Web Appendix for Weather and Infant Mortality in Africa

Masayuki Kudamatsu, Torsten Persson, and David Strömberg
IIES, Stockholm University

October 24, 2012

Section 2 – Weather data The CRU data set is problematic for exploiting variation within location over time due to its interpolation method. First, changes in the weather outcomes in a given location may be due to the availability of nearby weather station data over time. Second, if the closest weather station with available data is too far, a long-term average value is used. See Climate Research Unit (undated) for details.

We download the ERA-40 data from ECMWF's Meteorological Archival and Retrieval System.[1]

The satellite data used in the ERA-40 reanalysis mainly provide temperature and humidity. These variables are highly spatially correlated, and thus the spatial resolution of satellite data, which is coarse for the earlier periods, does not hugely affect the data quality of re-analysis.

Section 2 – Seasonal weather variations The Inter Tropical Convergence Zone (ITCZ) is where trade winds from the northeast and the southeast converge (Griffiths 1972). Due to the low pressures along the ITCZ, convectional thunderstorms form daily and dump large amounts of scattered afternoon rains. Over land, the ITCZ moves north and south with the seasons, following the hottest part of the continent, causing large rainfall variations between dry and wet periods in a typical year. This is illustrated in Figure A1, which shows average amounts of rainfall across Africa in the middle month of each quarter.

[1] We are grateful for Heiner Körnich for help in this process.

Section 3 – Reasons for heterogeneous infant-mortality risk Firstborn babies are believed to face a higher risk of death due to malaria in pregnancy than those of higher birth order, although this heterogeneity appears to be absent in low malaria transmission areas (McGregor 1984). Rogerson et al. (2000) and Walker-Abbey et al. (2005) find that teenage women are more likely to be infected with malaria during pregnancy in Malawi and in Cameroon, respectively. Infants born to mothers infected with HIV as well as malaria face higher risks of low birth weight (ter Kuile et al. 2004). In general, firstborn babies, female babies, and babies born by stunted mothers, face a particular risk of low birthweight (Kramer 1987), which makes it plausible that such babies might be particularly at risk in the wake of malaria shocks.

Section 3 – Cubic polynomials Specifically, we include the following terms to the right hand side of equation (2):

$$\rho_3^T T_{g,t}^3 + \rho_2^T T_{g,t}^2 + \rho_1^T T_{g,t} + \rho_3^P P_{g,t}^3 + \rho_2^P P_{g,t}^2 + \rho_1^P P_{g,t}$$

where $T_{g,t}$ and $P_{g,t}$ are the average temperature and the total rainfall, respectively, in grid cell g over the months $t-11$ to t. In subsequent analysis, we always refer to these terms as cubic polynomials in rainfall and in temperature.

Section 3 – Individual heterogeneity and after-birth shocks Following the medical literature summarized earlier in this Web Appendix, we have investigated if the impact of maternal malaria exposure is heterogeneous across different types of babies, mothers, or households. In particular, we have estimated extensions of our basic econometric specification in equation (2), where all right-hand side variables are interacted with indicators for female babies, firstborn babies, young mothers (under 18), stunted mothers (2 standard deviations below the median stature of the WHO Child Growth Standard by WHO Multicenter Growth Reference Study Group, 2006), and households living in regions with high HIV prevalence rate (10% or higher according to the DHS HIV test results conducted in the 2000s). We have also investigated the heterogeneous impact by the education level of the household (whether both the baby's mother and her husband went to school for more than 8 years) and by affluence of the household (owning a majority of the consumer durables listed in the survey questionnaire). In these regressions, we always split the sample between endemic and epidemic areas. However, we find no significant patterns of heterogeneity in the data, while we always continue to find a significant effect of malaria shocks in epidemic areas but no such effects in endemic areas. This lack of heterogeneity across mothers and households is a bit surprising given the clinical evidence cited above. Our failure to find heterogeneous impacts of malaria in pregnancy across parities in endemic areas, however, is consistent with Guyatt and Snow (2001), who report malaria in pregnancy doubles the risk of low birthweight across all parities as well as for first pregnancies. Mutabingwa et al. (2005) also find that infants born to women with malaria-infected placenta are susceptible to malaria infection even if they are of higher birth order.

For each child, we have focused on malaria shocks during the year before birth and we have seen that these shocks in utero have a significant effect on the likelihood of survival. Do malaria shocks after birth affect the probability that a child dies before age one, directly or indirectly through the health of the mother? To analyze this question, we have run regressions where malaria exposure during the first year of life – either month by month or the cumulated number of months with a positive malaria index – is added as its own term and as its interaction term with $z_{g,t}$ to the right hand side of equation (2).[2] We find no significant effects on infant mortality of in-life shocks neither in epidemic areas, nor in endemic areas. On the other hand, in-utero malaria exposure continues to exercise a significant effect on infant death in epidemic areas of similar magnitude as in our earlier estimates..

Section 4 – NDVI and growing seasons The growing season in a particular location depends on many factors other than the extent of rainfall, including soil qualities, crop types and the use of fertilizers. While some gridded information on these other factors exists, we take a convenient short cut to determine the relevant growing season for each of our DHS clusters, through measures of photosynthetic activity by remote sensing.

Photosynthesis is observable from a long distance, because growing plants reflect light at the infrared part of the spectrum and absorb light at the near-

[2]Malaria exposure in the n-th month after birth does not affect the survival of babies who died before turning n months. Therefore, including the 12-month exposure to malaria infection during the first year of life as a regressor to the whole sample will bias the estimated effects towards zero. To deal with this problem, for each n from 1 to 12, we restrict the sample to babies who survive at least the first n months after birth and use how many months are malarious during the n months after birth as a regressor.

red part of the spectrum. Therefore, ecologists often use data collected by satellites to measure plant growth through ongoing photosynthesis. This is known as the normalized difference vegetation index (NDVI). In the ecology and biology literature, the integral of NDVI values over the growing season is often used as a proxy for crop yields (e.g. Rasmussen 1992 for millet yields in Burkina Faso, Rasmussen 1997 and Rasmussen 1998 for millet yields in Senegal).

The map in Figure A2 shows the distribution of the average annual integrated NDVI across Africa, with bluer areas denoting areas with a low value – little photosynthetic activity over the year – and redder areas a high value. The two graphs in this figure plot observed NDVI values as the jagged thin curves over two years, 1982 and 1983, in two locations: one in Burkina Faso just at the boundary to Niger, and one in Tanzania just south of the Victoria Lake. In these graphs, the horizontal axis shows time measured in two-week periods; the vertical axis shows the NDVI value (multiplied by 10,000). Clearly, the peaks are much lower (note the different scales) for the Burkina Faso location than the Tanzania location, reflecting a lower amount of rainfall.

The two graphs in Figure A2 also demonstrate how the TIMESAT program works to determine the growing season from the NDVI time-series data. The TIMESAT program first produces smoothed (filtered) values of NDVI (shown as the thick curve in the graphs), where the smoothing is meant to eliminate temporary random fluctuations, for example, due to variations in cloud cover. The program then produces the times for the start and the end of the growing season defined as the time period in between 20% above one trough to 20% above the next, as shown by the points on the smooth curves in the figure. Notice that the duration of the growing season is much shorter in Burkina Faso than in Tanzania.

Section 4 – Data on crop prices We use crop-price data compiled by the USAID Famine Early Warning Systems Network (FEWS NET).[3] Specifically, we sample six major crops in Africa (maize, rice, wheat, cassava, millet, and sorghum) and markets in eight countries, for which we also have data for infant mortality (Burkina Faso, Ethiopia, Kenya, Malawi, Mali, Tanzania,

[3] We downloaded the data from `earlywarning.usgs.gov/adds` in October 2009. Since then, the FEWS NET has decided to discontinue the data distribution because, according to James Rowland at USGS, they are unable to keep the data up-to-date.

Uganda, and Zambia).[4] Price data is aggregated to the monthly frequency if the original data is daily or weekly. The geographic coordinate of each market in the data is obtained by searching the name of the market in the National Geospatial-Intelligence Agency's Geonames Search.[5] The ERA-40 weather data used in our infant-mortality analysis is then matched spatially with each market in ArcGIS 9.3 (the Spatial Join tool). Figure A3 shows the locations of 424 markets in the sample with the color indicating which climate zone (rainy or arid) the market belongs to. The sample period is from 1970 to 2002.

Section 4 – Empirical strategy for estimating price effects We estimate the following regression equation:

$$\ln p^p_{m,t} = \alpha^p_{m,s} + \beta^p_{c,y} + \gamma r_1^{g,t} + \delta D_{g,t} + \varepsilon^p_{m,t}$$

where $p^p_{m,t}$ is the price (in domestic currency units) of crop p in market m (located within grid cell g and in country c) in running month t (which is month s of year y), $r^1_{g,t}$ the total amount of rainfall during the previous completed growing season (corresponding to the first term on the right-hand side of equation (3)) in grid cell g, and $D_{g,t}$ the indicator for $r^1_{g,t}$ being two standard deviations below the location-specific mean.[6] We control for crop-by-market-by-month fixed effects, $\alpha^p_{m,s}$, so that the impact of weather is identified from year-to-year deviations from the average location-by-crop specific seasonal pattern. We also control for crop-by-country-by-year fixed effects, $\beta^p_{c,y}$, so as to take into account crop-by-country-specific non-parametric trends, as well as national price inflation and exchange-rate changes. The coefficients of interest, γ and δ, measure the percentage change in price due to a one-centimeter increase in growing-season rainfall and due to unusually low growing-season precipitation, respectively. We estimate this equation separately for rainy and arid areas with standard errors clustered at the ERA-40 grid-cell level.

[4]These six major crops account for 57% of calorie availability in Africa in 2000 according to FAO's Food Balance Sheets. In the price data, each crop has subcategories (in flour, dried, fresh, etc.). We treat each subcategory as a single crop when we create fixed effects. Therefore, there are more than 6 crops in the sample.

[5]The address is `geonames.nga.mil`. If the name of the market cannot be found, we use Global Gazetteer Version 2.1 (`www.fallingrain.com/world`), Wikipedia and then the Google search as the final resort.

[6]Note that $D_{g,t}$ is different from $d_{g,t}$ defined in equation (4). It is defined over $r_1^{g,t}$, not over $r_{g,t}$.

The results are displayed in Table 4.

Section 4 – Specification of regressions for heterogeneous effects To investigate the effects of birth seasons, we focus on the drought effects in arid areas and run the regression:

$$m_{i,c,x,t} = \sum_{k=0}^{3} \beta^k r_{g,t} \cdot q_{g,s}^k + \sum_{k=0}^{3} \gamma^k d_{g,t} \cdot q_{g,s}^k + \alpha_{c,s} + \alpha_{x,y} + \varepsilon_{i,c,x,t} \ ,$$

where $q_{g,s}^k$ is a dummy that equals one if calendar month s of date t falls within the k^{th} quarter since the beginning of the growing season in grid g ($k = 0$ for the quarter immediately before the growing season starts). Figure 7 plots estimated coefficients of the γ^ks and their 95% confidence intervals.

To test for different effects of household charateristics, we run the regression:

$$\begin{aligned} m_{i,c,x,t} = {} & \beta r_{g,t} + \beta^f r_{g,t} \cdot f_i + \gamma d_{g,t} + \gamma^f d_{g,t} \cdot f_i \\ & + \alpha_{c,s} + \alpha_{c,s}^f \cdot f_i + \alpha_{x,y} + \alpha_{x,y}^f \cdot f_i + \varepsilon_{i,c,x,t} \ , \end{aligned}$$

where f_i is an indicator of baby i's household type (agricultural or well-educated). Note that this specification allows cluster-by-month and country-by-year fixed effects to differ across different household types. Table 7 reports the estimated coefficients: β, β^f, γ, and γ^f.

Section 5 – Epidemiological estimates of infant deaths due to maternal malaria Steketee et al (2001) review 34 studies, only two of which attempt to estimate the contribution of malaria to infant mortality, namely Greenwood et al. (1992) and Steketee et al. (1996). Both studies are unable to directly estimate the effect because sample sizes are not large enough to estimate effects on such rare events as infant deaths. Instead, both studies first estimate the effect of malaria medicine on birth weight by using a randomized control trial, and then combine this estimate with the correlation coefficient between low birth weight and infant mortality based on observational data. In the words of Steketee et al. (2001), "No study has made the direct observation because the required sample size to make the observation is prohibitively large." This issue is of course even more pronounced in non-endemic areas, which explains the paucity of epidemiological studies there.

Specifically, Greenwood et al. (1992) study villages near the town of Farafenni on the north bank of the Gambia river and finds that malaria medication could reduce infant mortality by 6% (from 50 to 48 per 1000). Stekeete et al. (1996) study women enrolled in four antenatal clinics in a highly endemic malaria area of rural Malawi and estimate the effect of malaria to 3-5% of infant mortality, at an average infant mortality rate of 157 per 1000. Steketee et al. (2001) summarize these estimates and adjust the summary estimate to an upper limit of 8%, to account for effects through anemia. Scaling the effects by regional infant mortality rates, the paper finds that globally, 75,000 to 200,000 infant deaths in each year might be attributable to malaria infection during pregnancy.

Section 5 – Estimates of child malaria deaths based on verbal autopsy reprts Morris et al. (2003) use data from 38 verbal autopsy studies (14 in Africa) to establish the cause of death. The share of deaths in each study attributed to each of five causes (relative to the share attributed to pneumonia), is regressed on a number of characteristics. The ratio of malaria to pneumonia deaths, for example, is regressed on the estimated population at risk of malaria based on a model similar to ours (the MARA model), the proportion of births attended by a qualified professional, access to safe drinking water, and a South Asia dummy variable. The regression coefficients are then used to predict the distributions of under-5 deaths by cause for the entire region of Africa. Using this methodology, the study concludes that 24 percent of child deaths in sub-Saharan Africa are caused by malaria. Later studies such as Johnson et al. (2010) and Black et al. (2010) reach similar numbers based on a larger number of verbal autopsy studies. For Africa as a whole, these authors attribute 16% of under-5 deaths to malaria.

Section 5 – Estimating the number of births We start with the Gridded Population of the World version 3 (CIESIN and CIAT 2005) to obtain the population count in each 2.5 by 2.5 arc minute cells in 1990, 1995 and 2000. Next, we aggregate this number to the intersections of the ERA-40 cells and country boundaries. We then linearly interpolate the population at the intersection level for years in the 1990s. For years before 1990, we use the country-level population growth from the World Development Indicators

(WDI; World Bank 2012) to extrapolate population counts. Finally, we use the country-level crude birth rate from the WDI to obtain the number of births in each year and then aggregate it to the ERA-40 cell level.

Section 5 – Weather and population for 2081-2100 To obtain monthly mean temperature and total precipitation for the period of 2051-2100, we rely on the forecast made by the EC-EARTH climate model for the CMIP5 project.[7] The spatial resolution of data is based on the Gaussian Grid N80, approximately 1.125 by 1.125 degree.[8] To define malaria zones and climate zones for 2081-2100, we use all the 50 years of forecast.

For projecting the number of births in each 1.125 by 1.125 degree cell for the period of 2081-2100, we take the following steps similar to the one used for years 1981-2000. We first aggregate population counts in 2000, obtained from the Gridded Population of the World version 3 (CIESIN and CIAT 2005), to the intersections of the 1.125 by 1.125 degree cells and the country boundaries. We then obtain the country-level data on crude birth rates for 2081-2100 and population growth rates from 2000 to each year during 2081-2100, both based on United Nations Population Division (2011). Finally, we multiply the 2000 population counts with the population growth rates and the crude birth rates for each cell-country intersection and sum them up to the whole cell level. For the number of births during the 1981-2000 period used as the fixed population in columns (2) and (4) of Table 8, we follow the procedure in the previous subsection except that we use the 1.125 by 1.125 degree cells. Appendix Figure A4 shows the spatial distribution of changes in the number of births from the 1981-2000 period to the 2081-2100 period.

[7] The data can be found at the the Earth System Grid Federation website (pcmdi9.llnl.gov) by setting the search categories as: "CMIP5" for Project, "EC-EARTH" for Model, "rcp45" or "rcp85" for Experiment, "mon" for Time Frequency, "atmos" for Realm, and "r8i1p1" for Ensemble. The variable name is "tas" (near-surface air temperature) for temperature and "pr" for precipitation. We thank Heiner Körnich and Johannes Karlsson for helping us download the data and convert the data file format.

[8] In the Gaussian Grid, the latitude is not equally spaced. For creating the maps in this paper (Figures 9 and 10), however, we assume each grid cell is a square of 1.125 by 1.125 degree.

References

[1] Black, Robert E et al. 2010. "Global, Regional, and National Causes of Child Mortality in 2008: A Systematic Analysis." *Lancet*, 375(9730): 1969–1987.

[2] CIESIN and CIAT. 2005. Gridded Population of the World, Version 3 (GPWv3). Palisades, NY: Socioeconomic Data and Applications Center (SEDAC), Columbia University. Available at http://sedac.ciesin.columbia.edu/gpw.

[3] Climate Research Unit. Undated. "Advice for Time-Series Analysis." Available at: http://www.cru.uea.ac.uk/ timm/grid/ts-advice.html [Accessed August 20, 2010].

[4] Greenwood AM, et al. 1992. "Malaria chemoprophylaxis, birth weight and child survival." *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 86: 483–485.

[5] Griffiths, John F., ed. 1972. *Climates of Africa*. Amsterdam: Elsevier Publishing Company.

[6] Guyatt, Helen L., and Robert W. Snow. 2001. "Malaria in Pregnancy as an Indirect Cause of Infant Mortality in Sub-Saharan Africa." *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 95(6): 569-576.

[7] Kramer, Michael S. 1987. "Determinants of Low Birth Weight: Methodological Assessment and Meta-Analysis." *Bulletin of the World Health Organization*, 65(5): 663-737.

[8] Johnson, Hope L. et al. 2010, "Estimating the Distribution of Causes of Death among Children Age 1–59 months in Highmortality Countries with Incomplete Death Certification." *International Journal of Epidemiology*, 39: 1103–1114.

[9] McGregor, Ian A. 1984. "Epidemiology, Malaria and Pregnancy." *American Journal of Tropical Medicine and Hygiene*, 33(4): 517-525.

[10] Morris et al. 2003. "Predicting the Distribution of Under-five Deaths by Cause in Countries without Adequate Vital Registration Systems." *International Journal of Epidemiology*, 32: 1041–1051.

[11] Mutabingwa, Theonest K et al. 2005. "Maternal Malaria and Gravidity Interact to Modify Infant Susceptibility to Malaria." *Public Library of Science Medicine*, 2(12): e407.

[12] Rasmussen, Michael S. 1992. "Assessment of Millet Yields and Production in Northern Burkina Faso Using Integrated NDVI – from the AVHRR." *International Journal of Remote Sensing*, 13(18): 3431.

[13] Rasmussen, Michael S. 1997. "Operational Yield Forecast Using AVHRR-NDVI-Data: Reduction of Environmental and Inter-Annual Variability." *International Journal of Remote Sensing*, 18(5): 1059.

[14] Rasmussen, Michael S. 1998. "Developing Simple, Operational, Consistent NDVI-Vegetation Models by Applying Environmental and Climatic Information. Part II: Crop Yield Assessment." *International Journal of Remote Sensing*, 19(1): 119.

[15] Rogerson, Stephen J et al. 2000. "Malaria and Anemia in Antenatal Women in Blantyre, Malawi: A Twelve-Month Survey." *American Journal of Tropical Medicine and Hygiene*, 62(3): 335-340.

[16] Steketee Richard W et al. 1996. "Developing Effective Strategies for Malaria Prevention Programs for Pregnant African Women." *American Journal of Tropical Medicine and Hygiene* 55(suppl 1): 95–100.

[17] Steketee, Richard W et al. 2001. "The Burden of Malaria in Pregnancy in Malaria-Endemic Areas." *American Journal of Tropical Medicine and Hygiene*, 64(1 suppl): 28-35.

[18] ter Kuile, Feiko O. et al. 2004. "The Burden of Co-Infection with Human Immunodeficiency Virus Type 1 and Malaria in Pregnant Women in Sub-Saharan Africa." *American Journal of Tropical Medicine and Hygiene,* 71(2 suppl): 41-54.

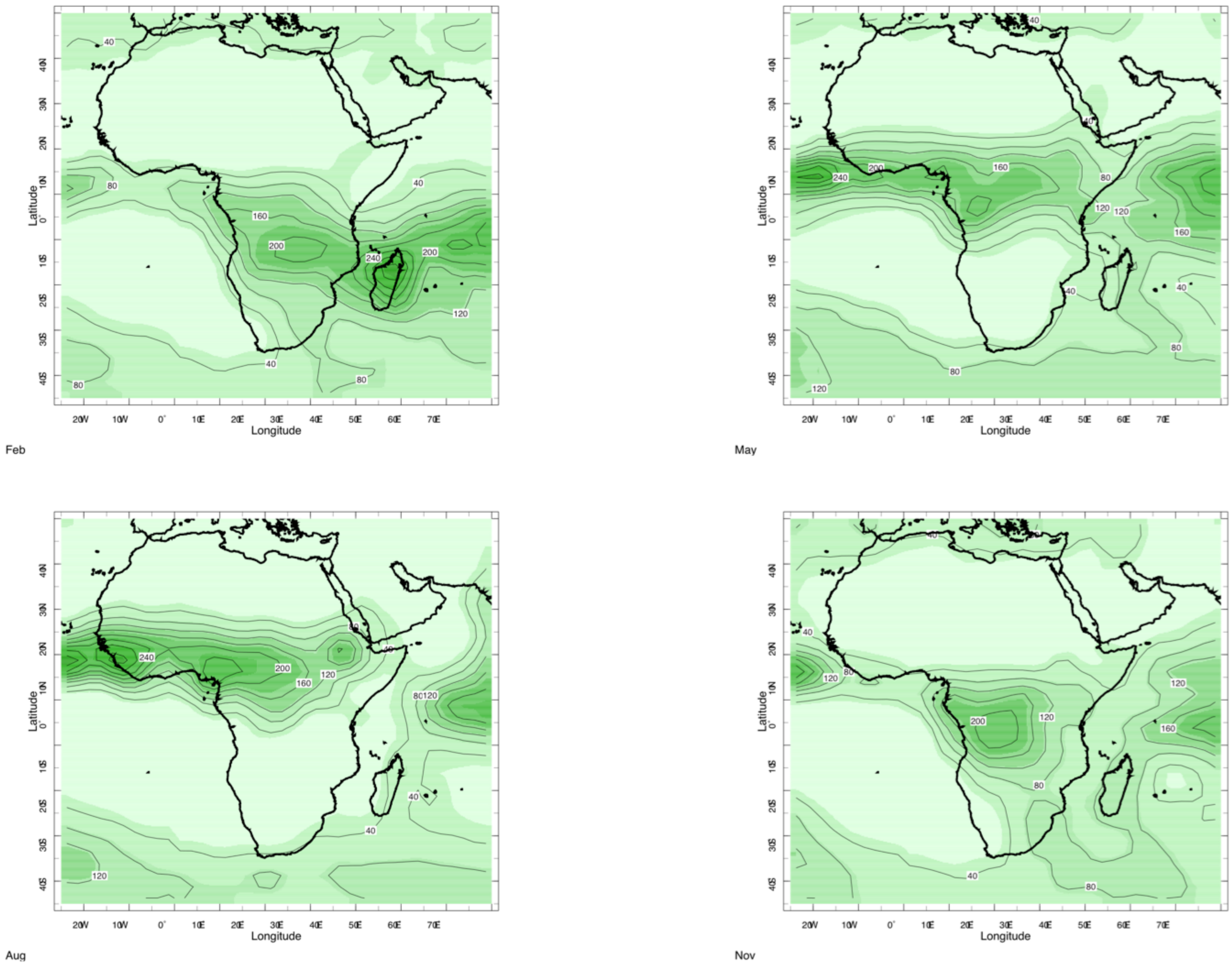
[19] United Nations Population Division. 2011. *World Population Prospects: The 2010 Revision.* Available at esa.un.org/unpd/wpp/unpp/panel_indicators.htm (accessed on 21 October 2012).

[20] Walker-Abbey, Annie et al. 2005. "Malaria in Pregnant Cameroonian Women: The Effect of Age and Gravidity on Submicroscopic and

Mixed-Species Infections and Multiple Parasite Genotypes." *American Journal of Tropical Medicine and Hygiene*, 72(3): 229-235.

[21] WHO Multicentre Growth Reference Study Group. 2006. *WHO Child Growth Standards: Length/Height-for-age, Weight-for-age, Weight-for-length, Weight-for-height and Body Mass Index-for-age: Methods and Development*. Geneva: World Health Organization.

[22] World Bank. 2012. World Development Indicators. Available at data.worldbank.org/data-catalog/world-development-indicators.



Appendix Figure A1 – Total monthly rainfall (in mm) in Africa for February, May, August, and November

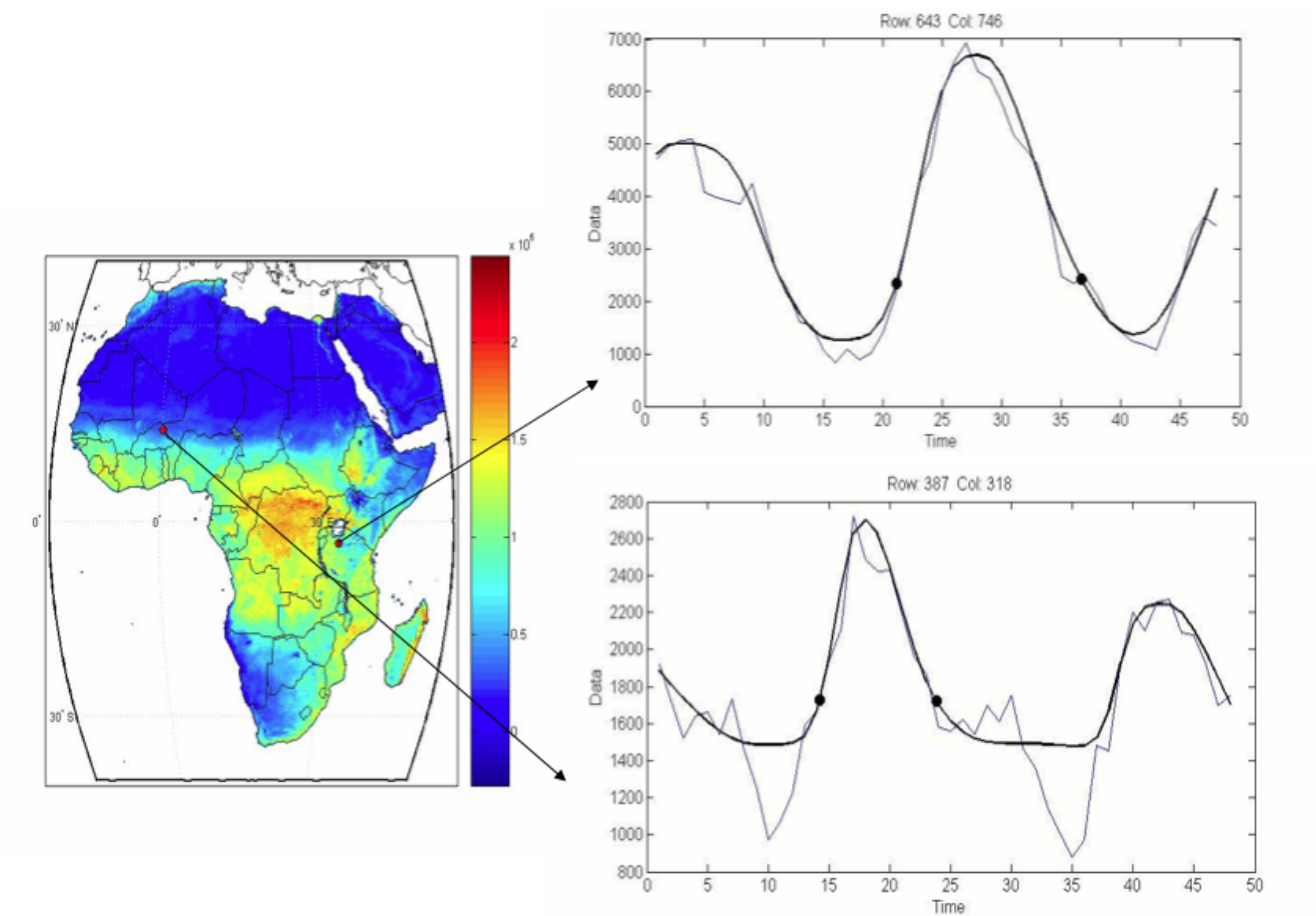
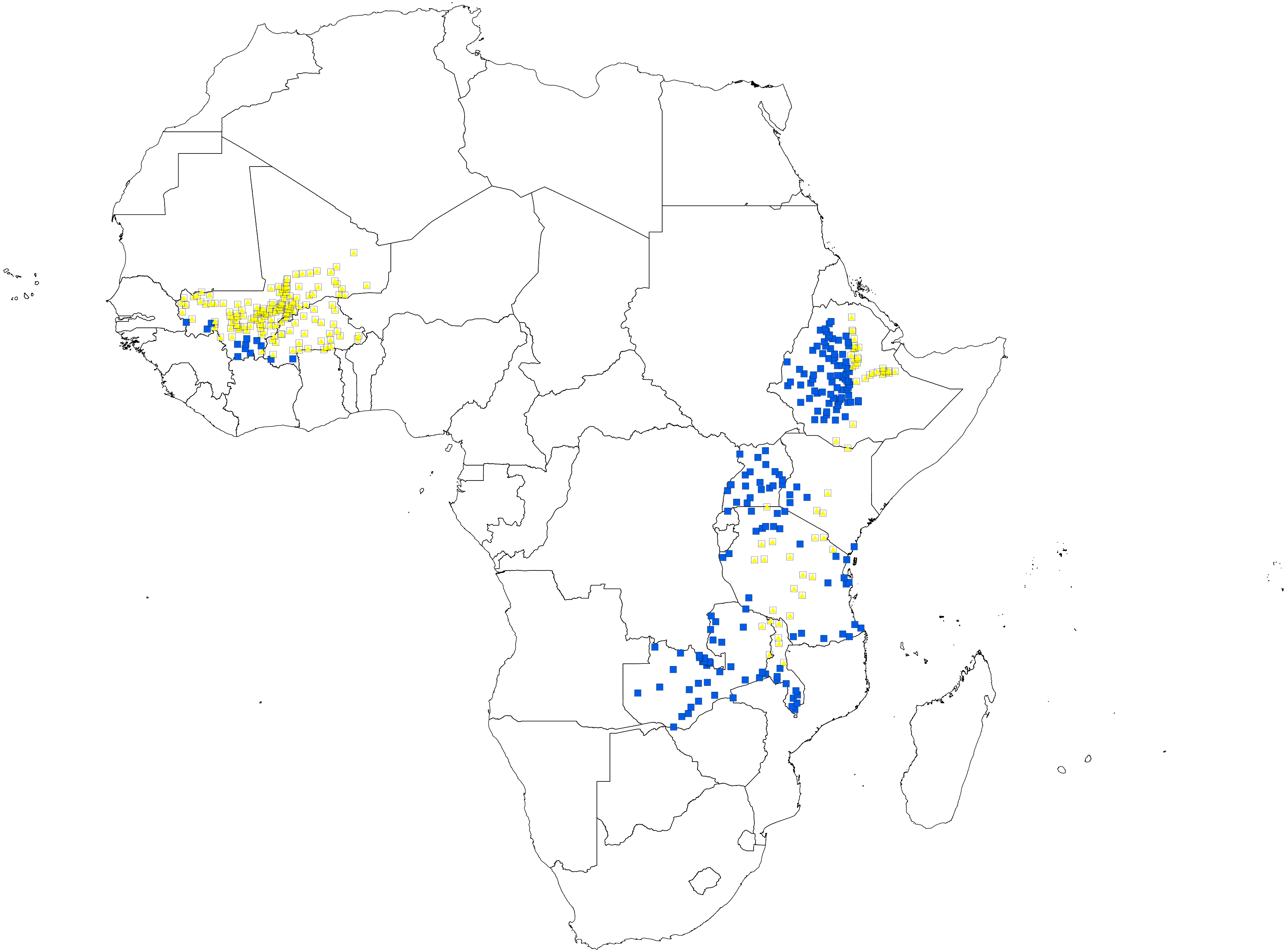
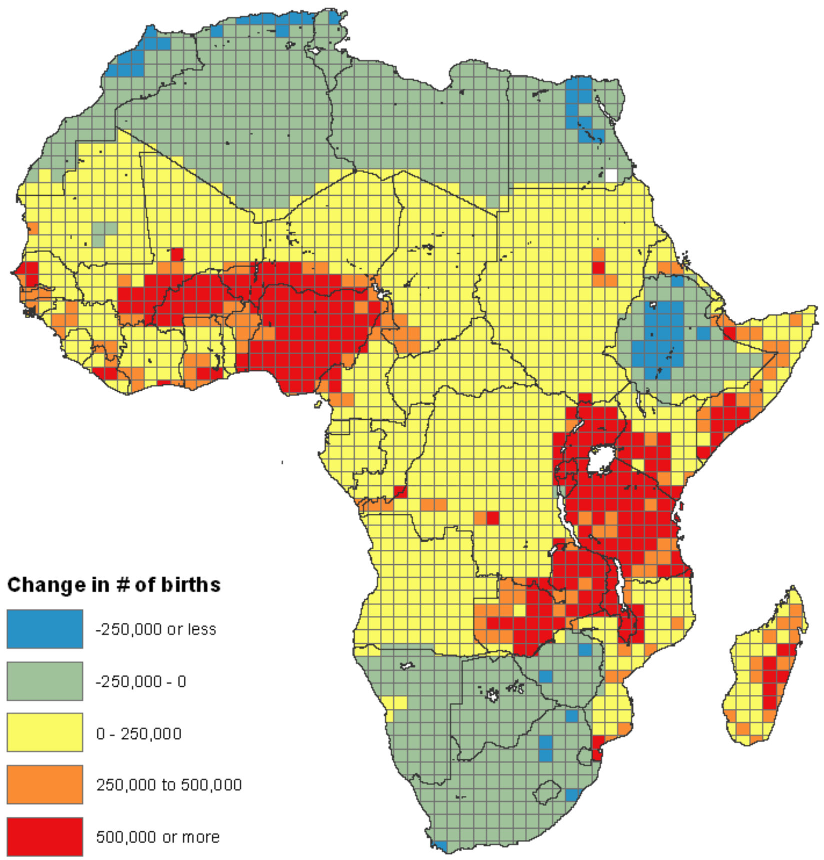


Figure A2 – Actual and fitted NDVI in Burkina Faso and Tanzania



Appendix Figure A3 – Crop markets in the Sample

Notes: Blue and yellow squares indicate markets in rainy and arid areas, respectively.



Appendix Figure A4 – Changes in the number of births from 1981-2000 to 2081-2100

Notes: The map shows the difference of the number of births in 2081-2100 from that in 1981-2000 so that a positive number indicate the 2081-2100 period has a higher number. See the Web Appendix for how the number of births in each period is estimated.