Earth System

Open Access

Science
Data

Discussions

Interactive comment on "Local models reveal greater spatial variation than global grids in an urban mosaic: Hong Kong climate, vegetation, and topography rasters" *by* Brett Morgan and Benoit Guénard

Brett Morgan and Benoit Guénard

brettmorgan2@gmail.com

Received and published: 26 March 2019

Dear Anonymous Referee 1, Thank you very much for reviewing the manuscript and providing your feedback and concerns. Below we provide point to point responses (AC) to your comments (RC), as well as changes in the manuscript (CM). Page and line numbers refer to those in the submitted manuscript. We also provide an attached pdf document showing tracked changes, new citations, figures, and an appendix added to the original manuscript.

On behalf of the authors,
Brett Morgan

RC - Referee comment   **AC** - Author comment   **CM** - Change in the manuscript

RC1.01 The manuscript is not sufficiently organized and confused with no novelty and explicit research question. There are many too short subsections, which should be merged. Methods are not much clear because details and relevant references have not been provided. Consequently, it is not much easy to follow results and discussion.
AC1.01 We agree that improving the clarity and organization of the manuscript is necessary, though challenging because of the large number of data inputs, outputs, and analyses. We have restructured and added to sections (especially Section 3 - Methods and Section 4 - Results and Discussion) to improve clarity. The novelty of the manuscript is the data itself, as stated on Page 1, Line 9: "To our knowledge, this is the first set of published environmental rasters specific to Hong Kong."; Page 4, Lines 1-3: "Therefore Hong Kong is in dire need of a comprehensive suite of accessible environmental GIS data, at a resolution finer than 1 km, suitable for species distribution modeling and other local applications. To this end, we developed new, 30 m resolution rasters of topography, NDVI, and interpolated climate variables for each month of the year."; and Page 10, Line 25: "This diverse set of 30 m resolution topography, climate, and remote sensing data include the first published interpolation of long-term climate averages specific to Hong Kong." Please see AC1.06 for our response regarding a research question. As most readers will likely use only parts of the provided data, we believe that retaining the subsections will help the reader quickly find information of relevance for the data they want to use. Lumping subsections together would likely add to the confusion mentioned.

RC1.02 The Authors have used data associated at support sizes very different. They should take into account the change of support.
AC1.02 We are uncertain what the reviewer means by "data associated at support sizes very different," and would appreciate further explanation. If the concern is that

input rasters used as model predictors were initially at different resolutions, higher resolution products were resampled to 30 m before model building (Page 4, Lines 15-16).

RC1.03 The title should be made more informative and effective.
AC1.03 We have reformulated the title to make it more informative and better reflect the focus of the manuscript. We welcome additional suggestions on how it could be improved.
CM1.03 Title: New 30 m resolution Hong Kong climate, vegetation, and topography rasters indicate greater spatial variation than global grids within an urban mosaic

RC1.04 The Abstract has not the required structure and does not summarize the whole manuscript. It should be organized better and explain clearly what was done, what was found and what are the main conclusions. Generally, the first sentence should provide briefly the rational of the topic being investigated.
AC1.04 We are not aware of abstract structure requirements that this abstract does not adhere to. In the ESSD manuscript preparation guidelines for authors, it is stated "The abstract should be intelligible to the general reader without reference to the text. After a brief introduction of the topic, the summary recapitulates the key points of the article and mentions possible directions for prospective research. Reference citations should not be included in this section, unless urgently required, and abbreviations should not be included without explanations. Please include the DOI(s) to the referenced data set(s) as well as the citation(s)."

RC1.05 Keywords are missing.
AC1.05 We would happily provide keywords, but we did not find a format for them in the Earth Systems Science Data LaTeX template, and published papers in ESSD do not have keywords.

RC1.06 The Introduction section is confused ant not sufficiently organized. Particularly, reading the title, one is expecting to find in the Introduction the presentation of what the title promises, but unfortunately it is not so. The Introduction should be improved and

the topic being investigated should be explained clearly. The novelty and objectives are missing. A manuscript to be considered a research paper, a research question must be clearly stated. In addition, the Authors should explain the gap in the topic being investigated and how their study fills such a gap.

AC1.06 We hope the changes in the title resolve the stated discrepancy in the introduction. Many of the missing elements (novelty, objectives, research gap, research question) are present in section 2 about the study area, which is meant to be an extension of the introduction. For example, the knowledge gap is that Hong Kong is lacking appropriate resolution data for local applications (Page 4, Line 1). The order of these elements could be rearranged, but it seems less logical to pose this research question and the objective of developing higher resolution rasters before introducing Hong Kong and the existing GIS data available for it. Alternatively, sections 1 and 2 (Introduction and Study Area) could be merged into a single large introduction section. However we believe keeping these sections separate allows the reader to more easily navigate to content of interest. We are skeptical that a central research question is necessary for this manuscript. Much scientific research is indeed hypothesis-driven, but in alignment with the title of this journal, Earth System Science Data, our project is data-driven. In the "About" section of the ESSD website, it is stated "Articles in the data section may pertain to the planning, instrumentation, and execution of experiments or collection of data. Any interpretation of data is outside the scope of regular articles." In agreement with this defined scope, our primary goal in writing this manuscript is to describe the development of the provided data, rather than answering a central question.

RC1.07 A well-organized Materials and Methods section is missing. The sections '2 Study area' and '3 Methods' should be included in a new Materials and Methods section which allows readers to follow the progress of the objectives in the manuscript and support results and discussion. In the methods, how data have been analysed and combined should be explained providing sufficient details. Particularly, the Authors should explain how they have taken into account the change of support problem to have all data associated to the same support size. Details and references on statistical

methods are missing.

AC1.07 We share your concerns on the methods section, which we have improved with various changes in structure, additional statistical details, and references throughout. Specifically we have better explained the meaning of each variable and the reasoning behind their development. We do not believe that merging the methods section with section 2, "Study Area," would be beneficial. Section 2 is largely descriptive and doesn't cover any of the materials (data sources) used in the analyses, so the content would be out of place in a Materials and Methods section. As said in AC1.02, we are unsure what is meant by support size, and we would appreciate further explanation.

RC1.08 Results and Discussion sections should be improved and supported by a new Materials and Methods section.

AC1.08 For the Materials and Methods sections, please refer to AC1.07. The results and discussion section has been modified to improve the clarity and content of the manuscript. This has included creation of section 4.5 "Limitations and next steps" and section 4.4 "Value and Utility," which discusses the results in consideration of how they will enable SDM and other environmental research in this important region.

RC1.09 Conclusions are poor: they should be improved and to show the improvement of our knowledge.

AC1.09 Thank you for this feedback, we agree that improved conclusions are desirable. We believe the improvement in our knowledge is summarized in the first sentence of the conclusions: "This diverse set of 30 m resolution topography, climate, and remote sensing data include the first published interpolation of long-term climate averages specific to Hong Kong."

Please also note the supplement to this comment:
https://www.earth-syst-sci-data-discuss.net/essd-2018-132/essd-2018-132-AC1-supplement.pdf

Printer-friendly version

Discussion paper

ESSDD

Interactive comment

Printer-friendly version

Discussion paper

ESSDD

Interactive
comment

Interactive comment on "Local models reveal greater spatial variation than global grids in an urban mosaic: Hong Kong climate, vegetation, and topography rasters" *by* Brett Morgan and Benoit Guénard

Brett Morgan and Benoit Guénard

brettmorgan2@gmail.com

Received and published: 26 March 2019

Dear Anonymous Referee 2,
Thank you very much for reviewing the manuscript and providing your feedback. Below
we provide point to point responses (AC) to your comments (RC), as well as changes in
the manuscript (CM). Page and line numbers refer to those in the submitted manuscript.
We also provide a pdf supplement showing tracked changes, new citations, figures, and
an appendix added to the original manuscript.

On behalf of the authors,
Brett Morgan

RC - Referee comment　　**AC** - Author comment　　**CM** - Change in the manuscript

RC2.01 This study aims to produce a high resolution (30 m raster) data set of climate and environmental variables for the Hong Kong region. Unfortunately, I find the manuscript to be confusing, showing an overall disconnection between sections. The manuscript focuses on a large but incomplete description of the variables included in the data set, and does not address the main conclusion stated in its title ("Local models reveal greater spatial variations than global grid in an urban mosaic").

AC2.01 Changes addressing these concerns have been made throughout the manuscript. In our modifications, we attempted to make the manuscript more cohesive, with more explanation of how the various data are related. This included completion of the methods section with more detailed descriptions of variables. We have added Figure S1, a flow chart which we hope illustrates connections between sections as well as the data provided. We have altered the title to make it more informative and better reflect the focus of the manuscript.

CM2.01 Title: <u>New 30 m resolution</u> Hong Kong climate, vegetation, and topography rasters <u>indicate greater spatial variation than global grids within an urban mosaic</u>

RC2.02 The introduction section discusses the application of "Species distribution modeling (SDM)" and how this type of analysis is affected by the spatial resolution of the environmental data employed. However, this introductory discussion seems to be irrelevant within the context of the manuscript, as SDM is rarely mentioned again throughout the text. Abbreviations such as NDVI are used throughout the abstract and introduction but are not explained until the later sections of the methods section.

AC2.02 We agree that the omission of meaningful discussion of SDM implications was an oversight. We have added this in Section 4.4 Value and Utility, including how the results will benefit SDM studies and why this improvement in our knowledge is much

needed. We have clarified the meaning of NDVI in the abstract and introduction.

CM2.02 Page 1, Line 7: The data include topographic variables, Normalized Difference Vegetation Index, and interpolated climate variables based on weather station observations.

Page 2, Line 21: For example, vegetation measures like the Normalized Difference Vegetation Index (NDVI) in fragmented forests are unlikely to be relevant if the grain size is much larger than the forest patch size, because each grid cell will be a single averaged value.

RC2.03 In the method section each of the topographic and climate variables, as well as remote sensing products are mentioned. However, it seems to me that each of the subsections focuses on irrelevant details, and there is no clear descriptive explanation of a) what these variables are? b) why were they chosen? and c) how were they processed?

AC2.03 We have modified the various methods subsections where this information was missing, adding more description of the meaning of each variable and stating that variables were chosen based on the availability of source data, as well as their expected utility in SDM research. We have added Appendix 1, which provides definitions of all climate and topography variables for easy reference. We have also added Figure S1, which shows the general raster workflow, helping explain how variables were processed.

AC2.03 Page 4, Line 8: The variables developed were selected based on their utility in environmental research, especially SDM, as well as the availability of appropriate source data.

RC2.04 The results and discussion section is also vague and difficult to read. There is no clear distinction between the validation data set/model and the novel data/model analysis produced by this study. The figures lack explanation within the main text, and it is hard to see how they convey the results of the study.

AC2.04 In the results and discussion section, we have reorganized much of the

content including merging research limitations into one section, and adding discussion of the results in light of potential SDM applications. We added sentences at the end of the paragraph describing cross-validation procedures, to clear up the distinction between the validation and final rasters produced. We have added additional figure references in the text, and we would welcome suggestions on how the figures could better convey the results.

CM2.04 Page 6, Line 17: This cross-validation procedure was used only to produce these validation measurements. The finalized monthly climate rasters described above were trained using all available data.

RC2.05 Overall, I believe this manuscript needs substantial revisions, and perhaps a reassessment of the scientific goals that it is trying to communicate.

AC2.05 We believe we have fixed the main problem, which was that the previous manuscript title was misleading regarding the main scientific goals. Other sections of the manuscript have been modified to reflect this, and more detail and explanation has been added to improve clarity.

Please also note the supplement to this comment:
https://www.earth-syst-sci-data-discuss.net/essd-2018-132/essd-2018-132-AC2-supplement.pdf

Interactive comment on Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-132, 2018.

Interactive comment on "Local models reveal greater spatial variation than global grids in an urban mosaic: Hong Kong climate, vegetation, and topography rasters" *by* Brett Morgan and Benoit Guénard

Brett Morgan and Benoit Guénard

brettmorgan2@gmail.com

Received and published: 26 March 2019

Dear Anonymous Referee 3,
Thank you very much for reviewing the manuscript and providing your feedback. Below we provide point to point responses (AC) to your comments (RC), as well as changes in the manuscript (CM). Page and line numbers refer to those in the submitted manuscript. We also provide a pdf supplement showing tracked changes, new citations, figures, and an appendix added to the original manuscript.

On behalf of the authors,
Brett Morgan

RC - Referee comment   **AC** - Author comment   **CM** - Change in the manuscript

RC3.01 Throughout section 4, you provide comments on how the dataset you have created could or should be improved. This is useful, but it also gives the impression that your dataset is not that good after all. It would be better to either a) clarify in the introduction and abstract that this work is simply a first pass, and that more needs to be done, or b) collect all of these comments in a separate section. Perhaps you can include them in section 4.4: limitations and next steps.
AC3.01 We have combined and condensed the discussion of data limitations into Section 4.4: Limitations and next steps. Thank you for the suggestion.

RC3.02 Data: I found it hard to quickly extract information about the datafiles from the figshare website. Can you reproduce Table 1 along with the data?
AC3.02 Thank you for this suggestion, we have added a file in the figshare repository equivalent to Table 1 of the manuscript, showing file descriptions, units, and raster summary statistics.

RC3.03 Why are you not providing the monthly data through figshare or the doi?
AC3.03 The thought was that having another 120 raster files in the repository would complicate finding and downloading the desired files, especially because we expect only the yearly summary layers would be of interest to most users. We decided to compress the monthly models into a single zip file now available in the repository. This way, they are available for those who are interested but avoid user confusion.
CM3.03 Page 11, Line 4: Individual monthly rasters for each of the 10 climate variables are available as a compressed zip file.

RC3.04 Many of the data files seem to be relatively binary: black or white. I'm not an expert in rasters so I might be missing something here, but how can I extract the

C2

high-resolution detail you are championing in the article?

AC3.04 If we understand correctly, the black and white you describe is referring to the file previews shown on the figshare website. Unfortunately figshare seems to have this standard rendering of raster files, displaying the preview as a binary image. To access the raster data directly and display it how you want, you would need to download the files and open them in GIS software, such as QGIS.

RC3.05 Page 1, line 4: 'variations' not 'variation'

AC3.05 We adopted this suggestion.

CM3.05 Further, these global datasets likely underestimate local climate variations because they do not incorporate locally relevant variables.

RC3.06 Page 3, line 22: are hill fires always human-induced?

AC3.06 As far as is known, yes. In Hong Kong, lightning only occurs during heavy rain, usually during the monsoonal summer. The vast majority of these fires happen during the dry winter, with spikes in frequency associated with holidays and religious practices where burning in hillside cemeteries is practiced. (see Chau, 1994: http://hub.hku.hk/handle/10722/34430)

RC3.07 Page 5, line 2: 'temperature buffers' not 'a temperature buffer'

AC3.07 We adopted this suggestion.

CM3.07 Water bodies adjacent to land areas can act as temperature buffers, contribute to evaporative cooling (Lookingbill and Urban, 2003), and influence precipitation patterns (Heiblum et al., 2011; Paiva et al., 2011); therefore considering their distribution is important for climatic predictions.

RC3.08 Page 5, line 22: can you provide a reference to Hong Kong's dense network of stations?

AC3.08 Yes, we have added a reference.

CM3.08 In contrast, interpolation in Hong Kong is benefitted by a relatively small geographic area and a quite dense network of weather data provided by dozens of perma-

nent weather stations (Hong Kong Observatory, 2018).

RC3.09 Page 5, line 27 and 28: I think you should add the word 'absolute' before the variables maximum and minimum temperature, to clarify that these are the highest and lowest temperatures recorded in each month.

AC3.09 We did consider adding the word "absolute" to these variables, but this might add to confusion about the meaning of the measurement. "Absolute maximum temperature" of a given month might normally refer to the highest temperature ever recorded in that month, but what we provide instead is data that represents the averaged absolute maximum values recorded over a period of 20 years. To ensure clarity of the meaning of each variable, we have added Appendix 1, which provides definitions for easy reference.

RC3.10 Page 6, line 10: why do you have high confidence in the long-term averaged weather station data?

AC3.10 The confidence is based on the good availability of measurements for averaging - a weather station's data was only included if at least 8 full years of measurements were available for use, and most stations had many more years than that threshold. So we can be fairly confident that the averages are good approximations of the true climate at each station.

CM3.10 This low lambda value was selected because of the relatively high confidence in the long-term averaged weather station values (based on at least 8 years of data).

RC3.11 Page 7, line 8: please refer to the resolution of the rasters as 30m, to be consistent.

AC3.11 We adopted this suggestion.

CM3.11 All rasters are provided at an identical 1 arc second (30 m) resolution and in the WGS84 geographic coordinate system.

RC3.12 Page 7, line 15: Aren't you only providing the rasters at one scale?

AC3.12 The rasters are all at one resolution (30 m), but the values of these topographic

C4

variables were calculated using buffers at multiple scales. For example, relative elevation at a given 30 m pixel will vary depending on the size of the surrounding area (in our case, a circle of a given radius) to which it is compared. We adjusted the wording to help clarify.

CM3.12 For this reason, we provide these rasters calculated at multiple buffer scales.

RC3.13 Page 8, line 5: Can you provide a brief explanation why the highest maximum temps are in inland valleys?

AC3.13 Yes, and we have also added a sentence about how the difference in maximum vs. minimum temperature patterns can be explained by urban heat island effects.

CM3.13 This pattern may be explained by urban heat retention: buildings act as heat sinks which absorb solar radiation during the day, and slowly release heat at night, causing increased minimum temperatures. The high maximum temperatures in inland valleys may be due to reduced air circulation in sheltered locations, and lack of complex vegetation or urban structures providing shade.

RC3.14 Page 8, line 15-16: Aren't you arguing in this study that your new dataset is high resolution? Consider rephrasing this sentence.

AC3.14 Not exactly. We do want to highlight that our rasters have a much higher resolution than similar datasets that are available globally. However, describing a raster simply as "high" or "low" resolution is quite arbitrary, as many 1 km datasets are described as "high resolution." For certain applications, 30 m resolution would be quite low. Assessment of urban microclimate is one of those cases, and we attempt to convey this.

RC3.15 Page 9, line 29: I would say 'our models', rather than 'The new models'
AC3.15 We adopted this suggestion.
CM3.15 Our models generally indicate greater spatial variation than Worldclim, with cool areas colder, warm areas hotter, and wet areas wetter.

Please also note the supplement to this comment:

https://www.earth-syst-sci-data-discuss.net/essd-2018-132/essd-2018-132-AC3-supplement.pdf

Interactive comment on Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-132, 2018.

ESSDD

Interactive
comment

Interactive comment on "Local models reveal greater spatial variation than global grids in an urban mosaic: Hong Kong climate, vegetation, and topography rasters" *by* Brett Morgan and Benoit Guénard

Brett Morgan and Benoit Guénard

brettmorgan2@gmail.com

Received and published: 26 March 2019

Dear Anonymous Referee 4,
Thank you very much for reviewing the manuscript and providing your feedback. Below we provide point to point responses (AC) to your comments (RC), as well as changes in the manuscript (CM). Page and line numbers refer to those in the submitted manuscript. We also provide a pdf supplement showing tracked changes, new citations, figures, and an appendix added to the original manuscript. This comment appears to have had some text encoding errors which we have left intact.

On behalf of the authors,
Brett Morgan

RC - Referee comment   **AC** - Author comment   **CM** - Change in the manuscript

RC4.01 The authors developed a very high-resolution (30m) gridded dataset of climate, NDVI, and topography for Hong Kong. The meteorological observations from weather stations are interpolated using thin plate spline model. The motivation for fine resolution dataset for Species Distribution Modeling (SDM) is clear and the final product of the study will be useful for SDM and other local applications, however, the manuscript lacks justification for the methodology used and meaningful evaluation of results. It seems to me that the construction of climate data at this high resolution is the novelty of the paper and the main finding (i.e. greater spatial variation in finer resolution data than the coarser) does not add anything new. The way method section is described is not clear; each variable is prepared separately and then they were used as inputs to the statistical model for the climatology interpolation? Why did the authors choose this method over others and how are the 6 predictors chosen? Also, the use of 'climate modeling' in the text is confusing as it usually refers to general circulation models or regional climate models, but the terminology is used for the spatial interpolation model. I recommend changing the title to something like "development of 30 m raster dataset of climate, vegetation, and topography for Hong Kong" and list specific comments below.

AC4.01 We agree that the main finding and novelty of this study is the higher resolution of the developed rasters. To reflect this focus and better represent the contents of the manuscript, we have modified the title as seen below. However, we believe that the findings of greater spatial variation in climate results is still a salient component of the study. Although this is a result that might be expected, it will have important consequences in projects that use this data, in particular for species distribution modelling for which changes in a few degrees can substantially modify final outcomes. Users should know that not only is the resolution different from products like WorldClim, but also that

the values are different. We also believe the increased variation indicates that global climate interpolation data exclude climate forcing factors that are relevant at smaller scales, which is an important result. We have made various changes and clarifications in the methods section, and provide a workflow schematic in Figure S1 to illustrate preparation of all of the rasters and variables. The thin plate spline methodology was used because of the availability of tools in the R environment to implement it, as well as its history of use in climate interpolation research, which we have now addressed in the manuscript. To select the six climate predictors, we searched the literature for what types of variables have been used in similar studies in the past, and then used those that we expected to have climatic effects at the geographic extent and scale of this study.

CM4.01 Title: <u>New 30 m resolution</u> Hong Kong climate, vegetation, and topography rasters indicate greater spatial variation than global grids within an urban mosaic
Page 5, Line 31: <u>Independent variables were selected by searching the literature for similar studies, and choosing predictors we expected to have an influence on climate at this regional scale.</u>

RC4.02 Gridded meteorological datasets have been generated using station observations and a variety of interpolation methods in the past. A flagship climate dataset may be the CRU climate data (New et al, 2002) which used thin plate spline technique, with functions of latitude, longitude, and elevation (and mean precipitation for precipitation coefficient of variation). The technique seems to be the standard in recently increasing number of global gridded climatological datasets with increasing spatial resolution (eg. WorldClim2, TerraClim). Additional spatial information that represent physical processes are required in order to resolve higher resolution. I understand a unique situation for Hong Kong for the small domain with dense station network, which may allow simplification compared to constructing global data, but it would be helpful to tie into existing gridded climatology data w.r.t. method of prediction. The paper may shed some lights on improving precipitation interpolation.

AC4.02 Thank you! We agree it is important to make clear that this methodology has

been used much in the past as well as recently in climate interpolation studies. We have added explanation in the methods to reflect this.

CM4.02 Page 5, Line 22: Here we use multiple linear regression to predict geographic climate patterns using weather station training points and raster covariates. This is followed by thin plate spline (TPS) interpolation (see Wahba, 1979) of the regression model residuals. TPS is a widely used approach in climate interpolation (e.g. New et al., 2002; Fick and Hijmans, 2017), which fits a curved surface to irregularly distributed points. This two-step interpolation (regression followed by TPS) was based on the approach of Meineri and Hylander (2017).

RC4.03 The stations should be indicated in the map of Hong Kong, Figure 1.

AC4.03 Adding the stations to Figure 1 would make it quite busy due to the density of stations in Hong Kong, so instead we plotted a new map as a supplemental figure showing the distribution of weather stations. It shows stations from which temperature or rainfall measurements are available. It is also used an opportunity to display elevation more clearly.

CM4.03 Added Figure S2.

RC4.04 Methods: I'm aware that R is a statistical package software. But what is the prediction model usedâA ÌȨṪ̈ĞlÌinear regression? Section 3.2, page 5 line 30- page 6 line 17 describes two-step process, which seems to be the main model (as referred to "our model", "local model", "new model"). Either moving section 3.2 to the first section, or giving an overview of the model before subsections begin, and streamlining the reference to the model will help clarify. Does water proximity include inland water bodies such as river, pond, and wetland? Could NDVI be included as a predictorâA ÌȨṪ̈Ğwouldn't it add more physical characteristics? Though annual mean or monthly climatology of NDVI, rather than instantaneous is suitable.

AC4.04 We have added an overview section as suggested (shown in CM4.02) that makes clear the modeling method is linear regression, with other relevant explanation. Water proximity does consider inland water bodies, most of which are artificial reser-

voirs. NDVI could be used as a predictor, but it is variable on a granular scale where neighboring pixels can have very different values. It is unlikely that climate variables would vary in this way, so we decided not to include NDVI as a predictor. Also, even if they are surrounded by vegetation, most weather stations are likely positioned on some type of structure that would bias local NDVI measurements.

CM4.04   Page   5,   Line   5:   <u>Second</u>,   water   proximity (including inland water bodies) was calculated as the percent surface land in the area surrounding a given pixel.

RC4.05 As NDVI data is the only remote sensing, physical variable that resolves 30-m, I think it's important to compile climatology. Authors admit that the index values vary seasonally (page 10, line 12), which seems to contradict with the statement earlier on the instantaneous NDVI being representative. With strong seasonality of rainfall pattern in Hong Kong from June to August, I'd expect NDVI would respond. Landsat data extends several decades, so I can't imagine there's not enough data to capture seasonal variation. If no data during monsoon season, dry and winter low and wet summer high would be useful.

AC4.05 We are unsure what is meant by "compile climatology," in relation to NDVI, and if our following response does not address this comment we would appreciate further explanation. We agree that discussion of the NDVI data may appear contradictory, and would like to emphasize that while the NDVI values may fluctuate, we expect that the overall geographic pattern of NDVI (highest in dense forests, lowest in urban centers) remains fairly consistent throughout the year. While precipitation is indeed strongly seasonal, the vegetation in Hong Kong is not seasonally deciduous, but evergreen, so changes in NDVI are unlikely to be drastic. Exceptions may include agricultural areas with rapid shifts associated with harvest cycles. The difficulty in acquiring suitable Landsat images for Hong Kong stems from several factors, as touched on within the manuscript. First, cloud cover is a hindrance, and overcast skies are especially common in the first half of the year. For the months of June-August, mean cloud cover (at the Hong Kong Airport weather station) can range from 65 to 80 percent. So

most Landsat images captured during this time are entirely obscured, and very few are cloud-free. Second, no Landsat image covers the entirety of Hong Kong: it is on the edge of the satellite path, and so generating a complete NDVI raster requires finding and merging two or more suitable images taken at a similar time of the year.

Upon checking Landsat 8 further, creating a dry season / winter NDVI layer would certainly be possible, but for calculating summer NDVI not enough cloud-free images are available. We will check Landsat 7, 5, and 4 databases as they are also available at 30 m resolution, but a concern with using older data that is that in comparisons some NDVI differences will be due to succession of vegetation and disturbance of lands due to continuous development over the years rather than seasonality.

RC4.06 Precipitation results (4.2.2): I don't understand the last sentence. GCM outputs can be a predictor? If you mean using dynamical models, neither the GCMs nor even higher-resolution regional weather forecast model can't resolve micrometeorlogy at 30-m. Downscaling dynamical model climatology is a possibility but it will be a whole new paper and I'm not sure if it's attainable for 30m with limited information at hand.

AC4.06 Thank you, yes it makes sense that circulation models wouldn't be helpful without first downscaling them, and indeed that seems like it would be a substantial effort outside the scope of this study.

CM4.06 The sentence has been removed.

RC4.07 Climate variables discussion (4.2.4): Though direct validation is not possible, temperature and precipitation could be evaluated qualitatively. Worldclim2 is average for 1970-2000 but your climatology is for 1998-2017, so it's not apple-to-apple comparison. Did you adjust Worldclim data? Could that be the reason for huge discrepancy in precipitation? TerraClimate data set is coarser at âĹij 4km but coves 1958-2015 (https://www.nature.com/articles/sdata2017191), so you could get closer climatology of 1998-2015 for the comparison. I would first check if the climatology agrees at station locations, then map out the differences at 4 km. For temperature, you can downscale the 4 km data to 30 m via elevation correction using constant lapse rates of

-6.5âŮęC/km (Willmott and Matsuura, 1995; Maurer et al., 2002) since you have 30 m elevation data which can easily be aggregated up to 4km. The downscaled temperature should provide similar features as the modeled results and physical range of differences to expect. Also, effects of predictors other than elevation would be shown where they differ. Precipitation is difficult to evaluate or even to predict as indicated in the text. Does Hong Kong have radar data?

AC4.07 In our comparisons with WorldClim, the primary goal was to assess the geographic differences between the two datasets. So if in choosing a dataset for comparison, there is a tradeoff between higher resolution rasters (WorldClim 2) and a more congruent temporal window (TerraClim), we prefer to use the dataset with higher resolution. We did not adjust the WorldClim values. We did consider testing model predictions (both our models and WorldClim) against actual station values, but it seems there would be an issue of testing with data that was used to train one model but not the other. As for downscaling 4 km TerraClimate data down to 30 m based on elevation only, we are unsure that would allow for a valid comparison. Temperature lapse rates would likely vary by geographic region, as well as the temperature variable under consideration. The Hong Kong Observatory does record radar data (shown here https://bit.ly/2FyS4Rj), but it doesn't seem historical radar data are available for download.

RC4.08 Skin temperature from Landsat could be another data to evaluate the heterogeneity of the modeled temperature. Though skin temperature is not exactly the same as in-situ 2-m air temperature, it is an observation based, independent data.

AC4.08 Thank you for this interesting proposition. Is skin temperature the same as thermal infrared images (Landsat provides two thermal infrared bands, but only at 100 m resolution)? It seems this could be quite biased by many factors, like the ground cover, sun intensity at the moment the image was taken, etc. While we were able to put together NDVI, we are no remote sensing experts, and so would need some more guidance on how validation using Landsat data might be accomplished.

RC4.09 My understanding is that bilinear interpolation is for coarser to finer spatial interpolation and for aggregating from finer to coarser, arithmetic or area weighted average is appropriate. I'm wondering if using bilinear to aggregate from 30m to 1 km (Figure 7 etc.) results in different 1km if arithmetic averaging is used.

AC4.09 We are not familiar with using arithmetic or area weighted average interpolation, but our understanding is that bilinear interpolation is generally appropriate to use for resampling continuous rasters. We think the different methods would likely produce different results, but then the challenge would be to evaluate which is actually superior to the other in terms of accuracy.

RC4.10 Next step: It is important to note what's missing and limited for future enhancement, but you should also encourage people to use this dataset. Isn't the dataset ready to use in SDM to address the issues raised in the introduction section? 30 m is remarkably high resolution and the entire raster data contain valuable information for many modeling studies and local management applications.

AC4.10 Thank you! We have consolidated limitations into one section (4.4), and have tried to be more positive about the opportunities offered by the newly developed rasters in our discussion.

Please also note the supplement to this comment:
https://www.earth-syst-sci-data-discuss.net/essd-2018-132/essd-2018-132-AC4-supplement.pdf

―――――――――――――――――――――

Interactive comment on Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-132, 2018.

~~Local models reveal greater spatial variation than global grids in an urban mosaic:~~ New 30 m resolution Hong Kong climate, vegetation, and topography rasters indicate greater spatial variation than global grids within an urban mosaic

Brett Morgan[1] and Benoit Guénard[1]

[1]School of Biological Sciences, The University of Hong Kong, Hong Kong SAR

Correspondence: Brett Morgan (brettmorgan2@gmail.com)

Abstract. The recent proliferation of high quality global gridded GIS datasets has spurred a renaissance of studies in many fields, particularly biogeography. However these data, often 1 km at the finest scale available, are too coarse for applications such as precise designation of conservation priority areas and species distribution modeling, or purposes outside of biology such as city planning and precision agriculture. Further, these global datasets likely underestimate local climate ~~variation~~ variations

because they do not incorporate locally relevant variables. Here we describe a comprehensive set of 30 m resolution rasters for Hong Kong, a small subtropical territory with highly variable terrain where intense anthropogenic disturbance meets a robust protected area system. The data include topographic variables, ~~NDVI~~Normalized Difference Vegetation Index, and interpolated climate variables based on weather station observations. We present validation statistics that convey each climate variable~~'~~'s reliability, and compare our results to a widely used global dataset, finding that our models consistently reflect greater climatic

variation. To our knowledge, this is the first set of published environmental rasters specific to Hong Kong. We hope this diverse suite of geographic data will facilitate future environmental and ecological studies in this region of the world, where a spatial understanding of rapid urbanization, introduced species pressure, and conservation efforts is critical. The dataset is accessible at https://figshare.com/s/3a5634e36e80dc33444c.

1 Introduction

Scale of analysis has long been considered a key concern in biogeographic research (Levin, 1992). Multiple types of scale are relevant to environmental data, including analysis grain, response grain, spatial structure, and study extent (Mertes and Jetz, 2018). Analysis grain, the minimum unit of spatial resolution in a spatial grid, is commonly referred to as a pixel or cell. In research that uses environmental raster data, the pixel size directly dictates the types of biogeographic questions that can be reasonably addressed.

This relationship between analysis grain and study suitability is complex, and higher resolutions are not always advantageous. For example, in global analyses excessively high resolution data would be computationally cumbersome and unnecessary if the goal is to characterize broad patterns. However as shown below, many studies have found notable benefits of higher

resolution climatic predictors. Unfortunately, regional analyses lacking local data are limited to using global datasets and the grain size at which they are available (e.g. Cheng and Bonebrake, 2017).

Species distribution modeling (SDM) is a common application of gridded environmental data, where the selected analysis grain has important consequences. In SDM, one or more geographic predictors are associated statistically with the location of known observations of a species. The resulting statistical model can be converted to a geographic model: a spatially continuous measure of species occurrence likelihood across the landscape of interest. SDMs are used for many applications, including predicting potential ranges of invasive species, characterizing ecological constraints on species ranges, discovering biodiversity, and planning protected areas (Peterson et al., 2011). The ~~manipulation~~ effects of SDM grain size manipulation is an active area of research. Below, we summarize findings on four main effects~~are summarized~~: estimated distribution size, inclusion of fine scale features, predictor variable selection, and model predictive ability.

Coarser environmental data consistently result in SDMs that predict larger areas of species presence (Connor et al., 2017; Franklin et al., 2013; Seo et al., 2009). Overestimation of SDMs is especially a concern for conservation purposes, where inferred size of suitable habitat is often used to inform extinction risk assessments. Mistakenly large calculated distributions could result in species that are assigned artificially low risk levels.

Coarse resolution predictors can cause SDMs to omit small, but important areas. Particularly of interest are microrefugia, climatically unique patches of land that can harbor rare species, and are especially important for conservation as species distributions respond to climate change (Dobrowski, 2010). Meineri and Hylander (~~2016~~2017) demonstrated that because high resolution climate models included such microrefugia, the resulting species distribution models predicted lower extinction rates for plant species than coarser predictors. Nezer et al. (2016) found that 10 m or 100 m resolution SDMs can reveal other distribution features invisible at lower resolutions (1 km): movement corridors, isolated habitat patches, geomorphologic features, and anthropogenic effects on distributions.

SDM scale can also affect which predictors are selected for model calculation. Certain predictors may be excluded in SDMs because they lack explanatory power at the chosen scale of analysis (Mertes and Jetz, 2017). For example, vegetation ~~indices like NDVI~~ measures like the Normalized Difference Vegetation Index (NDVI) in fragmented forests are unlikely to be relevant if the grain size is much larger than the forest patch size, because each grid cell will be a single averaged value. This means that coarse models might not only mischaracterize the distribution pattern itself, but they also may fail to explicate important environmental relationships that determine species occurrence. Indeed, Nezer et al. (2016) found that the most important predictors (vegetation, slope) in their highest resolution models (10 m) were "nearly meaningless" at 1 km resolution. Another study found similar differences in predictor importance related to variation in scale (Lasseur et al., 2006). Of course, predictor importance is always relative and thus is subject to which predictors are included in model building. Therefore this pattern is not expected to be observed in all studies, but should not be overlooked as a potential source of bias.

Last, any consistent effects of SDM grain size on the overall predictive ability of SDMs are unclear. The most commonly used measure of SDM performance is Area Under Curve (AUC), where a higher value indicates a greater ability to differentiate between area the species is present or absent. Some studies found increased SDM resolution resulted in increased AUC (Seo et al., 2009; Nezer et al., 2016), while others found no effect (Pradervand et al. 2014) or mixed effects depending on dataset

(Guisan et al., 2007). These studies used different species, predictors, scales, regions, and modeling algorithms, so further research is required to investigate any association between SDM grain size and AUC.

The above advantages of higher resolution environmental data in SDM may be dependent on project-specific factors, such as the quality of species records available and the goals of the research. For example, using environmental grids of a smaller grain size than the locational accuracy of the available species records is untenable. Additionally, stationary species (e.g. lichens) may be more strongly affected by local factors while highly mobile species (e.g. birds) may only be limited at broader scales. Indeed, it has been shown that plant (rather than bird or mammal) species models with highest locational accuracy were those most improved by higher resolution (Guisan et al., 2007). Lastly, the utility of fine grain environmental grids may depend on habitat; flat deserts likely have less biologically relevant fine-scale spatial variation compared to mountainous forests or subtropical areas fragmented by human activity, like Hong Kong.

In this study, a new series of rasters for Hong Kong are introduced particularly suited for SDM. The layers produced focus on long term climate averages, topography, and vegetation. We asked how the new 30 m scale rasters provide new information on climatic variables in Hong Kong, in comparison to a global dataset already available. We hypothesize that our new climate data will indicate greater variation (measured as raster standard deviation) in climate variables. The development of high-resolution environmental rasters is particularly important in tropical regions where species exhibit small distribution ranges (as predicted by Rapoport's Rule: Stevens, 1989) and where understanding interactions between organisms and their changing habitats is paramount.

2 Study area: Hong Kong

Geographic data of appropriate resolution is critically important for conducting research within the Hong Kong Special Administrative Region of China, because of its complex landscape. Hong Kong exhibits dramatically variable topography, fitting numerous small islands, dozens of mountain peaks over 500 m, 733 km of coastline, and a human population of over 7 million into a land area of only 1,104 km² (Fig. 1). Seasonally variable monsoon winds deliver equatorial heat and torrential precipitation in summer, while northerly winds carry chilly dry air from continental Asia during the winter (Dudgeon and Corlett, 1994). However, daily temperature fluctuations are attenuated by the surrounding South China Sea and Pearl River Estuary. Hong Kong's terrain typically exhibits a stark bifurcation between some of the most densely constructed areas in the world (Lau and Zhang, 2015) and steep, vegetated slopes. Uninhabited expanses are protected as part of 24 country parks and additional special areas that cover over 40% of the territory's land (Agriculture, Fisheries and Conservation Department, 2017). Even within these more natural areas, a strong disturbance gradient encompasses grasslands, shrublands, evergreen secondary forests, and old-growth *feng shui* woods that have been protected from deforestation. Historically Hong Kong has been largely stripped of its trees, and only since the end of World War II and later the establishment of the Country Park system have large swathes of forest begun to regenerate (Zhuang and Corlett, 1997). However this process is frequently reset by human-induced hill fires, which maintain predominantly upland areas as shrubland or grassland (Marafa and Chau, 1999). Hong Kong harbors several unique and restricted habitats, including mangroves in coastal areas and freshwater wetlands in the far northwest.

Hong Kong climate data is available within a variety of global gridded climate datasets (WorldClim 2 - Fick and Hijmans, 2017; MerraClim - Vega et al., 2017; CHELSA - Karger et al, 2017), but none of these have a resolution higher than 1 km. We suspect those global climate models underestimate variation in local climate values, even after consideration of the coarser scale. Local studies of Hong Kong meteorology have largely focused on characterizing and mitigating the effects of urbanization (e.g. Shi et al., 2018; Wang et al., 2017; Nichol et al., 2014; Liu and Zhang, 2011; Ng, 2009; Giridharan et al., 2004). Unfortunately, it appears the climate of Hong Kong's landscape as a whole has been given little notice, and we are unaware of long-term averaged climate rasters available for the region. Relevant studies that do exist include limited variables, and the data appear to be publicly unavailable. We are additionally unaware of Hong Kong data publicly available for vegetation indices such as NDVI, or topographic data other than elevation.

Therefore Hong Kong is in dire need of a comprehensive suite of accessible environmental GIS data, at a resolution finer than 1 km, suitable for species distribution modeling and other local applications. To this end, we developed new, 30 m resolution rasters of topography, NDVI, and 10 interpolated climate variables for each month of the year. We hypothesize that in addition to providing this finer resolution, our new climate data will indicate greater variation (measured as raster standard deviation) in climate variables than currently available global data products.

3   Methods

All data manipulation and geographic analyses were conducted in the R statistical computing environment (v3.3.2, R Core Team, 2016) using RStudio (v1.0.136, RStudio Team, 2015) unless otherwise noted. Analyses are divided into three broad categories of data products, detailed in the sections below: topographic variables, climate variables, and remote sensing variables. The variables developed were selected based on their utility in environmental research, especially SDM, as well as the availability of appropriate source data. An overview schematic of the data workflow is available in Figure S1.

3.1   Topographic variables

Data on the physical characteristics of Hong Kong's landmass were assembled from remote sensing inputs, crowdsourced coastline polygons, and a digital terrain model. The topographic variables developed are coastline, elevation, slope, aspect, terrain roughness, relative elevation, distance to coast, water proximity, and urbanicity.

3.1.1   Coastline

As reclamation of land from the ocean in Hong Kong is ongoing, obtaining current data for the coastline can be challenging. Natural coastline and reservoir vectors were downloaded from OpenStreetMap (2018) and merged in QGIS (v3.01, QGIS Development Team, 2018) to produce a shapefile of polygons representing Hong Kong land area as of January 2018. All output rasters were masked to this area.

4

3.1.2 Elevation, slope, aspect, and roughness

A 5 m resolution Hong Kong digital terrain model (Lands Department, 2017) was upscaled using bilinear resampling. The resulting 30 m DEM was used as the elevation data throughout the study. Four other topographic predictor layers were derived directly from this DEM: aspect, slope, aspect*slope, and a roughness index. These were calculated using the Hong Kong elevation raster with the terrain() function in the R raster package, using all 8 neighboring cells (queen case). Aspect was transformed from degrees to a measure of north-south exposure ("northness") by cos(aspect*pi/180).

3.1.3 Relative elevation

~~A~~ Relative elevation is a measure of the difference in elevation between the pixel of interest, and the lowest pixel within a given radius. A pixel on a mountain peak has a high relative elevation, while a pixel on a flat plain has a relative elevation of 0 (regardless of its elevation above or below sea level). A set of relative elevation layers for Hong Kong were calculated at multiple scales, following the moving window approach of Bennie et al. (2010). ~~They are a measure of the difference in elevation between the pixel of interest, and the lowest nearby pixel within a given radius.~~ The radii used were 60 m, 120 m, 240 m, 480 m, and 960 m. These ~~may be useful for a variety of purposes, but~~ layers are expected to be most applicable as measures of surface water drainage, and therefore soil moisture as well. Relative elevation has ~~also been used in climate modeling~~ been used as a covariate in climate interpolation as a proxy for cool air draining (Bennie et al., 2010; Ashcroft and Gollan, 2012), but ~~we do not include it~~ was not included here as a predictor ~~here~~ as Hong Kong lacks large valleys and other sheltered areas where this effect would be most relevant.

3.1.4 Distance to coast and water proximity

Water bodies adjacent to land areas can act as ~~a temperature buffer~~temperature buffers, contribute to evaporative cooling (Lookingbill and Urban, 2003), and influence precipitation patterns (Heiblum et al., 2011; Paiva et al., 2011); therefore considering their ~~distribution~~ presence is important for climatic predictions. Here, two different methods were used to quantify water body distribution in Hong Kong: distance to coast and water proximity. A distance to ~~coastline rasterwas produced,~~ coast raster, measured in meters, was produced using the distance() function in the raster package ~~and a~~ with the Hong Kong coastline shapefile ~~. However, because of the complexity of Hong Kong's coastline, it appears simple distance to coast may not be the best representation of water proximity for climate predictions. Therefore,~~ described in section 3.1.1. Distance to coast did not incorporate inland water bodies. Second, water proximity (including inland water bodies) was calculated as the percent surface land in the area surrounding a given pixel. A value of 1 means that the area within a given radius is entirely terrestrial, while 0 indicates it is entirely aquatic. Multiple water proximity rasters ~~at varying scales were also calculated ,~~were calculated with varying radii using a circular moving window approach ~~similar to that described in other climate interpolation studies (like that described by~~ Aalto et al. ~~,~~ (2017), to represent buffering processes at different scales. The radii used were 0.75 km, 1.5 km, 3 km, 6 km, and 12 km. ~~A value of 1 means that the area within the given radius is entirely terrestrial, while 0 indicates it is entirely aquatic.~~

3.1.5 Urbanicity

~~In~~ Urbanicity rasters were developed because in densely constructed areas, ~~the~~ urban heat island ~~effect is~~ effects are expected to influence temperatures (Nichol et al., 2013; Shi et al., 2018), and therefore urbanicity may be an important predictor in climate interpolation. High rise buildings can influence temperature by blocking wind, creating shade, acting as heat sinks, and producing thermal pollution. These effects are particularly relevant for this study, as some of Hong Kong~~'~~'s weather observation stations are adjacent to or inside urban centers. To quantify the distribution of developed area, we used a 30 m resolution dataset of percent impervious surface (Brown de Colstoun et al., 2017), which we expect to strongly correlate with urban development. ~~However~~ For use in climate predictions this data was smoothed using a Gaussian moving window, because bulk air temperature is not expected to vary at a granular (30 m) scale~~, this data was smoothed using a Gaussian moving window at three~~. at three buffer scales (sigma = 10, 50, 100)~~to create 'urbanicity' layers .~~, using the focalWeight() and focal() functions in the raster R package, where type = 'Gauss'. The resulting 'urbanicity' layers were later used as climate predictors. In these rasters, completely impervious locations have a value of 100, while vegetated areas are 0.

3.2 Climate variables

Climate interpolators are often faced with the challenge of estimating climate parameters over a large area ~~with~~ using sparse weather station observations, at least in part of the region considered (e.g. Hu et al., 2016). In contrast, interpolation in Hong Kong is benefitted by a relatively small geographic area and a quite dense network of weather data provided by dozens of permanent weather stations (Hong Kong Observatory, 2018; see Figure S2). Here we use multiple linear regression to predict geographic climate patterns using weather station training points and raster covariates. This is followed by thin plate spline ~~interpolation,~~ (TPS) interpolation (see Wahba, 1979) of the regression model residuals. TPS is a widely used approach in climate interpolation (e.g. New et al., 2002; Fick and Hijmans, 2017), which fits a curved surface to irregularly distributed points. This two-step interpolation (regression followed by TPS) was based on the approach of Meineri and Hylander (2017).

Weather station observation data and geographic coordinates were downloaded from the web portal of the Hong Kong Observatory (2018). As the goal was to produce a representation of long-term but modern climate, measurements over 20 years (1998 to 2017) were included. To ensure averages were reliable, weather stations were only included for interpolation of each variable if at least 8 years of complete data were available within the 20 year window. The minimum number of stations used for each model is provided in Table 2. Monthly observations of ten variables were obtained: maximum temperature, mean daily maximum temperature, mean daily temperature, mean daily minimum temperature, minimum temperature, mean dew point, mean relative humidity, mean wind speed, mean air pressure, and total rainfall.

~~Spatial climate modeling~~ Climate interpolation consisted of two main steps. First, a ~~generalized~~ linear model was built for each climate variable for each month of the year. ~~Six topographic climate predictors were~~ Independent variables were selected by searching the literature for similar studies, and choosing predictors we expected to have an influence on climate at this regional scale. When necessary, each predictor was statistically transformed to approach a normal distribution. The six topographic predictors used as model building candidates were: elevation, log-transformed distance to coast, exponentially

transformed fine and coarse water proximity, log-transformed urbanicity (sigma = 50), and 'northness' - the cross product of aspect and slope. The water proximity layers were products of ~~combining multiple scales~~ additively combining multiple scale rasters into fewer predictors: fine water proximity was the sum of 0.75 km, 1.5 km, 3 km scale rasters, while coarse was the sum of 6 km, and 12 km. ~~All~~ The six model predictors were tested for collinearity and no problems were found. Linear models were built using the lm() R function. All predictors were initially included, then using the step() function, pared down in each regression model using stepwise bidirectional selection based on AIC, using 4 degrees of freedom as a penalty to make predictor selection stricter than the default. The resulting regression model was used to calculate a climate value at each grid cell based on a linear relationship with the selected predictors.

Second, to adjust for local variation in climate that is not associated with topography, the linear model residuals at each station were calculated and interpolated using the thin plate spline approach implemented in the fields R package. The lambda smoothing parameter, which determines how closely the fitted surface matches input values, was set to 0.01. ~~A fairly~~ This low lambda value was selected because of the relatively high confidence in the long-term averaged weather station values (based on at least 8 years of data). This effectively produces a smoothed layer of local deviation from the linear model, which was used to additively adjust the results of the linear ~~models~~ model predictions and produce finalized climate rasters.

We measured the spatial predictive ability of models using ten-fold ~~cross validation~~ cross-validation (Dobesch et al., 2007). In each validation round, 10% of weather stations were reserved as a test dataset and the remainder were used for training. Average root mean squared error of the test data subset from the final model prediction was used as an error measurement. To normalize these error measures across the climate variables, we adjusted them as a percentage of the standard deviation of the initial weather station values measured. This cross-validation procedure was used only to produce these validation measurements. The finalized monthly climate rasters described above were trained using all available data.

The ~~monthly models~~ finalized monthly rasters were then summarized into ~~raster~~ layers that characterize yearly climatic means and variation. These include 19 "bioclimatic" variables using the biovars() function in the dismo R package (Hijmans et al., 2017), which are specifically suited for species distribution modeling and other ecological purposes. This also allows our data to be compared with other climate data products that use the same calculations. Because those calculations only use rainfall and average daily maximum and minimum temperatures in each month, we also produced yearly average layers of dewpoint, relative humidity, mean daily temperature, air pressure, and wind speed. Also provided are layers of highest and lowest average monthly extreme temperatures, and their difference (extreme temperature annual range). These two variables characterize temperature extremes experienced in a given location better than the bioclimatic variables.

For comparison with global climate data products, we resampled bioclimatic variables to the same (1 km) resolution as WorldClim using bilinear interpolation. Only pixels present in both data products were used for comparisons.

3.3   Remote sensing data

Normalized difference vegetation index (NDVI) is a common metric of vegetation presence and density derived from satellite imagery. To calculate ~~normalized difference vegetation index (NDVI), Landsat~~ NDVI, Landsat 8 images (U.S. Geological Survey, 2018) of Hong Kong were obtained. We downloaded one image from March 2016 that covers much of Hong Kong

except for the far eastern areas, and is free of clouds. This was supplemented with an image from March 2018 after adjustment, so that all land areas of the region were included. NDVI calculations were completed using the standard equation:

$$NDVI = (NIR - Red)/(NIR + Red) \tag{1}$$

Where NIR is near-infrared (Landsat band 5: 0.851 to 0.879 µm) and Red is visible red radiation (Landsat band 4: 0.636 to
0.673 µm). The resulting NDVI value varies between 1 and -1, where higher values correspond with denser vegetation.

4 Results and discussion

Results of this environmental analysis of Hong Kong include 48 rasters and one vector file. All rasters are provided at an identical 1 arc second (~~0.03 km~~30 m) resolution and in the WGS84 geographic coordinate system. Summary values and filenames are provided in the data repository.

4.1 Topographic variables

Distance to coast results show that approximately 42% of Hong Kong's land area is within 1 km of the coastline. However it is apparent that inland areas often feature steep inclines, as half of Hong Kong's land is above 84 m elevation.

For variables like relative elevation, urbanicity, and water proximity, the ideal scale of raster calculation is dependent on the desired effect to be captured, and perhaps other characteristics of the landscape in question. For this reason, we provide ~~the rasters at multiple~~ these rasters calculated at multiple buffer scales.

Urbanicity results show that the majority of land in Hong Kong is not near urban areas, as the median raster ~~value is~~ values are below 4% urban at all scales calculated ~~. It is also apparent that inferring urban development from impervious surface is not ideal, as sometimes bare soil or rock are sensed as impervious. Also, there is little ability for such a measure to differentiate between a dense urban core of high-rises, and large paved areas (such as parking lots or airports). Unfortunately, accessible data on the geographic distribution of the urban environment in Hong Kong is limited. For climate modeling, an urbanicity measure that takes into account building height or population density at a 30 m or finer scale could be preferable.~~ (Table 1). This shows that although Hong Kong has extremely dense urban cores, most of its mountainous terrain is unpopulated.

4.2 Climate variables

Minimally, 32,024 monthly weather station measurements over 20 years (1998 to 2017) were used to construct climate models for all months and variables at finer resolution compared to global datasets (Fig. 2). High weather station density and availability of data on multiple candidate topographic climate-forcing factors allowed for high confidence in many climate variable models, especially those related to temperature (Figs. 3, 4). The climate interpolation results include monthly models of ten variables including temperature, precipitation, and humidity, making a total of 120 individual models produced (monthly models of three

temperature variables are shown in Fig. 5). For all variables, the predictors included in monthly models are displayed in Figure 6, and the number of stations with data included is in Table 2.

4.2.1 Temperature

Temperature was found to vary considerably across Hong Kong, with more than 6ºC difference in mean annual temperature between the highest mountain peaks (>900 m, <18ºC) and some low-lying urbanized areas (>24ºC). While mean and minimum temperature are highest in urban areas, maximum temperature shows a different pattern with a maximum in inland valleys in the northern New Territories. ~~The high~~ This pattern may be explained by urban heat retention: buildings act as heat sinks which absorb solar radiation during the day, and slowly release heat at night, causing increased minimum temperatures (see Oke, 1982). The high maximum temperatures in inland valleys may be due to reduced air circulation in sheltered locations, and lack of complex vegetation or urban structures providing shade. The high accuracy of temperature models (Figs. 3, 4) is likely due to a strong association with elevation; elevation was by far the most commonly included predictor for temperature models (Fig. 6). Urbanicity was important for mean and minimum temperature, but not maximum temperature. Water proximity and coast distance were differentially included depending on the variable, while aspect*slope rarely had an effect. ~~A few special notes on the limitations of predicted temperature values in dense urban areas: temperature predictions of our landscape-level climate models will not reflect the high spatial variation in temperature found in urban microclimates. Although the manned Kowloon HKO weather station is inside a densely populated area, as pointed out by Nichol and To (2012) it is still in a small parklike area surrounded by trees, and therefore is not representative of the most densely urbanized areas of Hong Kong. Other stations in urban areas are similarly near green spaces or otherwise open areas. Higher resolution (say 5 m or 1 m) studies of urban thermal distributions would strongly benefit from analysis of wind patterns, building height, thermal pollution, and other factors (e.g. Shi et al., 2018). Therefore granular, ground-level temperatures in urban areas are likely substantially different than the broader air temperature values our models provide. One area of particular interest is the Hong Kong Airport, a massive area reclaimed from the ocean, north of Lantau Island. The weather station here has the highest urbanicity value, because the airport is mostly impervious surface. However it lacks properties of truly urbanized areas, with no permanent population lacking typical urban morphology. Therefore climate variables in this area may be biased, especially for variables like wind speed that would be affected by the presence of tall buildings. The airport weather station often had the highest mean temperatures recorded, perhaps indicating that extensive impervious surface is more important than wind blockage or thermal pollution for maintaining high temperatures.~~

4.2.2 Rainfall

In our models, the highest annual rainfall (bio12) areas in Hong Kong (>2500 mm annually) are inland and at high elevations, presumably because of condensation from humid air as it passes over mountains. Areas near the coast, particularly small outlying islands and the eastern coast in Lung Kwu Tan receive the lowest amount of annual rainfall (<1600 mm). Precipitation of driest month (bio14) was uniformly low, ranging from 20 to 40 mm, but the relative pattern of high and low precipitation areas remained similar. The most commonly included model predictor was fine-scale water proximity (Figure 6). Elevation

was predictive for 5 out of 12 months, but few other topographic predictors were useful. Seasonality of rainfall in Hong Kong is strong. Averaged across all locations, 52% of total yearly rainfall was recorded in three months (June through August). ~~Although rainfall~~ Rainfall models were informed by more weather stations than any other climate variable (Table 2), but they have the highest relative standard error (Fig. 3) and therefore the lowest accuracy. Because they are influenced by both global and locally variable wind patterns, precipitation distributions are notoriously difficult to predict, especially in urban areas (Cristiano et al., 2017). ~~Our relatively poor results may be explained by this, as well as lack of appropriate local predictors. We did not explore the possibility of using global circulation models as predictors because we expected that they would affect areas of Hong Kong equally, but perhaps they would be beneficial.~~

4.2.3   Dew point, humidity, pressure, and wind speed

Dew point exhibits a similar pattern to other temperature variables, with mean annual dew point ranging from 15.5ºC at mountain peaks to around 19ºC on small islands and lower areas. Mean annual relative humidity reaches a maximum of about 90% at Tai Mo Shan, while many urban areas in Kowloon, Tuen Mun, and Yuen Long are between 70 and 75%. Surprisingly, mean annual air pressure has a positive correlation with elevation; the highest values (reaching 1014 hPa) are at mountain peaks, and particularly low values (as low as 1012.5 hPa) in coastal areas of southern and western Hong Kong. Mean annual wind speed is also strongly associated with elevation, with mean annual values above 30 km/h on Lantau Island mountain peaks, down to below 5 km/h in interior low elevation areas of the New Territories.

4.2.4   ~~Climate variables general discussion~~ Comparisons with global climate data

~~Similar to other climate interpolation studies, bias in the physical locations of automatic weather stations may be of concern. Weather stations are often intentionally placed in flat, open areas with the goal of measuring weather that is relevant to a broad geographic area, rather than locations that may experience unique local climate. It may be for this reason that Slope*Aspect was infrequently useful for model construction, as few stations are on steep slopes. Elevational distribution of stations may also be a source of bias. Although a weather station operates at the highest point in Hong Kong (Tai Mo Shan, 955 m), there are only two other stations above 600 m.~~

~~We used cross-validation to measure the spatial predictive ability of the climate models. However, this method is only able to test models against locations where weather stations are present, and validation based on an independently collected dataset would be ideal. One common validation method is to use weather data loggers placed across elevational and land-use gradients (Meineri and Hylander, 2017). Such an approach would allow for explicit testing and comparing predictiveness of climate products for different areas of Hong Kong.~~

Our new climate models are compared with a recent global climate dataset to identify differences in predictions of Hong Kong climate values (Fig. 7). ~~Worldclim~~ WorldClim 2 was produced using a similar interpolation approach with regression modeling and thin plate spline interpolation, but also included satellite-derived covariates in addition to topography (Fick and Hijmans, 2017). Because ~~Worldclim~~ WorldClim incorporates vast amounts of data from multiple databases covering over-lapping geographic and political entities, it is difficult to ascertain exactly which individual weather stations were included,

and we were unable to determine whether any Hong Kong weather stations were included or if the datasets are completely independent. However, the model predictions differ substantially (Figs. 1, 7; Table 3). ~~The new~~ Our models generally indicate greater spatial variation than ~~Worldclim~~WorldClim, with cool areas colder, warm areas hotter, and wet areas wetter. For example in average low temperature of coldest month (bio6), high elevation areas could be more than 2ºC lower, and urban areas

5   more than 2ºC higher than ~~Worldclim~~ WorldClim indicates (Fig. 7a). To further quantify differences in values between these two datasets, for each of the 19 bioclimatic variables we calculated the standard deviation of raster values (Table 3). All of our interpolated climate rasters had a higher standard deviation than their ~~Worldclim~~ WorldClim 2 counterparts. These results suggest that unless global climate models increase in resolution and accuracy, regional models will remain critical for local applications.

10   **4.3 Remote sensing variable**

The NDVI data represents vegetation quality and density based on two merged satellite images, both in March of their respective years. Although this is only an instantaneous representation of NDVI, we expect it to correlate strongly with ~~relative vegetation density as well as unvegetated areas~~ the spatial pattern of vegetation density throughout the year. ~~NDVI~~ Certain plant species shed and regenerate their leaves during specific months ranging from winter through mid-summer, but Hong

15   Kong's woody vegetation is overall evergreen (Dudgeon and Corlett, 1994), so seasonal changes in NDVI are not expected to be drastic. NDVI values above 0.4 include Hong Kong's densest forests, while unvegetated or urbanized areas are well below 0.1. The densest vegetation (> 0.4 NDVI) in Hong Kong tends to be on slopes between 100 m and 400 m elevation (Fig. 8), and is distributed between Hong Kong Island, Lantau Island, and the New Territories. One exception is the verdant mangrove forests, at sea level. The patchy distribution of high density vegetation likely reflects the effects of historical deforestation. The

20   largest patches are found on the southeastern slopes of Tai To Yan in the New Territories. The relative distribution of NDVI classes along Hong Kong's elevational gradient is shown in Figure 8. ~~Vegetation index values vary depending on time of year and with recent weather conditions, so much more future work could be done to characterize how this measure~~ Future work could determine to what extent NDVI changes over time, in response to seasonality or recent weather. The limiting factor is the availability of data of adequate temporal resolution, as many satellite images of Hong Kong are obscured by cloud cover or

25   degraded by poor air quality.

4.4 ~~Next steps~~Value and Utility

~~Important gaps in Hong Kong geographic data remain. Finer grain, microclimatic variables (Lembrechts~~ This new data will benefit environmental research, and specifically SDM studies, in two main ways. First, it will enable finer scale analyses than previously possible. For SDM, this means improved detection of climatic microrefugia (Meineri and Hylander, 2017), and the

30   ability to differentiate between human altered habitat and natural areas. Rampant development and a shifting climate make this knowledge of local species persistence more important than ever. Additionally, this is especially relevant in Hong Kong where topography varies dramatically, and where urban areas form a complex mosaic with undeveloped expanses.

Second, we provide a diverse array rasters derived from multiple independent data sources, but in a single resolution and format to facilitate further analysis and synthesis of meaning. For SDM, these diverse layers have distinct advantages over datasets that only contain climate data. Compared to climate data alone, using diverse predictors including topographic characteristics have been shown to be important variables for accurate SDM results, such as predicting the spread of invasive species in new ranges (Peterson and Nakazasa, 2008). However benefits of non-climate data may only be evident in finer scale SDMs (Luoto et al., ~~2018)based on remote sensing data and temperature loggers could be a next step for local climatemodeling.~~ 2007).

Finally, such high quality, diverse geographic data is especially uncommon in tropical regions, where improved knowledge for environmental research and biological conservation is most needed. According to Rapoport's Rule, tropical species are more likely to have smaller distributions (Stevens, 1989), and therefore future execution of local SDM studies to understand their ranges are particularly important.

4.5 Limitations and next steps

Here we outline how shortfalls of the data presented may be improved in the future. First, though we inferred Hong Kong's pattern of urban development from impervious surface data, this is less than ideal because in addition to concrete, bare soil or rock are sensed as impervious. Also, it cannot differentiate dense urban cores of high-rises from large paved areas. For climate modeling, an urbanicity measure that considers building height or population density at a 30 m or finer scale could be preferable.

Second, while our temperature rasters should accurately represent air temperature in open areas, they do not reflect the high spatial variation in temperature found in urban microclimates. Forexample, although the manned Kowloon HKO weather station is inside a densely populated area, as pointed out by Nichol and To (2012) it is still in a small parklike area surrounded by trees, and therefore is not representative of the most densely urbanized areas of Hong Kong. Other stations in urban areas are similarly near green spaces or otherwise open areas. Higher resolution (say 5 m or 1 m) studies of urban thermal distributions would strongly benefit from analysis of wind patterns, building height, thermal pollution, and other factors (e.g. Shi et al., 2018). Therefore granular, ground-level temperatures in urban areas are likely substantially different than the broader air temperature values our models provide.

Similar to other climate interpolation studies, bias in the physical locations of automatic weather stations may be of concern. Weather stations are often intentionally placed in flat, open areas with the goal of measuring weather that is relevant to a broad geographic area, rather than locations that may experience unique local climate. It may be for this reason that Slope*Aspect was infrequently useful for model construction, as few stations are on steep slopes. Elevational distribution of stations may also be a source of bias; although a weather station operates at the highest point in Hong Kong (Tai Mo Shan, 955 m), there are only two other stations above 600 m.

Finally, while we used cross-validation to measure the spatial predictive ability of the climate models, this method is only able to test models against locations where weather stations are present; validation based on an independently collected dataset would be ideal. One common validation method is to use weather data loggers placed across elevational and land-use gradients

(Meineri and Hylander, 2017). Such an approach would allow for explicit testing and comparing predictiveness of climate products for different areas of Hong Kong.

Important gaps in Hong Kong geographic data remain. Models projecting future climate scenarios would enable biodiversity change predictions, ~~and~~ with additional variables like cloud cover and solar radiation ~~would be~~ useful. A discrete classification of habitat type would be useful for ecological research~~. A series of studies in the past, commissioned by the Hong Kong government, classified Hong Kong's terrestrial habitats (e.g. Environmental Resources Management, 2009), however the results of these very thorough remote sensing studies are not accessible. Quality~~, and quality soil type data is lacking~~, and as mentioned before a better index of population density or built-up area would be welcomed~~. Availability of such data for Hong Kong would complement the ~~results~~ findings of this project, ~~and benefit geographic studies in Hong Kong~~which significantly advance our understanding of geographic heterogeneity in this complex tropical region.

5 Conclusions

This diverse set of 30 m resolution topography, climate, and remote sensing data include the first published interpolation of long-term climate averages specific to Hong Kong. Our findings suggest that global interpolated climate datasets are limited by their resolution, and underestimate local climate variability. Therefore the availability of such local data will remain critically important for the forseeable future. This new data will allow for a new generation of studies in Hong Kong, and enable connections between environmental data and biotic patterns at a much finer scale than previously possible. Aside from clear uses in conservation, ecological and biogeographic research, we also expect this freely accessible dataset to be broadly applicable for many sectors, including tourism, hydrology, recreation, agriculture, mapmaking, and real estate.

6 Data availability

GeoTIFF raster and shapefile documents can be downloaded from figshare: https://doi.org/10.6084/m9.figshare.6791276. A document in the repository includes file names, descriptions, and summary statistics for all provided rasters. Individual monthly rasters for each of the 10 climate variables are available as a compressed zip file.

References

Aalto, J., Riihimäki, H., Meineri, E., Hylander, K., and Luoto, M.: Revealing topoclimatic heterogeneity using meteorological station data. Int. J. Climatol., 37, 544-556, 2017.

Agriculture, Fisheries and Conservation Department, The Government of the Hong Kong Special Administrative Region: Annual Report 2016-2017, Country and Marine Parks, available at https://www.afcd.gov.hk/misc/download/annualreport2017/en/country-parks.html, 2017.

Ashcroft, M. B., and Gollan, J. R.: Fine-resolution (25 m) topoclimatic grids of near-surface (5 cm) extreme temperatures and humidities across various habitats in a large (200× 300 km) and diverse region. Int. J. Climatol., 32, 2134-2148, 2012.

Bennie, J. J., Wiltshire, A. J., Joyce, A. N., Clark, D., Lloyd, A. R., Adamson, J., Parr, T., Baxter, R., and Huntley, B.: Characterising inter-annual variation in the spatial pattern of thermal microclimate in a UK upland using a combined empirical–physical model. Agr. Forest Meteorol., 150, 12-19, 2010.

Cheng, W., and Bonebrake, T. C.: Conservation effectiveness of protected areas for Hong Kong butterflies declines under climate change. J. Insect Conserv., 21, 599-606, 2017.

Brown de Colstoun, E. C., C. Huang, P. Wang, J. C. Tilton, B. Tan. J. Phillips, S. Niemczura, P.-Y. Ling, and Wolfe, R.E.: Global Man-made Impervious Surface (GMIS) Dataset from Landsat, NASA Socioeconomic Data and Applications Center (SEDAC), Palisades, NY, 2017.

Connor, T., Hull, V., Viña, A., Shortridge, A., Tang, Y., Zhang, J., Wang, F., and Liu, J.: Effects of grain size and niche breadth on species distribution modeling, Ecography, 41, 1270-1282, 2018.

Cristiano, E., Veldhuis, M. C. T., and Giesen, N. V. D.: Spatial and temporal variability of rainfall and their effects on hydrological response in urban areas–a review. Hydrol. Earth Syst. Sc., 21, 3859-3878, 2017.

Dobesch, H., Dumolard, P., and Dyras, I. (Eds.).: Spatial interpolation for climate data: the use of GIS in climatology and meteorology. John Wiley and Sons, 2013.

Dobrowski, S. Z.: A climatic basis for microrefugia: the influence of terrain on climate, Glob. Change Biol., 17, 1022-1035, 2010.

Dudgeon, D., and Corlett, R.: Hills and streams: an ecology of Hong Kong. Hong Kong University Press, Hong Kong, 1994.

Environmental Resources Management: 2008 Update of Terrestrial Habitat Mapping and Ranking based on Conservation Value. Report to Sustainable Development Division, 2009.

Fick, S. E., and Hijmans, R. J.: WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol., 37, 4302-4315, 2017.

Franklin, J., Davis, F. W., Ikegami, M., Syphard, A. D., Flint, L. E., Flint, A. L., and Hannah, L.: Modeling plant species distributions under future climates: how fine scale do climate projections need to be?, Glob. Change Biol., 19, 473-483, 2013.

Giridharan, R., Ganesan, S., and Lau, S. S. Y.: Daytime urban heat island effect in high-rise and high-density residential developments in Hong Kong. Energ. Buildings, 36, 525-534, 2004.

Guisan, A., Graham, C. H., Elith, J., Huettmann, F., and NCEAS Species Distribution Modelling Group.: Sensitivity of predictive species distribution models to change in grain size. Divers. Distrib., 13, 332-340, 2007.

Heiblum, R. H., Koren, I., and Altaratz, O.: Analyzing coastal precipitation using TRMM observations. Atmos. Chem. Phys., 11, 13201-13217, 2011.

Hijmans, R. J., Phillips, S., Leathwick, J., Elith, J., and Hijmans, M. R. J.: Package 'dismo'. Circles, 9, 2017.

Hu, Z., Hu, Q., Zhang, C., Chen, X., and Li, Q.: Evaluation of reanalysis, spatially interpolated and satellite remotely sensed precipitation data sets in central Asia. J. Geophysical Res.: Atmospheres, 121, 5648-5663, 2016.

Hong Kong Observatory, hko.gov.hk, last accessed 20 September 2018.

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmerman, N.E., Linder, H. P., and Kessler, M.: Climatologies at high resolution for the earth's land surface areas. Scientific Data, 4, 170122, 2017.

Lands Department, The Government of the Hong Kong Special Administrative Region: Digital Terrain Model (DTM), available at https://data.gov.hk/en-data/dataset/hk-landsd-openmap-5m-grid-dtm, 2017.

Lassueur, T., Joost, S., and Randin, C. F.: Very high resolution digital elevation models: Do they improve models of plant species distribution?. Ecol. Model., 198, 139-153, 2006.

Lau, S. S., and Zhang, Q.: Genesis of a vertical city in Hong Kong. Int. J. High-Rise Build., 4, 117-125, 2015.

Lembrechts, J., Nijs, I., and Lenoir, J.: Incorporating microclimate into species distribution models. Ecography, In Press, 2018.

Levin, S. A.: The problem of pattern and scale in ecology: the Robert H. MacArthur award lecture, Ecology, 73, 1943-1967, 1992.

Liu, L., and Zhang, Y.: Urban heat island analysis using the Landsat TM data and ASTER data: A case study in Hong Kong. Remote Sensing, 3, 1535-1552, 2011.

Lookingbill, T. R., and Urban, D. L.: Spatial estimation of air temperature differences for landscape-scale studies in montane environments. Agr. Forest Meteorol., 114, 141-151, 2003.

Luoto, M., Virkkala, R., and Heikkinen R. K.: The role of land cover in bioclimatic models depends on spatial resolution. Global Ecol. Biogeogr., 16, 34-42, 2007.

Marafa, L. M., and Chau, K. C.: Effect of hill fire on upland soil in Hong Kong. Forest Ecol. Manag., 120, 97-104, 1999.

Meineri, E., and Hylander, K.: Fine-grain, large-domain climate models based on climate station and comprehensive topographic information improve microrefugia detection. Ecography, 40, 1003-1013, 2017.

Mertes, K., and Jetz, W.: Disentangling scale dependencies in species environmental niches and distributions, Ecography, 41, 1604-1615, 2018.

New, M., Lister, D., Hulme, M., and Makin, I.: A high-resolution data set of surface climate over global land areas. Clim. Res., 21, 1-24, 2002.

Nezer, O., Bar-David, S., Gueta, T., and Carmel, Y.: High-resolution species-distribution model based on systematic sampling and indirect observations. Biodivers. Conserv., 26, 421-437, 2017.

Ng, E.: Wind and heat environment in densely built urban areas in Hong Kong. Glob. Environ. Res., 13, 169-178, 2009.

Nichol, J., Hang, T. P., and Ng, E.: Temperature projection in a tropical city using remote sensing and dynamic modeling. Clim. Dynam., 42, 2921-2929, 2014.

Nichol, J. E., and To, P. H.: Temporal characteristics of thermal satellite images for urban heat stress and heat island mapping. ISPRS J. Photogramm., 74, 153-162, 2012.

Oke, T. R.: The energetic basis of the urban heat island. Quart. J. R. Met. Soc., 108, 1-24, 1982.

OpenStreetMap, www.openstreetmap.org, last access: 13 Febrauary 2018.

Paiva, R. C. D., Buarque, D. C., Clarke, R. T., Collischonn, W., and Allasia, D. G.: Reduced precipitation over large water bodies in the Brazilian Amazon shown from TRMM data. Geophys. Res. Lett., 38, 2011.

Peterson, A. T., and Nakazawa, Y.: Environmental data sets matter in ecological niche modelling: an example with Solenopsis invicta and Solenopsis richteri. Global Ecol. Biogeogr., 17, 135-144, 2008.

Peterson, A. T., Soberón, J., Pearson, R. G., Anderson, R. P., Martínez-Meyer, E., Nakamura, M., and Araújo, M. B.: Ecological niches and geographic distributions, Princeton University Press, Princeton, NJ, 2011.

Pradervand, J. N., Dubuis, A., Pellissier, L., Guisan, A., and Randin, C.: Very high resolution environmental predictors in species distribution models: Moving beyond topography?. Prog. Phys. Geog., 38, 79-96, 2014.

5 QGIS Development Team: QGIS Geographic Information System, Open Source Geospatial Foundation Project, http://qgis.osgeo.org, 2018.

R Core Team: R: A language and environment for statistical computing, 2016.

RStudio Team: RStudio: integrated development for R, RStudio, Inc., Boston, MA, http://www.rstudio.com, 42, 2015.

Seo, C., Thorne, J. H., Hannah, L., and Thuiller, W.: Scale effects in species distribution models: implications for conservation planning under climate change, Biol. Letters, 5, 39-43, 2009.

10 Shi, Y., Katzschner, L., and Ng, E.: Modelling the fine-scale spatiotemporal pattern of urban heat island effect using land use regression approach in a megacity. Sci. Total Environ., 618, 891-904, 2018.

Stevens, G.: The latitudinal gradient in geographical range: how so many species coexist in the tropics. Amer. Nat., 133, 240-256, 1989.

U.S. Geological Survey: Landsat-8, https://doi.org/10.5066/F71835S6, 2018.

Wahba, G.: How to Smooth Curves and Surfaces with Splines and Cross-Validation. Ft. Belvoir: Defense Technical Information Center, 1979.

15 Wang, R., Ren, C., Xu, Y., Lau, K. K. L., and Shi, Y.: Mapping the local climate zones of urban areas by GIS-based and WUDAPT methods: A case study of Hong Kong. Urban Clim., 24, 567-576, 2018.

Vega, G. C., Pertierra, L. R., and Olalla-Tárraga, M. Á.: MERRAclim, a high-resolution global dataset of remotely sensed bioclimatic variables for ecological modelling. Scientific Data, 4, 170078, 2017.

20 Zhuang, X. Y., and Corlett, R. T.: Forest and forest succession in Hong Kong, China. J. Trop. Ecol., 13, 857-866, 1997.
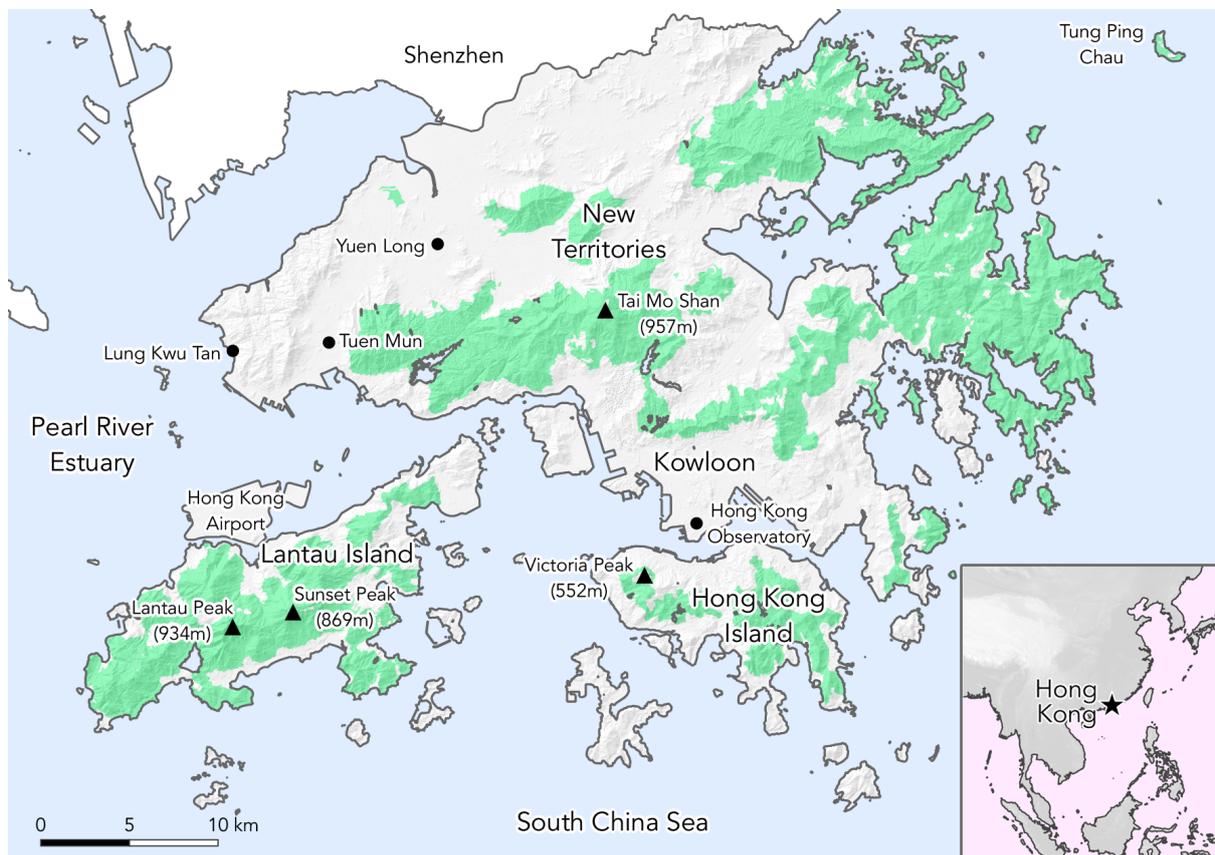
Figure 1. Hong Kong geography. The three highest peaks in the territory, as well as the highest point on Hong Kong Island are marked. Areas protected as Country Parks are highlighted in green.
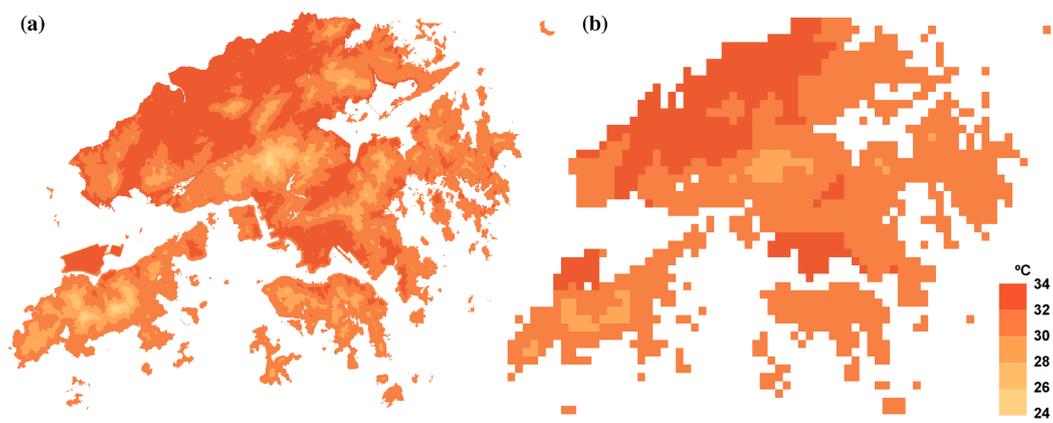
Figure 2. Comparison of average high of warmest month (bio5) model results for Hong Kong. (a) is from our newly interpolated climate models at 30 m resolution, while (b) is 1 km resolution data available as part of WorldClim 2 (Fick and Hijmans, 2017). Not only is the resolution markedly improved, but also the temperature values are more varied, for instance on the large southern islands.
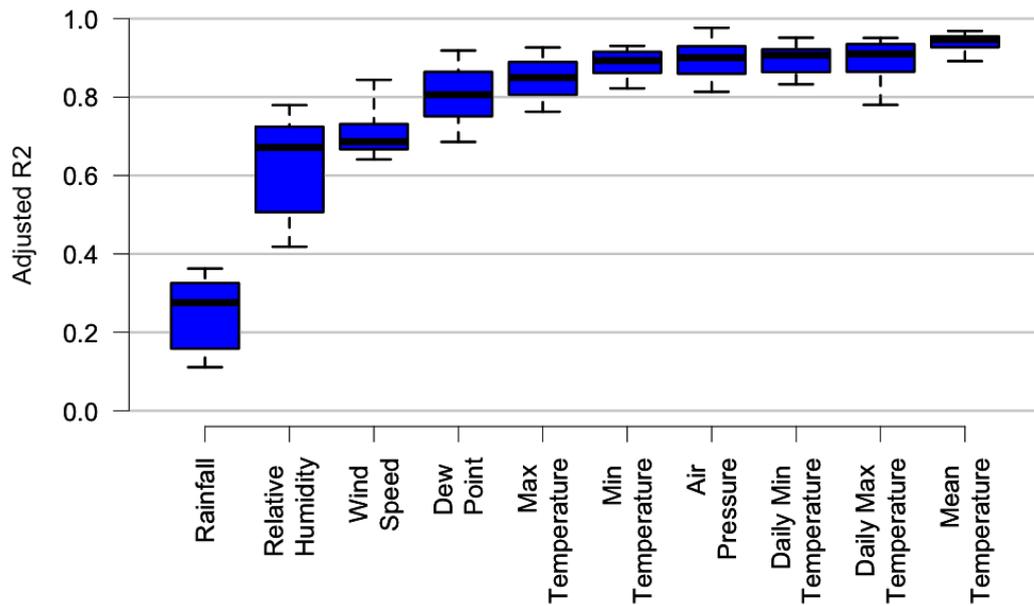
Figure 3. Adjusted r2 values of initial (pre-spline) regression models. Each boxplot includes 12 points, one for each monthly model. Temperature variation, especially mean temperature, was best explained by linear modeling, while rainfall was predicted the most poorly.
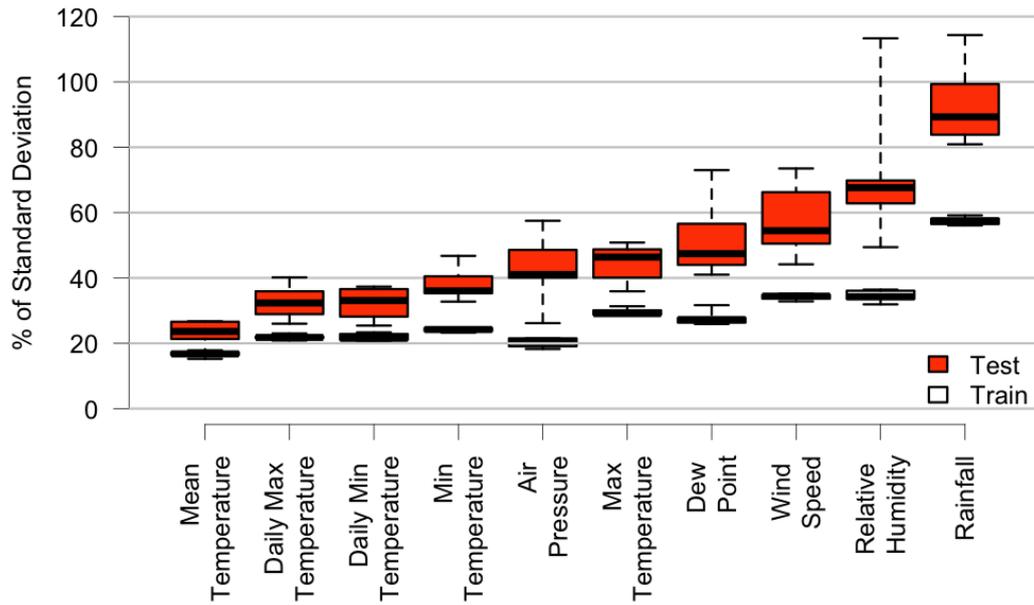
Figure 4. Relative magnitude of training and testing dataset errors, from 10 validation rounds of climate variable modeling. A value of 100 indicates for that climate model, that the average difference between the value recorded at a given weather station and the value predicted by the model at that location, is equal to the standard deviation of the initial set of all values recorded at all weather stations for that climate variable.
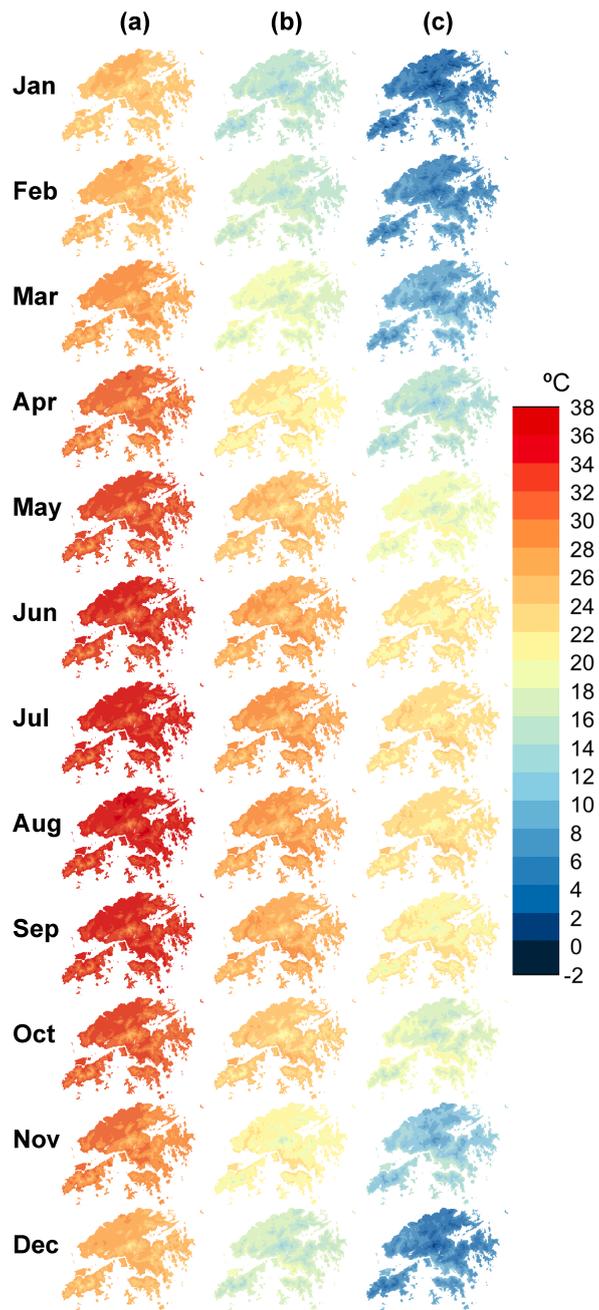
Figure 5. Model results for three of ten interpolated climate variables. (a) Maximum temperature, (b) Mean temperature, and (c) Minimum temperature.
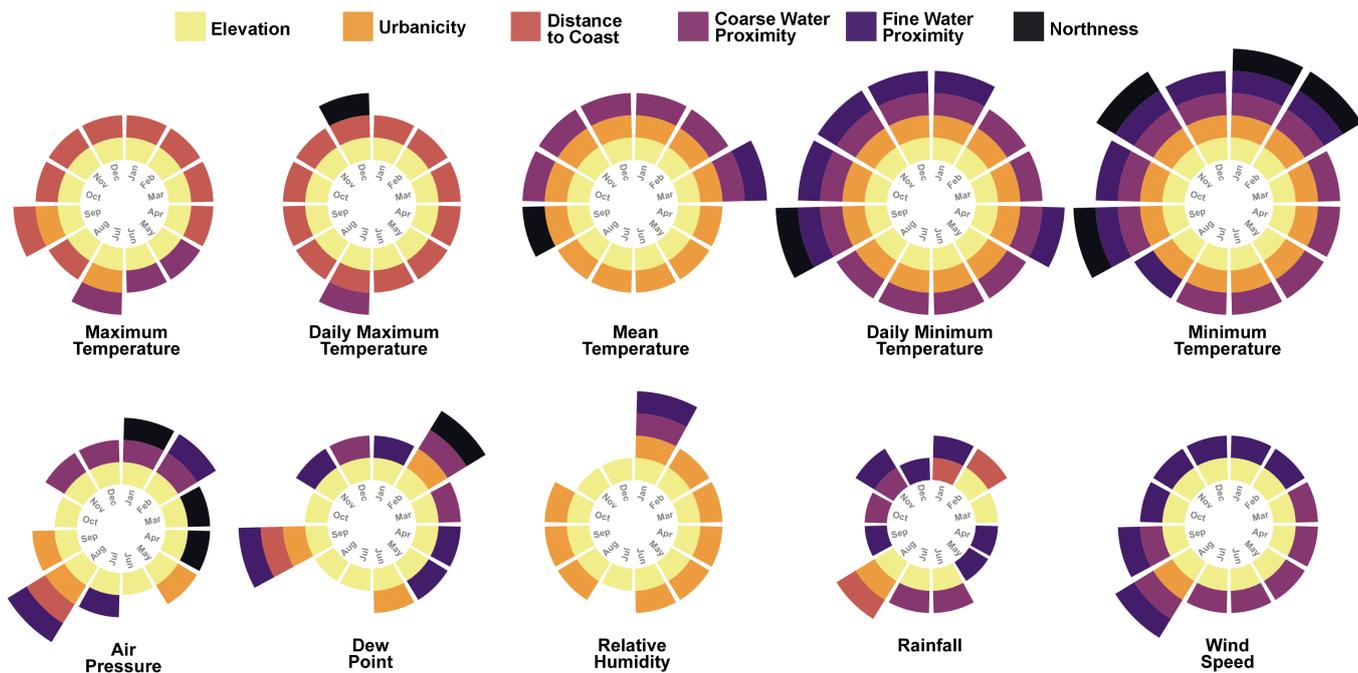
Figure 6. Regression predictors included in monthly models for 10 climate variables. Each predictor is represented by a different color. Minimum and mean temperature variables were most predictable, consistently including elevation and urbanicity. Rainfall patterns were most difficult, with the fewest predictors included.
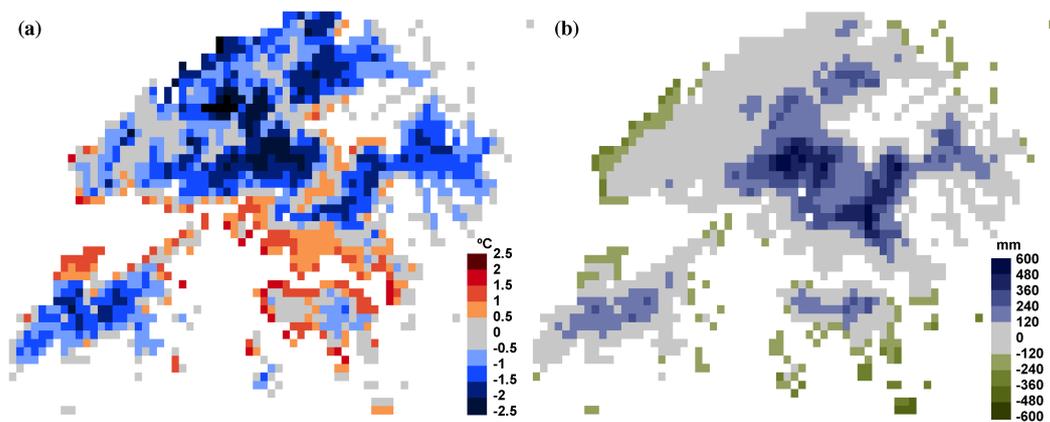
Figure 7. Differences between results of this study and Worldclim 2 (Fick and Hijmans, 2017) values. (a) is average low temperature of coldest month (bio6), with red where the local model is warmer than WorldClim, and blue is colder. (b) shows annual precipitation (~~bio2~~bio12), with blue where the local model predicts more rainfall than WorldClim, and tan is less rainfall. Our model results were resampled to 1 km resolution using bilinear interpolation to allow for these comparisons.
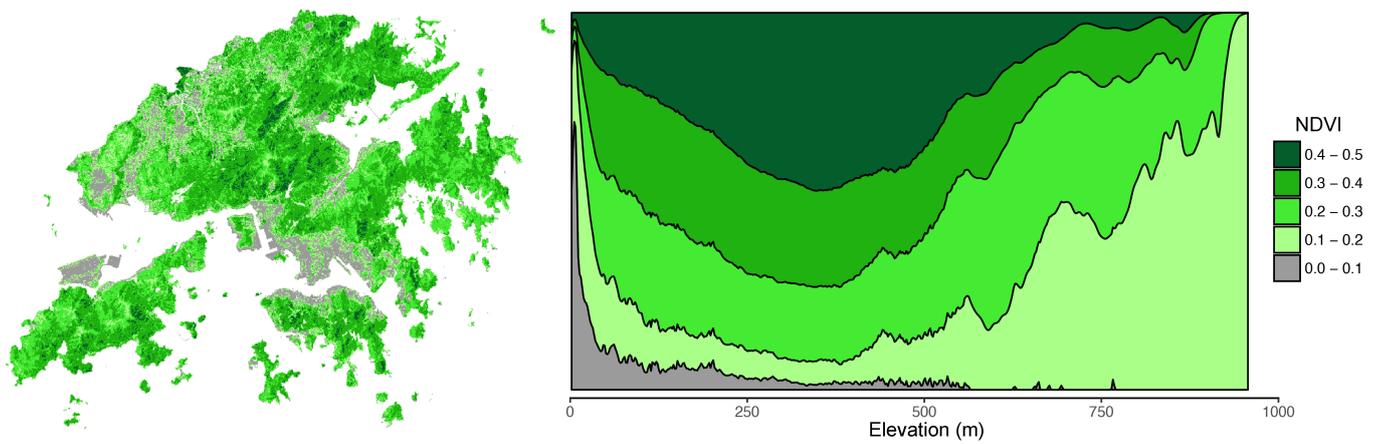
Figure 8. NDVI class composition over Hong Kong's elevational range. The majority of land area near sea level is below NDVI 0.1, while Hong Kong's highest elevation areas are between 0.1 and 0.2, indicating short vegetation. The elevation range with proportionally the most dense vegetation (0.4 to 0.5 NDVI) is 300 to 400 m.
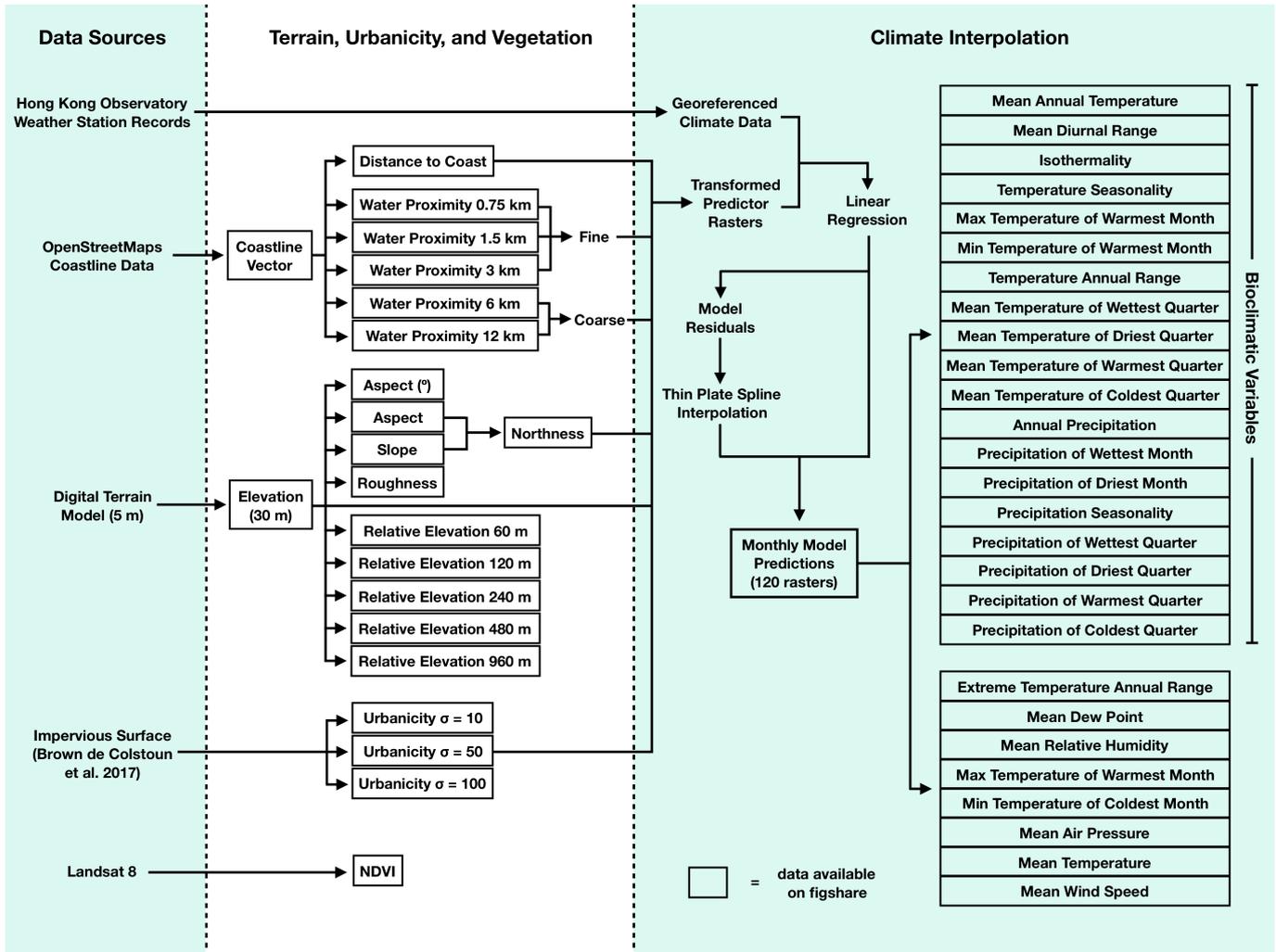
Figure S1. Schematic of data products and the sources that informed them. Items enclosed in a box represent the files available for download from the figshare repository.

Figure S2. Permanent weather stations operated by the Hong Kong Observatory. Symbols indicate what type of data is available from each station: temperature, rainfall, or both.

Table 1. Raster product descriptions, units, and 5th, 50th, and 95th percentile values.

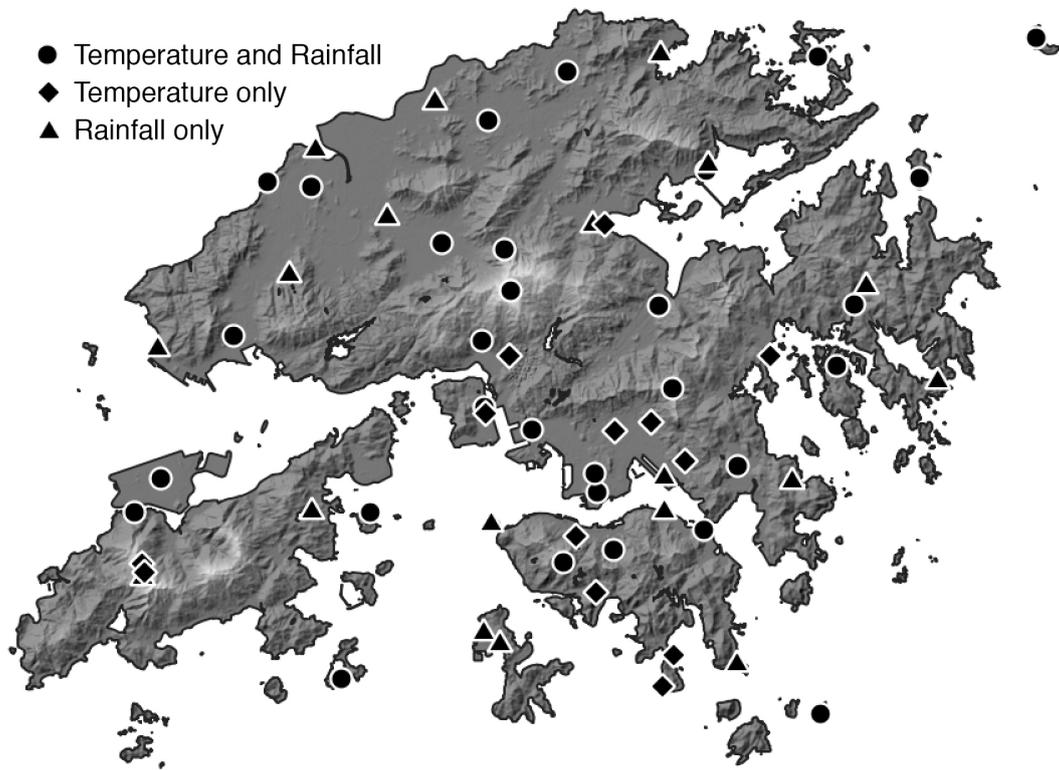| Description	Unit	5%	50%	95%	Filename
Aspect (Northness)	index	-0.99	0	0.99	aspect.tif
Aspect (Degree)	º	18	180	341	aspect_degree.tif
Slope	º	0	17	33	slope.tif
Terrain Roughness	index	0.33	24.95	50.67	rough.tif
Elevation	m	5	84	407	elevation.tif
Aspect * Slope	index	-23.5	0	23.58	aspect_x_slope.tif
Distance to Coast	m	68	1349	6186	waterdist.tif
Relative Elevation (60 m radius)	m	0	16	37	relelev60.tif
Relative Elevation (120 m radius)	m	0	28	69	relelev120.tif
Relative Elevation (240 m radius)	m	2	46	124	relelev240.tif
Relative Elevation (480 m radius)	m	2	64	208	relelev480.tif
Relative Elevation (960 m radius)	m	3	76	308	relelev960.tif
Water Proximity (0.75 km radius)	proportion	0.52	1	1	water25.tif
Water Proximity (1.5 km radius)	proportion	0.4	0.98	1	water50.tif
Water Proximity (3 km radius)	proportion	0.33	0.88	1	water100.tif
Water Proximity (6 km radius)	proportion	0.31	0.74	1	water200.tif
Water Proximity (12 km radius)	proportion	0.27	0.66	0.94	water400.tif
Annual Mean Temperature	ºC	20.8	22.9	24	biovars_t_1.tif
Mean Diurnal Range (Mean (max temp-min temp))	ºC	4.9	6.2	7.7	biovars_t_2.tif
Isothermality (bio2/bio7) (* 100)	index	27.4	31.9	35.6	biovars_t_3.tif
Temperature Seasonality (standard deviation *100)	index	467	496	512	biovars_t_4.tif
Average High Temperature of Warmest Month	ºC	28.9	31.5	32.8	biovars_t_5.tif
Average Low Temperature of Coldest Month	ºC	9.5	11.7	13.9	biovars_t_6.tif
Temperature Annual Range (bio5-bio6)	ºC	17.7	19.6	21.6	biovars_t_7.tif
Mean Temperature of Wettest Quarter	ºC	25.8	27.8	29.2	biovars_t_8.tif
Mean Temperature of Driest Quarter	ºC	14.4	16.3	17.4	biovars_t_9.tif
Mean Temperature of Warmest Quarter	ºC	25.9	28.2	29.2	biovars_t_10.tif
Mean Temperature of Coldest Quarter	ºC	14.4	16.3	17.4	biovars_t_11.tif
Annual Precipitation	mm	1738	2079	2415	biovars_t_12.tif
Precipitation of Wettest Month	mm	345	425	521	biovars_t_13.tif
Precipitation of Driest Month	mm	25	32	35	biovars_t_14.tif
Precipitation Seasonality (Coefficient of Variation)	index	78.7	82.8	86	biovars_t_15.tif

Table 1. Continued.

Precipitation of Wettest Quarter	mm	883	1085	1276	biovars_t_16.tif
Precipitation of Driest Quarter	mm	86	104	112	biovars_t_17.tif
Precipitation of Warmest Quarter	mm	814	1054	1260	biovars_t_18.tif
Precipitation of Coldest Quarter	mm	86	104	112	biovars_t_19.tif
Extreme Temperature Annual Range	℃	26.3	29	32.1	avars_annual_range.tif
Annual Mean Dew Point	℃	17.3	18.4	19.1	avars_dewp_mean.tif
Annual Mean Relative Humidity	%	75.4	80.4	84.9	avars_humid_mean.tif
Maximum Temperature of Warmest Month	℃	32.3	35	36.2	avars_max_tmax.tif
Minimum Temperature of Coldest Month	℃	2.4	5.6	8.6	avars_min_tmin.tif
Annual Mean Air Pressure	hPa	1012.5	1012.8	1013.4	avars_press_mean.tif
Actual Annual Mean Temperature	℃	20.3	22.4	23.6	avars_tmean_mean.tif
Annual Mean Wind Speed	km/h	5.4	11.6	19.2	avars_windsp_mean.tif
Urbanicity (sigma = 10)	%	0	0	68.9	urbanicity_gauss10.tif
Urbanicity (sigma = 50)	%	0	1.5	56	urbanicity_gauss50.tif
Urbanicity (sigma = 100)	%	0	3.3	50.1	urbanicity_gauss100.tif
Normalized Difference Vegetation Index (NDVI)	index	0.05	0.29	0.39	hk_ndvi.tif
Hong Kong Coastline and Reservoirs	-	-	-	-	HK_border.shp

Table 2. Number of weather stations that contributed data for each climate model.

	press	tmax	mtmax	tmean	mtmin	tmin	dewp	humid	prec	windsp
Jan	17	39	39	38	39	38	23	23	40	28
Feb	17	40	40	39	40	39	25	25	41	28
Mar	17	39	39	38	39	39	25	25	40	28
Apr	18	39	39	37	39	39	24	24	41	29
May	17	39	39	39	39	39	24	24	41	27
Jun	16	38	38	37	38	38	24	24	42	27
Jul	17	37	37	37	37	37	24	24	41	28
Aug	17	39	39	39	39	39	25	25	40	27
Sep	16	40	40	38	40	40	25	25	41	27
Oct	18	42	42	42	42	42	26	26	43	29
Nov	18	42	42	41	42	42	26	26	43	29
Dec	18	43	43	42	43	42	25	25	44	29

Table 3. Comparison of variation in bioclimatic variable rasters. All new rasters are more variable than their corresponding Worldclim 2 layers. Increased standard deviation ranges from 1.4x to 3.4x.

	Local Model SD	Worldclim 2 SD	Ratio
bio 1	1	0.5	1.9
bio 2	0.8	0.3	3
bio 3	2.5	0.7	3.4
bio 4	14.6	10.2	1.4
bio 5	1.2	0.7	1.7
bio 6	1.3	0.5	2.8
bio 7	1.2	0.6	2
bio 8	1.1	0.6	1.8
bio 9	0.9	0.5	1.9
bio 10	1.1	0.6	1.7
bio 11	0.9	0.5	1.9
bio 12	204.4	95.4	2.1
bio 13	52.9	21.5	2.5
bio 14	3.1	1.6	1.9
bio 15	2.2	1.1	2
bio 16	119.9	54.2	2.2
bio 17	8.2	4.1	2
bio 18	136.2	67.9	2
bio 19	8.2	5.5	1.5

Appendix A: Glossary of variable definitions

Maximum temperature   the highest temperature observed within a month

Mean daily maximum temperature   the mean of all daily high temperatures within a month

Mean daily temperature   the mean of all temperatures within a month

5   **Mean daily minimum temperature**   the mean of all daily low temperatures within a month

Minimum temperature   the lowest temperature observed within a month

Mean dew point   the mean of all dew point observations within a month

Mean relative humidity   the mean of all relative humidity observations within a month

Mean wind speed   the mean of all wind speed observations within a month

10   **Mean air pressure**   the mean of all air pressure observations within a month

Rainfall   the total of all rain recorded within a month

Relative elevation   the difference in elevation between the pixel of interest, and the lowest pixel within a given radius

Distance to coast   geometric distance between the pixel of interest and the nearest oceanic coastline

Water proximity   percent of area that is terrestrial within a given radius of the pixel of interest

15   **NDVI**   Normalized Difference Vegetation Index

Urbanicity   measure of area that is impervious surface within a given radius of the pixel of interest