GP <> Collider Hackathon: Extreme Temperature Index Development
March 31, 2019, Asheville, NC

Overview:
Global Parametrics (GP) uses geophysical models to provide developing economies with financial instruments to mitigate or recover from extreme events. Financial instruments come in many forms, but GP's are typically index-based risk transfer mechanisms, meaning that a payment is triggered if a predefined index is exceeded. For example, if rainfall exceeds 200mm in a 5 days period in Addis Ababa, if the center of tropical cyclone above category 3 approaches within 100km of Antananarivo, or if soil moisture anomaly drops below the 10th percentile in cropland regions of Mali.

For this hackathon, participants will extract modeled weather data for approximately 1,600 populated places from GP's project server. These data will serve as inputs to an Extreme Temperature Index (ETI) application, to be developed as the primary objective of the hackathon. Participants will be provided a 30-minute briefing on the models and data systems by a GP engineer, before being set loose on the data archive.

Objective:
Develop an application that implements ETI models to assess historical and forecast extreme temperature events. Using historical data for each location, ETI determine thresholds that can serve as exceedance triggers for payment, humanitarian response, or other remediation strategies to be developed by each team. Participants will be provided with a document of existing ETI methodologies (see Appendix A), but are encouraged to consider other techniques to determine extreme temperatures. The ETI methodology should consider both extreme heat and cold and preferably provide a confidence measure (e.g., probability) of occurrence, given the current state of weather.

Desired Features
- Data tables displaying relevant inputs (e.g., location, historical statistics) and corresponding outputs (e.g., current ETI, forecast ETI, ETI trigger/threshold)
- Spatial display of the ETI
- A prototype, interactive graphical user interface
- A short presentation overviewing the employed ETI methodology, including supporting data and statistics
- All code/documents committed to the GitHub under your team's branch (e.g., https://github.com/TheCollider/gp_collider_hackathon_repo/tree/team_x)

Deliverables:
No specific deliverables are dictated, instead participants are encouraged to interpret the desired features using the diversity of backgrounds represented on their team. All teams will present their findings and pitch their ETI application at the conclusion of the hackathon.

Notes:

- Participants are encouraged (but not limited) to use the data provided by GP
- Participants are encouraged to implement multiple ETI techniques and generate a probabilistic view of ETI
- Any model requires "ground truthing" to evaluate its prediction skill. Keep in mind how each module of your index and the index as a whole can be validated.

Technical Materials:

For this hackathon, your team will need internet connected computers, the ability to access a PostgreSQL database (or download and import large CSV files), a data analysis program, and a mapping program. Recommended utilities are:

- Data analysis utility: R/R-Studio, Python, Matlab
- PostgreSQL utility: psql, PgAdmin
- Mapping utility: QGIS, ArcGIS, Leaflet, Google EE

Data:

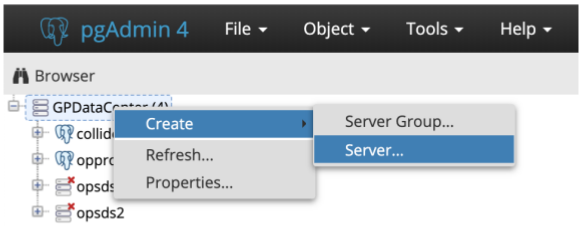
GP will provide a combination of exposure and climatological data. The exposure is comprised of a database of 1,600 urban areas around the world with populations ≥250,000. The climatological data consists of two reanalysis datasets, CFSR/v2 and JRA55, and one forecast dataset, CFSv2. Each climatology dataset contains the maximum and minimum daily temperature, among several other variables (see Appendix B). The data is available through GP's PostgreSQL database and in CSV format. Participants are encouraged to consider data in addition to those provided. For example, key contributors to heat stress include relative humidity and demographic conditions, which are not provided.

Data in GP's PostgreSQL database can be accessed through PostgreSQL utilities, such as PgAdmin4 and psql, or by using the GP taRpan R package.

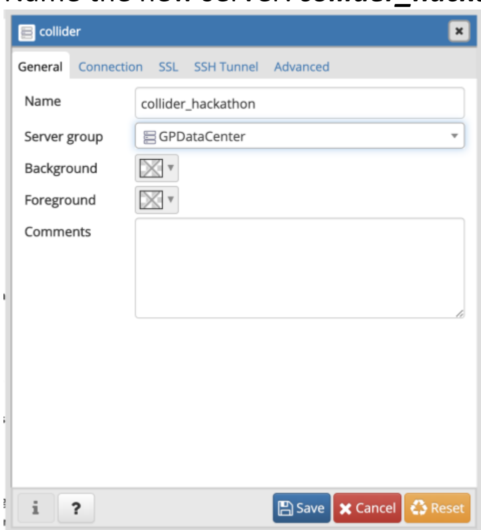
Data - access using pgAdmin 4 or psql:

Access to GP's data center through [pgAdmin 4] is setup by installing the utility and creating a new server connection, as follows:

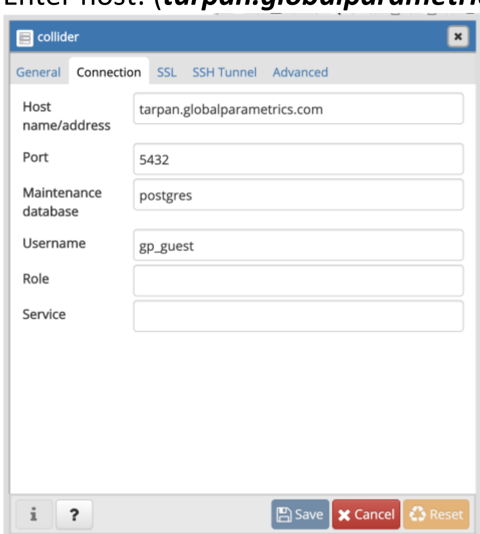
1. Connect a new server



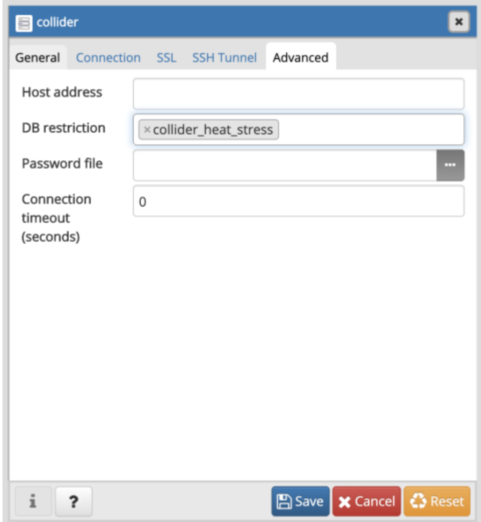
2. Name the new server: ***collider_hackathon***



3. Enter host: (***tarpan.globalparametrics.com***) and username: (***gp_guest***)



4. Specify the database for access: ***collider_heat_stress***



If you prefer to access PostgreSQL from the terminal, you can access the database by issuing the following command:

```
psql -h tarpan.globalparametrics.com -d collider_heat_stress -U gp_guest
```

<u>Data - access using R/RStudio</u>:

The ***taRpan*** R package takes care of the PostgreSQL queries and user configuration for you and ultimately makes accessing the data easier…if you're an R user. A readonly version of ***taRpan*** can be installed using the following commands:

```
install.packages('remotes')
library(remotes)
install_github('GlobalParametrics/taRpan_readonly')
```

Once installed, issue the command `library('taRpan')` to load the package

This package has numerous functions, but probably the most useful for the purposes of the hackathon are `tarpan2_get_table()` and `tarpan_model_data()`, which can be queried as follows:

```
tarpan2_get_table(con = 'collider_heat_stress',
table_name = 'project_tbl')
```

This function returns a data frame with the city location and population data, including key fields, such as location name (fname), Geonames ID (geo_id), and administrative regions, that can be used as geography inputs to the `tarpan_model_data()` function. The `tarpan_model_data()` can be used to extract climatology data for one or more locations. An example follows:

```
tarpan_model_data(dbname = 'collider_heat_stress',
geography = 'Charlotte', model = 'cfs2_forecast',
variable = 'max_temperature_gmtday',start = '2019-03-12',
end = '2019-03-31')
```

Data - access using CSV:

Data is also available to be downloaded in CSV format using the following link, when substituted with the 'FileName' in the table below. Use any utility you see fit (e.g., wget, R::download.file), but note the size of each file and carefully consider the datasets you require prior to downloading.

https://gpdev1.file.core.windows.net/gpfiles/collider_heat_stress/**{FileName}**?sv=2018-03-28&ss=f&srt=co&sp=rwdlc&se=2019-04-03T19:40:56Z&st=2019-03-30T11:40:56Z&spr=https,http&sig=oD1QiM%2FeC0RSuwuG1T4iJrIbiVU1jF%2Fq%2BTgZS44gfQ4%3D

Model	Type	Variable	Bias Corr.	Size	FileName
CFSR/v2	Hindcast	Max Temp (K)	N	2.1GB	collider_cfs2_hindcast_max_temp.csv
CFSR/v2	Hindcast	Max Temp (K)	Y	2.1GB	collider_cfs2_hindcast_max_temp_resmap.csv
CFSR/v2	Hindcast	Min Temp (K)	N	2.1GB	collider_cfs2_hindcast_min_temp.csv
CFSR/v2	Hindcast	Min Temp (K)	Y	2.1GB	collider_cfs2_hindcast_min_temp_resmap.csv
CFSR/v2	Hindcast	Max Wind (m/s)	N	2.1GB	collider_cfs2_hindcast_wind.csv
CFSR/v2	Hindcast	Max Wind (m/s)	Y	2.1GB	collider_cfs2_hindcast_wind_resmap.csv
CFSR/v2	Hindcast	Max Wind (m/s)	N	2.1GB	collider_cfs2_hindcast_precip.csv
CFSR/v2	Hindcast	Tot. Precip. (mm)	Y	2.0GB	collider_cfs2_hindcast_precip_resmap.csv
CFSR/v2	Hindcast	Tot. Precip. (mm)	N	2.0GB	collider_cfs2_hindcast_wind_temp.csv
JRA55	Hindcast	Max Temp (K)	N	2.8GB	collider_jra55_hindcast_max_temp.csv
JRA55	Hindcast	Min Temp (K)	N	2.8GB	collider_jra55_hindcast_min_temp.csv
JRA55	Hindcast	Max Wind (m/s)	N	2.8GB	collider_jra55_hindcast_wind.csv
JRA55	Hindcast	Tot. Precip. (mm)	N	2.8GB	collider_jra55_hindcast_precip.csv
CFSR/v2	Forecast	Max Temp (K)	N	21GB	collider_cfs2_forecast_max_temp.csv
CFSR/v2	Forecast	Min Temp (K)	N	21GB	collider_cfs2_forecast_min_temp.csv
Geonames	Exposure	Location	N/A	0.3MB	collider_project_tbl.csv
Geonames	Exposure	Xref	N/A	0.2MB	collider_project_xref.csv

Appendix A:
Many extreme temperature indices (ETIs), as proposed in this workshop, are currently operational in locations around the world. What differentiates the work here from prior studies is an effort to develop generalized, global ETIs that are effective for the purpose of triggering ex-post or ex-ante financing to affected locations.

Among the many reference materials available via a search, we recommend you refer to the following documents to guide the development of your index and application:

1. Heatwaves and Health: Guidance on Warning-System Development
2. Global predictability of temperature extremes
3. Comparing meteorological and perceived heat wave events (see attached)

These documents provide useful context for the problem at hand and overview work that has already been completed by experts in the field. Of note, are previously developed temperature-based indices that can serve as components of your algorithm and application (Table 1).

Table 1 Summary of meteorological definitions of extreme temperature considered in this study.

Label	*Description*	*Locations Used*	*Citation*
AvgRel4	$\geq$ 4 days with average daily temperature > 90th percentile	North Africa	Fontaine et al (2013)
AvgRel2(mid)	$\geq$ 2 days with average daily temperature > 95th percentile	USA,	Anderson and Bell (2010),
AvgRel3+	period of sequential days > 81st percentile, with $\geq$ 3 days > 97.5th percentile, with average temperature also > 97.5th percentile	China	Peng et al (2010)
MaxAbs2	$\geq$ 2 days with max daily temperature > 35°C, between 15th of June and 15th of September	Shanghai	Tan et al (2006)
MaxAbs2(d/n)	$\geq$ 2 days with daytime max temperature > 30°C and nighttime max temperature > 15°C	UK	UK Met Office (2015)
AvgRel2(high)	$\geq$ 2 days with average daily temperature > 98th percentile	USA	Kent et al (2013)
AvgRel2(low)	$\geq$ 2 days with average daily temperature > 90th percentile	USA	Kent et al (2013)
MaxAbs3	$\geq$ 3 days with max daily temperature > 35°C	Australia	Hansen et al (2008)
MaxRel1(low)	> 5°C anomaly in daily max temperature, when climatological value is $\leq$ 40°C OR > 4°C anomaly in daily max temperature, when climatological value is > 40°C OR daily max temperature > 45°C	India	Indian Met. Dept. (2015)
MaxRel1(high)	> 7°C anomaly in daily max temperature, when climatological value is $\leq$ 40°C OR > 6°C anomaly in daily max, when climatological value is > 40°C OR daily max temperature > 45°C	India	Indian Met. Dept. (2015)
MaxRel6	$\geq$ 6 days with max daily temperature > 90th percentile, using base peried of 1960-1990, and 5 day window for calculating percentile	—	Karl et al (1999)

Excerpted from 'Comparing Meteorological and Perceived Heat Wave Events'

Humans are not only impacted by temperature itself, but also meteorological (e.g., humidity, wind – or lack thereof), demographic (e.g., poverty, age, gender), and human physiology (e.g., adaptation, ailments) conditions. A more thorough treatment of these conditions is provided in "*Heatwaves and Health: Guidance on Warning-System Development*", but participants are encouraged to think broadly about the factors that contribute to extreme temperatures and how these may be condensed into a consistent index.

Appendix B:

Exposure metadata:

Variable	Description
fname	friendly-name
sub_iso	iso3 or classification not specified in admin region
geo_id	unique id
admin0_country_name	Admin Level 0 (e.g., country)
admin1_state_name	Admin Level 1 (e.g.,state or province)
admin2_county_name	Admin Level 2 (e.g., county)
admin3_name	Admin Level 3 (e.g., zip code)
admin4_name	Admin Level 4 (e.g., census block)
area	Sq. km
asciiname	standard search name
continent	2 letter continent code
currency	local currency
data_provider	data provided (e.g., Geonames, NaturalEarth, Landscan)
elevation	meters AMSL
feature_class	lassification of assets at location
geometry	WKT format, support for POINT MULTIPOINT POLYGON
geoname_id	original geoname id when published
hname	hashname (if applicable)
iso3	3-Letter country code
label	name for plotting and labeling <20 characters
land_cover	MODIS landcover type
landscan_id	original landscan id when published
level	Estimation of scale (e.g. Admin0, Admin1 etc,)
naturalearth_id	orginal natural earth id when published
op_code	batch processing number code
population	Population, as specified by Geonames
publisher_sku	Publisher name
timezone	GMT offset - not verified
wiki_url	wikipedia page that cross ref location. no guarantees
branch_book	weighting value of location (if applicable)
ag_season	start of agricultural season (if applicable)
cci_start	crop calendar start date (if applicable)
cci_end	crop calendar end date (if applicable)
region	classification not specified in admin region

General circulation model (GCM) climatology metadata:

Variable	Description
gid	unique id, corresponds to 'geo_id' in the exposure
model_grid	GCM model grid id
initial_time	time at which values are generated
valid_time	time at which values are valid
model	GCM model name
val	GCM grid value
event_id	Name of event (equals model for GCMs)

CFSR/v2 climatology variables:

Variable	Description	Unit
soilq_005cm	Soil moisture at a depth of 5cm	%
soilq_020cm	Soil moisture at a depth of 20cm	%
soilq_070cm	Soil moisture at a depth of 70cm	%
soilq_150cm	Soil moisture at a depth of 150cm	%
max_temperature_gmtday	Maximum daily temperature	K
min_temperature_gmtday	time at which values are generated	K
total_precipitation_gmtday	time at which values are valid	mm
peak_wind_speed_gmtday	GCM model name	m/s

CFSv2 forecast variables:

Variable	Description	Unit
soilq_005cm	Soil moisture at a depth of 5cm	%
soilq_020cm	Soil moisture at a depth of 20cm	%
soilq_070cm	Soil moisture at a depth of 70cm	%
soilq_150cm	Soil moisture at a depth of 150cm	%
max_temperature_gmtday	Maximum daily temperature	K
min_temperature_gmtday	time at which values are generated	K

JRA55 climatology variables:

Variable	Description	Unit
soil_moisture_pct	Soil moisture volume	%
max_temperature_gmtday	Maximum daily temperature	K
min_temperature_gmtday	time at which values are generated	K
total_precipitation_gmtday	time at which values are valid	mm
peak_wind_speed_gmtday	GCM model name	m/s