

# SCIENTIFIC REPORTS

OPEN

## Spatiotemporal Co-existence of Female Thyroid and Breast Cancers in Hangzhou, China

Xufeng Fei<sup>1</sup>, George Christakos<sup>2,3</sup>, Zhaohan Lou<sup>2</sup>, Yanjun Ren<sup>4</sup>, Qingmin Liu<sup>4</sup> & Jiaping Wu<sup>2</sup>

Received: 12 April 2016

Accepted: 02 June 2016

Published: 24 June 2016

Thyroid and breast cancers (TC, BC) are common female malignant tumors worldwide. Studies suggest that TC patients have a higher BC risk, and vice versa. However, it has not been investigated quantitatively if there is an association between the space-time TC and BC incidence distributions at the population level. This work aims to answer this question. 5358 TC and 8784 BC (female) cases were diagnosed in Hangzhou (China, 2008–2012). Pearson and Spearman rank correlation coefficients of the TC and BC incidences were high, and their patterns were geographically similar. The spatiotemporal co-existence of TC and BC distributions was investigated using the integrative disease predictability (IDP) criterion: if TC-BC association is part of the disease mapping knowledge bases, it should yield improved space-time incidence predictions. Improved TC (BC) incidence predictions were generated when integrating both TC and BC data than when using only TC (BC) data. IDP consistently demonstrated the spatiotemporal co-existence of TC and BC distributions throughout Hangzhou (2008–2012), which means that when the population experiences high incidences of one kind of cancer attention should be paid to the other kind of cancer too. The strength of TC-BC association was measured by the IDP coefficients and incidence prediction accuracy.

*Breast cancer* (BC) is the most common malignant tumor among women, especially for women aged between 40 and 60 years old<sup>1–4</sup>. The distribution of BC shows different degrees of variability, at the local or the global scale<sup>4,5</sup>. The BC incidence in developed countries is about 3 times higher than in developing countries, like China<sup>6</sup>. It has been estimated that more than 1.6 million of new BC cases were diagnosed in 2012 worldwide causing over a half million deaths<sup>7</sup>. Many previous studies have discussed the BC risk factors, such as body mass index (BMI), family history of BC, dietary patterns, socioeconomic status, smoking and alcohol consumption<sup>8–13</sup>. Among them, the reproductive, menstrual and hormonal factors (such as parity rate, pregnancy, age at menarche, menstrual cycle regularity, menopausal status, exogenous hormone use and hormone replacement therapy) have been found to have significant impacts on BC risk<sup>14–17</sup>.

*Thyroid cancer* (TC), on the other hand, is the most common endocrine malignant tumor, and its incidence has displayed a sharp increase during the last few decades<sup>18</sup>. Established TC pathogenic factors are ionizing radiation<sup>19</sup> and benign proliferative thyroid disease<sup>20</sup>. Other potential factors include vitamins and minerals, diet, fish and shellfish consumption, vegetable consumption, smoking, alcohol intake, ethnic background, genetic factors, higher BMI, and diabetes<sup>21–29</sup>. Some studies have considered the relationship between iodine intake and TC risk, but their results are rather inconclusive<sup>20,30,31</sup>. Endocrine disrupting compounds such as polychlorinated biphenyls (PCBs), polybrominated diphenyl ethers (PBDEs) and pesticides, which can disturb thyroid function, have been linked to carcinogenesis through a variety of mechanisms<sup>32–34</sup>. TC is gender-biased, i.e., it is more common among females than among males with a female to male ratio of about 3 to 1 in the general population<sup>35</sup>. TC and BC have similar age-specific incidence rates: women between 40–60 years old have the highest TC incidence rate, which is also the case of BC incidence. None of the risk factors mentioned above can account for the TC incidence disparity among different genders. Furthermore, it has been suggested that improved detection and the use of thyroid imaging (leading to an increasing number of detected small papillary cancer cases) may be responsible for the fast increase of reported TC incidences<sup>36–38</sup>. However, not all studies agree with this explanation, arguing,

<sup>1</sup>College of Environmental and Resource Sciences, Zhejiang University, Hangzhou, China. <sup>2</sup>Institute of Islands and Coastal Ecosystems, Zhejiang University, Zhoushan, China. <sup>3</sup>Department of Geography, San Diego State University, San Diego, CA, USA. <sup>4</sup>Hangzhou Center for Disease Control and Prevention, Hangzhou, China. Correspondence and requests for materials should be addressed to G.C. (email: gchristakos@zju.edu.cn) or J.W. (email: jw67@zju.edu.cn)

instead, that the increasing number of all-sizes TC observed may be due to two coexisting reasons: increased detection, and environmental risk factors<sup>38,39</sup>.

In light of the above studies, it is obvious that BC and TC have certain risk factors in common (genes, lifestyle, diet habits, over-diagnosed and hormonal, menstrual and reproductive factors), in which case one may justifiably hypothesize that the TC and BC distributions are co-existent. Indeed, several researchers have suggested that due to common risk factors there may exist a strong association between TC and BC incidence<sup>40,41</sup>.

Generally, the rigorous investigation of potential associations between diseases is a very complicated issue, leading to several criteria for testing possible associations, including deterministic criteria (i.e., necessary and sufficient associations) and stochastic criteria (concerned with factors that raise the chances of certain associations)<sup>3,32,42,43</sup>. The deterministic viewpoint of disease association may be valid only if all the relevant physical and biological laws linking the diseases are known in detail (then, in principle, causality could be established by analyzing the ways in which these laws are formulated and the ways they are verified or rejected by experimentation). Although very desirable, such an approach is rarely materialized in health practice.

Instead, most studies are actually concerned with stochastic disease associations at the population level. While stochastic associations do not usually imply necessary and sufficient causation criteria, they offer useful insight into such a complicated situation of considerable public health significance<sup>32,44</sup>. Studies of disease associations in human populations involve a variety of techniques, including visual comparisons of the patterns exhibited by the disease factors, statistical analysis of population disease occurrences across a geographic area, multi-perspective analysis, and geographical information systems techniques<sup>30,45–49</sup>. Yet, these techniques may suffer certain drawbacks when seeking to identify disease interactions and causation<sup>41,43,50</sup>.

In this work, by taking into consideration that the assessment of the spatiotemporal variation of disease incidence could improve our understanding of TC-BC associations at the population level, the combination of the *Bayesian maximum entropy* (BME) theory<sup>51</sup> and the *integrative disease predictability* (IDP) criterion<sup>52,53</sup> provides a powerful quantitative approach to study the spatiotemporal co-existence of TC and BC in Hangzhou city. Specifically, the IDP criterion was used to investigate the spatiotemporal co-existence of TC and BC incidence distributions and assess quantitatively the strength of the TC-BC association in terms of the predictability of the corresponding BME-generated cancer maps.

## Materials and Methods

**TC and BC Incidence Data.** Newly diagnosed female TC cases (International Classification of Disease: ICD-10 code C73) and female BC cases (International Classification of Disease: ICD-10 code C50) in the city of Hangzhou (HZ) from Jan 1, 2008 to Dec 31, 2012 were obtained from the Hangzhou center for Disease Control and Prevention (CDC). Detailed information about TC and BC in the study area can be found in the relevant literature<sup>54,55</sup>. Briefly, the CDC confirmed 5358 female TC patients and 8784 female BC patients in 8.70 million person-years. According to the detailed residential information available, all cases were allocated to 200 Hangzhou townships. The indirectly standardized ratio of female TC over female BC incidences<sup>56</sup> was calculated based on female township-level age-specific population data and female TC and BC incidences of different age categories over China available by the Chinese National Cancer Center<sup>4,57</sup>. In order to reduce unstable incidence estimation at the township level (which could be caused by small populations) the 5-year averages of TC and BC incidences were calculated and used to assess the co-existence between the two kinds of cancers (the person-years for each township were estimated by multiplying their 2010 female population by 5).

**The Integrative Disease Predictability (IDP) Method.** Modeling the association between TC and BC incidence distributions involves considerable uncertainty due to limited knowledge of the underlying mechanism varying across space-time (epistemic uncertainty) and the complexity of the phenomenon itself in multiple space-time scales (ontic uncertainty). However, if the randomly varying TC and BC incidence distributions are modeled as *spatiotemporal random fields*,  $S/TRF$ <sup>58</sup>, it is possible to evaluate the statistical space-time properties of these distributions. Therefore, in this work the TC and BC distributions were mathematically represented as the  $S/TRF$

$$TC_p \text{ and } BC_p, \quad (1)$$

respectively, where  $p = (s, t)$  is a space-time vector denoting spatial location  $s$  and time instant  $t$ . The space-time domain of the diseases is a continuum, in which space represents the order of co-existence and time represents the order of successive existence of TC and BC. Randomness manifests itself as an ensemble of possible realizations regarding the TC and BC incidence distributions. Methodologically, a disease random field is then a collection of realizations (possibilities) of the cancer distribution, where the probability that each one of these possible realizations occurs is expressed by the corresponding probability law of the disease (see, e.g.,  $f(BC)$  in Eqs (7, 8) below). As it turns out, incidence distributions are well represented by theoretical probability laws (Gaussian or non-Gaussian), and this permits us to calculate various space-time properties of these distributions with reasonable accuracy.

In the random field theory context, the IDP approach<sup>52,53,59</sup> considers the study of the TC-BC co-existence across space-time as a knowledge synthesis affair that leads to an association model integrating different kinds of cancer data in order to draw sound inferences about the health situation in Hangzhou city. Since the TC-BC association is an uncertain affair, mainstream deterministic logic is of little use. Stochastic inference, on the other hand, by definition has the flexibility for a realistic description of disease associations. The IDP approach postulates that the existence of a TC-BC association should lead to actual BC prediction errors  $e_{TC_p, BC_p}(BC_p)$  at a set of “control” points  $p' \neq p$  when data about both TC and BC are combined that are consistently smaller than the BC prediction errors  $e_{BC_p}(BC_p)$  when only BC data are used, i.e.,

$$e_{TC_p, BC_p}(BC_{p'}) < e_{BC_p}(BC_{p'}), \quad (2)$$

where space-time “control” points are locations-times where BC data are available and can be compared with BC incidence predictions.

The predictability criterion of Eq. (2) may take many forms. A possible form is provided by the *IDP coefficient* defined as the prediction error difference<sup>53</sup>

$$\beta_{BC_{p'}} = e_{TC_p, BC_p}(BC_{p'}) - e_{BC_p}(BC_{p'}). \quad (3)$$

Then, a consistently negative  $\beta_{BC_{p'}}$  map supports an association between BC-TC incidence distributions. The concept underlying Eq. (3) is straightforward: if BC incidence predictability is consistently improved across the space-time disease domain by using valid information about the TC incidence distribution, this is an indication of a TC-BC association. Given the current state of understanding of fundamental processes (biological, genetic etc.), we may not be able to explain this association in rigorous scientific terms, but the TC-BC association nevertheless exists and cannot be ignored. In the present study, for cross-verification purposes we applied the IDP criterion to assess the predictability of both kinds of cancer, i.e., in addition to  $\beta_{BC_{p'}}$ , the IDP coefficient

$$\beta_{TC_{p'}} = e_{BC_p, TC_p}(TC_{p'}) - e_{TC_p}(TC_{p'}) \quad (4)$$

was also calculated, where now  $e_{BC_p, TC_p}(TC_{p'})$  vs.  $e_{TC_p}(TC_{p'})$  denote, respectively, that information about both TC and BC incidences vs. information only about TC incidence was taken into account when TC predictions at the selected control points  $p'$  were generated. Specifically, when one of the two cancer types (TC or BC) was used as hard (i.e., exact) data to calculate the prior incidence probability density functions, the other type (BC or TC) was used as soft (i.e., uncertain) data combined with hard data to predict cancer incidence across space-time. Providing that there are no strong effects due to hidden confounding factors, an improved prediction of the BC (resp. TC) incidence distribution obtained from the combination of TC and BC data as compared to the prediction obtained merely from BC (resp. TC) data supports the existence of an association between the TC and BC incidence distributions.

In practice, the prediction errors at the set of control points  $p'$ , i.e.,  $e_{BC_p}(BC_{p'})$ ,  $e_{TC_p}(TC_{p'})$ ,  $e_{TC_p, BC_p}(BC_{p'})$ , and  $e_{BC_p, TC_p}(TC_{p'})$ , were calculated using scalar and vector BME techniques<sup>60</sup>; i.e.,

$$e_{BC_p}(BC_{p'}) = |BC_{p'} - \hat{BC}_{p'}|, \quad (5)$$

where  $BC_{p'}$  and  $\hat{BC}_{p'}$  denote, respectively, the actual and scalar BME-predicted BC incidences at the control points  $p'$  using only BC data at points  $p$ ; and

$$e_{TC_p, BC_p}(BC_{p'}) = |BC_{p'} - \hat{BC}_{p'}^{\cap}|, \quad (6)$$

where  $\hat{BC}_{p'}^{\cap}$  now denote the vector BME-predicted BC incidences at  $p'$  using both BC and TC data at points  $p$  (and, similarly for  $e_{TC_p}(TC_{p'})$  and  $e_{BC_p, TC_p}(TC_{p'})$ ).

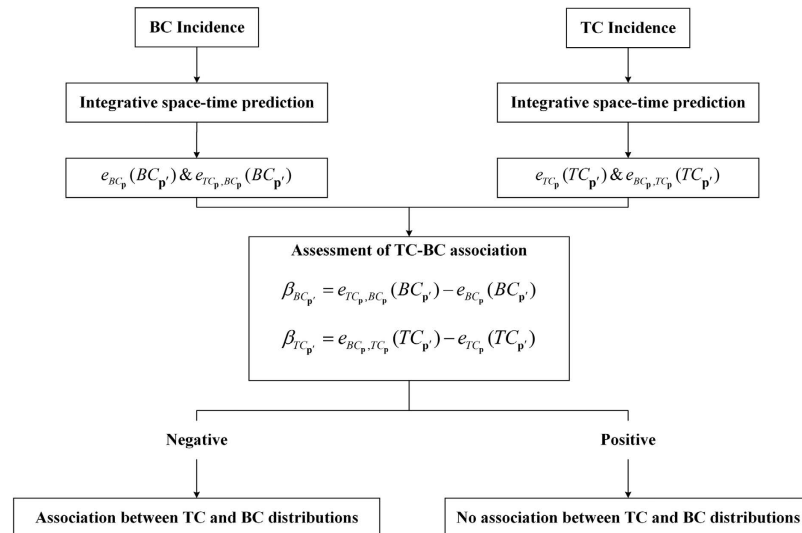
The basic set of BME equations used in the calculation of the  $\hat{BC}_{p'}$  and  $\hat{BC}_{p'}^{\cap}$  space-time predictions are<sup>51,53</sup>,

$$\int dG(\mathbf{g} - \bar{\mathbf{g}})e^{\mu \cdot \mathbf{g}} = 0 \quad (7)$$

$$\int dS e^{\mu \cdot \mathbf{g}} - Af(BC) = 0, \quad (8)$$

where  $\mathbf{g}$  is a vector of functions expressing mathematically the available core knowledge base  $G$ , including spatio-temporal dependence models, epidemic laws, and scientific theories;  $\bar{\mathbf{g}}$  denotes the mean value of  $\mathbf{g}$ ;  $S$  represents the available site-specific (Hangzhou city) knowledge base about the diseases as described earlier;  $\mu$  is a vector of coefficients representing the relative importance of each  $\mathbf{g}$ -function ( $\mu \cdot \mathbf{g}$  denotes the inner product of the vectors  $\mathbf{g}$  and  $\mu$ , which are both functions of space-time); and  $A$  is a normalization parameter. Eqs (7, 8) can be solved with respect to the BC probability law  $f(BC)$  at all disease mapping points of interest (i.e., space-time points at which BC predictions are sought). Software libraries have been developed dealing with the solution of Eqs (7, 8) in real world conditions, including BMElib, SEKS-GUI, Quantum BME, and StarBME<sup>60</sup>. Eqs (7, 8) are valid in terms of TC and have been used for the calculation of the  $\hat{TC}_{p'}$  and  $\hat{TC}_{p'}^{\cap}$  space-time predictions. More technical details concerning the BME equations above can be found in the relevant literature.

BME techniques are successfully used in many scientific disciplines<sup>61-67</sup>, and among their attractive features are the following<sup>51,60,68</sup>: they offer a complete disease spread characterization in terms of spatiotemporal random field theory; they account for various sources of knowledge, core and case-specific (scientific models, probabilistic evidence, empirical relationships); they do not require any assumption regarding the shape of the underlying probability law (hence, non-Gaussian laws are automatically incorporated); they lead to nonlinear predictors, in general, and can obtain mainstream predictors (e.g., kriging) as their limiting cases; they can model



**Figure 1.** An outline of the IDP approach.

heterogeneous (spatially non-homogeneous/temporally non-stationary) disease data; and they can perform vector disease mapping (i.e., involving several diseases simultaneously), as is the case with TC and BC incidences in the present study. More details concerning BME theory and techniques can be found in the relevant literature.

The IDP criterion can provide a useful approach to estimate quantitatively the overall and local strength of TC-BC associations in a region during the time period of interest on the basis of successful space-time cancer incidence predictions (the outline of the IDP approach is given in Fig. 1). In the present Hangzhou study:

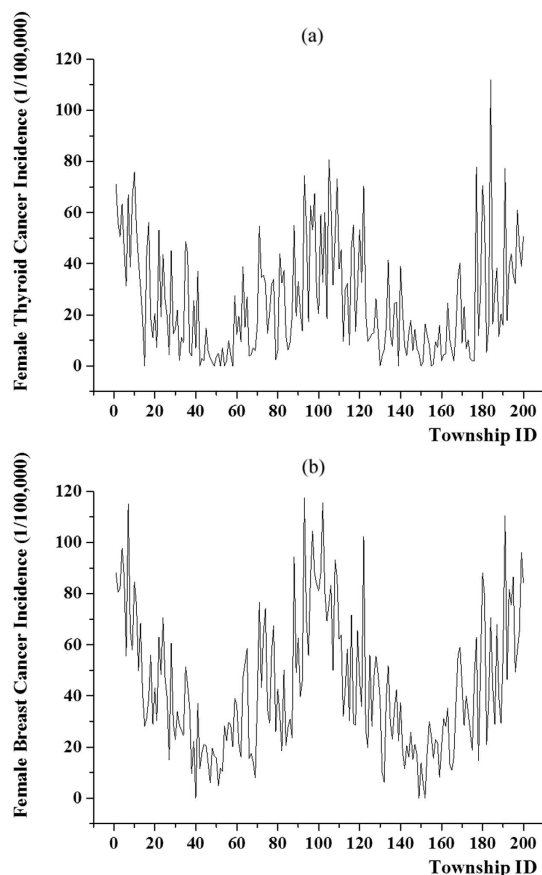
- The 200 township level incidence data were considered in a 10-fold validation framework in which the data were divided into 10 distinct subsets (each subset included 1/10th of the total dataset). Then, each one of these (1/10th) subsets was, in turn, selected as the validation set, whereas the remaining subset (consisting of 9/10th of the data) served as the training subset providing the hard data used to estimate incidence rates at the validation points.
- TC (resp. BC) incidence was predicted separately based on its own incidence (hard) data and the TC (resp. BC) prediction error was recorded as  $e_{TC_{p'}}(BC_{p'})$  (resp.  $e_{BC_{p'}}(TC_{p'})$ ).
- Soft incidence data were introduced in the selected validation subset in the form of probability density functions (pdf) constructed by first developing a linear regression model to relate TC and BC incidences (which showed a high linear correlation), and then using the model's means and variances to construct the TC pdf (resp. BC pdf).
- The TC (resp. BC) incidence was predicted based on its own and the other cancer type data, i.e., when predicting TC (resp. BC) incidence, TC (resp. BC) incidence data were used as hard data (this is the 9/10<sup>th</sup>-data subset above) and BC (resp. TC) data were used as soft data (this is the 1/10<sup>th</sup>-validation subset) in the form of TC pdf (resp. BC pdf) obtained in (c) above; and the prediction error was recorded as  $e_{TC_{p}, BC_{p'}}(BC_{p'})$  (resp.  $e_{BC_{p}, TC_{p'}}(TC_{p'})$ ).
- Finally the association between TC and BC incidences could be assessed quantitatively on the basis of  $\beta_{BC_{p'}}$  and  $\beta_{TC_{p'}}$ . As noted earlier, the TC-BC association at the population level is supported if the  $\beta_{BC_{p'}}$  and  $\beta_{TC_{p'}}$  are found to be consistently negative across the space-time study domain.

Computationally, all the data processing and mapping operations were carried out with ArcMap 9.3<sup>69</sup>, whereas BME space-time TC and BC incidence predictions were generated using the SEKS-GUI v1. 0.8 software<sup>60</sup>.

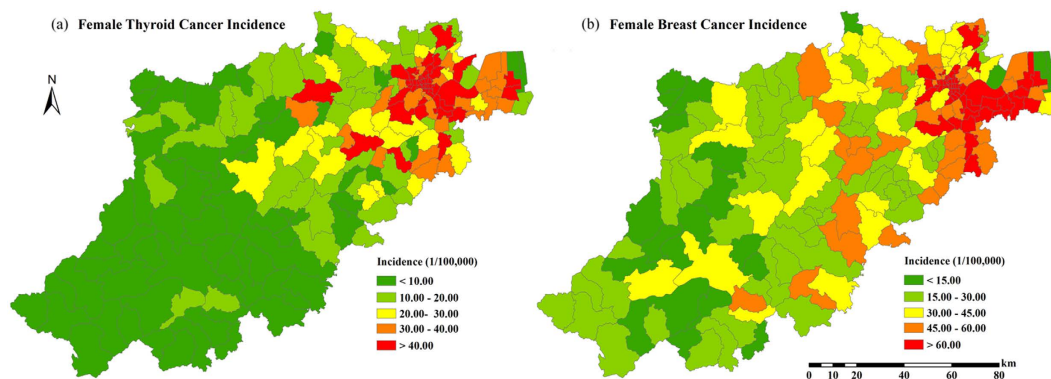
## Results

The present work is, in a sense, an ecological study conducted using data at the township level. Naturally, two limited cases were noticed in townships with small populations: the TC and BC incidences would be zero if no cancer patient was recorded during a specific year, and extremely high when for a small population base even a small number of patients would lead to high incidence. As noted earlier, to overcome these computational issues the 5-year averaged TC and BC incidences were calculated to produce more stable incidence estimates. Accordingly, Fig. 2a displays the 5-year averaged profile of TC incidence for the 200 Hangzhou townships considered, and Fig. 2b that of BC incidence. Visual comparison of Fig. 2a,b clearly suggested the existence of a strong correlation between TC and BC incidences. Also, the Pearson and Spearman rank correlation coefficients of the 5-year averaged TC and BC incidences were found to be 0.797 ( $p < 0.01$ ) and 0.808 ( $p < 0.01$ ), respectively, at the township level.

In Fig. 3a the geographical TC incidence distribution averaged during the 5-years (2008–2012) exhibited considerable spatial variability across the Hangzhou area: townships located in the southwestern parts of the



**Figure 2.** 5-year average female thyroid cancer (a) and breast cancer (b) incidence in each township.

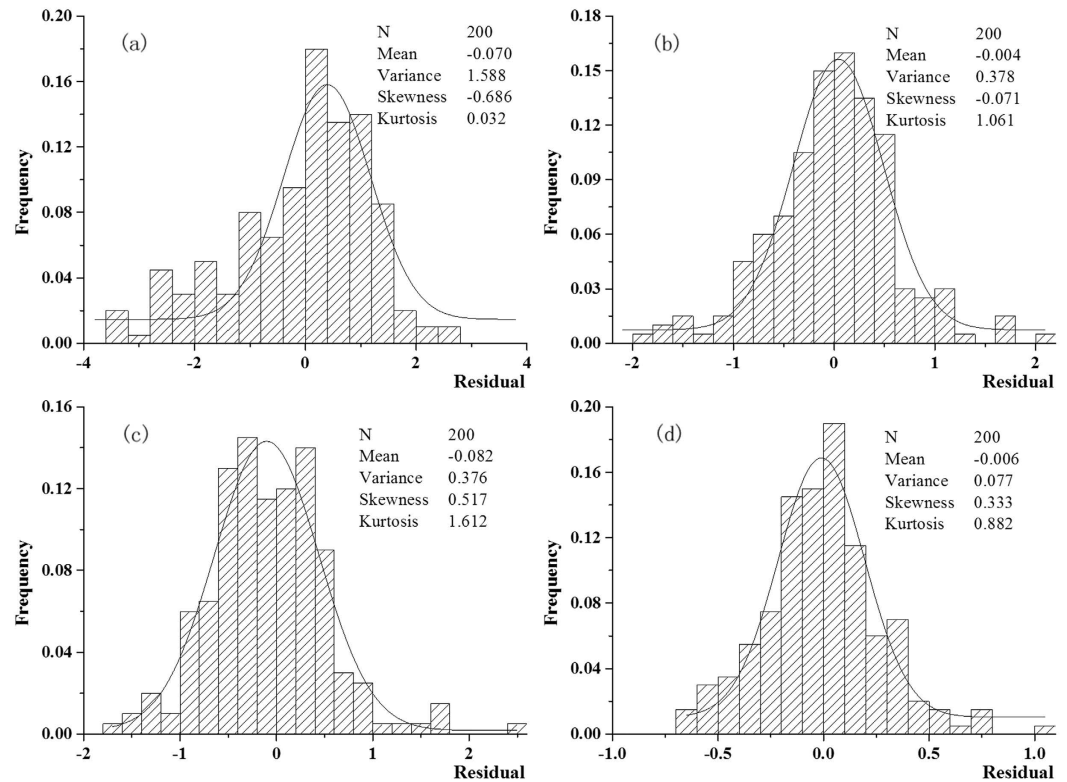


**Figure 3.** The distribution of 5-year average female thyroid/breast cancer incidence in Hangzhou. (2008–2012). Created by ArcMap 9.3.1 <http://www.esri.com/software/arcgis/eval-help/arcgis-931>.

region revealed the lowest TC incidence (less than 10 per 100,000 people), whereas townships located in the northeastern parts experienced the highest incidence (more than 40 patients per 100,000 people). The pattern of geographical BC distribution averaged during the 5-years (Fig. 3b) was similar to that of TC: the southwest areas showed the lowest incidence (less than 15 patients per 100,000 people), whereas the northeast areas showed the highest incidence (more than 60 patients per 100,000 people).

Figure 4 displays the frequency graphs of incidence prediction errors derived by the BME technique in the following cases: (a) only TC incidence values were used as hard data to predict TC incidence; (b) TC incidence were used as hard data and BC incidence information was used as soft data to predict TC incidence; (c) only BC incidence values were used as hard data to predict BC incidence; and (d) BC incidence values were used as hard data and TC incidence information was used as soft data to predict BC incidence.



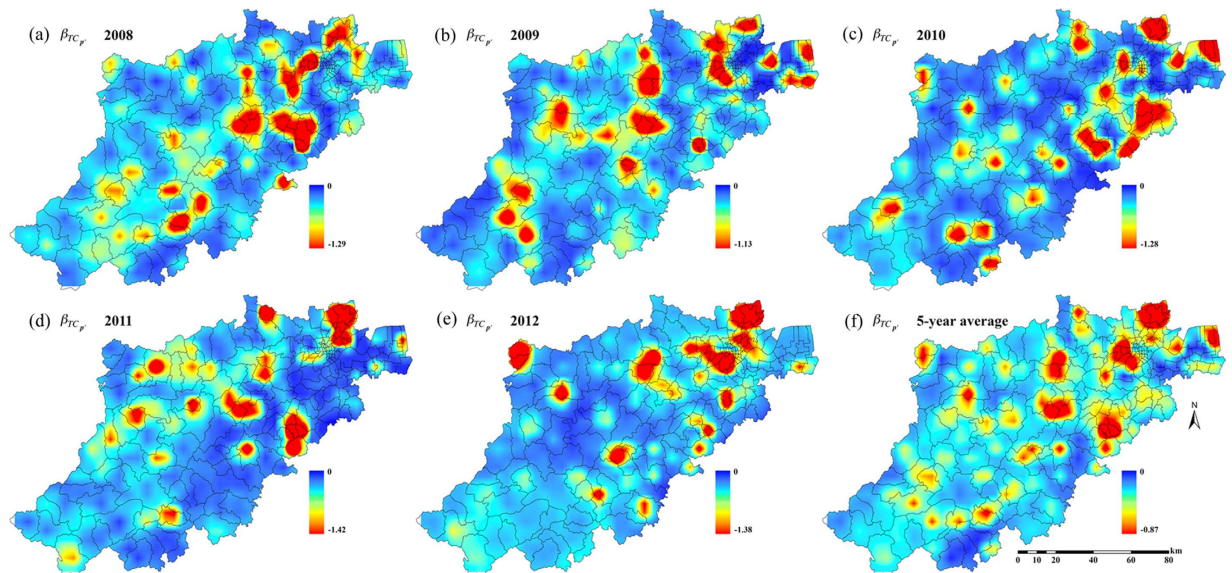


**Figure 4.** The distribution of prediction errors through Bayesian maximum entropy techniques. (a) only TC incidence were used as hard data to predict TC incidence, (b) TC incidence were used as hard data and BC incidence were used as soft data to predict TC incidence, (c) only BC incidence were used as hard data to predict BC incidence and (d) BC incidence were used as hard data and TC incidence were used as soft data to predict BC incidence.

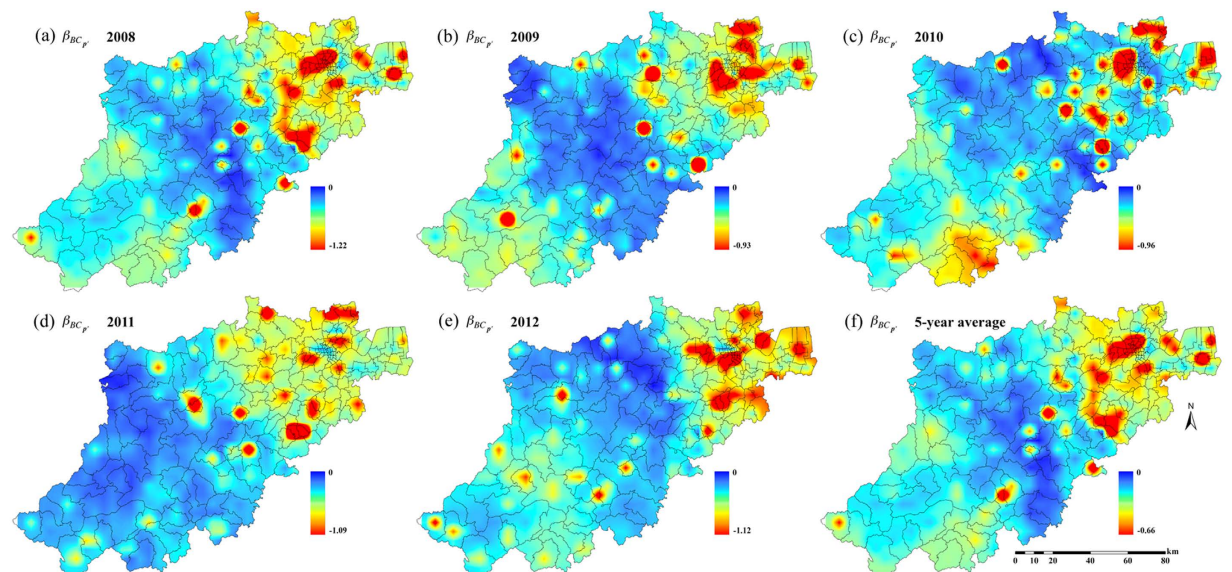
In particular, the predictive TC incidence map corresponding to the prediction error frequency graph of Fig. 4b had a smaller root mean square error, (RMSE = 5.21) than that corresponding to Fig. 4a (RMSE = 16.27). And, the TC prediction error graph of case *b* was more Gaussian-like than that of case *a*. The comparison of cases *a* vs. *b* clearly shows that the incorporation of soft BC incidence data provided an improved TC incidence prediction compared to that using hard TC incidence data only. Similarly, the frequency graph of Fig. 4d displayed a smaller RMSE (0.74) than that of Fig. 4c (RMSE = 5.78). Also, the TC prediction error graph of case *d* was more Gaussian-like than that of case *c*. Again, the comparison of cases *c* vs. *d* demonstrated that the incorporation of soft TC incidence data provided better BC incidence predictions compared to that using only hard BC incidence data.

Figures 5 and 6 show the distribution of  $\beta_{BC_p}$  and  $\beta_{TC_p}$  calculated by Eqs (3) and (4). Specifically, in Fig. 5a–e we see the maps of the coefficient  $\beta_{TC_p}$  in Hangzhou during the years 2008–2012, whereas Fig. 6a–e show the maps of the coefficient  $\beta_{BC_p}$  in Hangzhou for the same time period. Note that the values of  $\beta_{TC_p}$  and  $\beta_{BC_p}$  were consistently negative across the Hangzhou area during the 5-year period, which, according to the IDP criterion, implies the presence of a “TC incidence–BC incidence” association at the local township population level during the 5-year period. Since the IDP coefficients provide a measure of the geographical variation of the contribution of the TC–BC incidence association to incidence prediction, the higher (in absolute value) the coefficient  $\beta_{BC_p}$  (resp.  $\beta_{TC_p}$ ) in an area, the higher the contribution of the TC–BC association in BC (resp. TC) incidence prediction in this area. This contribution varies across space and time, and it can be different for the two cancer types. For example, Fig. 5a reveals that the contribution of the TC–BC association in TC incidence prediction is higher in parts of southwestern Hangzhou, whereas in Fig. 6a the association’s contribution in BC incidence prediction is higher in parts of northeastern Hangzhou. Figures 5f and 6f show the 5-years averaged maps of  $\beta_{TC_p}$  and  $\beta_{BC_p}$ , respectively.

Interestingly, the geographical variations of the  $\beta_{TC_p}$  and  $\beta_{BC_p}$  distributions of each one of the five years considered (Figs 5a–e and 6a–e) showed some small differences, which were due to corresponding differences in the disease prediction error maps ( $e_{TC_p, BC_p}(BC_p)$  vs.  $e_{BC_p, TC_p}(TC_p)$ ), which, in turn, were due to the different TC and BC datasets as well as to the different space-time covariances,  $c_{TC}(|h|, \tau)$  and  $c_{BC}(|h|, \tau)$ , used in TC and BC incidence prediction. Indeed, the different shapes of the two space-time incidence covariances can be clearly seen in Fig. 7a,b (the spatial TC correlation range is much longer than the BC correlation range, implying stronger geographical dependence between TC incidences than between BC incidences, the TC incidence variance is



**Figure 5. Maps of  $\beta_{TC_{p'}}$ .** The  $e_{BC_{p'} TC_{p'}}(TC_{p'})$  denotes TC incidence prediction errors obtained using TC incidence as hard data and BC incidence as soft data; and the  $e_{TC_{p'}}(TC_{p'})$  denotes TC incidence prediction errors obtained using TC incidence as hard data only. Created by ArcMap 9.3.1 <http://www.esri.com/software/arcgis/eval-help/arcgis-931>.

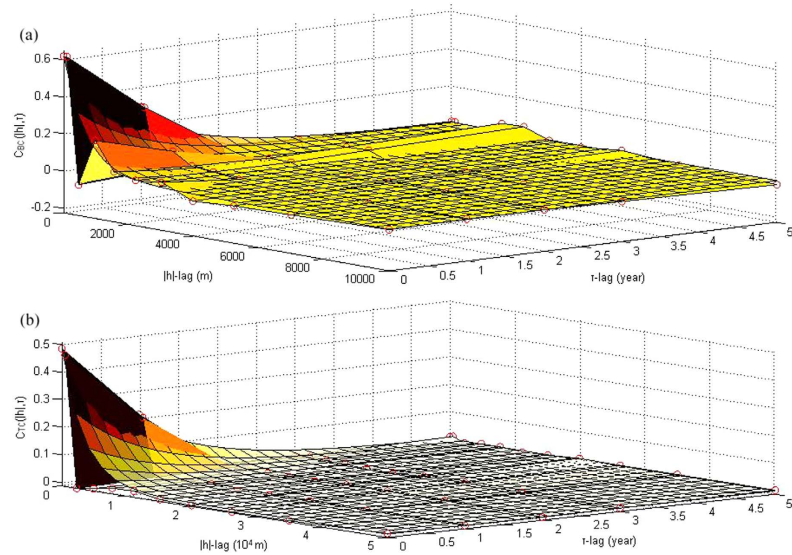


**Figure 6. Maps of  $\beta_{BC_{p'}}$ .** The  $e_{TC_{p'} BC_{p'}}(BC_{p'})$  denotes BC incidence prediction errors obtained using BC incidence as hard data and TC incidence as soft data; and the  $e_{BC_{p'}}(BC_{p'})$  denotes BC incidence prediction errors obtained using BC incidence as hard data only. Created by ArcMap 9.3.1 <http://www.esri.com/software/arcgis/eval-help/arcgis-931>.

smaller than that of BC incidence, implying less local TC variation than BC variation, etc.). These differences between the incidence covariances can affect the results of BME prediction of the TC and BC incidences across space-time.

Moreover, the global numerical values of the IDP coefficients, i.e. the averages of the  $\beta_{TC_{p'}}$  and  $\beta_{BC_{p'}}$  coefficients for the five maps of Figs 5a–e and 6a–e, can be calculated by, respectively,

$$\beta_{TC} = \frac{1}{5|A|} \sum_{t'=2008}^{2012} \int_A ds' \beta_{TC_{p'}} = -0.17, \quad (9)$$



**Figure 7.** Plots of the space-time covariances of the TC incidence and BC incidence distributions, **(a)**  $c_{TC}(|h|, \tau)$ , and **(b)**  $c_{BC}(|h|, \tau)$ , respectively. Circles denote experimental (sample) covariance values, whereas continuous (color) surfaces and continuous lines denote the fitted theoretical models.

$$\beta_{BC} = \frac{1}{5|A|} \sum_{t'=2008}^{2012} \int_A ds' \beta_{BCp'} = -0.18, \quad (10)$$

where  $p' = (s', t')$ ,  $A$  denotes the Hangzhou area, and  $|A|$  is the magnitude of  $A$  (a total of 1000 space-time points were used in the calculation of the averages in Eqs (9, 10)). The  $\beta_{TC}$  and  $\beta_{BC}$  values, which provide a quantitative assessment of the magnitude of the TC-BC association, were found to be practically identical, as should be expected, since both coefficients refer to the same TC-BC association. The result of Eqs (9, 10) further support the spatiotemporal co-existence of TC and BC incidence distributions at the population level, according to the IDP criterion, and are in agreement with some earlier findings suggesting the co-existence of TC and BC diseases<sup>17,39–41</sup>.

## Conclusions-Discussion

Since BC and TC are two of the most common cancers in the female populations, it is important to study whether the two cancers co-exist at the administrative level. The quantitative assessment of this association over a geographical region and during a specified time period is an important but difficult endeavor. Indeed, epidemiologic research concerning the spatiotemporal co-existence of diseases involves a number of traditional methods (visual comparisons of disease patterns, statistical analysis of disease occurrences across a geographical region, multi-perspective analysis, and geographical information systems techniques) that have been found to have certain drawbacks when seeking to identify disease interactions and causation. For example, previous studies have used correlation analysis to assess any similarities and differences between BC and prostate cancer (PC) mortalities in Spain, but they lacked the ability to assess BC-PC associations at specific locations and time periods<sup>7</sup>. Analogical, cluster maps and correlation analysis were also used to compare the spatial trends of BC and PC incidences in the United States<sup>55</sup>. Although the results could determine whether or not the two cancers are “high-high” or “low-low” correlated in specific geographic areas, they still lacked the ability to assess local correlations quantitatively.

In the present work, the implementation of the IDP criterion to evaluate the spatiotemporal co-existence of BC and TC incidence distributions offered new insight into disease associations in Hangzhou city during a 5-year time period. The theoretical support of the IDP criterion is twofold: mathematically, the combined structural and uncertain components of TC and BC incidence distributions are represented by the spatiotemporal random field theory; and, logically, the potential TC-BC associations are assessed in terms of consistent improvements in the incidence predictability of one disease resulting by the synthesis of information sources about both diseases. It should be noticed that the IDP approach can gain additional support from the fact that in health sciences one finds an interesting connection between prediction and causation: occasionally, the same epistemic and empirical conditions that generate sound predictions of a disease are used to obtain meaningful causal explanations of it, as well<sup>43</sup>.

A significant positive correlation was observed between the TC and BC incidence distributions in Hangzhou city during 2008–2012. Subsequently, the magnitude of the association between space-time TC and BC distributions, at the global and local scales, was quantitatively assessed in terms of the IDP coefficients. The results supported the spatiotemporal co-existence of TC and BC incidence distributions at the population level. They



also showed that the Hangzhou population had a significantly higher risk to develop TC when an increased BC incidence had been observed in the area during the 5-year time period, and vice versa.

The spatiotemporal co-existence of TC and BC could be partially attributed to common risk factors. First, as is well-known, hormonal, menstrual and reproductive factors (early age at menarche, later age pregnancy and menopause, increased exogenous hormone use and hormone replacement therapy etc.) have a strong effect on BC risk, and certain studies have hypothesized that the same factors may play a role in TC etiology, as well. These studies, however, are not consistent: for example, a case-control study in the Middle East has suggested that reproductive factors may influence or contribute to female TC risk (although hormone use was not associated with female TC risk)<sup>70</sup>, whereas a prospective study of Norway women during 1961–1989 suggested that there was no strong association between reproductive factors and TC risk<sup>56</sup>. In recent years, a population-base, case-control TC study among young women in France pointed out that reproductive and hormonal factors leading to estrogen exposure can increase TC risk<sup>39</sup>, but the European Prospective Investigation into cancer and Nutrition (EPIC) cohort study suggested that hormonal use, reproductive and menstrual factors have no strong associations with TC etiology<sup>18</sup>. In the present study, female BC incidence showed a similar geographical pattern with female TC incidence, e.g., the incidences of both diseases gradually increased from the southwestern to the northeastern part of Hangzhou. It was also established that higher (lower) BC incidence was co-existence with higher (lower) TC incidence, and the TC-BC incidence associations were consistent across the study area. These findings supported the hypothesis that hormonal, menstrual and reproductive factors are associated with TC risk.

Second, TC incidence revealed a sharp increase in recent decades, however this increase has been detected mainly in small (<2 cm), well-differentiated papillary cancers, whereas TC mortality did not show a significantly increasing trend. Therefore, improved detection of subclinical diseases that are indolent and would never cause symptoms (overdiagnosis) due to the increasing use of more sensitive diagnostic methods could, at least partially, explain the increasing TC incidence<sup>34,71</sup>. In previous ecological studies, the socioeconomic status was used as a surrogate of access to health care (e.g., health insurance coverage) and to diagnostic technologies (e.g., imaging and fine-needle aspiration)<sup>72,73</sup>. It has been found that better socioeconomic status (representing better health care and diagnostic technology accessibility) was positively correlated with TC risk, and this finding is also valid in the BC case. In fact, many studies have discussed the association between socioeconomic status and BC risk<sup>2,11,54,74</sup>. Higher socioeconomic status (assessed in terms of higher income, education level and less unemployment) represents better health awareness and health care accessibility that could lead to BC be diagnosed at an early stage, thus leading to positive correlations between BC and socioeconomic status. The co-existing TC and BC cancer types were characterized by  $\beta_{TC_{p'}}$  and  $\beta_{BC_{p'}}$  distributions that were both spatially heterogeneous at the township level (Figs 5 and 6). Compared to previous Hangzhou studies<sup>54</sup>, the stronger TC-BC co-existence was observed mainly in subdistrict areas with the highest socioeconomic status. Better health care, diagnostic technology accessibility and over-diagnosis assessment might strengthen the co-existence of TC and BC, which could be also the focus of future works.

Third, there are some other common TC and BC risk factors that may be responsible for the spatial coexistence of the two cancers, such as urban-rural disparity, higher BMI, smoking and alcohol consumption, exogenous estrogen exposure (PCBs, PBDEs, estradiol, oestrone). The TC and BC pathogeneses are very complex. Although BME and IDP could control for confounding factors, due to the ecological design of the present study and data limitations we did not control for these factors. Other study limitations included the lack of specific information about TC sub-types (papillary thyroid carcinoma, follicular thyroid carcinoma etc.) and BC sub-types (estrogen receptor positive –ER+, estrogen receptor negative –ER– etc.). As there are different etiologies for each cancer sub-type, the lack of sub-type information may create some problems in disease etiology. Although townships were sufficiently populated for the present study purposes (so that, e.g., hard incidence data can be considered in the Hangzhou study with high confidence), some techniques (including BME<sup>64</sup> and Poisson kriging<sup>75</sup>) can model uncertain incidence information, thus improving incidence estimation accuracy. This is a possibility to be explored in a future work.

## References

1. International Agency for Research on Cancer, GLOBOCAN 2008. Cancer incidence, mortality and prevalence worldwide in 2008 (2013). <http://globocan.iarc.fr>.
2. Fei, X., Wu, J., Kong, Z. & Christakos, G. Urban-rural disparity of breast cancer and socioeconomic risk factors in China. *PLoS One*. **10**(2), e0117572–e0117572 (2014).
3. Muir, C. S. International variation in high-risk populations. In *Persons at High Risk of Cancer. An Approach to Cancer Etiology and Control*: 293–305. Fraumeni J. F. Jr. (ed.). Acad. Press, New York, NY. (1975).
4. Tan, J., Su, Y. & Fan, Z. eds. *The Atlas of Endemic Diseases and Their Environments in the People's Republic of China* (1st ed.). Science Press, Beijing, China. (1989).
5. Christakos, G. & Lai, J. A study of the breast cancer dynamics in North Carolina. *Social Science & Medicine D*. **45**(10), 1503–1517 (1997).
6. Pantavos, A. *et al.* Total dietary antioxidant capacity, individual antioxidant intake and breast cancer risk: The Rotterdam study. *Int J Cancer*. **136**(9), 2178–2186 (2014).
7. López-Abente, G., Mispireta, S. & Pollán, M. Breast and prostate cancer: an analysis of common epidemiological features in mortality trends in Spain. *BMC Cancer*. **14**(1), 874 (2014).
8. Cheraghi, Z., Poorolajal, J., Hashem, T., Esmailnasab, N. & Irani, A. D. Effect of body mass index on breast cancer during premenopausal and postmenopausal periods: a meta-analysis. *PLoS One*. **7**(12), e51446 (2012).
9. Assi, V. *et al.* A case-control study to assess the impact of mammographic density on breast cancer risk in women aged 40–49 at intermediate familial risk. *Int J Cancer*. **136**(10), 2378–2387 (2014).
10. Couto, E. *et al.* Mediterranean dietary pattern and risk of breast cancer. *PLoS One*. **8**(2), e55374 (2013).
11. Adams, J., White, M. & Forman, D. Are there socioeconomic gradients in stage and grade of breast cancer at diagnosis? Cross sectional analysis of UK cancer registry data. *Brit Med J*. **329**(7458), 142 (2004).
12. Catsburg, C., Miller, A. B. & Rohan, T. E. Active cigarette smoking and risk of breast cancer. *Int J Cancer*. **136**(9), 2204–2209 (2014).

13. Pezzotti, A. *et al.* The mitochondrial A10398G polymorphism, interaction with alcohol consumption, and breast cancer risk. *PLoS One*. **4**(4), e5356 (2009).
14. Pike, M. C., Krailo, M. D., Henderson, B. E., Casagrande, J. T. & Hoel, D. G. 'Hormonal' risk factors, 'breast tissue age' and the age-incidence of breast cancer. *Nature*. **303**(30), 767–770 (1983).
15. LaCroix, A. Z. & Burke, W. Breast cancer and hormone replacement therapy. *Lancet*. **350**(9084), 1042–1043 (1997).
16. Collaborative Group on Hormonal Factors in Breast Cancer. Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data on 53 297 women with breast cancer and 100 239 women without breast cancer from 54 epidemiological studies. *Lancet*. **347**(9017), 1713–1727 (1996).
17. Marchbanks, P. A. *et al.* Oral contraceptives and the risk of breast cancer. *New Engl J Med*. **346**(26), 2025–2032 (2002).
18. Zamora-Ros, R. *et al.* Reproductive and menstrual factors and risk of differentiated thyroid carcinoma: The EPIC study. *Int J Cancer*. **136**(5), 1218–1227 (2015).
19. Hatch, M., Ron, E., Bouville, A., Zablotska, L. & Howe, G. The Chernobyl disaster: Cancer following the accident at the Chernobyl nuclear power plant. *Epidemiol Rev*. **27**, 56–66 (2005).
20. Kolonel, L. N., Hankin, J. H., Wilkens, L. R., Fukunaga, F. H. & Hinds, M. W. An Epidemiologic-Study of Thyroid-Cancer in Hawaii. *Cancer Cause Control*. **1**, 223–234 (1990).
21. Zhang, L. R., Sawka, A. M., Adams, L., Hatfield, N. & Hung, R. J. Vitamin and mineral supplements and thyroid cancer: a systematic review. *Eur J Cancer Prev*. **22**, 158–168 (2013).
22. Soloway, L. E. *et al.* Thyroid cancer incidence in highly observant Jewish neighborhoods in Metropolitan New York City. *Thyroid*. **21**, 1255–1261 (2011).
23. Mack, W. J., Preston-Martin, S., Bernstein, L., Qian, D. & Xiang, M. Reproductive and hormonal risk factors for thyroid cancer in Los Angeles County females. *Cancer Epidemiol Biomarkers Prev*. **8**, 991–997 (1999).
24. Allen, N. E. *et al.* Moderate alcohol intake and cancer incidence in women. *J Natl Cancer Inst*. **101**, 296–305 (2009).
25. Chen, S. T., Hsueh, C., Chiou, W. K. & Lin, J. D. Disease-specific mortality and secondary primary cancer in well-differentiated thyroid cancer with type 2 diabetes mellitus. *PLoS One*. **8**(1), e51179 (2013).
26. Zhu, H., Xi, Q., Liu, L., Wang, J. & Gu, M. Quantitative Assessment of Common Genetic Variants on FOXE1 and Differentiated Thyroid Cancer Risk. *PLoS One*. **9**(1), e87332 (2014).
27. Dal Maso, L., Bosetti, C., La Vecchia, C. & Franceschi, S. Risk factors for thyroid cancer: an epidemiological review focused on nutritional factors. *Cancer Cause Control*. **20**, 75–86 (2009).
28. Cléro, É. *et al.* Dietary iodine and thyroid cancer risk in French Polynesia: A case-control study. *Thyroid*. **22**, 422–429 (2012).
29. Peterson, E., De, P. & Nuttall, R. BMI, Diet and Female Reproductive Factors as Risks for Thyroid Cancer: A Systematic Review. *PLoS One*. **7**(1), e29177 (2012).
30. Tan, W. Y. *Stochastic Models for Carcinogenesis* (Vol. 116). CRC Press. (1991).
31. Freeman, L. E. B. *et al.* Atrazine and Cancer Incidence Among Pesticide Applicators in the Agricultural Health Study (1994–2007). *Environ Health Perspect*. **119**, 1253–1259 (2011).
32. Solan, S. *et al.* Cancer incidence in World Trade Center rescue and recovery workers, 2001–2008. *Environ Health Perspect*. **121**, 699–704 (2013).
33. Leux, C. & Guenel, P. Risk factors of thyroid tumors: Role of environmental and occupational exposures to chemical pollutants. *Rev Epidemiol Sante*. **58**, 359–367 (2010).
34. Davies, L. & Welch, H. G. Increasing incidence of thyroid cancer in the United States, 1973–2002. *JAMA-J Am Med Assoc*. **295**(18), 2164–2167 (2006).
35. Guan, H. *et al.* Association of high iodine intake with the T1799A BRAF mutation in papillary thyroid cancer. *J Clin Endocrinol Metab*. **94**(5), 1612–1617 (2009).
36. Vigneri, R., Malandrino, P. & Vigneri, P. The changing epidemiology of thyroid cancer: why is incidence increasing? *Curr Opin Oncol*. **27**(1), 1–7 (2015).
37. Chen, A. Y., Jemal, A. & Ward, E. M. Increasing incidence of differentiated thyroid cancer in the United States, 1988–2005. *Cancer*. **115**(16), 3801–3807 (2009).
38. Fei, X., Yang, D., Kong, Z., Lou, Z. & Wu, J. Thyroid cancer incidence in China between 2005 and 2009. *Stoch Environ Res Risk Assess*. **28**, 1075–1082 (2014).
39. Xhaard, C. *et al.* Menstrual and reproductive factors in the risk of differentiated thyroid carcinoma in young women in France: A Population-based case-control study. *Am J Epidemiol*. **180**(10), 1007–1017 (2014).
40. Chen, A. Y. *et al.* The development of breast carcinoma in women with thyroid carcinoma. *Cancer*. **92**(2), 225–231 (2001).
41. Joseph, K. R., Edirimanne, S. & Eslick, G. D. The association between breast cancer and thyroid cancer: a meta-analysis. *Breast Cancer Res Tr*. **152**(1), 173–181 (2015).
42. Salmon, W. C. *Causality and explanation* (pp. 320–9). Oxford Univ Press, (1998).
43. Christakos, G., Wang, J. & Wu, J. *Stochastic Medical Reasoning and Environmental Health Exposure*. Imperial College Press, (2014).
44. Rothman, K. J., Greenland, S. & Lash, T. L. eds. *Modern Epidemiology*. Lippincott Williams & Wilkins. (2008).
45. Mayer, J. D. The role of spatial analysis and geographic data in the detection of disease causation. *Soc. Sci. Med*. **17**(16), 1213–1221 (1983).
46. Glatte, E. Atlas of cancer incidence in Norway 1970–1979. *Recent Results in Cancer Research*. **114**, 216–226 (1989).
47. Blot, W. J. & McLaughlin, J. K. Geographic patterns of breast cancer among American women. *Jour. Nat. Cancer Inst*. **87**(24), 1819–1820 (1995).
48. Sturgeon, S. R. *et al.* Variation in mortality from breast cancer among white women in the U.S. *J. Natl. Cancer Inst*. **87**(24), 1846–1853 (1995).
49. Kolovos, A., Skupin, A., Jerrett, M. & Christakos, G. Multi-perspective analysis and spatiotemporal mapping of air pollution sensor data. *Environ. Sci. and Technol*. **44**(17), 6738–6744 (2010).
50. Susser, M. What is a cause and how do we know one? *Am J Epidemiol*. **133**, 635–648 (1991).
51. Christakos, G. *Modern Spatiotemporal Geostatistics*. Oxford Acad Press, New York, NY (2000).
52. Christakos, G. & Serre, M. L. Spatiotemporal analysis of environmental exposure-health effect associations. *J Expo Anal Env Epid*. **10**(2), 168–187 (2000).
53. Christakos, G. *Integrative Problem-Solving in a Time of Decadence*. Springer Science & Business Media. (2010).
54. Fei, X. *et al.* A geographic analysis about the spatiotemporal pattern of breast cancer in Hangzhou from 2008 to 2012. *PLoS One*. **11**(1), e0147866 (2016).
55. Mandal, R., St-Hilaire, S., Kie, J. G. & Derryberry, D. Spatial trends of breast and prostate cancers in the United States between 2000 and 2005. *Int J Health Geogr*. **8**(1), 1 (2009).
56. Akslen, L. A., Nilssen, S. & Kvåle, G. Reproductive factors and risk of thyroid cancer. A prospective study of 63,090 women from Norway. *Brit J Cancer*. **65**(5), 772 (1992).
57. Hao, J. & Chen, W. Q. *Chinese cancer registry annual report*. Beijing: Military Medical Science Press, 17, pp 44–47 (2012).
58. Christakos, G. & Hristopoulos, D. T. *Stochastic Environmental Health Modelling*. Kluwer Academic Publ, Boston, MA. (1998).
59. Christakos, G. & Serre, M. L. BME analysis of spatiotemporal particulate matter distributions in North Carolina. *Atmos Environ*. **34**(20), 3393–3406 (2000).

60. Yu, H. L. *et al.* Interactive spatiotemporal modelling of health systems: the SEKS-GUI framework. *Stoch Environ Research and Risk Assess.* **21**(5), 555–572 (2007).
61. D'Or, D. & Bogaert, P. Continuous valued map reconstruction with the Bayesian Maximum Entropy. *Geoderma.* **112**, 169–178 (2003).
62. Akita, Y., Carter, G. & Serre, M. L. Spatiotemporal non-attainment assessment of surface water tetrachloroethene in New Jersey. *J of Environmental Quality.* **36**, 508–520 (2007).
63. Puangthongthub, S., Wangwongwatana, S., Kamens, R. M. & Serre, M. L. Modeling the space/time distribution of particulate matter in Thailand and optimizing its monitoring network. *Atmospheric Environment.* **41**, 7788–7805 (2007).
64. Lee, S. J., Balling, R. & Gober, P. Bayesian maximum entropy mapping and the soft data problem in urban climate research. *Annals of the Assoc of Amer Geographers.* **98**, 309–322 (2008).
65. Coulliette, A. D., Money, E. S., Serre, M. L. & Noble, R. T. Space/time analysis of fecal pollution and rainfall in an eastern North Carolina estuary. *Environ Science & Technology.* **43**, 3728–3735 (2009).
66. Gummer, B. *The Scouring Angel-The Black Death in the British Isles.* The Bodley Head, London, UK (2009).
67. Nazelle, A. D., Arunachalam, S. & Serre, M. L. Bayesian maximum entropy integration of ozone observations and model predictions: An application for attainment demonstration in North Carolina. *Environ Science and Technology.* **44**, 5707–5713 (2010).
68. Armstrong, M. *Basic Linear Geostatistics.* Springer Science & Business Media (1998).
69. ESRI, E. ArcMap 9.3. ESRI® ArcGIS, 2008, 9, 1999–2008 (2008).
70. Memon, A., Darif, M., Al-Saleh, K. & Suresh, A. Epidemiology of reproductive and hormonal factors in thyroid cancer: Evidence from a case-control study in the Middle East. *Int J Cancer.* **97**(1), 82–89 (2002).
71. Davies, L. & Welch, H. G. Current thyroid cancer trends in the United States. *J Am Med Assoc (JAMA).* **140**(4), 317–322 (2014).
72. Morris, L. G., Sikora, A. G., Tosteson, T. D. & Davies, L. The increasing incidence of thyroid cancer: the influence of access to care. *Thyroid.* **23**(7), 885–891 (2013).
73. Li, N., Du, X. L., Reitzel, L. R., Xu, L. & Sturgis, E. M. Impact of enhanced detection on the increase in thyroid cancer incidence in the United States: review of incidence trends by socioeconomic status within the surveillance, epidemiology, and end results registry, 1980–2008. *Thyroid.* **23**(1), 103–110 (2013).
74. Hall, S. E. & Holman, C. D. J. Inequalities in breast cancer reconstructive surgery according to social and locational status in Western Australia. *Eur J Surg Oncol.* **29**(6), 519–525 (2003).
75. Goovaerts, P. Geostatistical analysis of disease data: estimation of cancer mortality risk from empirical frequencies using Poisson kriging. *Int J Health Geogr.* **4**(1), 1 (2005).

## Acknowledgements

This work was supported partially by the Ministry of Science and Technology, China (2015DFA01410) and the China Scholarship Council (201506320036).

## Author Contributions

X.F., G.C. and J.W. designed the study; Q.L. and Y.R. collected the Data; X.F., G.C. and Z.L. analyzed the results; X.F. and G.C. wrote the main manuscript text. All authors reviewed the manuscript.

## Additional Information

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Fei, X. *et al.* Spatiotemporal Co-existence of Female Thyroid and Breast Cancers in Hangzhou, China. *Sci. Rep.* **6**, 28524; doi: 10.1038/srep28524 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>