

## Studying the global distribution of infectious diseases using GIS and RS

David J. Rogers and Sarah E. Randolph

West Nile virus, severe acute respiratory syndrome and monkeypox are infectious diseases that have recently been introduced into areas far from their region of origin. The greatest risk of new diseases comes from zoonoses — pathogens that circulate among wild animals and are occasionally transferred to humans by intermediate invertebrate hosts or vectors that are sensitive to climatic conditions. Analytical tools that are based on geographical information systems and that can incorporate remotely sensed information about the environment offer the potential to define the limiting conditions for any disease in its native region for which there are at least some distribution data. The direction, intensity or likelihood of its spread to new regions could then be predicted, potentially allowing disease early-warning systems to be developed.

West Nile virus first appeared in the New World in 1999, in New York, USA. By the end of 2002, this virus had spread to Canada and 44 states of the USA (plus the District of Columbia) and, in 2002 alone, caused clinical infections in more than 4,000 people and 284 deaths<sup>1</sup>.

Severe acute respiratory syndrome (SARS) was first recognized in China in mid-November 2002. By mid-June 2003, it had affected more than 8,000 people, killing 790, in 33 countries of all the inhabited continents of the world<sup>2</sup>.

Monkeypox, which is related to smallpox, is known historically only in Africa, but was

(probably) imported into Texas, USA, in early April 2003, in a shipment of Gambian rats that were destined for the pet trade. Transfer to native prairie dogs, which were in turn shipped to other pet traders or were displayed and handled in pet fairs, carried monkeypox to Wisconsin, Indiana, Illinois, Missouri, Kansas and Ohio, and resulted in a total of 71 human infections in these states by early July 2003 (REF. 3).

In all of these examples, little or nothing is known about the natural circulation of the pathogens concerned within their countries of origin, or about the precise method, or likelihood, of their transfer to new areas. For every 'new' or 'emerging' disease of this type, there could be many others that are carried by the same mechanisms, but which then fail to establish in new countries for reasons that are also unclear.

These examples, and many more like them, show that the modern equivalent of the spread of cholera from the Broad Street pump to Londoners in the 1840s (REF. 4) could now encompass the world and its population. 'Small-world' effects increasingly 'connect' geographically distant places<sup>5</sup> — affecting not only knowledge, trade and tourism, but also infectious diseases. The immediate questions to be answered are which diseases will be involved in such transfers, from which sources, to which destinations and by what routes; and what will be the chances of the diseases establishing and spreading from any point of introduction?

A geographical information system (GIS), which can hold both the disease data and any other information within the same

geographical framework, has the analytical power to help answer these questions. When the diseases in question are known, or suspected, to have environmental risk factors, the addition of remotely sensed (RS) environmental data to the GIS greatly enhances its explanatory power<sup>6</sup>. This article explains how using RS data in a GIS can help us understand the spatio-temporal dynamics of a wide range of disease systems, especially, but not exclusively, those with environmental correlates.

### Steps to understanding epidemiology

To predict where diseases might spread, it is necessary to understand their epidemiology in the regions in which they have been historically recorded. First, we must identify the pieces of the puzzle — the pathogen, its vertebrate hosts and the routes of transmission between hosts. Second, the patterns of each disease must be recorded — both the distribution in space and changes with time. An essential part of this second step is an appreciation of the environment in which transmission is taking place. Third, we need to understand the dynamic processes of transmission that ultimately determine the patterns that are observed. Given this baseline knowledge, we might finally be able to estimate the likelihood of pathogen spread to, and establishment in, new areas.

**Diagnosis.** The first task is essentially one of diagnosis, a task that is not always simple in complex natural biological systems. Identifying the pieces of the puzzle can lead to an expectation of the resulting patterns, but it can be difficult to anticipate the number of important pieces that are involved. For example, it might have been expected that West Nile virus in North America would be transmitted by only a few native American mosquito species (the vectors). A pathogen in a new continent is unlikely to find many competent vectors, and many flaviviruses (the family to which West Nile virus belongs) typically have only one

Box 1 | **Satellites for epidemiology**

Satellites provide global data on environmental conditions at spatial resolutions that cannot be achieved by meteorological networks. Satellite sensors detect reflected visible or infra-red radiation within the 0.3–14  $\mu\text{m}$  region of the electromagnetic spectrum, but design constraints dictate certain combinations of spatial, temporal and spectral resolutions, and therefore different uses of data from different satellite types.

- *Earth-observing satellites.* These produce data with spatial resolutions (that is, pixel sizes) of 1–4 m (the Ikonos 2 satellite, which has four multispectral channels plus one panchromatic channel), 2.5–20 m (the Satellite pour l'Observation de la Terre (SPOT)); four multispectral channels, 1 panchromatic channel and 4 coarser resolution channels of 1.1 km), 30–120 m (the Landsat 1–5 satellites; up to seven channels) or 15–60 m (the Landsat 7 satellite; seven multispectral channels plus one panchromatic channel), but low repeat frequencies of only 11 (Ikonos), 26 (SPOT) or 16 (Landsat) days.
- *Orbiting oceanographic or atmospheric satellites.* These satellites have lower spatial resolutions, as low as 1.1 km (the National Oceanographic and Atmospheric Administration Advanced Very High Resolution Radiometer (NOAA AVHRR); six channels), but they produce two images per day of the entire Earth's surface. This satellite series has given us a ~20-year set of continuous observations, but has now been effectively replaced by the Terra and Aqua spacecraft with their Moderate Resolution and Imaging Spectrometer (MODIS; 36 channels with 0.25–1 km resolution, and a 1–2 day return time) and, in the case of the Terra satellites, the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER; 14 channels with 15 m–90 m resolution, 16-day return time; see [Terra satellite](#) in the Online links).
- *Geostationary weather satellites.* These satellites maintain a constant position relative to the earth, and give spatial resolutions of 1–8 km (the Geostationary Operational Environmental Satellite (GOES) for the Americas; five channels) or 2.5–5 km (the Meteosats 4–6 for Europe and Africa; three channels) and two images of the entire Earth half-disk each hour. The Meteosats 4–6 will be superseded by the recently commissioned Meteosat Second Generation (MSG) satellite with its Spinning Enhanced Visible and Infrared Imager (SEVIRI; 12 channels, half-disk images every 15 minutes; see the [European Space Agency](#) in the Online links).

Atmospheric contamination, such as clouds, commonly obscures images, so low-frequency images provide only occasional snapshots that cannot record the important seasonal determinants of pathogen transmission rates. By contrast, high-frequency images can be combined to produce relatively cloud-free monthly 'maximum-value composites'<sup>39</sup>, that are of greater use in studying dynamic epidemiological processes.

The processed satellite sensor data can be used directly, but are often combined to produce indices that are related to ground-based variables relevant to epidemiological events. These include land surface temperature (LST) data, which are related to the Earth's surface temperature; the normalized difference vegetation index (NDVI), which is related to plant photosynthetic activity and is a useful correlate of ground moisture conditions; near-surface air temperature (T<sub>v</sub>X); and cold cloud duration (CCD) correlated with rainfall (all reviewed in REF. 39).

principal vector species and relatively few additional species of any transmission importance in any region<sup>7</sup>. By August 2003, however, West Nile virus, its RNA or antigens had been detected in 43 mosquito species in the United States<sup>8</sup> (although this does not prove full transmission competence). Had this large number of potential vector species been known or predicted in advance, the rapid spread of West Nile virus in North America would not have been so surprising.

*Disease patterns in space and time — statistical models.* The second set of problems — that of the distribution, incidence or prevalence of a disease — is usually tackled using statistical techniques. As empirical disease data are

typically sparse, even for well-established, perennial diseases, researchers seek to make predictive maps to fill in the gaps using richer environmental data. All available information is gathered — for example, records of the known presence or intensity of the disease from areas as geographically extensive, and covering as wide a range of natural environments, as possible. This information is stored within a GIS that is, at its core, little more than a database that also records the geographical location of each observation<sup>9</sup>. In addition, the GIS must contain the environmental and other data in the same geographical framework, thereby allowing correlations to be established between these data and the disease data<sup>10</sup>.

Whilst the database origins allow all the usual calculations of summary or averaged data, the geographical essence of GIS allows them to be used to reveal the spatio-temporal structure of disease cases. For example, disease clusters — the geographical co-occurrence of cases that indicates local transmission of infectious diseases or the presence of environmental determinants of non-infectious diseases — are common. Childhood arthritis is not traditionally thought of as infectious, but the recognition of a cluster of cases in Old Lyme, Connecticut in the late 1970s<sup>11</sup> (albeit without the benefit of a GIS) ultimately led to the discovery of tick-borne Lyme borreliosis<sup>12</sup>, the most widespread and prevalent vector-borne infection in the northern temperate world.

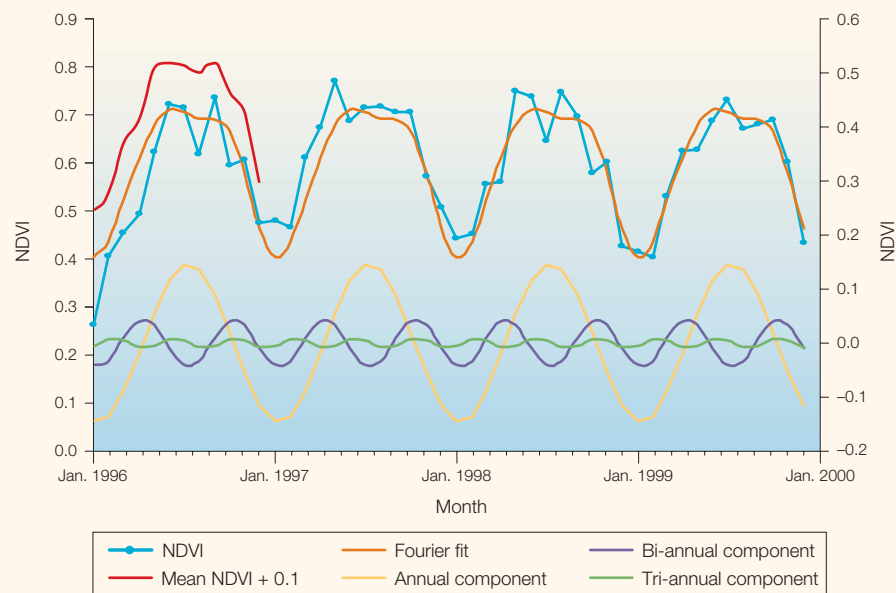
*Satellite imagery for monitoring environments.* Satellite imagery (BOX 1) is a powerful component of modern disease GISs. Satellite sensors provide data from which information about rainfall, temperature, humidity and vegetation conditions at the Earth's surface can be derived. These conditions are crucial for the indirect transmission of pathogens by vectors or intermediate hosts, such as insects, ticks, snails or rodents<sup>6,13,14</sup>. Environmental conditions might also be important for any directly transmitted (for example, host-to-host) pathogens that must survive for any period of time outside the host. Bovine tuberculosis outbreaks in cattle herds are thought by many to be caused by contamination from infected wild animals, and high-risk areas have been predicted accurately using seasonal features of atmospheric humidity and air temperature<sup>15</sup>. This result might offer some clues about the precise route of transmission, which is still unknown.

*Seasonal changes and risk maps.* For many pathogens, and especially for those with intermediate hosts, transmission is seasonal. Furthermore, the continuing existence of the pathogen depends not just on conditions during the transmission season (usually spring through to autumn in temperate regions, and the rainy season in tropical regions), but also on conditions during winter or the dry season — these conditions can determine the survival of the pathogen or its vector from one year to the next. Satellite images with high spatial resolution, such as Landsat imagery, are not recorded sufficiently often to capture the full details of seasonal cycles; their infrequent images often miss important periods.

Satellite imagery at high temporal resolution (BOX 1), however, can produce clear monthly pictures (although at the expense of

spatial resolution), but the result is large volumes of data, which often show strong serial correlations that affect the power of statistical analyses. To reduce the volume of data and remove these serial correlations without losing the biologically meaningful signals, a technique of time-series analysis that was invented by the French mathematician Joseph Fourier (1768–1830) is used. Fourier solved a problem in calculus that had defeated Newton and subsequent generations of mathematicians, by showing that a complex time series can always be expressed as the sum of a series of sine curves with different amplitudes, frequencies and phases (that is, timings) around a characteristic mean. Using Fourier's techniques, therefore, it is possible to extract information about the annual, bi-annual and tri-annual cycles of rainfall, temperature and other parameters that characterize the natural environments of diseases from the multi-temporal satellite data (FIG. 1) (it is this shift from the time to the frequency domain that removes the serial correlation in the satellite data). The output of temporal Fourier analysis is a set of orthogonal (that is, uncorrelated) variables that capture the seasonality that is of interest in epidemiology<sup>6,13,16</sup>, and these variables can therefore be used to classify habitats and describe vector and pathogen distributions. For disease systems, Fourier variables are the environmental equivalent of the genes of individual pathogens, and whole Fourier-processed images (FIG. 2) that capture all the interactive space–time features of a habitat can be likened to the organismal genome.

In a GIS, the statistical relationships that are established between the disease and environmental data sets are applied at the full spatial resolution of the latter, richer data sets, to produce a 'risk map' (FIG. 3). This effectively shows the similarity of environmental conditions in unsurveyed places to environmental conditions in which the disease has been recorded as being either present or absent. This similarity is usually expressed as the probability with which each area on the ground (corresponding to one picture element, or 'pixel', of the satellite imagery) belongs to the class of areas that are known to contain the disease. Errors in risk maps arise for several reasons; sometimes the input data are out-of-date, or simply wrong; sometimes the explanatory variables that are used are inappropriate; and sometimes the model itself is wrong. Only false-negatives (that is, false predictions of absence) are clear indicators of an incorrect or inappropriate model, and much can be learnt by investigating why these arise. Even when none of these applies, risk maps often indicate

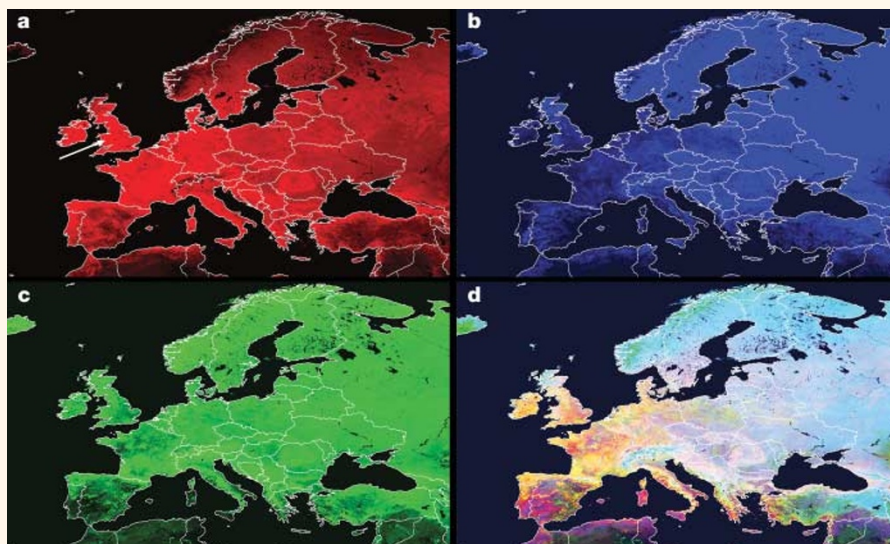


**Figure 1 | Normalized difference vegetation index signal from a point in central Wales, UK.** The monthly normalized difference vegetation index (NDVI) signal from a point in central Wales, UK (3.58°W, 52.34°N) for 1996–1999 is shown in light blue, the mean monthly signal for this period is shown in red (displaced vertically by 0.1 for clarity), the temporal Fourier fit to these data is shown in orange, and the annual, bi-annual and tri-annual components of this fit are shown in yellow, purple and green, respectively (right-hand scale). The temporal Fourier fit describes the average annual cycle at this site well.

larger areas as 'at risk' than are known to be affected by the disease at present (that is, many false-positives), because an organism or disease will not occupy all 'suitable' habitats. False-positives also highlight an important application of predictive risk maps — to warn health agencies of the potential spread of a disease

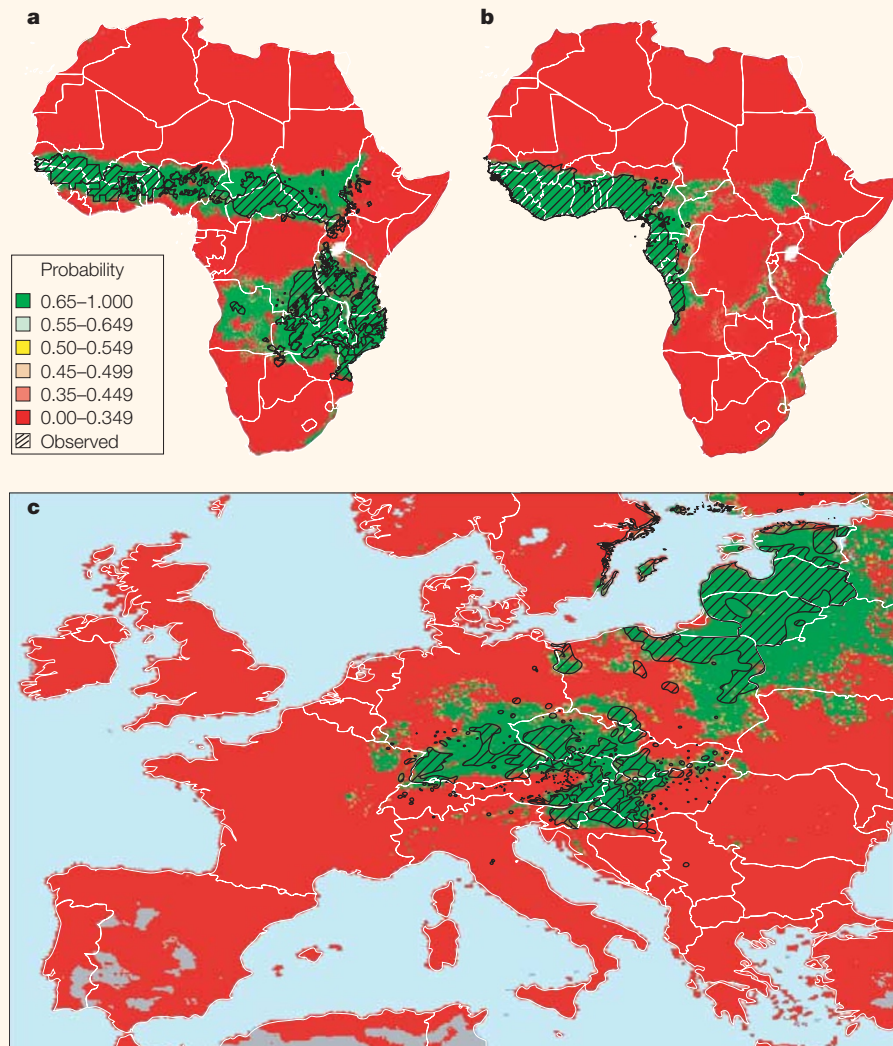
into these areas<sup>17</sup>. The significant increases in the incidence of tick-borne encephalitis (TBE) over the past two decades have indeed been accompanied by new records in many supposedly 'false-positive' regions<sup>18–21</sup>.

With every powerful technique there is a disadvantage. The ability of GISs to produce



**Figure 2 | Normalized difference vegetation index data (NDVI) Fourier images of Europe.** **a** | The mean is shown in red. **b** | The annual amplitude is shown in blue. **c** | The annual phase (that is, the timing of the annual peak) is shown in green. **d** | All three signals are shown together, which shows how this method of analysis captures habitat seasonality across Europe (notice, for example, that the mean is generally higher in western Europe, with the exception of many parts of Spain, but the annual amplitude is higher in eastern Europe). The white arrow in the mean image points to the site from which the data for FIG. 1 were obtained.





**Figure 3 | Results of statistical modelling of the distribution of vectors and disease using selected temporal Fourier-processed images as predictor variables. a** | The distribution of *Glossina morsitans* (three subspecies) throughout Africa. **b** | The distribution of *Glossina palpalis* (two subspecies) in West Africa. **c** | The distribution of tick-borne encephalitis in Europe (historical mapped data is shown cross-hatched in black). The predicted habitat suitability for the vectors or disease are on a probability scale from zero (red) to 1 (green) (see inset for legend); **(a)** 91% correct with 7% false-positives (that is, false predictions of presence) and 2% false-negatives (that is, false predictions of absence); sensitivity 0.96, specificity 0.857; **(b)** 96% correct with 3% false-positives and 1% false-negatives; sensitivity 0.97, specificity 0.94; **(c)** 90.5% correct with 8.1% false-positives, 1.4% false-negatives; sensitivity 0.966, specificity 0.865. Modified with permission from REFS 40,41 © (2000) Elsevier Science.

multiple, derived layers of data often compounds any errors that are present in each contributory variable. This is especially likely when no attempt is made to verify these derived layers through field observations (for example, checking vegetation classifications that are based on satellite data), and the effects are especially severe when the errors are systematic. The production of visually appealing, even statistically sound, results that do not reveal anything useful about either pattern or process is perhaps the greatest danger facing newcomers to this powerful technology.

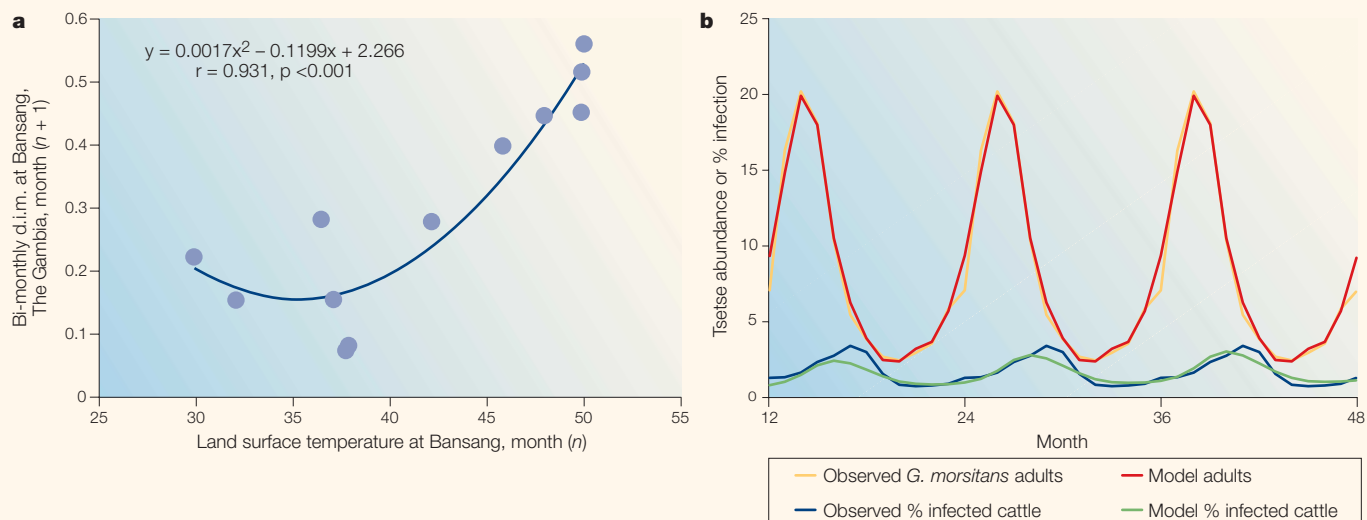
**Longer-term changes.** With historical, geo-referenced disease records over several (preferably many) years, time-series analysis can be applied to make predictions about the future of the disease. Long-term records of disease are relatively rare, and usually apply to point or small-area locations. The problem with statistical approaches (for example, various auto-regressive techniques) to these types of data is that they assume that the transmission dynamics of disease in the future will be similar to those of the past. In reality, pathogen–host systems are always evolving and their environments are changing, which

makes the future unpredictable using such statistical approaches.

### Infection processes in space and time

Satellite imagery at high temporal resolution is also a powerful tool in the third step of understanding epidemiology using dynamic biological or process-based models. These models capture details of pathogen transmission in a series of mathematical equations that describe the birth and death of organisms, and transmission processes, directly. The same remotely sensed data that are used to describe the patterns of infection in statistical models can also be used to explain transmission in biological models. For example, satellite data have been correlated with key behavioural or demographic rates of the vectors (FIG. 4a) of indirectly transmitted diseases, thereby identifying the conditions that are best for pathogen transmission. A model of a tsetse-fly population, which is driven by satellite-derived variables, captures 93% of the observed seasonal pattern (FIG. 4b). Given the extra layers of complexity that are involved in pathogen transmission by vectors, it is not surprising that the model of trypanosomiasis seasonal prevalence, which is driven, in turn, by this tsetse model, gives a less good fit (62%) to the disease data (FIG. 4b). In this case, the model assumed that cattle supply only 30% of the blood meals for the tsetse, but are the only significant source of trypanosome infections for the flies, when in reality wildlife species also play a part in transmission<sup>22</sup>. This emphasizes our first point — the importance of identifying all the pieces of the puzzle before we can make reliable predictions.

The persistence of an infectious agent can also be predicted from the widely used  $R_0$  equation, which arises from rearranging the transmission equations to determine whether, at very low prevalence in an entirely susceptible population, an infection will increase or decline to extinction<sup>23</sup>. Given that six out of the seven parameters and variables of the  $R_0$  equation for vector-borne pathogens are determined not by the vertebrate host but by the vector<sup>24</sup>, which is highly susceptible to climatic influences, it is possible to produce  $R_0$  maps for vector-borne diseases using satellite imagery; however, the first attempts have not yet been validated. A map describing the relative transmission risk of the directly transmitted foot and mouth disease in the United Kingdom revealed considerable spatial heterogeneity<sup>25</sup>, but the mapping was not based on any environmental variables that might



**Figure 4 | Biological models driven by remotely sensed data.** **a** | The relationship between the mortality rate of the tsetse *Glossina morsitans submorsitans* at Bansang, The Gambia, West Africa, and the satellite-derived land surface temperature for the previous month. Mortality is the two-point (previous, present month) smoothed density-independent mortality (d.i.m.) that is calculated from regular trapped samples of flies. **b** | A least-squares-fitted biological tsetse population model (red line), which is driven by satellite data by the relationship shown in **(a)**, describes 93% of the variance of the observed mean monthly field data (yellow line). The fitted tsetse model was then used in a simple disease transmission model (lower, green line) and describes 62% of the variance of the observed field data (blue line). Field data are the average monthly values from the sample sites, and are repeated over a 3-year period in the figure for clarity. Modified with permission from REF. 40 © (2000) Elsevier Science.

help explain this spatial pattern. A map of fully parameterized  $R_0$  values would show potential sources ( $R_0 \geq 1$ ) and sinks ( $R_0 < 1$ ) of infection, and would be of obvious benefit to control efforts, because only sources of infection require control, and only to levels where they also become sinks.

#### Climate and global spread of disease

Taken on a global scale, these same ideas can identify potential sources of infection for new disease outbreaks elsewhere. As there are so few global disease data sets, we are generally forced to use one or other statistical approach to disease mapping. Climatic similarities between areas that are, at present, disease-affected and disease-free, can be used as a guide to the environmental permissiveness of these latter areas for the diseases in question. Like risk mapping, however, climate matching is a probabilistic exercise; similarity of climates does not guarantee similarity of diseases, but simply makes it more likely, all other things being equal. The present-day match of the UK climate to global climates (FIG. 5) indicates that UK inhabitants might expect to share climatically determined diseases not just with parts of mainland Europe, but also with China and North America. Given suitable connections by trade or travel<sup>26</sup>, if only one of these climatically matched regions is a source of emergence of any new disease — as China is for new strains of influenza — all of these regions could potentially suffer from it.

Much attention has been focused recently on the impacts of past and future climate changes on diseases<sup>27,28</sup>. To identify causality rather than coincidence in past events as a prerequisite for future predictions, we must have evidence of significant changes both in the climatic drivers of infection and in the incidence of the diseases themselves, at the same times, in the same places and in the right ‘direction’ according to our understanding of how climate affects transmission. The step increase in the incidence of TBE in Sweden from 1983 to 1986 (REF. 29), for example, preceded the sudden increase in temperature at the end of the 1980s (REF. 21), showing that milder winters and springs cannot have been the causal factors, as claimed<sup>30</sup>. The analytical power in a GIS is the obvious way<sup>9</sup> to explore which of many disparate factors — biological and non-biological — might be causal for spatially heterogeneous changes in disease incidence on continental scales, as has been recorded for TBE in Europe<sup>31</sup> and malaria in Africa<sup>32</sup>.

To predict future impacts, future climate scenarios must be obtained at a sufficiently high spatial resolution to be useful, together with good models of climate and disease linkages. Unfortunately, at present, the spatial resolution of both past and future climate data sets is inadequate; the first is owing to historically inadequate meteorological station coverage in many (mostly

tropical) areas of the world<sup>33</sup>, and the second is a consequence of the computing requirements of global circulation models<sup>34</sup>. We can do little about the historical climatic data sets, and computing power is not predicted to increase sufficiently rapidly to address the second problem in the near future. Climatically speaking, we know neither where we are coming from, nor where we are headed, with anything like the precision that is required to predict spatially variable, climatically sensitive diseases.

#### Conclusions

With today’s global connectivity, humans are now the new vectors of infectious agents, travelling, or trading over, greater distances at faster speeds than can be achieved unaided by any pathogen or its vectors. The consequent increasingly urgent need to develop global disease early-warning systems (DEWS) is fortunately matched by rapidly developing tools: RS imagery within a GIS now offers greater power for describing, explaining and predicting epidemiological phenomena than ever before. Furthermore, although global routes of infection are increasingly common, global information flow has increased in parallel, thanks to the internet, and information can be transmitted even faster than any pathogen. In the most recent example, once the international community became aware of SARS, the response in laboratories and health agencies

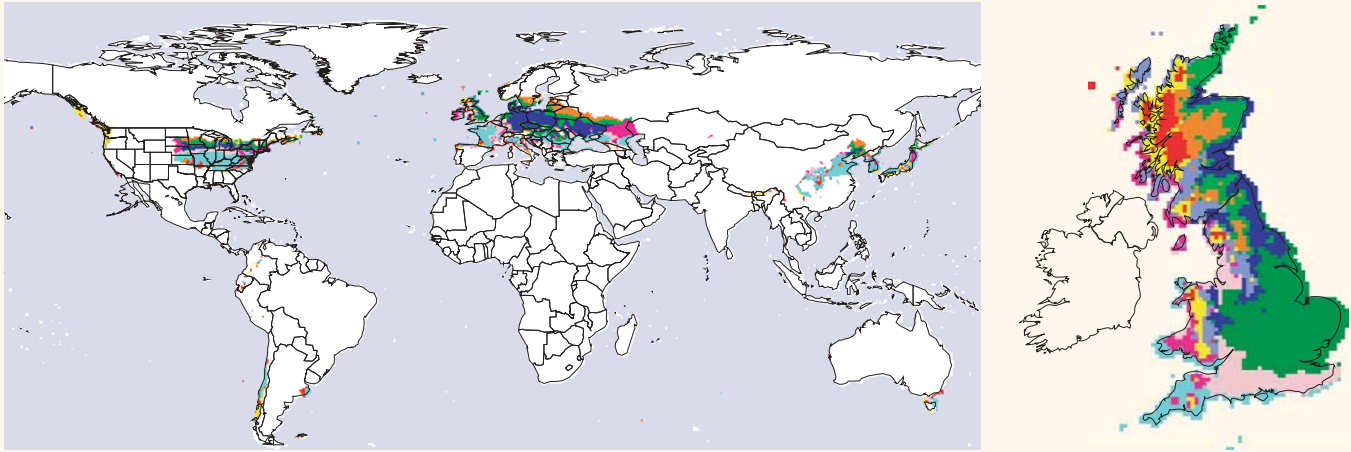


Figure 5 | **The use of meteorological data averages to define ten climate zones in Great Britain.** The mean, maxima and minima of temperature, rainfall and vapour pressure measurements from gridded meteorological data averages for 1961–1990 (REF. 33) were clustered within an image-processing system (ERDAS Imagine © Leica Geosystems, ISODATA clustering) to define ten climate zones in Great Britain (see map on right, arbitrary colours). The global map shows how matches to these climate zones occur in Europe, North America and China, with smaller area matches elsewhere. White indicates no significant matching. Reproduced with permission from REF. 42 © (2002). Crown copyright material is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland.

around the world — which resulted in an unprecedented rapid identification of the agent, its routes of transmission and the networks of connectivity — apparently achieved a (temporary?) halt to its progress within much less than a year. By contrast, however, although the response to the introduction of West Nile virus in the United States was equally impressive (a delay of less than 2 months between the first (unrecognized) human case in New York and identification of the virus), the spread of this disease has not been stopped, and it must now be regarded as endemic throughout much of the United States. The differences between these two diseases can be attributed to many factors, but all will be embedded within one or other parameter or variable of the transmission and  $R_0$  equations.

Finally, we highlight two key issues in epidemic disease prediction that have pervaded this review — space and time. Although epidemiological surveillance networks, or global satellite systems, constrain the collection of data to particular intervals of time (the reporting period) or space (the pixel size), diseases do not respect such artificial units. We must develop methods that allow us to move seamlessly across scales of time and space as we analyse epidemics that occur throughout the spectrum from small isolated communities to the global population. Classical time-series analysis of disease data, or Fourier analyses of temporal satellite series, assume a constancy (mathematical stationarity) in the disease system or the environment that is rarely shown in practice. Wavelet techniques, which essentially divide

time into smaller units, allow unit-to-unit variation of disease–environment interactions<sup>35,36</sup>; and other advanced techniques allow continuous variation in these interactions<sup>37</sup>. Similar techniques can be applied in the spatial domain to satellite imagery. The enormous range of pixel sizes that are available in commercial satellite imagery (from 0.6 m to >1 km) challenges us to develop methods to identify the appropriate data resolutions, and to integrate these data into spatial units that are relevant to disease transmission, such as the movement range of an infected individual for a directly transmitted pathogen, or the flight range of an insect for a vector-borne pathogen. Again, wavelet and other techniques might help here, by enabling us to look at the environment from the perspective of disease systems that are driven more by biological realism than by technological wizardry. A unified space–time theory of epidemics, in which a three-dimensional space–time pixel replaces a two-dimensional one as the basic recording unit, is still some way off<sup>38</sup>, but GIS and RS technologies are the vehicles that, potentially, will get us there the fastest. All we have to decide is in which direction to go.

David J. Rogers is at the TALA Research Group, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

Sarah E. Randolph is at the Oxford Tick Research Group, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

e-mails: david.rogers@zoo.ox.ac.uk; sarah.randolph@zoo.ox.ac.uk

doi:10.1038/nrmicro776

1. CDC. West Nile Virus 2002 case count. [online], (cited 14 Oct 2003), <http://www.cdc.gov/ncidod/dvbid/westnile/surv&controlCaseCount02.htm> (2003).
2. Pearson, H., Clarke, T., Abbott, A., Knight, J. & Cyranoski, D. SARS What have we learned? *Nature* **424**, 121–126 (2003).
3. Zacharyczuk, C. Monkeypox detected in the Americas. [online], (cited 14 Oct 2003), <http://www.infectiousdiseasesnews.com> (2003).
4. Snow, J. Cholera and the water supply in the southern districts of London. *BMJ* **2**, 864 (1857).
5. Watts, D. J. *Small Worlds* (Princeton Univ. Press, New Jersey, 1999).
6. Hay, S. I., Randolph, S. E. & Rogers, D. J. (eds) *Remote Sensing and Geographical Information Systems in Epidemiology* (Academic Press, London, 2000).
7. Monath, T. P. (ed.) *The Arboviruses: Epidemiology and Ecology* (CRC Press, Boca Raton, Florida, 1989).
8. CDC Division of Vector-Borne Infectious Diseases: West Nile Virus. Mosquitoes. [online], (cited 14 Oct 2003), <http://www.cdc.gov/ncidod/dvbid/westnile/mosquitoSpecies.htm> (2003).
9. Elliott, P., Wakefield, J. C., Best, N. G. & Briggs, D. J. (eds) *Spatial Epidemiology: Methods and Applications* (Oxford Univ. Press, Oxford, 2000).
10. Kistemann, T., Dangendorf, F. & Schweikart, J. New perspectives on the use of Geographical Information Systems (GIS) in environmental health sciences. *Int. J. Hyg. Environ. Health* **205**, 169–181 (2002).
11. Steere, A. C. et al. Lyme arthritis: an epidemic of oligoarticular arthritis in children and adults in three Connecticut communities. *Arthritis Rheum.* **20**, 7–17 (1977).
12. Burgdorfer, W. et al. Lyme disease — a tick-borne spirochetosis? *Science* **220**, 321–322 (1982).
13. Hay, S. I., Tucker, C. J., Rogers, D. J. & Packer, M. J. Remotely sensed surrogates of meteorological data for the study of the distribution and abundance of arthropod vectors of disease. *Ann. Trop. Med. Parasitol.* **90**, 1–19 (1996).
14. Brooker, S. Schistosomes, snails and satellites. *Acta Tropica* **82**, 207–214 (2002).
15. Wint, G. R. W. et al. Mapping bovine TB in the UK using remotely sensed and environmental data. *Trends Microbiol.* **10**, 441–444 (2002).
16. Rogers, D. J. & Williams, B. G. in *35th Symposium of the British Ecological Society with the Society for Conservation Biology* (eds Edwards, P. J., May, R. & Webb, N. R.) 247–271 (Blackwell Scientific Publications, Oxford, 1993).
17. Guerra, M. et al. Predicting the risk of Lyme disease: habitat suitability for *Ixodes scapularis* in the North Central United States. *Emerg. Infect. Dis.* **8**, 289–297 (2002).
18. Danielová, V. & Benes, C. [Increased incidence of tickborne encephalitis in the Czech Republic] (in Czech). *Prakticky Lekar* **77**, 580–583 (1997).



19. Cleaver, H. Tick-borne illness increasing throughout Germany [online], (cited 14 Oct 2003), <[http://www.promedmail.org/pls/askus/f?p=2400:1202:1242085055564242893::NO::F2400\\_P1202\\_CHECK\\_DISPLAY,F2400\\_P1202\\_PUB\\_MAIL\\_ID:X,17759](http://www.promedmail.org/pls/askus/f?p=2400:1202:1242085055564242893::NO::F2400_P1202_CHECK_DISPLAY,F2400_P1202_PUB_MAIL_ID:X,17759)> (2002).
20. Distribution of TBE in Europe [online], (cited 14 Oct 2003), <[http://baxter.pixelwings.com/sixcms/detail.php?template\\_id=68&id=311&\\_l=2&\\_c1\\_id=372&\\_c2\\_id=474&\\_c3\\_id=&\\_p=&\\_s\\_id=302](http://baxter.pixelwings.com/sixcms/detail.php?template_id=68&id=311&_l=2&_c1_id=372&_c2_id=474&_c3_id=&_p=&_s_id=302)> (Baxter Vaccine AG, Vienna, 2003).
21. Randolph, S. E. Evidence that climate change has caused 'emergence' of tick-borne diseases in Europe? *Int. J. Med. Microbiol.* (in the press).
22. Hoare, C. *The Trypanosomes of Mammals* (Blackwell Scientific Publications, Oxford, 1972).
23. Anderson, R. M. (ed.) *Population Dynamics of Infectious Diseases: Theory and Applications* (Chapman & Hall, London, 1982).
24. Rogers, D. J., Randolph, S. E., Snow, R. W. & Hay, S. I. Satellite imagery in the study and forecast of malaria. *Nature* **415**, 710–715 (2002).
25. Ferguson, N. M., Donnelly, C. A. & Anderson, R. M. Transmission intensity and impact of control policies on the foot and mouth epidemic in Great Britain. *Nature* **413**, 542–548 (2001).
26. Longini, I. M. Jr. A mathematical model for predicting the geographic spread of new infectious agents. *Math. Biosci.* **90**, 367–383 (1988).
27. Rogers, D. J. & Randolph, S. E. The global spread of malaria in a future, warmer world. *Science* **289**, 1763–1766 (2000).
28. Randolph, S. E. & Rogers, D. J. Fragile transmission cycles of tick-borne encephalitis virus may be disrupted by predicted climate change. *Proc. R. Soc. Lond. B* **267**, 1741–1744 (2000).
29. International Scientific Working Group on Tick-borne Encephalitis. [online], (cited 14 Oct 2003), <<http://www.tbe-info.com/epidemiology/index.html>> (2002).
30. Lindgren, E. & Gustafson, R. Tick-borne encephalitis in Sweden and climate change. *Lancet* **358**, 16–18 (2001).
31. Randolph, S. E. The shifting landscape of tick-borne zoonoses: tick-borne encephalitis and Lyme borreliosis in Europe. *Phil. Trans. R. Soc. B* **356**, 1045–1056 (2001).
32. Hay, S. I. *et al.* Hot topic or hot air? Climate change and malaria resurgence in African highlands. *Trends Parasitol.* **18**, 530–534 (2002).
33. New, M., Hulme, M. & Jones, P. D. Representing twentieth century space-time climate variability. Part I: development of a 1961–90 mean monthly terrestrial climatology. *J. Climate* **12**, 829–856 (1999).
34. North, G. R. Review of global warming: the hard science by L. D. Danny Harvey. *Climatic Change* **49**, 493–497 (2001).
35. Mallat, S. *A Wavelet Tour of Signal Processing*. (Academic Press, San Diego, 1999).
36. Grenfell, B. T., Bjornstad, O. N. & Kappey, J. Travelling waves and spatial hierarchies in measles epidemics. *Nature* **414**, 716–723 (2001).
37. Bendat, J. S. & Piersol, A. G. *Random Data: Analysis and Measurement Procedures* (John Wiley & Sons, New York, 1986).
38. Ott, T. & Swiaczny, F. *Time-Integrative Geographic Information Systems: Management and Analysis of Spatio-Temporal Data* (Springer, Berlin, 2001).
39. Hay, S. I. An overview of remote sensing and geodesy for epidemiology and public health application. *Adv. Parasitol.* **47**, 1–35 (2000).
40. Rogers, D. J. Satellites, space, time and the African trypanosomiasis. *Adv. Parasitol.* **47**, 130–173 (2000).
41. Randolph, S. E. Ticks and tick-borne disease systems in space and from space. *Adv. Parasitol.* **47**, 217–243 (2000).
42. Rogers, D. J., Randolph, S. E., Lindsay, S. & Thomas, C. in *Health Effects of Climate Change in the UK* (ed. Maynard, R. L.) (UK Department of Health, London, 2002).

## Competing interests statement

The authors declare that they have no competing financial interests.

 **Online links**
**FURTHER INFORMATION**

European Space Agency: <http://www.esa.int>

Oxford Tick Research Group:

<http://users.ox.ac.uk/~zool0048/>

TALA Research Group: <http://www.tala.ox.ac.uk/>

Terra satellite: <http://terra.nasa.gov>

Access to this interactive links box is free online.

## INNOVATION

# Clostridia in cancer therapy

Nigel P. Minton

During the past decade, the search for an effective system for the selective delivery of high therapeutic doses of anti-cancer agents to tumours has explored a variety of ingenious and increasingly complex biological systems. These systems are most often based on gene therapy and use viral vectors as the delivery vehicle. Invariably, such systems have been found wanting with respect to a lack of tumour specificity, poor levels of transgene expression and inefficient distribution of the vector throughout the tumour mass. By contrast, the ability of intravenously injected clostridial spores to infiltrate, then selectively germinate in, the hypoxic regions of solid tumours seems to be a totally natural phenomenon, which requires no fundamental alterations and is exquisitely specific.

In many instances, traditional cancer treatments (surgery, radiotherapy and chemotherapy) are reasonably successful in controlling the disease. There are, however, many tumours and circumstances where these approaches are ineffective. Consequently, there is intensive activity around the world to develop alternative approaches. A fundamental requirement of any new anti-cancer therapy is the facility to subject tumour cells to a toxic agent and at the same time exclude normal healthy tissues from such exposure. Despite their conceptual simplicity, the derivation of such therapies is proving extremely challenging.

Over the past two decades, a significant amount of effort has focused on attempting to localize protein-based anti-cancer agents specifically to the site of tumours. The resulting therapeutic effect is dependent on the class of protein that is used. These range from proteins that, to a certain extent, have a direct effect (for example, immunogens, toxins and cytokines), to enzymes that are able to generate a cytotoxic drug from an innocuous precursor, such as prodrug-converting enzymes.

In all cases, these agents have been shown to be highly effective in killing tumour cells in studies using appropriate cell lines. The challenge has been to bring about their selective delivery to the tumour. Initially, strategies were devised that sought to target the protein itself, most prominently through its fusion to

an antibody raised against an antigen 'uniquely' present on tumour cells<sup>1</sup>. As we have entered the gene therapy era however, the focus has switched to the delivery of the gene that encodes the therapeutic protein — predominantly through the use of eukaryotic viral vectors<sup>2,3</sup>.

There are several inherent difficulties with the use of viral vectors. They exhibit a general lack of specificity for tumours and, once they have been delivered to the site of the tumour (for example, by direct injection), are relatively inefficient at being distributed throughout the tumour mass. Strategies have been devised to overcome these deficiencies, and progress is being made in several areas, including altering the viral tropism<sup>4</sup> and engineering the viruses such that they replicate selectively in tumour cells<sup>5,6</sup>. Alternatively, the possibility of developing non-viral delivery systems, based on liposomes or polyethyl glycol derivatives, for example, is being explored<sup>7,8</sup>.

Irrespective of the delivery system used, tumour specificity can also be elicited by placing the introduced therapeutic gene under the control of promoter elements that are only active in the tumour environment, thereby mitigating against production of the therapeutic agent in inadvertently transduced healthy tissues<sup>9–11</sup>. One of the most widely investigated strategies has been to incorporate hypoxia-regulatory elements (HREs) into the promoters that are used<sup>10</sup>. HREs are the binding sites for the hypoxia-inducible transcription factor 1 (HIF-1), which is overproduced under hypoxic (a reduced level of tissue oxygen tension) conditions<sup>12,13</sup>.

Anti-cancer strategies that are based on HREs are feasible because in most solid tumours the oxygen levels are reduced, as a consequence of poor vasculature. Although hypoxia leads to resistance to radiotherapeutic and chemotherapeutic agents, and an increased predisposition for tumour metastases<sup>14</sup>, these conditions also present an ideal environment for the growth of obligately anaerobic bacteria. Indeed, the colonization of colonic tumours by bacteria, such as *Clostridium* sp., can often represent the first indication that a patient has a cancerous growth. Experiments have shown that although intravenously injected