-temporal distribution of diseases when official data are lacking. Nevertheless, information from unofficial sources needs to be verified and confirmed.

In this study, we used internet-based data retrieved from the EIOS system to model the risk of Crimean-Congo haemorrhagic fever (CCHF) occurrence in the European Region. The objectives were (i) to assess the capacity of the EIOS system to strengthen data collection for

neglected diseases of veterinary and public health importance and (ii) to evaluate the use of the EIOS data for improving the understanding of the geographic extents of diseases and their level of risk. CCHF is a tick-borne disease caused by the arbovirus Crimean-Congo haemorrhagic fever virus (CCHFV; family *Nairoviridae*) transmitted by ticks of the genus *Hyalomma* [17-20]. A previous study attempted to model the risk of CCHF infection in humans using reported cases in literature [21]. However, thus far, no study has used epidemic intelligence tools investigating publicly available sources to create a database of occurrence points with the view of building a disease risk map.

## Methods

### Geographical setting and definitions

The study area is represented by the territories of 52 countries and territories within the European Region below ~60° North latitude (see Supplementary Material A, Table S1 for a list of the countries and territories considered in this study).

An occurrence point was defined as a geographical point where a case of CCHF in humans or an isolation of CCHFV from ticks was detected during the timeframe considered. This latitude was defined considering the gaps in satellite normalised difference vegetation index (NDVI) coverage at regions near the poles because of the long-lasting accumulation of snow or clouds, as well as the stable presence of the CCHF vector (ticks of genus *Hyalomma*) [22,23].

### Data sources and approach

Data on CCHF occurrence (January 2012–March 2022) were retrieved from the EIOS system using a specific category ('CCHF category') that implemented an algorithm to detect mentions of CCHF cases in humans or vectors. In practice, a list of predefined keywords

**TABLE**

Variables collated for use in the spatial model on Crimean-Congo haemorrhagic fever occurrence in 52 countries and territories within the European Region, 2012–2022

Definition	Code	Source
Annual mean temperature	WC_bio1	WorldClim [27]
Mean diurnal range (mean of monthly (max temperature – min temperature))	WC_bio2	WorldClim [27]
Isothermality (bio2/bio7) (x 100)	WC_bio3	WorldClim [27]
Temperature seasonality (standard deviation x 100)	WC_bio4	WorldClim [27]
Max temperature of the warmest month	WC_bio5	WorldClim [27]
Min temperature of the coldest month	WC_bio6	WorldClim [27]
Temperature annual range (bio5 – bio6)	WC_bio7	WorldClim [27]
Mean temperature of the wettest quarter	WC_bio8	WorldClim [27]
Mean temperature of the driest quarter	WC_bio9	WorldClim [27]
Mean temperature of the warmest quarter	WC_bio10	WorldClim [27]
Mean temperature of the coldest quarter	WC_bio11	WorldClim [27]
Annual precipitation	WC_bio12	WorldClim [27]
Precipitation of the wettest month	WC_bio13	WorldClim [27]
Precipitation of the driest month	WC_bio14	WorldClim [27]
Precipitation seasonality (coefficient of variation)	WC_bio15	WorldClim [27]
Precipitation of the wettest quarter	WC_bio16	WorldClim [27]
Precipitation of the driest quarter	WC_bio17	WorldClim [27]
Precipitation of the warmest quarter	WC_bio18	WorldClim [27]
Precipitation of the coldest quarter	WC_bio19	WorldClim [27]
Thornthwaite aridity index: index of the degree of water deficit below water need	ER_aridityIndexThornthwaite	ENVIREM dataset [28]
A metric of relative wetness and aridity	ER_climaticMoistureIndex	ENVIREM dataset [28]
Average temperature of the warmest month – average temperature of the coldest month	ER_continentality	ENVIREM dataset [28]
Sum of mean monthly temperature for months with mean temperature greater than 0°C multiplied by number of days	ER_growingDegDayso	ENVIREM dataset [28]
Sum of mean monthly temperature for months with mean temperature greater than 5°C multiplied by number of days	ER_growingDegDays5	ENVIREM dataset [28]
Count of the number of months with mean temperature greater than 10°C	ER_monthCountByTemp10	ENVIREM dataset [28]
Compensated thermicity index: sum of mean annual temperature, min temperature of the coldest month, max temperature of the coldest month x 10, with compensations for better comparability across the globe	ER_thermicityIndex	ENVIREM dataset [28]
SAGA GIS topographic wetness index	ER_topoWet	ENVIREM dataset [28]
Terrain roughness index	ER_tri	ENVIREM dataset [28]
Principal component 1 of NDVI <sup>a</sup> time series	ndvi.pca1	Copernicus Global Land Service [29]
Principal component 2 of NDVI <sup>a</sup> time series	ndvi.pca2	Copernicus Global Land Service [29]
Principal component 3 of NDVI <sup>a</sup> time series	ndvi.pca3	Copernicus Global Land Service [29]
Principal component 4 of NDVI <sup>a</sup> time series	ndvi.pca4	Copernicus Global Land Service [29]

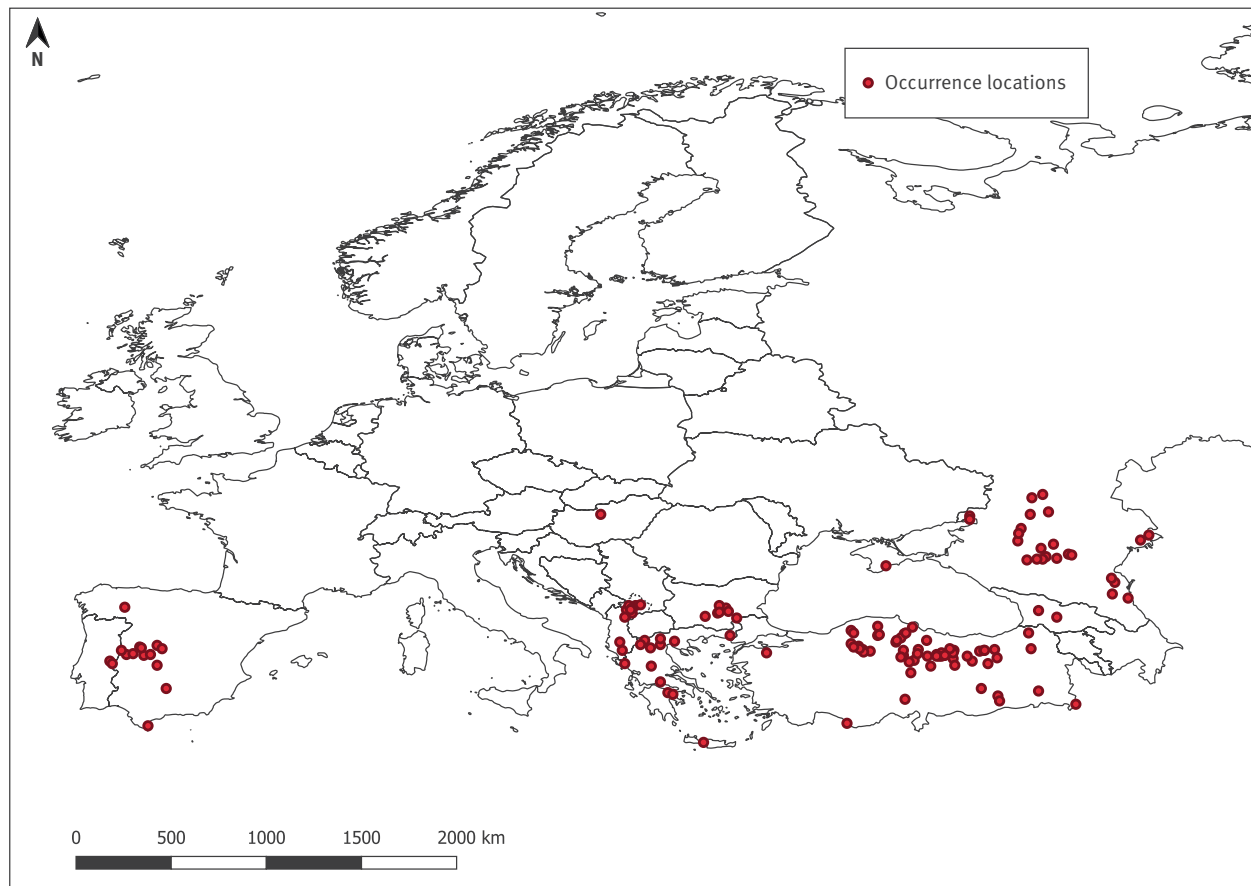
max: maximum; min: minimum; NDVI: normalised difference vegetation index.

<sup>a</sup> NDVI is an indicator of the greenness of the biomes.

A list of the countries and territories included can be found in Supplementary Material A Table S1 A.

**FIGURE 1**

Occurrence locations of Crimean-Congo haemorrhagic fever cases in humans and Crimean-Congo haemorrhagic fever virus isolates from ticks, European Region, 2012–2022 (n = 141 before spatial thinning)



and keyword combinations in different languages was used (see Supplementary Material B for an example of keywords in English). The EIOS system collates information from a broad range of sources including news media, social media and expert networks and news aggregator websites (e.g. ProMed, HealthMap and the Global Public Health Intelligence Network (GPHIN)). The EIOS system is constantly updated, collecting ‘close to real-time’ events. Data included in this study have incorporated the information available up to 2 March 2022. Since EIOS only started a consistent data retrieval in November 2017, we coupled these data with the information on CCHF cases during the period 2012–17 obtained from scientific literature.

The approach was based on the following steps: (i) building a filter in EIOS to detect all the news items falling under ‘CCHF category’, (ii) retrieving and double checking the geographical coordinates detected in the news items by EIOS, (iii) assessing the reliability of the data, (iv) complementing the occurrence data extracted from EIOS with the information retrieved from scientific literature, (v) retrieving and processing the rasters used as predictors and (vi) building a

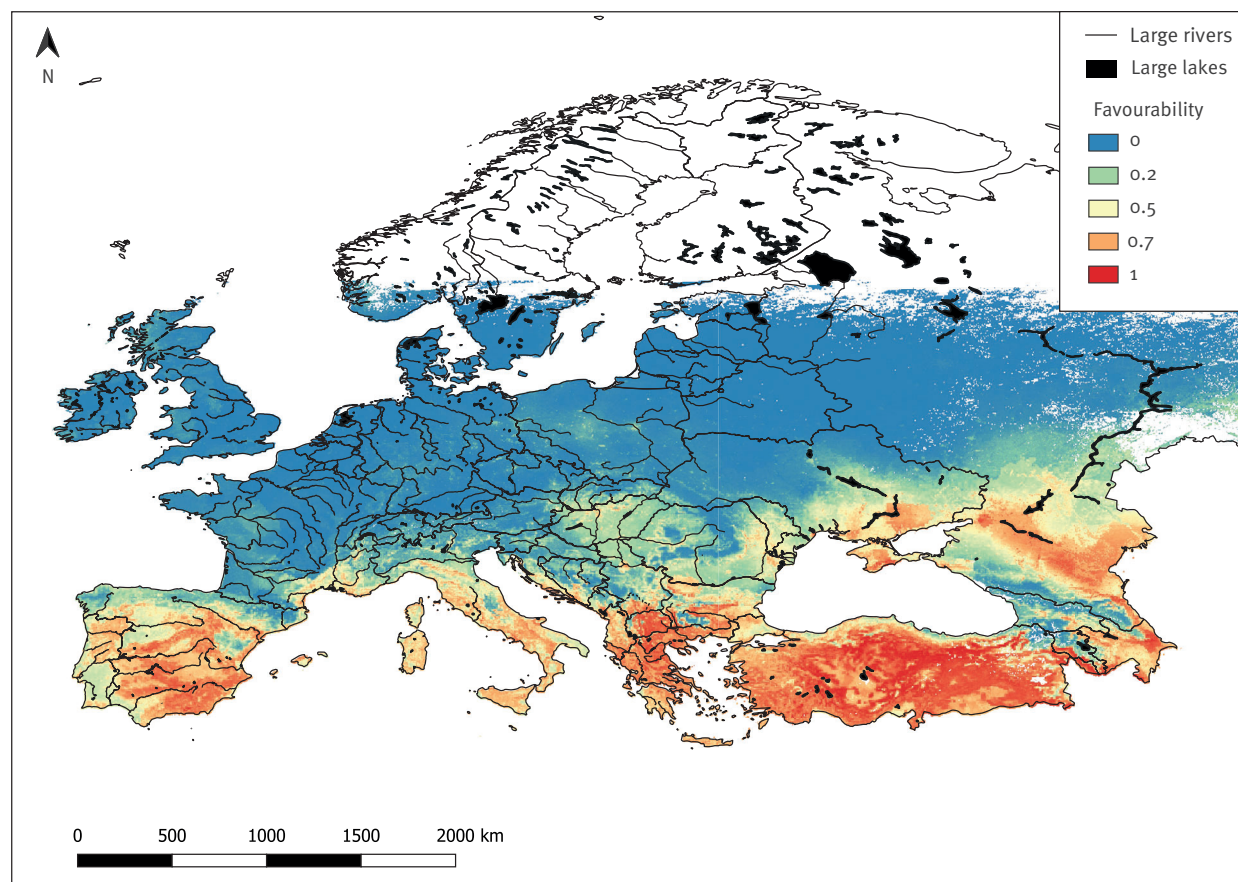
spatial model. Details of the methodology applied are provided hereunder.

### Filter criteria and data processing

Firstly, a filter was created in EIOS to identify and pin all news items that mentioned CCHF cases in humans or vectors to the ‘CCHF category’. Secondly, a spatial selection was applied, and a target area was drawn on the interface map provided by the system to include only the information relevant to the European Region. Lastly, each piece of news, along with its ID (assigned by EIOS) and geographical coordinates of all the mentioned locations, was extracted and exported to Excel. If a news item contained information about more than one location of the disease, multiple rows were created in the export. Each news item was thoroughly checked manually by the authors (AF, JCS and PT) to assess (i) the credibility and reliability of the website, (ii) the content of the news item, (iii) the accuracy and reliability of the source, (iv) confirmation of the same news item by more than one source and (v) strength of evidence for CCHF presence in that location. To do so, the map on geographic distribution of CCHF developed by WHO showing the areas with active CCHF circulation (virological and serological evidence) was consulted as

**FIGURE 2**

Favourability predictions for Crimean-Congo haemorrhagic fever occurrence, European Region below ~60° North latitude, 2012–2022



one of the criteria to check and validate the information retrieved [24]. News items mentioning the locations where the cases were hospitalised rather than the locations of the infection were discarded. Similarly, we did not consider news items where it was not possible to distinguish the infection site from the hospital location.

Given the limitations of the EIOS system before 2017, the dataset was completed by including information on CCHF cases retrieved from studies published after 2012 (mentioned in a recent meta-analysis [25]). To better define the territorial entities, two columns were added to the database, one containing the GeoNames Feature Code retrieved from GeoNames website [4] and another containing the information on the area size (expressed as km<sup>2</sup>) of the territorial entity and retrieved from Wikipedia. After removal of duplicates (news items and information from scientific literature referring to more cases occurring in the same location), administrative divisions greater than 5,000 km<sup>2</sup> were not included in the analysis to reduce the risk of bias in the model. If we had included large areas of presence, the predictors would have needed to be scaled at a lower resolution. The use of highly resolved predictors is indeed an

important prerequisite of spatial modelling since the detail of disease response to the environments may not be captured at a coarser resolution.

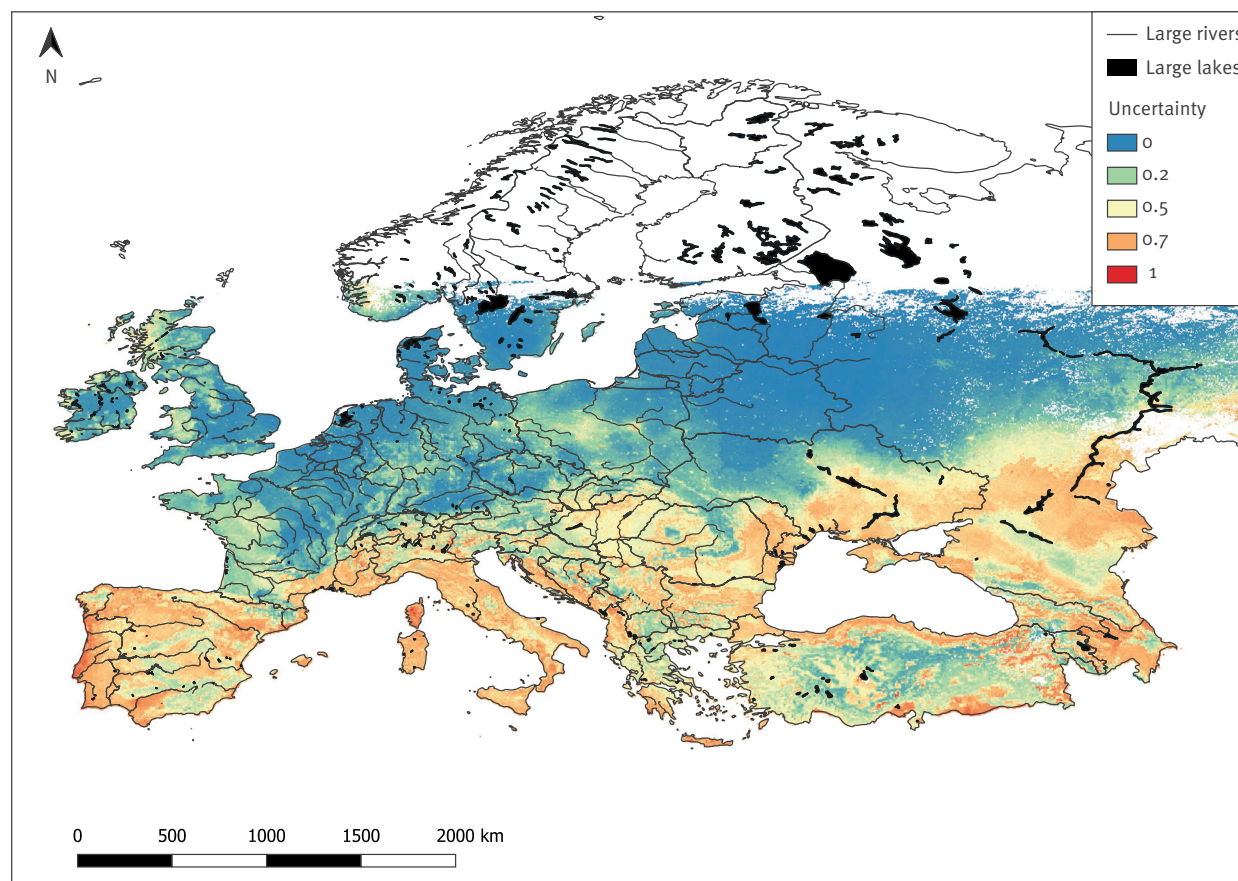
### Predictor variables

In this study, the term pixel is used to define the level of detail (cell size or spatial resolution) of the rasters used as predictors in the model. Thirty-three explanatory variables were initially considered to build the model (Table). We used the functions in the ‘sdmpredictors’ package [26] to access WorldClim [27] and ENVIREM [28] online datasets. For the NDVI, we downloaded the 10-daily data PROBAV\_V2.2.1 (1 Jan 2016–31 Dec 2020) from Copernicus Global Land Service [29]. One-month composites of NDVI were prepared through the method of the maximum pixel value to obtain the largest area without gaps in pixels. Principal component analysis (PCA) was run on the monthly composites using the `prcomp()` function to reduce the number of NDVI rasters. The first four components explaining 0.91% of the total variance were added to the set of explanatory variables (see Supplementary Material A, Figure S1 A for an analysis of the first four principal components). All the predictors were rescaled at a resolution of 0.044645 x 0.044645 decimal degrees



**FIGURE 3**

Uncertainty predictions<sup>a</sup> for Crimean-Congo haemorrhagic fever occurrence, European Region below ~60° North latitude, 2012–2022



<sup>a</sup> Posterior width, 95% credible intervals.

(corresponding to around 4 × 4 km in metric units), aligned and reprojected using the same coordinate reference system (World Geodetic System 1984 (WGS84) EPSG:4326).

### Occurrence points and pseudo-absence data

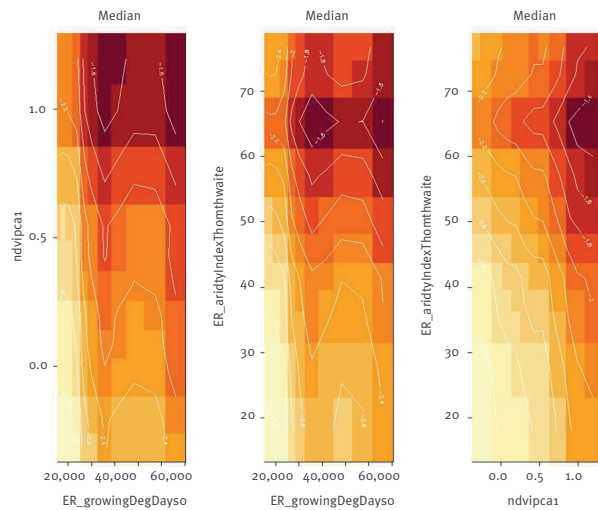
A spatial thinning procedure selecting only one presence within each pixel of the predictor variables was performed with `gridRecords()` function in `fuzzySim` package version 3.6 [30] to reduce both spatial bias and spatial autocorrelation. This function was used to obtain unique presences and absences from occurrence data (expressed as x and y coordinates) at the spatial resolution of the predictors. As absence points were not supplied, all pixels without any occurrence points were returned as pseudo-absences. Nevertheless, a random sample of the pseudo-absence pixels was selected to obtain a 1:10 ‘occurrence: pseudo-absence ratio’ in the model. This was done to avoid an excessive number of pseudo-absences compared with the number of occurrences [30].

### Spatial modelling

We built a Bayesian additive regression tree (BART) model using the `embarcadero` package version 1.2.0.1003 [31]. Under the Bayesian framework, the initial expected value (prior) is updated by the model to get the posterior distribution (predicted probabilities). Thus, for each pixel, a posterior distribution of predicted probabilities with the associated 95% credible intervals (CI) is obtained. The model was fit with all the variables to select the most important ones. The automated stepwise reduction algorithm with 50 iterations and 10 trees was used to eliminate the variables with the lowest importance and obtain the model with the lowest root mean square error (RMSE). Afterwards, the model was rerun using only the variables selected in the previous step. Variable importance plot, predictions and response plots were created with the in-built functions of the package. Finally, the probability maps were converted to favourability maps using `modEVA` package version 3.0 [32]. The percentage of pixels at different levels of risk and uncertainty were computed reclassifying the pixels values into three classes, i.e.

**FIGURE 4**

Bidimensional partial dependence plots of three variables<sup>a</sup> showing the interactions between pairs of variables influencing Crimean-Congo haemorrhagic fever occurrence probability, European Region below ~60° North latitude, 2012–2022



<sup>a</sup> Bidimensional plots for *ndvi.pca1* (principal component 1 of NDVI time series), *ER\_growingDegDayso* (sum of mean monthly temperature for months with mean temperature greater than 0°C multiplied by number of days) and *ER\_aridityIndexThornthwaite* (Thornthwaite aridity index: index of the degree of water deficit below water need). NDVI is an indicator of the greenness of the biomes.

The occurrence probability is represented by the colour progression: from white (lowest probability) to dark red (highest probability).

low ( $0 \leq x < 0.33$ ), medium ( $0.33 \leq x < 0.66$ ), and high ( $0.66 \leq x$ ), with QGIS version 3.22.4 [33].

Model performance was assessed by the area under the receiver operating characteristic curve (AUC), which measures overall discrimination capacity and threshold-dependent metrics and assesses how well the model distinguishes presence from absence. All the analyses were performed R software version 4.1.2 [34].

## Results

Three hundred and sixty-five news items published between 24 July 2012 to 22 March 2022 were retrieved from the EIOS system. These comprised 1,387 locations. Fifty-one additional locations were compiled from the meta-analysis. After removing the EIOS duplicates, the irrelevant data (news items not specifically referring to human cases or virus isolation from ticks) and uncertain information (news items without clear indication of the localisation of human cases/virus isolation from ticks), we obtained 172 records, of which further 31 were discarded as referring to administrative divisions with areas greater than 5,000 km<sup>2</sup>. Thus, 141 occurrence points were initially considered (Figure 1).

Following the spatial thinning, the final occurrence dataset comprised 136 pixels with occurrence and 1,360 pixels with pseudo-absence of CCHF. The model with the lowest average RMSE selected through the automated procedure retained the following variables in order of importance: *ER\_growingDegDayso*, *ndvi.pca1*, *ER\_aridityIndexThornthwaite*, *WC\_bio9*, *ER\_continentality*, *WC\_bio5*, *ER\_tri*, and *WC\_bio11* (see Supplementary Material A, Figure S2 A and S3 A for the variable selection and variable importance plots respectively). Only the first principal component of the NDVI was selected. This captures the major element of variability in the NDVI time series. The model performed well in terms of accuracy (AUC=0.95) (see Supplementary Material A, Figure S4 A for the model diagnostics), correct classification rate (CCR=0.80), correct prediction of presences (sensitivity/recall=0.99), correct prediction of absences (specificity=0.79), true skill statistic (sTSS=0.89), Cohen's kappa (skappa=0.69), but the positive predictive value was low (precision=0.31) (see Supplementary Material A, Figure S5 A for the model evaluation metrics). Figure 2 shows that the majority of CCHF risk is found in the Mediterranean basin and in the areas bordering the Black Sea. Considering the whole study area, the percentage of pixels at high risk was 16%, while 15% and 69% were at medium and low risk, respectively. There was a general decreasing risk trend from south to north across the entire European Region.

The uncertainty map highlights that some areas of predicted favourability are characterised by considerable uncertainty, presumably because of the lack of CCHF records in those zones (Figure 3). Nevertheless, the majority of the study area is characterised by a low level of uncertainty (57% of pixels), while only a smaller percentage of pixels are characterised by either medium (37%) or high (6%) levels of uncertainty.

Considering the variables retained, the model found a positive association between all temperature-related variables and the probability of CCHF occurrence, with an increased risk in warmer and drier areas (see Supplementary Material A, Figure S6 A for the partial dependence plots). In the model, there was no predictor that alone showed a marked and significant effect, but rather the combination of all variables influenced the risk of CCHF across the European Region. Considering the bidimensional partial dependence plots for the first three selected variables (Figure 4), the occurrence probability is remarkably high in those areas characterised by both high values of *ER\_growingDegDayso* and *ndvi.pca1*, both high values of *ER\_growingDegDayso* and *ER\_aridityIndexThornthwaite*, and both high values of *ndvi.pca1* and *ER\_aridityIndexThornthwaite*.

## Discussion

In this study, we used internet-based surveillance data retrieved from the EIOS system to model the risk of CCHF occurrence within the European Region. The

study underlines the effectiveness of the EIOS system in detecting, retrieving and compiling information from multiple sources and how these data can be used to build a disease risk map. CCHF was considered a good test case for several reasons. Firstly, the disease, which is endemic in many regions such as Africa, Asia, Eastern Europe and the Middle East, has given rise to many concerns over its spread outside the current geographical range [18]. Secondly, it is a tick-borne disease, and thus environmental variables play a pivotal role in shaping its geographical distribution and consequently its occurrence, can be well predicted through spatial modelling [19]. Thirdly, it is a neglected tropical disease posing a risk to human health [20]. Results of the model highlight that the distribution of CCHF is likely to be geographically broader than expected, and mainly driven by temperature/climate-related predictors. Not only does the model predict areas where CCHF is already established (e.g. Eastern Europe and the Black Sea region), but also those areas at risk of further CCHF expansion, given their favourable environmental conditions (e.g. Italy, France). Of note, during the revision of this work, author Fanelli and colleagues found seropositive cattle in Italy [35], confirming model predictions. It is important to stress that the presence probability predictions were converted into favourability predictions. The former are considered as the likelihood of finding the disease in a given environment depending on the probability of finding the disease given how rare or common it is (i.e. prevalence), whereas the latter do not necessarily reflect the locations where the disease is predicted to be present, but rather favourable areas for its occurrence, although the disease may be currently absent [36]. It is not surprising that the model has a low precision, meaning that it predicts positive pixels where we do not have any actual observations of human cases or virus isolation from ticks. These are represented by the locations which are likely to be favourable for CCHF occurrence, but where the disease or the circulation of the virus have yet to be recorded.

Another important consideration highlighted by this study is related to the countries where the disease seems to have a clustered distribution for which a low favourability and high uncertainty were found. One example is Georgia and its vicinity, where the presence points retrieved from the news items were limited to the plain areas of Khashuri and Tskaltubo. The low favourability and high uncertainty found in the areas close to the plain is not surprising given the complex orographic and very variable landscape-climatic conditions of the area. The low favourability is likely to be due to the mountain areas (Caucasus mountains) with lower temperatures and mostly unfavourable environmental conditions.

The overall percentage of areas at high favourability is quite low, highlighting that the risk is still low for most European countries. Interestingly, our predictions are in line with the risk estimates of authors Fanelli and Buonavoglia in an earlier publication [17]. In their

assessment, a low risk of CCHFV entry and exposure in large part of Western Europe and a medium risk for France and Italy was described. The presence of the *Hyalomma marginatum* resident population, which is the main vector of the virus, is the primary factor influencing the probability of the disease spread once the virus has been introduced in CCHF-free countries [17]. In our predictive map, the risk is high for almost the entire Italian peninsula, whereas France has the highest favourability along its Mediterranean coast. In line with this, Italy has a widespread population of *H. marginatum*, whereas France has only a hotspot area in the southern part of the country [22].

Fernández-Ruiz and Estrada-Peña [19] have recently assessed the expansion of suitable areas for *H. marginatum*, indicating that the Mediterranean countries, south-east central Europe and the southern Balkans are most likely to see future spread of the tick. Similarly, in our study, the most important variables retained by the model define the most favourable areas for the spread and colonisation of *H. marginatum*. It is important to derive the ecological meaning of the predictors retained, which are likely to play an important role in determining the suitability of areas for tick development. A relevant example is the sum of mean monthly temperature for months with mean temperature greater than 0°C multiplied by number of days, which may define the critical accumulated temperature necessary for tick spread and colonisation [19]. Another important predictor in our model is the NDVI which, at high values, contributes, along with the temperature variables, to increasing the probability of CCHF occurrence, which may be explained by the necessity of adequate relative humidity for the vector development [23].

In this study, we did not include any data on CCHF serology in humans or animals, which are unlikely to be reported in near real-time news item publications, as are the data of the EIOS system. The information on serological evidence of CCHF comes mainly from scientific literature and may be useful to complement the picture of the disease distribution [16]. It is worth mentioning that CCHFV can circulate unnoticed in numerous wild and domestic animals that are asymptomatic carriers of the virus and may act as reservoirs [37]. Cuadrado-Matías et al. [38] mapped the risk of exposure to CCHFV in Spain using serological data on red deer as an indicator of the transmission risk from infected ticks. Compared with their risk map, our predictions show a more extensive distribution of favourable environment in the Iberian Peninsula for CCHF occurrence. In particular, they found a higher risk mainly in the south-western part of the country, whereas our predictions indicate that the risk is high also along the Spain's east coast. This difference is likely to be due to the different modelling approaches adopted as well as the predictors used. Indeed, not only did Cuadrado-Matías et al. [38] use environmental variables, but also host population predictors regarding



both wild and domestic animals, which were found to be significant factors behind the risk of exposure to CCHFV. Unfortunately, this information, which could have improved our predictions, is missing or outdated at the European level. With reference to the distribution and density of domestic animals, some data are available through the Gridded Livestock of the World (GLW3) from the Food and Agriculture Organization of the United Nations (FAO), but the dataset has two major limitations: the distribution data refer to 2010 and their spatial resolution is around 10 km (coarser than our predictions) [39]. Wildlife data are available for most of the European countries, but they are not collected with consistent and homogeneous criteria, making it very difficult to use for modelling purposes [40].

It is important to highlight that the model was run on the full dataset (resubstitution). Indeed, it was not possible to set aside some data for testing given the small amount of presence points. Data splitting would have withdrawn a lot of information from the model. This should be kept in mind when considering the results of this study. Other limitations include the category definition for CCHF available in EIOS, which might have missed some news items reporting CCHFV from ticks. Despite these limitations, we feel that our model is of great value since it provides an updated picture of the risk of CCHF at the regional level, with a low level of uncertainty associated with the majority of the pixels. Our approach is unique in its kind, as we demonstrated the potential of internet-based data accessed through the EIOS system to create a disease risk map. In fact, although the traditional use of the EIOS system is to early detect potential public health threats through an event-based surveillance [41-43], their complex algorithms can be also used to automatically extract the geographical coordinates of outbreak locations, which may, in turn, be used to build spatial risk models. These have important consequences for health-related decision-making and planning effective actions in target areas [44-49].

## Conclusions

The practical implications of using epidemic intelligence tools are considerable in terms of helping design effective disease control strategies. However, they require a reasonable effort to develop adequate and unbiased algorithms, to select accurate and relevant sources, and to extract the exact geographical coordinates of disease outbreaks.

As recently mentioned in a paper on the ‘costs and benefits of primary prevention of zoonotic pandemics’ [50], current plans to address future pandemic catastrophes still mainly consider the approach of ‘detecting and containing emerging zoonotic threats’ [51]. However, much more must be done on the prevention side. Reacting to a spillover/pandemic/epidemic event after it occurs can be, in fact, much more costly than working to prevent it. In this sense, the case study presented here may constitute a model to

combine epidemic intelligence tools and advanced analytical approaches to detect and assess changes in disease distribution, allowing prevention or early mitigation of disease events of veterinary and public health importance.

## Ethical statement

Ethical approval was not required for this study as it includes information freely available in the public domain (news). This research does not contain any studies with animals or humans performed by any of the authors.

## Disclaimer

The views and opinions expressed in this article are those personal of the authors and do not necessarily reflect the official policy or position of the World Health Organization (WHO) and World Organisation for Animal Health (WOAH). Assumptions made within the analysis are not reflective of the position of the WHO and WOAH.

## Acknowledgements

The authors would like to acknowledge the work of Montserrat Arroyo (WOAH DDG) to revise the article prior to submission.

## Conflict of interest

None declared.

## Authors' contributions

Angela Fanelli conceptualized the study, investigated the study, contributed to methodology, curated the data, involved in formal analysis, and wrote—original draft preparation. Paolo Tizzani supervised the study, contributed to methodology, curated the data, reviewed and edited it. Johannes Christof Schnitzler and Alastair Donachie curated the data, reviewed and edited it. Marco De Nardi, Ilaria Capua, Gianvito Lanave, Paula Caceres-Soto and Domenico Buonavoglia reviewed, edited and validated the study.

## References

1. World Health Organization (WHO). Early detection, verification, assessment and communication. Geneva: WHO. [Accessed: 24 May 2022]. Available from: <https://www.who.int/initiatives/eios>
2. Spagnolo L, Abdelmalik P, Doherty B, Fabbri M, Ferrer M, Osato C, et al. Integration of the Epidemic Intelligence from Open Sources (EIOS) system and the INFORM suite: enhancing early warning with contextual data for informed decision making. Luxembourg: Publications Office of the European Union; 2020. Available from: <https://data.europa.eu/doi/10.2760/958918>
3. Ehrmann M, Jacquet G, Steinberger R. Multilingual entity name variants and titles as Linked Data. *Semant Web*. 2016;8(2):283-95. <https://doi.org/10.3233/SW-160228>
4. GeoNames. [Accessed: 24 May 2022]. Available from: <http://www.geonames.org>
5. Balajee SA, Salyer SJ, Greene-Cramer B, Sadek M, Mounts AW. The practice of event-based surveillance: concept and methods. *Glob Secur Health Sci Policy*. 2021;6(1):1-9. <https://doi.org/10.1080/23779497.2020.1848444>
6. Caceres P, Tizzani P, Ntsama F, Mora R. The World Organisation for Animal Health: notification of animal diseases. *Rev Sci Tech*. 2020;39(1):289-97. <https://doi.org/10.20506/rst.39.1.3082> PMID: 32729558

7. Cárdenas L, Awada L, Tizzani P, Cáceres P, Casal J. Characterization and evolution of countries affected by bovine brucellosis (1996-2014). *Transbound Emerg Dis.* 2019;66(3):1280-90. <https://doi.org/10.1111/tbed.13144> PMID: 30739391
8. Fanelli A, Buonavoglia D, Pleite CMC, Tizzani P. Paratuberculosis at European scale: an overview from 2010 to 2017. *Vet Ital.* 2020;56(1):13-21. PMID: 32343090
9. Fanelli A, Tizzani P. Spatial and temporal analysis of varroosis from 2005 to 2018. *Res Vet Sci.* 2020;131:215-21. <https://doi.org/10.1016/j.rvsc.2020.04.017> PMID: 32408232
10. Meske M, Fanelli A, Rocha F, Awada L, Soto PC, Mapitse N, et al. Evolution of rabies in South America and inter-species dynamics (2009-2018). *Trop Med Infect Dis.* 2021;6(2):98. <https://doi.org/10.3390/tropicalmed6020098> PMID: 34207822
11. Fanelli A, Tizzani P, Buonavoglia D. Crimean-Congo Haemorrhagic Fever (CCHF) in animals: Global characterization and evolution from 2006 to 2019. *Transbound Emerg Dis.* 2022;69(3):1556-67. <https://doi.org/10.1111/tbed.14120> PMID: 33949116
12. Fanelli A, Galgano M, Sposato A, Buonavoglia D. Assessment of Paratuberculosis international official reporting in Europe using the information supplied to the WOAHP by National Veterinary Services. *Vet Ital.* 2022;58(2). <https://doi.org/10.12834/VetItl.2625.16709.3> PMID: 36586107
13. World Health Organization (WHO). Global Health Observatory. Geneva: WHO. [Accessed: 24 May 2022]. Available from: <https://www.who.int/data/gho>
14. World Health Organization (WHO). Global Influenza Programme. Geneva: WHO. [Accessed: 24 May 2022]. Available from: <https://www.who.int/teams/global-influenza-programme/surveillance-and-monitoring>
15. World Health Organization (WHO). Disease Outbreak News. Geneva: WHO. [Accessed: 24 May 2022]. Available from: <https://www.who.int/emergencies/disease-outbreak-news>
16. Fanelli A, Awada L, Cáceres-Soto P, Diaz F, Grillo T, Gizo I, et al. Sensitivity of an international notification system for wildlife diseases: A case study using the OIE-WAHIS data on tularemia. *Zoonoses Public Health.* 2022;69(4):286-94. <https://doi.org/10.1111/zph.12916> PMID: 35092712
17. Fanelli A, Buonavoglia D. Risk of Crimean Congo haemorrhagic fever virus (CCHF) introduction and spread in CCHF-free countries in southern and Western Europe: A semi-quantitative risk assessment. *One Health.* 2021;13:100290. <https://doi.org/10.1016/j.onehlt.2021.100290> PMID: 34307823
18. Maltezos HC, Andonova L, Andraghetti R, Bouloy M, Ergonul O, Jongejan F, et al. Crimean-Congo hemorrhagic fever in Europe: current situation calls for preparedness. *Euro Surveill.* 2010;15(10):19504. <https://doi.org/10.2807/ese.15.10.19504-en> PMID: 20403306
19. Fernández-Ruiz N, Estrada-Peña A. Towards new horizons: climate trends in Europe increase the environmental suitability for permanent populations of *Hyalomma marginatum* (Ixodidae). *Pathogens.* 2021;10(2):95. <https://doi.org/10.3390/pathogens10020095> PMID: 33494140
20. Greene L, Uwishema O, Nicholas A, Kapoor A, Berjaoui C, Adamolekun E, et al. Crimean-Congo haemorrhagic fever during the COVID-19 pandemic in Africa: Efforts, recommendations and challenges at hand. *Afr J Emerg Med.* 2022;12(2):117-20. <https://doi.org/10.1016/j.afjem.2022.02.004> PMID: 35223387
21. Messina JP, Pigott DM, Golding N, Duda KA, Brownstein JS, Weiss DJ, et al. The global distribution of Crimean-Congo hemorrhagic fever. *Trans R Soc Trop Med Hyg.* 2015;109(8):503-13. <https://doi.org/10.1093/trstmh/trv050> PMID: 26142451
22. European Centre for Disease Prevention and Control (ECDC). *Hyalomma marginatum* - current known distribution: March 2021. Stockholm: ECDC; 2021. Available from: <https://www.ecdc.europa.eu/en/publications-data/hyalomma-marginatum-current-known-distribution-march-2021>
23. Estrada-Peña A, Estrada-Sánchez A, de la Fuente J. A global set of Fourier-transformed remotely sensed covariates for the description of abiotic niche in epidemiological studies of tick vector species. *Parasit Vectors.* 2014;7(1):302. <https://doi.org/10.1186/1756-3305-7-302> PMID: 24984933
24. World Health Organization (WHO). Geographic distribution of Crimean-Congo Haemorrhagic Fever. Geneva: WHO; 2017. Available from [https://www.who.int/health-topics/crimean-congo-haemorrhagic-fever#tab=tab\\_1](https://www.who.int/health-topics/crimean-congo-haemorrhagic-fever#tab=tab_1)
25. Belobo JTE, Kenmoe S, Kengne-Nde C, Emoh CPD, Bowo-Ngandji A, Tchatchouang S, et al. Worldwide epidemiology of Crimean-Congo Hemorrhagic Fever Virus in humans, ticks and other animal species, a systematic review and meta-analysis. *PLoS Negl Trop Dis.* 2021;15(4):e0009299. <https://doi.org/10.1371/journal.pntd.0009299> PMID: 33886556
26. Bosch S, Fernandez S. *sdm*predictors: Species Distribution Modelling Predictor Datasets. R package version 0.212. 2022. <https://CRANR-project.org/package=sdmpredictors>
27. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol.* 2005;25(15):1965-78. <https://doi.org/10.1002/joc.1276>
28. Title PO, Bemmels JB. ENVIREM: an expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling. *Ecography.* 2018;41(2):291-307. <https://doi.org/10.1111/ecog.02880>
29. Copernicus Global Land Service. Normalized Difference Vegetation Index. Brussels: European Commission Joint Research Centre (JRC). [Accessed: 24 Mar 2022]. Available from: <https://land.copernicus.eu/global/products/ndvi>
30. Báez JC, Barbosa AM, Pascual P, Ramos ML, Abascal F. Ensemble modeling of the potential distribution of the whale shark in the Atlantic Ocean. *Ecol Evol.* 2019;10(1):175-84. <https://doi.org/10.1002/ece3.5884> PMID: 31988721
31. Carlson CJ. embarcadero: Species distribution modelling with Bayesian additive regression trees in R. *Methods Ecol Evol.* 2020;11(7):850-8. <https://doi.org/10.1111/2041-210X.13389>
32. Márcia Barbosa A, Real R, Muñoz A-R, Brown JA. New measures for assessing model equilibrium and prediction mismatch in species distribution models. *Divers Distrib.* 2013;19(10):1333-8. <https://doi.org/10.1111/ddi.12100>
33. QGIS Development Team. QGIS: a free and open source geographic information system. Open Source Geospatial Foundation; 2022. Available from: <http://qgis.osgeo.org>
34. R Core Team. A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna: R Foundation for Statistical Computing; 2021. Available from: <https://www.R-project.org>
35. Fanelli A, Buonavoglia D, Lanave G, Monaco F, Quaranta V, Catanzariti R, et al. First serological evidence of Crimean-Congo haemorrhagic fever virus in transhumant bovines in Italy. *Transbound Emerg Dis.* 2022;69(6):4022-7. <https://doi.org/10.1111/tbed.14710> PMID: 36150076
36. Acevedo P, Real R. Favourability: concept, distinctive characteristics and potential usefulness. *Naturwissenschaften.* 2012;99(7):515-22. <https://doi.org/10.1007/s00114-012-0926-0> PMID: 22660474
37. Whitehouse CA. Crimean-Congo hemorrhagic fever. *Antiviral Res.* 2004;64(3):145-60. <https://doi.org/10.1016/j.antiviral.2004.08.001> PMID: 15550268
38. Cuadrado-Matías R, Cardoso B, Sas MA, García-Bocanegra I, Schuster I, González-Barrio D, et al. Red deer reveal spatial risks of Crimean-Congo haemorrhagic fever virus infection. *Transbound Emerg Dis.* 2022;69(4):e630-45. <https://doi.org/10.1111/tbed.14385> PMID: 34739746
39. Gilbert M, Nicolas G, Cinardi G, Van Boeckel TP, Vanwambeke SO, Wint GRW, et al. Global distribution data for cattle, buffaloes, horses, sheep, goats, pigs, chickens and ducks in 2010. *Sci Data.* 2018;5(1):180227. <https://doi.org/10.1038/sdata.2018.227> PMID: 30375994
40. Grignolio S, Apollonio M, Brivio F, Vicente J, Acevedo P, Palencia P, et al. Guidance on estimation of abundance and density data of wild ruminant population: methods, challenges, possibilities. *EFSA Support Publ.* 2020;17(6):EN1876. <https://doi.org/10.2903/sp.efsa.2020.EN-1876>
41. Williams GS, Impouma B, Mboussou F, Lee TMH, Ogundiran O, Okot C, et al. Implementing epidemic intelligence in the WHO African region for early detection and response to acute public health events. *Epidemiol Infect.* 2021;149:e261. <https://doi.org/10.1017/S095026882100114X> PMID: 33985609
42. Ibrahim NK. Epidemiologic surveillance for controlling Covid-19 pandemic: types, challenges and implications. *J Infect Public Health.* 2020;13(11):1630-8. <https://doi.org/10.1016/j.jiph.2020.07.019> PMID: 32855090
43. Yan SJ, Chughtai AA, Macintyre CR. Utility and potential of rapid epidemic intelligence from internet-based sources. *Int J Infect Dis.* 2017;63:77-87. <https://doi.org/10.1016/j.ijid.2017.07.020> PMID: 28765076
44. Iacopelli F, Fanelli A, Tizzani P, Berriatua E, Prieto P, Martínez-Carrasco C, et al. Spatio-temporal patterns of sarcoptic mange in red deer and Iberian ibex in a multi-host natural park. *Res Vet Sci.* 2020;128:224-9. <https://doi.org/10.1016/j.rvsc.2019.11.014> PMID: 31837510
45. Fanelli A, Tizzani P, Belleau E. Gastrointestinal parasitic infestation in the Rock ptarmigan *Lagopus muta* in the French Alps and French Pyrenees based on long-term sampling (1987-2018). *Parasitology.* 2020;147(7):828-34. <https://doi.org/10.1017/S0031182020000517> PMID: 32234098
46. Sanchis-Monsonís G, Fanelli A, Tizzani P, Martínez-Carrasco C. First epidemiological data on *Spirocerca vulpis* in the red fox:

- A parasite of clustered geographical distribution. *Vet Parasitol Reg Stud Rep.* 2019;18:100338. <https://doi.org/10.1016/j.vprsr.2019.100338> PMID: 31796180
47. Johnson EE, Escobar LE, Zambrana-Torrel C. An ecological framework for modeling the geography of disease transmission. *Trends Ecol Evol.* 2019;34(7):655-68. <https://doi.org/10.1016/j.tree.2019.03.004> PMID: 31078330
  48. Sanchis-Monsonís G, Fanelli A, Martínez-Carrasco C, Tizzani P. The typical cestodes of the red fox in eastern areas of the Iberian Peninsula have a grouped distribution. *Vet Parasitol.* 2020;283:109168. <https://doi.org/10.1016/j.vetpar.2020.109168> PMID: 32585533
  49. Ortega N, Fanelli A, Serrano A, Martínez-Carrasco C, Escribano F, Tizzani P, et al. Salmonella seroprevalence in wild boar from Southeast Spain depends on host population density. *Res Vet Sci.* 2020;132:400-3. <https://doi.org/10.1016/j.rvsc.2020.07.026> PMID: 32763568
  50. Bernstein AS, Ando AW, Loch-Temzelides T, Vale MM, Li BV, Li H, et al. The costs and benefits of primary prevention of zoonotic pandemics. *Sci Adv.* 2022;8(5):eabl4183. <https://doi.org/10.1126/sciadv.abl4183> PMID: 35119921
  51. Elias C, Nkengasong JN, Qadri F. Emerging infectious diseases - learning from the past and looking to the future. *N Engl J Med.* 2021;384(13):1181-4. <https://doi.org/10.1056/NEJMp2034517> PMID: 33793147

### License, supplementary material and copyright

This is an open-access article distributed under the terms of the Creative Commons Attribution (CC BY 4.0) Licence. You may share and adapt the material, but must give appropriate credit to the source, provide a link to the licence and indicate if changes were made.

Any supplementary material referenced in the article can be found in the online version.

This article is copyright of the authors or their affiliated institutions, 2023.